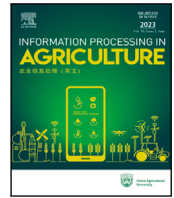




Contents lists available at ScienceDirect

## Information Processing in Agriculture

journal homepage: [www.keaipublishing.com/en/journals/information-processing-in-agriculture/](http://www.keaipublishing.com/en/journals/information-processing-in-agriculture/)

# Machine learning for asymptomatic Ratoon stunting disease detection with freely available satellite based multispectral imaging

Ethan Kane Waters<sup>a,b,c</sup> ,\* , Carla Chia-Ming Chen<sup>b,c,d</sup> , Mostafa Rahimi Azghadi<sup>b,c,d</sup> 

<sup>a</sup> Department of Primary Industries, 203 Tor Street, Toowoomba, 4350, QLD, Australia

<sup>b</sup> College of Science and Engineering, James Cook University, Townsville, 4818, QLD, Australia

<sup>c</sup> Agriculture Technology and Adoption Centre, James Cook University, Townsville, 4814, QLD, Australia

<sup>d</sup> Centre for AI and Data Science Innovation, James Cook University, Townsville, 4814, QLD, Australia

## ARTICLE INFO

## Keywords:

Sugarcane  
Health monitoring system  
Remote sensing  
Satellite-based spectroscopy  
Machine learning  
Vegetation indices  
Disease detection

## ABSTRACT

Disease detection in sugarcane, particularly the identification of asymptomatic infectious diseases such as Ratoon Stunting Disease (RSD), is critical for effective crop management. This study employed various machine learning techniques to detect the presence of RSD in different sugarcane varieties, using vegetation indices derived from freely available satellite-based spectral data. Our results show that the Support Vector Machine with a Radial Basis Function Kernel (SVM-RBF) was the most effective algorithm, achieving classification accuracy between 85.64% and 96.55%, depending on the variety. Gradient Boosting and Random Forest also demonstrated high performance achieving accuracy between 83.33% to 96.55%, while Logistic Regression and Quadratic Discriminant Analysis showed variable results across different varieties. The inclusion of sugarcane variety and vegetation indices was important in the detection of RSD. This agreed with what was identified in the current literature. Our study highlights the potential of satellite-based remote sensing as a cost-effective and efficient method for large-scale sugarcane disease detection alternative to traditional manual laboratory testing methods.

## 1. Introduction

Disease presents a formidable hurdle to maximising yield in the sugarcane industry with Ratoon stunting disease (RSD) emerging as a key contributor globally [1]. The bacterium *Leifsonia xyli subsp. xyli* (Lxx) is the primary causal agent of RSD, which is primarily propagated with contaminated cutting implements [2]. RSD infection can result in substantial yield losses (up to 60%), depending on sugarcane variety and water availability during growth periods [1,3–5]. This results in pronounced economic repercussions, with an estimated annual loss of \$25 million observed across 87,000 hectares monitored in Australia in 2019 [6]. The lack of external symptoms [1,4,5,7] has made the detection and the management of RSD a significant challenge. It relies heavily on laboratory diagnostic techniques such as Polymerase Chain Reaction (PCR), Quantitative Polymerase Chain Reaction (qPCR), Loop-Mediated Isothermal Amplification (LAMP), or Leaf Sheet Biopsy Quantitative Polymerase Chain Reaction (LSB-qPCR) for RSD detection [8–11]. Collecting the field samples required for these techniques is both time-consuming and costly, especially for large-area detection. This underscores the need to develop a more efficient method for large-scale detection of RSD.

Previous research has demonstrated success in using spectroscopy to detect disease and pests in sugarcane by leveraging subtle differences often imperceptible to the naked eye [12]. Spectroscopy measures the electromagnetic radiation reflected from an object, providing insights into its chemical composition, molecular structure, and physical properties [13] that may indicate the presence of disease. However, most of the studies in the literature utilised handheld spectrometers [12,14–19], which are challenging to scale up for large field detection. Hence, recent studies have shifted towards drone or satellite-based observations [20–25], acknowledging the need for larger-scale diagnostic methods. Despite this, no studies to date have yet utilised satellite-based multispectral imaging for disease detection in sugarcane. Although three previous studies have explored the diagnosis of asymptomatic sugarcane diseases with machine learning and spectroscopy, both relied on handheld spectrometers [12,14,15]. No current research has applied large-scale remote sensing techniques to detect asymptomatic diseases in sugarcane.

Freely available satellite-based remote sensing offers a cost-effective and efficient alternative to the traditional, resource-intensive methods of identifying and managing disease in sugarcane [27]. In particular,

\* Corresponding author at: Department of Primary Industries, 203 Tor Street, Toowoomba, 4350, QLD, Australia.

E-mail addresses: [ethan.waters@dpi.qld.gov.au](mailto:ethan.waters@dpi.qld.gov.au) (E.K. Waters), [carla.ewels@jcu.edu.au](mailto:carla.ewels@jcu.edu.au) (C.C.-M. Chen), [mostafa.rahimiazghadi@jcu.edu.au](mailto:mostafa.rahimiazghadi@jcu.edu.au) (M.R. Azghadi).

<https://doi.org/10.1016/j.inpa.2025.12.001>

Received 24 October 2024; Received in revised form 19 October 2025; Accepted 1 December 2025

Available online 1 December 2025

2214-3173/© 2025 The Authors. Published by Elsevier B.V. on behalf of China Agricultural University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Table 1**  
Current literature for disease and pest detection in sugarcane with ML and spectroscopy.

Reference	Health condition	Spectroscopy	Category	ML algorithm	Best classification accuracy
Apan et al. [21,26]	Orange Rust	Hyperspectral	Satellite	LDA	96.90%
Moriya et al. [20]	Mosaic	Hyperspectral	Drone	SID	92.50%
Grisham et al. [15]	SCYLV	Hyperspectral	Handheld Spectrometer	LDA	73%
Narmilan et al. [22]	White Leaf Disease	Multispectral	Drone	RF, DT, KNN, XGB	92%, 91%, 92%, 92%
Simões and Rios do Amaral [23]	Orange & Brown Rust	Multispectral	Drone	RF, KNN, SVM	90%, 90%, 90%; 86%, 83%, 88%
Ong et al. [16]	Brown Stripe & Ring Spot	Hyperspectral	Handheld Spectrometer	RF, SVM, NB	95%, 85%, 77%
Johansen et al. [24]	Cane Grub	Multispectral	Satellite	GEOBIA	79%
Johansen et al. [25]	Cane Grub	Multispectral	Satellite	GEOBIA	98.7%
Bao et al. [12,14]	Smut & Mosaic	Hyperspectral	Handheld Spectrometer	CNN	>90%
Vargas et al. [17]	Diatraea saccharalis	Hyperspectral & Multispectral	Handheld Spectrometer & Satellite	N/A	79.8% & 85.5%
Abdel-Rahman et al. [18]	Thrips	Hyperspectral	Handheld Spectrometer	N/A	N/A
Soca-Muñoz et al. [19]	Orange & Brown Rust	Hyperspectral & Multispectral	Handheld Spectrometer & Drone	N/A	N/A

free publicly accessible multispectral satellite data can alleviate the financial burden of purchasing expensive spectrometers images, and promote the widespread adoption of this advanced technology.

Spectroscopy data can be converted into a ‘Vegetation Index’ to accentuate the important characteristics of vegetation for the specific task [28,29]. The suitability of a vegetation index for a particular application depends on both the spectrometer and the project objective [28]. The majority of the sugarcane disease and pest detection studies [17,19,21–25] leverage vegetation indices to improve performance. This was particularly prominent in the study by Apan et al. [21] which developed several new vegetation indices for sugarcane disease detection with a focus on water and vegetation stress.

The application of machine learning (ML) algorithms to spectroscopy data has emerged as a promising approach to aid disease management in sugarcane (Table 1). A wide range of algorithms has been employed in this field, including XGBoost (XGB), Random Forest (RF), Decision Trees (DT), K-Nearest Neighbours (KNN), Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), and Neural Networks (NN) [12,15,16,21,22,22,23,26]. Despite the diversity in approaches, only a handful of studies have systematically compared ML performance [16,22,23]. Among the algorithms, RF is the most common and consistently achieves high classification accuracy [16,22,23]. Similarly, SVM has demonstrated strong performance, achieving promising results [22,23]. However, despite the variety of methods applied, the limited exploration of each method is insufficient to draw robust conclusions about the relative effectiveness of these algorithms. This highlights the need for continued comprehensive comparisons across multiple datasets and conditions to better understand their relative effectiveness in sugarcane disease management. Additionally, few studies have accounted for the impact of different sugarcane varieties on model performance, warranting further exploration given the increased variation introduced by classifying several varieties [15,18,23].

The primary objective of this research is to evaluate the effectiveness of various machine learning algorithms in classifying asymptomatic RSD in multiple varieties of sugarcane with freely available multispectral satellite data. Additionally, the study seeks to identify which vegetation indices are most important for diagnosing RSD. To the best of our knowledge, this study represents the first effort to classify an asymptomatic disease in sugarcane with satellite-based spectroscopy and is the first instance of diagnosing RSD with machine learning and spectroscopy [27].

## 2. Methodology

### 2.1. Overall process summary

The project comprises three stages: ground-truth data collection, data preprocessing, and ML development. In stage one, RSD data is collected from sugarcane blocks. For stage two, sentinel-2 images with pre-applied atmospheric correction undergo preprocessing before pixelwise values are extracted and labelled from raw spectral bands and vegetation indices. In the third stage, the dataset is divided into an 80:20 train-test split, with 10-fold validation applied to the training set for hyperparameter tuning. The top-performing models are subsequently evaluated on the test set by performing bootstrapping with 5000 samples. Permutation testing is then performed to ensure model performance is significantly better than a null distribution. The steps are detailed in the following sections.

### 2.2. Study location & ground truth data collection

Sampling to develop the ground truth dataset for this study was conducted by trained field agronomists at Herbert Cane Productivity Services Limited (HCPSL) between February and March 2022 across 76 sugarcane blocks in the Herbert region of Queensland, Australia. The dataset details the RSD status and variety of the sugarcane in each of the 76 blocks, which are comprised of five sugarcane varieties, Q200, Q208, Q240, Q253 and SRA14.

Two types of sugarcane blocks were sampled; farmer-owned blocks and HCPSL-owned seed production blocks. HCPSL-owned seed production blocks exist to provide farmers the opportunity to purchase disease-free sugarcane seeds due to strict hygiene and disease testing requirements. HCPSL followed their standard RSD sampling protocol for both types of blocks. In seed production plots samples from a sugarcane stick was collected every 20 m in a grid pattern. In Farmer-owned blocks, samples from twelve sugarcane sticks were collected from the four-corners of the block, and four from randomly selected locations within the block. The xylem extraction method was employed, in which juice from each sugarcane stick was extracted into a vial using pressure generated by an air compressor. The extracted juice from all samples was then frozen and subsequently analysed with qPCR testing at Sugar Research Australia (SRA) laboratories. The entire block

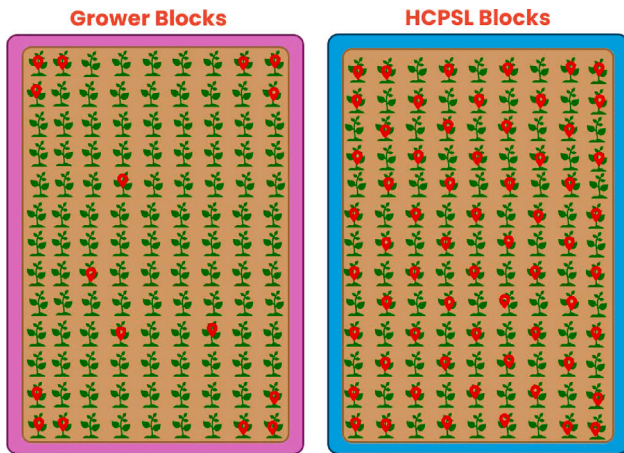


Fig. 1. Visual representation of the field sampling for both block types where the red marker indicates a sampling location.

is classified as RSD Positive if any sample within that block returned as RSD positive, and only labelled RSD Negative if all samples return a negative result. A visualisation of the field sampling method for the two block types is shown in Fig. 1.

To establish reliable ground truth data for model training and validation, RSD negative samples were sourced exclusively from HCPSL seed cane blocks that returned negative qPCR test results, while RSD positive samples were sourced from grower blocks that returned positive qPCR test results. The more intensive sampling protocols implemented by HCPSL staff to meet strict disease control requirements led to different sampling densities between the two block types. However, this does not introduce bias into the dataset. The study labels data at a block level and consequently does not analyse within block variation which could be influenced by different sampling densities. Additionally, RSD Negative samples were only drawn from HCPSL seed blocks with the higher sampling density to increase confidence that these blocks were truly absent of RSD. This conservative approach to labelling negative samples is appropriate given the asymptomatic and contagious nature of RSD, where undetected infections pose a risk of mislabelling. In contrast, RSD positive samples were drawn only from grower blocks where qPCR testing confirmed disease presence. While these blocks were sampled less intensively, this does not compromise the validity of the positive labels, as only confirmed cases were included. It is possible that some infected blocks were missed due to lower sampling density and therefore excluded from the dataset entirely. Overall, the differing sampling strategies were intentionally structured to minimise label noise and enhance the reliability of both positive and negative class assignments.

A shapefile corresponding to the sampled blocks with the sugarcane variety and RSD status determined in the laboratory testing was produced by HCPSL GIS officer Rod Nielson. However, due to a nondisclosure agreement with the farmers, the shapefile and precise maps of the sampled areas cannot be published.

### 2.3. Multispectral satellite image acquisition and data labelling

This study utilised the European Space Agency's (ESA) Sentinel-2 satellite series, which offers freely available products for the wavelengths necessary to calculate the vegetation indices listed in Table 3. "Level-2A" products are provided for all required bands at a 20 m resolution and have had atmospheric correction applied by ESA with the Sen2Cor process [30]. To streamline data processing, a Python script was developed as a pipeline to extract all spectral bands from the "Level-2A" satellite image data products provided by Sentinel-2 via the "Sentinel Hub Process API".

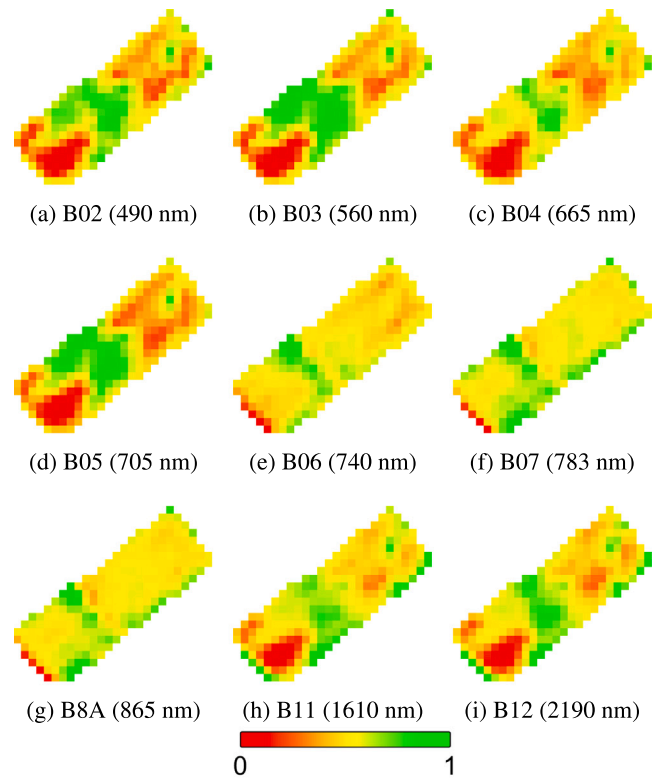


Fig. 2. Sentinel-2 multispectral bands for an example block, with each panel showing a different band. Pixel values have been normalised to highlight within-field variation.

Table 2

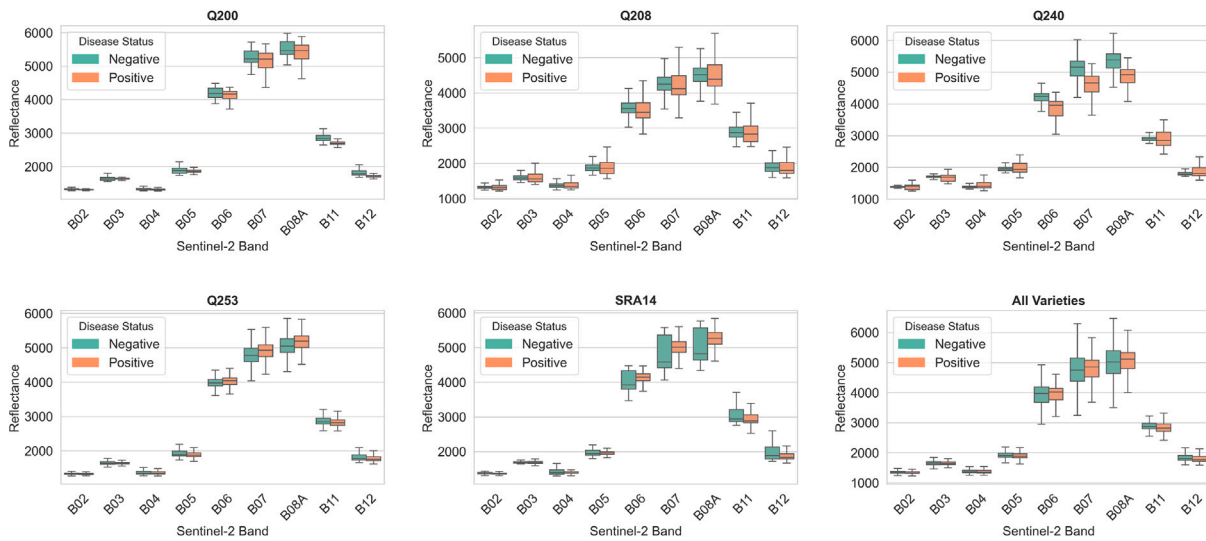
Ground truth pixel observations at a 20 m resolution for RSD detection.

Variety	RSD positive	RSD negative	Total
Q200	145	389	534
Q208	869	649	1518
Q240	766	573	1339
Q253	886	1769	2655
SRA14	88	89	177
	2754	3469	6223

The multispectral data used in the study were captured by Sentinel-2 on 27th February 2022, as it had the least cloud coverage during the sampling period. Since RSD is transmitted through contaminated tools during planting or harvesting, and the study period was during the plant growth period, we assumed that the infection status remained unchanged during the study period. Only one day of image data was used in this study.

The Sentinel-2 data processing pipeline used the QGIS Python API to transform the coordinate reference system (CRS) of the Sentinel-2 products to match the CRS of the ground-truth shapefile representing the farms of interest. The ground-truth data was then used to extract the 76 sampled blocks from the Sentinel-2 imagery, ensuring that each pixel within the block geometries was labelled with both disease status and variety. Fig. 2 shows an example block across Sentinel-2 bands, highlighting within-field variation captured in the multispectral data. Each labelled pixel within the shapefile geometries was subsequently utilised as a separate observation. Table 2 summarises the number of ground-truth pixel observations at 20 m resolution for each sugarcane variety.

The distributions of reflectance values across Sentinel-2 bands vary with both disease status and variety. Fig. 3 illustrates the spectral response of each variety. For example, in variety Q240, RSD-positive pixels generally exhibit lower reflectance in the red-edge and near-infrared



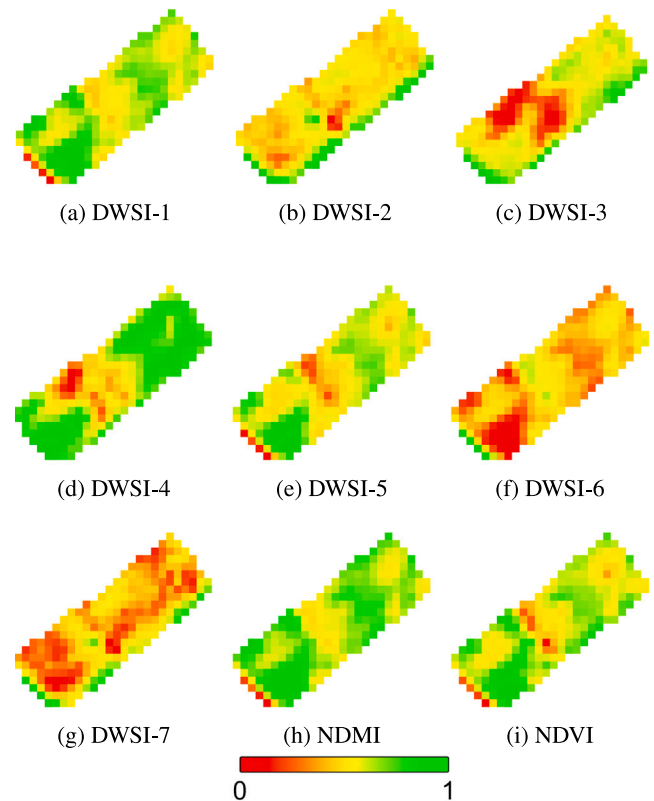
**Fig. 3.** Reflectance distributions across Sentinel-2 bands for five sugarcane varieties and an overall panel. Outliers are omitted for clarity. Healthy and diseased crops are colour-coded (green = healthy, orange = diseased).

regions (bands 6–8) compared with RSD-negative pixels, whereas varieties such as SRA14 and Q253 show higher median reflectance for RSD-positive pixels.

#### 2.4. Vegetation indices

The causal bacteria of RSD infects the xylem vessels responsible for water transport in sugarcane, reducing water absorption and resulting in vegetation stress [1,2]. To detect these internal symptoms, vegetation indices were selected that measure key indicators of plant health, as shown in Table 3. Several indices were chosen to assess vegetation moisture content [21,31], with the intention highlighting reduced water uptake caused by RSD infection of the xylem. Others focus on assessing symptoms of plant stress [26,32,33], indicated by changes in photosynthetic activity or chlorophyll levels. In addition to vegetation indices, previous research has demonstrated that the raw Sentinel-2's bands in the Short-wave Infrared Region (SWIR) are sensitive to moisture absorption in vegetation [34–36], while bands in the Near-Infrared (NIR) region have been previously used as indicators of vegetation stress [37], making both potential predictors for RSD. Together, the indices and bands provide a comprehensive net to capture the expected physiological impacts of RSD on sugarcane.

To optimise class separation, strategic modifications of vegetation indices were explored. The alteration of existing indices was guided by first identifying statistically significant bands for diagnosing RSD with an independent t-test (significance level = 0.05) for each variety and spectral band, with Bonferroni's correction applied to adjust for type I error. Although statistically significant bands varied between varieties, bands B08 A and B12 were consistent across most varieties, supporting the connection between vegetation moisture and stress as key indicators for diagnosing RSD. With the identification of statistically significant bands, opportunities arise to modify key vegetation indices to enhance class separation. Among the most relevant indices are the Disease Water Stress Indices (DWSIs), which have proven effective in detecting sugarcane disease through satellite-based remote sensing [26]. The DWSIs were designed to target water absorption at R1660 (Sentinel-2's B11). Incorporating both B11 and B12, given their vegetation moisture absorption sensitivity [34–36] and B12's statistical significance, led to modified versions of these indices for Sentinel-2, DWSI 6 and 7. Additionally, DWSI 7 was further modified to include the red-edge band, B5, as the red-edge region is sensitive to stress-related changes in vegetation and has previously been effective for vegetation classification [38,39]. This modification resulted in the creation of DWSI 8,



**Fig. 4.** Nine vegetation indices for a sample block that are historically significant for vegetation monitoring or were found to be important for detecting RSD. Pixel values have been normalised to highlight within-field variation.

designed to better capture stress-related signals in vegetation as a result of RSD.

The spectral bands of the satellite image captured on the 27th February 2022 were utilised to compute the 19 vegetation indices listed in Table 3 with the QGIS Python API for each pixel in the dataset. A visualisation of several key vegetation indices for an example block is shown in Fig. 4, illustrating patterns relevant to sugarcane disease detection.

**Table 3**  
Vegetation indices utilised in this study for RSD detection.

Reference	Vegetation index	Formula
Rouse et al. [33]	Normalised Difference Vegetation Index (NDVI)	$\frac{NIR-RED}{NIR+RED}$
Kaufman and Tanre [40]	Atmospherically Resistant Vegetation Index (ARVI)	$\frac{NIR-(RED-BLUE)}{NIR+(RED-BLUE)}$
Genc et al. [41]	Simple ratio index (SRI)	$\frac{NIR}{RED}$
Merzlyak et al. [32]	Plant Senescence Reflectance index (PSRI)	$\frac{R_{800}-R_{500}}{R_{750}}$
[42]	Ratio Vegetation Index (RVI)	$\frac{RED}{NIR}$
McFeeters [43]	Normalised difference Water index (NDWI)	$\frac{GREEN-NIR}{GREEN+NIR}$
Gao [31]	Normalised Difference Moisture Index (NDMI)	$\frac{NIR-SWIR}{NIR+SWIR}$
Tucker [44]	Normalised Green Red Difference Index (NGRDI)	$\frac{GREEN-RED}{GREEN+RED}$
Gitelson et al. [45]	Visible Atmospherically Resistant Index (VARI)	$\frac{GREEN-RED}{GREEN+RED-BLUE}$
Datt [46]	Simple Ratio 860/550	$\frac{R_{860}}{R_{550}}$
Apan et al. [21,26]	Disease-Water Stress Index 1 (DWSI-1)	$\frac{R_{800}}{R_{1660}}$
Apan et al. [21,26]	Disease-Water Stress Index 2 (DWSI-2)	$\frac{R_{1660}}{R_{550}}$
Apan et al. [21,26]	Disease-Water Stress Index 3 (DWSI-3)	$\frac{R_{1660}}{R_{680}}$
Apan et al. [21,26]	Disease-Water Stress Index 4 (DWSI-4)	$\frac{R_{550}}{R_{680}}$
Apan et al. [21,26]	Disease-Water Stress Index 5 (DWSI-5)	$\frac{R_{800}+R_{550}}{R_{1660}+R_{680}}$
Rouse et al. [33]	Green Blue NDVI (GBNDVI)	$\frac{NIR-(Blue+Green)}{NIR+Blue+Green}$
This Study	Disease-Water Stress Index 6 (DWSI-6)	$\frac{B_{12}+B_{11}}{B_{8A}}$
This Study	Disease-Water Stress Index 7 (DWSI-7)	$\frac{B_{12}+B_{11}}{B_4}$
This Study	Disease-Water Stress Index 8 (DWSI-8)	$\frac{B_{12}+B_{11}}{B_4+B_5}$

**Notes:** NIR = Near Infrared Band; RED = Red Band; GREEN = Green Band; BLUE = Blue Band; SWIR = Shortwave Infrared Band.  $R_x$  refers to reflectance at wavelength x nm. Sentinel-2 spectral bands: B2 = Blue (490 nm), B3 = Green (560 nm), B4 = Red (665 nm), B5 = Red Edge 1 (705 nm), B8 = NIR (842 nm), B8A = Narrow NIR (865 nm), B11 = SWIR 1 (1610 nm), B12 = SWIR 2 (2190 nm).

## 2.5. Machine learning algorithm development

In this study, we compared five machine learning algorithms to predict the RSD status of a pixel. These are Random Forest (RF), Logistic Regression (LR), Quadratic Discriminant Analysis (QDA), Gradient Boosting (GB) and Radial Basis Function Support Vector Machine (SVM-RBF). Tree-based methods, such as RF and GB, were included due to their previous success in similar remote sensing studies. These models are particularly effective for handling high-dimensional data, non-linearity, and interactions among features, making them well-suited for RSD detection. SVM-RBF kernel was included due to its ability to effectively model complex non-linear relationships, which are often present in multispectral data. LR was included due to its simplicity and interpretability. LR is useful for assessing whether the problem could be adequately solved with linear methods without resorting to more complex models. QDA was tested as it is capable of modelling quadratic decision boundaries, which can be useful for capturing more complex class separations that may exist in the data.

Given that the sugarcane variety can be potentially unknown to the users, two sets of models were developed, one for when the sugarcane variety is known and the other when its unknown. The final dataset seen in Table 2 is largely unbalance, therefore, the dataset was down-sampled to achieve equal class distribution, to mitigate potential bias.

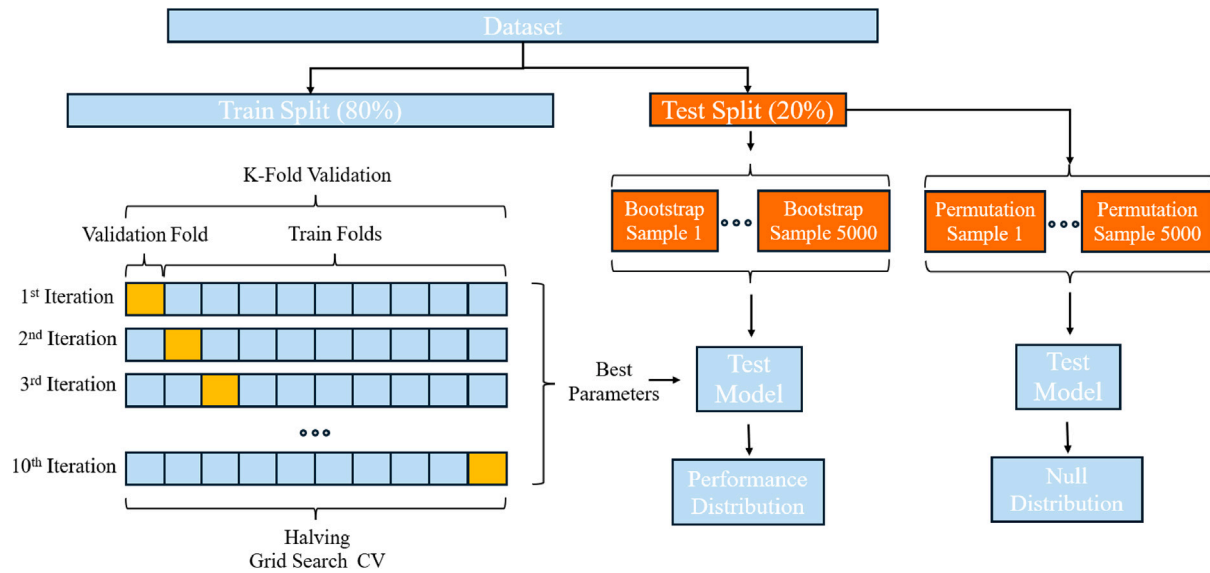
The dataset was partitioned into 80% for training and 20% for testing. Standard scaling was applied to the entire dataset, with the scaling parameters fitted exclusively on the training set. Hyperparameters were tuned using 10-fold cross-validation with ‘‘Halving Grid Search’’ on the 80% training set. HalvingGridSearchCV is a resource-efficient version of GridSearchCV from the scikit-learn library for hyperparameter tuning. HalvingGridSearchCV is an implementation of the successive halving algorithm, introduced by Jamieson and Talwalkar [47], which starts with many candidate parameter combinations and progressively eliminates the least promising ones, allocating more resources as the

number of iterations increases [48]. Hyperparameter tuning was performed with the default arguments of HalvingGridSearchCV, meaning the elimination factor (factor = 3), resources (resource = ‘n\_samples’), and maximum resources (max\_resources = ‘auto’) were used as defined in the scikit-learn documentation. Table 4 provides a summary of the search space explored.

The model’s performance was evaluated by bootstrapping the unseen test set with 5000 bootstrap samples, generating distributions for the chosen performance metrics, accuracy, precision, and recall [49]. Accuracy provides an overall measure of correct predictions, while precision and recall offer insights into the model’s ability to correctly identify positive cases and avoid false positives, respectively. In the context of large-scale disease detection, evaluating false positives and false negatives is critical to ensuring an appropriate balance is achieved. High false positive rates could lead to unnecessary interventions by agronomists, wasting time, diverting resources, and reducing the economic benefit of the system. After model evaluation, each machine learning algorithm was trained with 10-fold cross-validation on the entire dataset, to find the most appropriate hyperparameters for the entire dataset, after the performance has been determined. A visualisation of this machine learning workflow is shown in Fig. 5.

The Gini importance of the final RF and GB models was extracted to identify the most important vegetation indices and Sentinel-2 multi-spectral bands for RSD detection. Additionally, to indicate important features for RSD detection, independent t-tests (significance level = 0.05) were performed for each vegetation index, applying Bonferroni’s correction to adjust for type I error. Although the results are included in Fig. 7 and Tables 5–7, this approach does not account for correlations among the indices or interactions between them. Therefore, the findings should be interpreted with caution.

To assess the significance of the ML algorithms’ ability to detect RSD, a permutation test was conducted. The pixel labels were shuffled, and the models were retrained using the previously determined hyperparameters. This process was repeated 1000 times to generate a null distribution of accuracy, reflecting model performance when there



**Fig. 5.** Visual guide to the machine learning workflow. The dataset was split into training and testing sets. The training set underwent 10-fold cross-validation to select the best model. Model performance was then assessed by bootstrapping the test set 5000 times, followed by permutation testing.

**Table 4**  
Summary of hyperparameters search space utilised in halving grid search.

Model	Parameter	Search space
SVM	C	0.1, 1, 10, 100, 1000, 10000, 100000, 1000000
	gamma	0.001, 0.01, 0.1, 1, 10, 100
RF	n_estimators max_depth	50, 100, 200, 500, 1000, 1500 None, 10, 15, 20, 30
GB	n_estimators learning_rate max_depth	50, 100, 200, 500, 1000, 1500 0.01, 0.1, 0.2, 0.3 3, 5, 7, 10
LR	penalty solver gamma	'l1', 'l2' 'liblinear', 'saga' 0.001, 0.01, 0.1, 1, 10, 100
QDA	N/A	N/A

is no true association between the features and disease status. By comparing the original performance distribution to the null distribution, the statistical significance of the observed model performance can be evaluated.

## 2.6. Software, libraries & experimental environment

This study was conducted with several software tools and libraries for data processing, analysis, and model development. QGIS 3.36.0 was employed for examining satellite imagery and utilising its Python API to calculate vegetation indices, perform spatial joins, and extract tabular data from the spatial layers. Visual Studio Code version 1.89.1 was utilised as the programming IDE. Python 3.9.18 and a list of its libraries were utilised, including the following: matplotlib 3.7.2, seaborn 0.12.2, numpy 1.24.3, pandas 2.0.3, scikit-learn 1.3.0 and scipy 1.11.1.

All experiments were conducted on a system running Microsoft Windows 11 Home. The hardware configuration included an Intel(R) Core (TM) i7-8700 CPU operating at 3.20 GHz with 6 cores and 12 logical processors, supported by 16 GB of installed physical memory.

## 3. Results

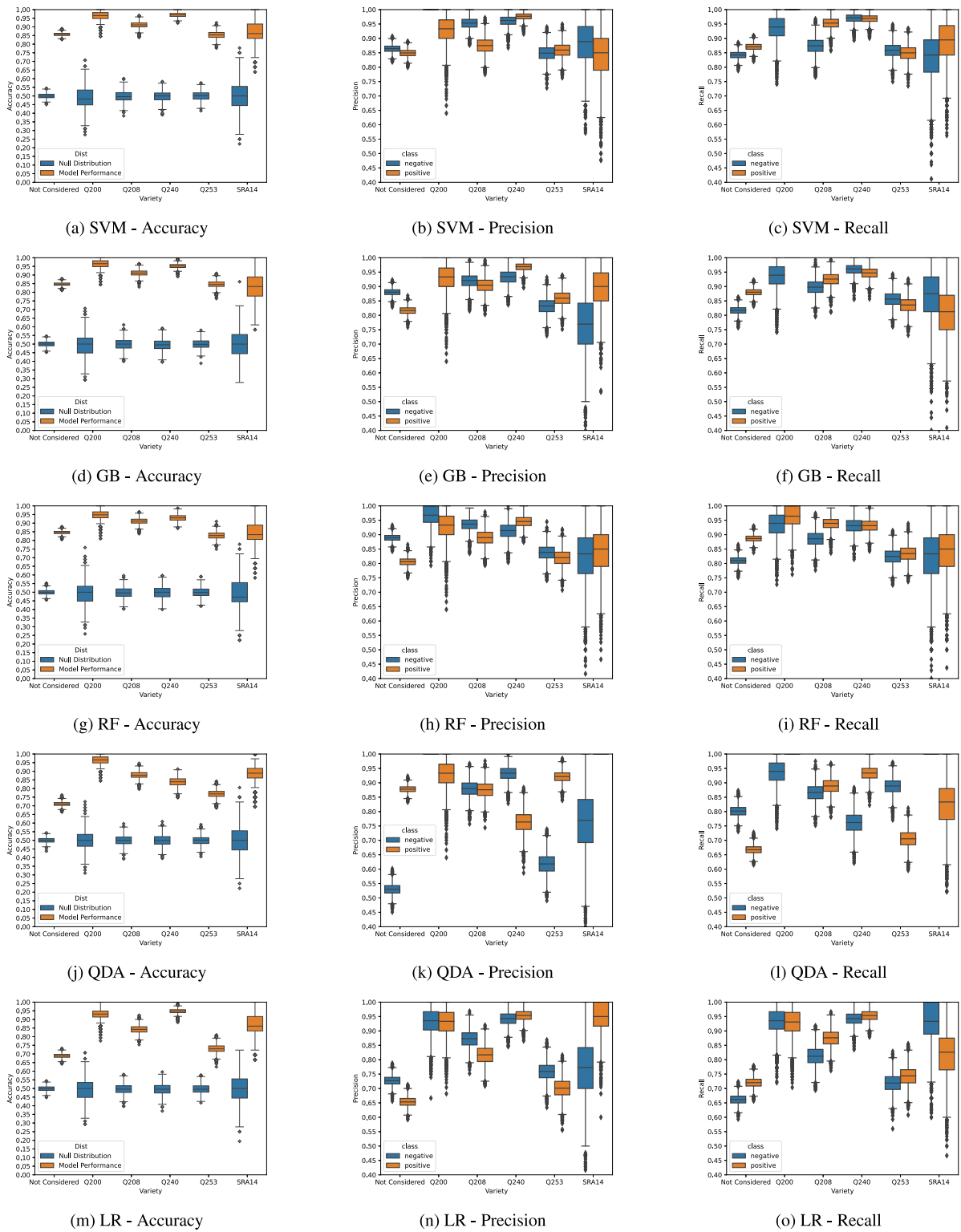
The classification and permutation testing results (Fig. 6) reveal a clear separation between the null distribution and the model accuracy distribution for all machine learning algorithms and varieties, with the

exception of SRA14. This demonstrates that the models significantly outperformed random chance. In contrast, the overlap of distributions for SRA14 across all models suggests that its performance was not substantially different from what would be expected by chance, indicating a weaker predictive power compared to models for the other varieties.

Three machine learning algorithms demonstrated over 80% accuracy across all varieties for RSD detection. SVM-RBF performed the best overall achieving 85.64% and 96.55% accuracy depending on the variety (Fig. 6). SVM-RBF achieved the highest classification accuracy when variety was not considered (85.64%) and for several specific varieties including Q208 (91.15%), Q240 (96.96%), and Q253 (85.35%). Similarly, GB and RF demonstrated consistently high performance across all varieties achieving 83.33% to 94.83% and 83.33% to 96.55% accuracy respectively, depending on the variety. All three models performed particularly well for Q200, Q208 and Q240 achieving between 91.15% and 96.55% accuracy depending on the specific variety and model. GB and RF tied for the highest median overall accuracy on Q208, though both models demonstrated less balanced performance in terms of precision and recall when compared to SVM-RBF. Additionally, GB matched the best median overall accuracy for Q200. A complete table of results for all models and varieties can be seen in Table 5.

In contrast, LR and QDA showed varying performance depending on the specific variety. These two models performed particularly poorly for variety Q253 and underperformed compared to the other models on Q240. However, they excelled with other varieties. QDA, in particular, emerged as the top-performing model for variety SRA14, achieving a median overall accuracy of 88.89%. Additionally, for Q200, QDA, SVM-RBF, and GB all tied as the best-performing models, each yielding identical classification reports with a median overall accuracy of 96.55%.

Variety-specific models for all machine learning algorithms performed as well as, or better than, the variety-agnostic models. Models specific to Q200, Q208 and Q240 outperformed the variety-agnostic model, whereas those specific to Q253 and SRA14 showed no significant improvement. RSD in varieties Q200 and Q240 was consistently classified with higher accuracy compared to the other varieties. Conversely, Q253 consistently showed lower classification accuracy, while SRA14 exhibited significant variability in performance. Notably, a closer analysis of positive class precision revealed that, in 4 out of 5 algorithms, varieties Q200 and Q208, or the variety-agnostic model, yielded lower precision for the positive class.



**Fig. 6.** Bootstrapping distribution of accuracy, precision, and recall for different ML algorithms diagnosing RSD in several sugarcane varieties. The accuracy distribution is compared with a null distribution generated with permutations to indicate significance of model performance for that specific variety.

The hyperparameters and overall accuracies of the final models trained on the entire dataset are reported in Table 6 now that an appropriate estimate of performance has been determined. The outcomes largely align with the previously observed accuracies, which is expected if the test set utilised was representative of the sample distribution. In this final evaluation, there were no ties for best performance across varieties. SVM-RBF maintained its position as the top-performing model

for most varieties and QDA continued to outperform other models for SRA14.

The feature importance for diagnosing RSD revealed some variation between the GB and RF classifiers, however, both generally agreed upon several key findings. Notably, when sugarcane variety was not considered, both classifiers identified DWSI-6 as the most influential feature, and identified DWSI-7, DWSI-2, and spectral bands 2, 5, 6,

**Table 5**  
Median classification metrics of bootstrap distributions for ML algorithms.

Model	Variety	Class	Precision	Recall	Accuracy
SVM: RBF	Q200	Positive	93.33%	100%	96.55%
		Negative	100%	93.39%	
	Q208	Positive	87.5%	95.31%	91.15%
		Negative	95.31%	87.41%	
	Q240	Positive	97.69%	96.95%	96.96%
		Negative	96.22%	97.14%	
	Q253	Positive	85.95%	84.97%	85.35%
Negative		84.86%	85.88%		
SRA14	Positive	85.00%	89.47%	86.11%	
	Negative	88.88%	84.21%		
Not Considered	Positive	84.94%	87.06%	85.64%	
	Negative	86.45%	84.24%		
GB	Q200	Positive	93.33%	100.00%	96.55%
		Negative	100.00%	93.39%	
	Q208	Positive	90.48%	92.59%	91.15%
		Negative	92.06%	89.78%	
	Q240	Positive	96.90%	94.74%	95.22%
		Negative	93.39%	96.12%	
	Q253	Positive	85.96%	83.59%	84.51%
Negative		83.23%	85.63%		
SRA14	Positive	90.00%	81.25%	83.33%	
	Negative	76.92%	87.5%		
Not Considered	Positive	81.67%	88.03%	84.73%	
	Negative	88.05%	81.71%		
RF	Q200	Positive	93.33%	96.43%	94.83%
		Negative	96.77%	93.94%	
	Q208	Positive	88.98%	93.94%	91.15%
		Negative	93.66%	88.57%	
	Q240	Positive	94.57%	93.13%	93.04%
		Negative	91.43%	93.13%	
	Q253	Positive	82.01%	83.43%	82.82%
Negative		83.82%	82.42%		
SRA14	Positive	85.00%	85.00%	83.33%	
	Negative	83.33%	83.33%		
Not Considered	Positive	80.58%	88.67%	84.59%	
	Negative	88.92%	80.99%		
QDA	Q200	Positive	93.33%	100.00%	96.55%
		Negative	100.00%	93.39%	
	Q208	Positive	87.59%	88.89%	87.69%
		Negative	88.00%	86.61%	
	Q240	Positive	76.38%	93.39%	83.91%
		Negative	93.33%	76.19%	
	Q253	Positive	92.18%	70.54%	76.90%
Negative		61.80%	88.89%		
SRA14	Positive	100.00%	83.33%	88.89%	
	Negative	76.92%	100.00%		
Not Considered	Positive	87.79%	66.74%	71.01%	
	Negative	52.99%	80.17%		
LR	Q200	Positive	93.33%	93.10%	93.10%
		Negative	93.55%	93.55%	
	Q208	Positive	81.69%	87.60%	84.23%
		Negative	87.29%	81.20%	
	Q240	Positive	95.39%	95.35%	94.78%
		Negative	94.29%	94.29%	
	Q253	Positive	70.12%	74.37%	72.96%
Negative		75.90%	71.88%		
SRA14	Positive	95.00%	82.61%	86.11%	
	Negative	77.27%	93.33%		
Not Considered	Positive	65.33%	72.03%	68.90%	
	Negative	72.73%	66.13%		

**Table 6**

Optimal hyper-parameters and mean classification accuracy for different machine learning algorithms and varieties determined with 10-fold cross-validation across the entire dataset. 'lr' = 'learning rate', 'md' = 'max depth'.

Model	Variety	Optimal hyper-parameters	Accuracy
SVM	Q200	'C': 1000, 'gamma': 0.001	99.17%
	Q208	'C': 1000, 'gamma': 0.01	90.14%
	Q240	'C': 10, 'gamma': 0.1	96.72%
	Q253	'C': 10, 'gamma': 1	85.39%
	SRA14	'C': 1000, 'gamma': 0.01	89.02%
	Not Considered	'C': 10, 'gamma': 0.1	84.59%
GB	Q200	'lr': 0.1, 'md': 3, 'estimators': 200	95.00%
	Q208	'lr': 0.2, 'md': 3, 'estimators': 500	88.68%
	Q240	'lr': 0.3, 'md': 3, 'estimators': 1000	95.17%
	Q253	'lr': 0.1, 'md': 5, 'estimators': 500	82.46%
	SRA14	'lr': 0.3, 'md': 3, 'estimators': 500	87.04%
	Not Considered	'lr': 0.1, 'md': 5, 'estimators': 1500	84.17%
RF	Q200	'md': None, 'estimators': 200	95.83%
	Q208	'md': 30, 'estimators': 1000	89.75%
	Q240	'md': 30, 'estimators': 200	94.81%
	Q253	'md': 20, 'estimators': 1000	85.19%
	SRA14	'md': None, 'estimators': 200	88.79%
	Not Considered	'md': 20, 'estimators': 200	84.26%
QDA	Q200	N/A	97.24%
	Q208	N/A	89.13%
	Q240	N/A	87.09%
	Q253	N/A	76.58%
	SRA14	N/A	92.71%
	Not Considered	N/A	69.46%
LR	Q200	'C': 100, 'penalty': 'l1', 'solver': 'liblinear'	93.33%
	Q208	'C': 100, 'penalty': 'l1', 'solver': 'liblinear'	84.31%
	Q240	'C': 100, 'penalty': 'l1', 'solver': 'liblinear'	95.24%
	Q253	'C': 100, 'penalty': 'l1', 'solver': 'liblinear'	74.34%
	SRA14	'C': 100, 'penalty': 'l1', 'solver': 'liblinear'	82.19%
	Not Considered	'C': 100, 'penalty': 'l1', 'solver': 'liblinear'	68.73%

and 11 among the top 10 most important features. Analysing feature importance for RSD diagnosis with separate models for each sugarcane variety indicated variations between both varieties and classifiers. No single feature appeared in the top 10 importance rankings across all varieties for both classifiers. However, across a majority of instances DWSI-6, DWSI-3, Band 2, 5, 6, 7 and 11 were the most frequently occurring features in the top 10 importance. A complete description of feature importance can be seen in Fig. 7. The t-test to identify statistically significant vegetation indices generally supports similar findings. A summary of these statistical significances is available in Table 7.

#### 4. Discussion

This study demonstrates that ML algorithms can effectively classify RSD across several varieties with freely available satellite-based multispectral data. Consistent with previous studies, RF, SVM, and GB variants demonstrated strong performance [16,22,23]. Notably, the SVM with RBF kernel slightly outperformed both RF and GB in this context. This may be attributed to the SVM's ability to effectively model complex non-linear decision boundaries in high-dimensional settings with limited training samples. Under these conditions tree-based methods may be less effective at capturing complex relationships [50]. Additionally, the underlying structure of the disease-related spectral data may be more effectively separated by a smooth, non-linear decision boundary, potentially contributing to the stronger performance observed with SVM using a RBF kernel. While previous sugarcane disease detection studies have primarily focused on drone-based spectroscopy [20,22,23], handheld spectroscopy [15,16] or hyperspectral satellite data [21,26], our findings extend this knowledge by demonstrating comparable, if not superior, effectiveness of classifying a sugarcane disease with multispectral satellite data. For example, Simões and

**Table 7**

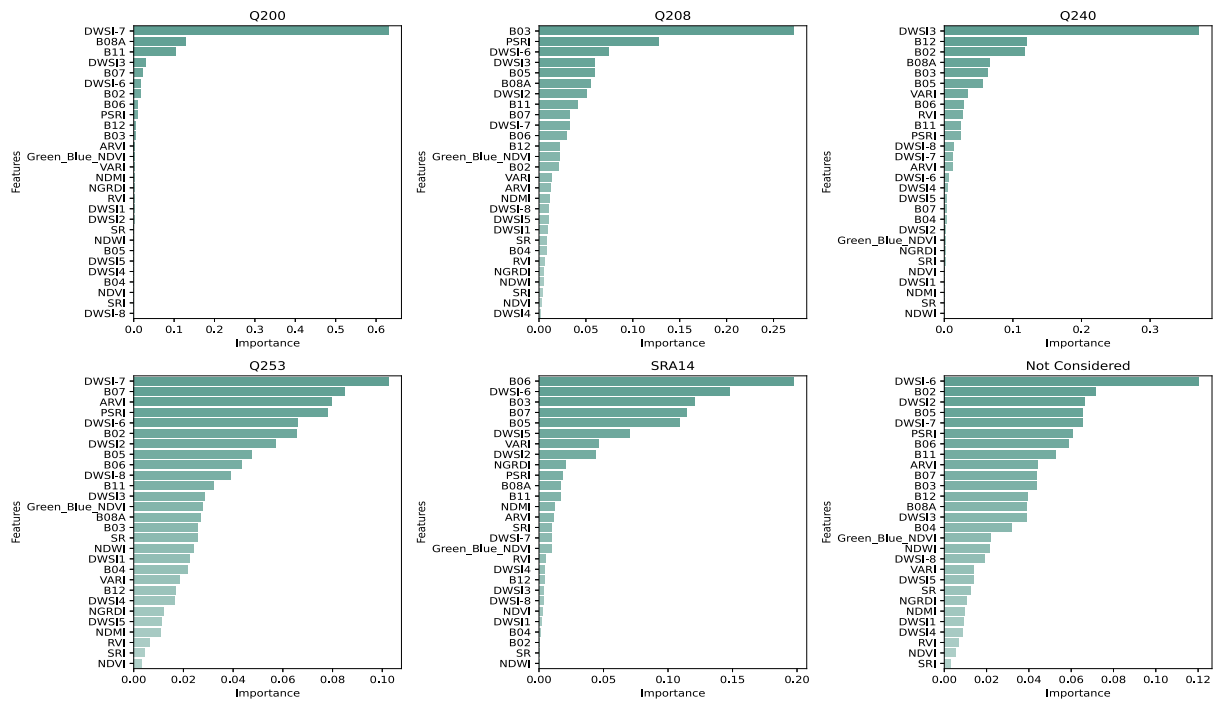
Summary of statistical significance from t-tests comparing healthy and diseased samples across different sugarcane varieties. ': adjusted p-value > 0.05, '\*\*': 0.05 >= adjusted p-value > 0.01, '\*\*\*': adjusted p-value <= 0.01.

Vegetation index	Q200	Q208	Q240	Q253	SRA14
ARVI	.	.	**	**	.
DWSI1	.	*	**	**	.
DWSI2	**	.	.	**	.
DWSI3	**	**	**	.	**
DWSI4	.	.	**	**	.
DWSI5	.	.	**	**	.
NDMI	.	.	**	**	.
NDVI	.	.	**	**	.
NDWI	.	.	**	**	.
NGRDI	.	.	**	**	.
PSRI	.	**	**	.	.
RVI	.	.	**	**	.
SRI	.	.	**	**	.
SR	.	.	**	**	.
VARI	.	.	**	**	.
Green Blue NDVI	.	.	**	**	.
DWSI 6	**	**	**	.	*
DWSI 7	**	**	**	.	.
DWSI 8	.	.	**	**	.

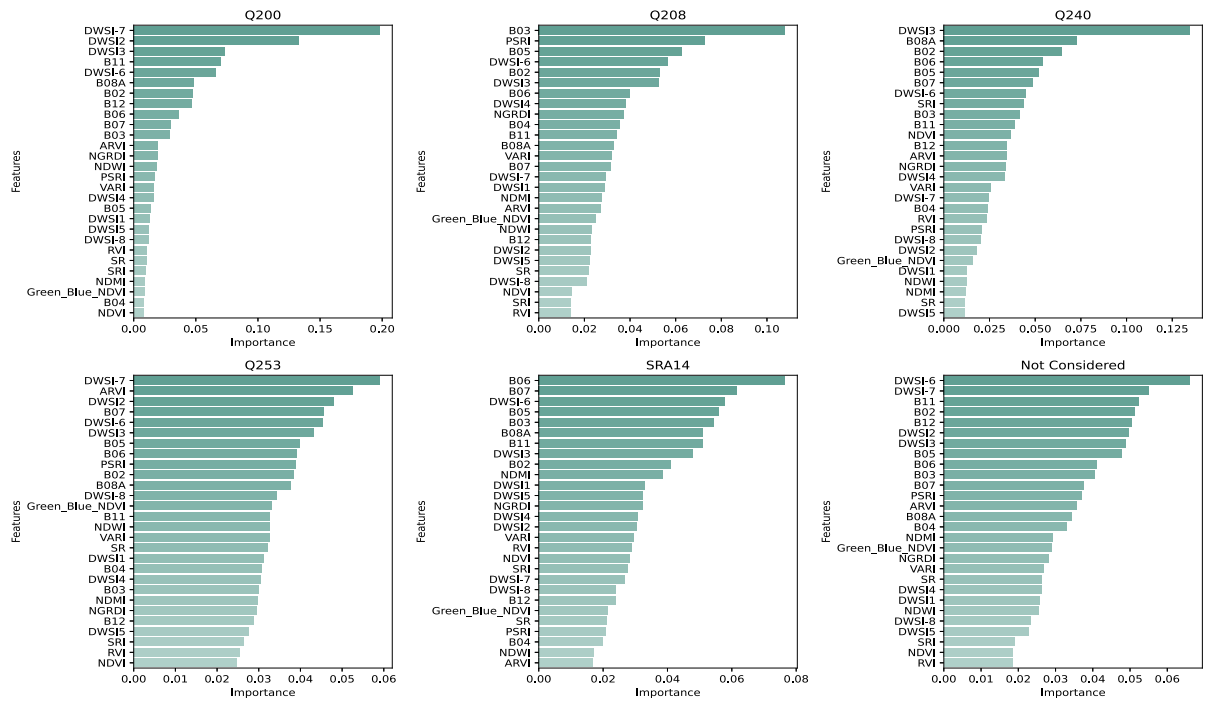
Rios do Amaral [23] reported classification accuracies between 86% and 90% for orange and brown rust with multispectral drone data. In comparison, our study achieved accuracies ranging from 83% and 97% accuracy for RF and SVM, depending on the variety, with freely available multispectral data. This demonstrates that our approach, utilising cost-effective large-scale satellite data, can achieve comparable or superior performance. Moreover, our study outperformed the only previous study detecting asymptomatic sugarcane disease [15]. This represents a significant advancement in early-stage disease detection and large-scale monitoring, where timely intervention is crucial for mitigating potential crop yield loss.

Vegetation Indices (VI) and spectral bands sensitive to vegetation moisture and stress, were consistently the most influential features, reinforcing the established understanding that RSD diminishes water uptake and heightens plant stress [1,3–5]. Consistent importance of the features DWSI-7, DWSI-6, DWSI-3, Band 5, 6 and 7, clearly reflects this. The trend was particularly pronounced in the variety-agnostic models, potentially due to the variation introduced by the different sugarcane varieties, forcing the model to focus on the most generalised disease signals common across all varieties. In contrast, feature importance varied among variety-specific models, reflecting the distinct biophysical characteristics of each variety and their unique responses to disease presence. The importance of DWSI-7 and DWSI-6 across the different models highlights that adapting existing DWSI to include the other statistically significant moisture absorption band contributed to the success of these models. Interestingly, Band 2 which is centred in the blue part of the visible spectrum was considered to be consistently important, despite RSD being considered as an asymptomatic disease universally by experts. While water and stress-related indices proved valuable, popular general health indicators like the Normalised Difference Vegetation Index (NDVI) performed poorly, potentially due to the asymptomatic nature of RSD.

Across all machine learning algorithms, variety-specific models were as or more effective than variety-agnostic ones. Performance of RSD detection varied between sugarcane varieties potentially arising from the physical differences between the varieties or external factors correlated with variety selection. Variables such as soil type, rainfall, and land topology may influence variety choice and introduce variability in the observed spectral reflectance. Consequently, future studies should incorporate soil data, rainfall, and land topology as predictors to assess their impact on RSD diagnosis and determine whether these factors explain variation that affects class separability. Notably, variety



(a) GB classifier



(b) RF classifier

Fig. 7. Feature importance of different sugarcane varieties.

SRA14 typically exhibited lower classification rates and larger variation, compared to other varieties with model performance that was not significantly greater than random chance. This may be attributed to a smaller sample size than the other varieties (see Table 2) and perhaps future studies with a larger dataset will find a significant performance difference.

Despite the high classification accuracies achieved by the models, there were several limitations of the study. Positive class precision was consistently lower than negative class precision for models classifying

RSD in Q200 and Q208, indicating a higher rate of false positives across the algorithms. This may have arisen from the assumption that RSD infection in a block was uniformly distributed by labelling all pixels within a block with the overall disease status of that block. This could result in regions with little or no RSD being mislabelled as ‘RSD Positive’, thereby affecting model training. Future studies should avoid this assumption where possible and implement higher-resolution sampling methods to evaluate local infection patterns more accurately.

Nevertheless, the 20 m satellite spatial resolution is sufficient for identifying infected blocks and supporting management decisions. Further increases in resolution are unlikely to meaningfully influence decision-making processes for practical RSD management. However, data fusion approaches combining satellite and drone imagery to improve spatial resolution may still offer value for other sugarcane diseases or cropping systems. Interestingly, this observation coincides with the only two varieties identified by Sugar Research Australia as partially resistant to RSD [7,51]. Partial resistance might reduce the spread of RSD sufficiently to create patches of non-infected areas within otherwise infected blocks. However, to verify this hypothesis, further research with higher resolution sampling methods would be necessary.

This study focused on developing models and a prototype system for RSD detection within the Herbert region in February 2022. Future work should assess the influence of other environmental and agronomic predictors that could not be accounted for in this study to ensure broader applicability in the industry. This includes exploring the impacts of factors such as temperature, humidity, sunlight duration, flowering, other sugarcane varieties, simultaneous disease infections, and precipitation on the performance of classification models. Furthermore, RSD detection should be trialled at different growth stages of sugarcane using time-series data to improve its applicability to the industry. Given their proven success in previous time-series classification tasks, exploring the use of techniques such as recurrent neural networks or transformers for this application represents a promising direction for future research. Additionally, comparing various principal component analysis algorithms could further optimise vegetation index characteristics while addressing and mitigating issues such as overfitting.

## 5. Conclusion

In this study, we demonstrated that RSD can be classified for several varieties with ML algorithms from freely available satellite-based spectroscopy data. The best-performing machine learning algorithm was Support Vector Machine with a Radial Basis Kernel achieving between 85.64% and 96.55% accuracy depending on the variety. The performance of all machine learning algorithms was found to be significantly better than the null distribution for all varieties except SRA14. In line with current literature, the inclusion of sugarcane variety information and vegetation indices improved classification rates of disease. Additionally, several varieties were found to be more challenging to classify for RSD compared to others. These promising initial results, coupled with the efficiency of classifying 76 blocks within minutes as opposed to months, indicates the potential benefits of implementing a large-scale health monitoring system using satellites and machine learning. Given that RSD is one of the most impactful diseases in sugarcane cultivation, this research significantly advances sugarcane disease management, with substantial practical applications. Providing farmers with a more accurate understanding of where RSD is present enables them to alter harvester paths to minimise disease spread, and source alternative clean, disease-free seeds from other blocks, or reliable suppliers, to ensure healthier crop cycles. Additionally, this will allow the implementation of strict hygiene protocols on infected blocks to prevent the disease from contaminating others. These applications indicate the potential of this research to support sustainable and efficient sugarcane cultivation.

## CRedit authorship contribution statement

**Ethan Kane Waters:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Carla Chia-Ming Chen:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Mostafa Rahimi Azghadi:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

## Ethics statement

Not applicable: This manuscript does not include human or animal research.

## Declaration of generative AI in scientific writing

The manuscript was not authored with the assistance of artificial intelligence (AI).

## Funding

This work was supported by Australia's Economic Accelerator Seed grant provided by the Australian Government, Department of Education.

## 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.

## Acknowledgements

We extend our sincere appreciation to HCPSL for their invaluable contribution in collecting and providing the ground-truth data essential for this study. Special thanks to Lawrence Di Bella, Rod Nielson, Adam Royle and Rhiannan Harragon from HCPSL for their industry knowledge and support in the ground truth data collection process.

## References

- [1] Chakraborty M, Soda N, Strachan S, Ngo CN, Bhuiyan SA, Shiddiky MJA, Ford R. Ratoon stunting disease of sugarcane: A review emphasizing detection strategies and challenges. *Phytopathology*® 2024;114(1):7–20. <http://dx.doi.org/10.1094/PHYTO-05-23-0181-RVW>, PMID: 37530477.
- [2] Davis MJ, Gillaspie AG, Vidaver AK, Harris RW. Clavibacter: a new genus containing some phytopathogenic coryneform bacteria, including clavibacter xyli subsp. xyli sp. nov., subsp. nov. and clavibacter xyli subsp. cynodontis subsp. nov., pathogens that cause ratoon stunting disease of sugarcane and bermudagrass stunting disease†. *Int J Syst Evol Microbiol* 1984;34(2):107–17. <http://dx.doi.org/10.1099/00207713-34-2-107>.
- [3] Bailey R, Bechet G. Effect of ratoon stunting disease on the yield and components of yield of sugarcane under rainfed conditions. *Proc South Afr Sugar Technol Assoc* 1986;60:204–10.
- [4] Davis MJ, Bailey RA. A guide to sugarcane diseases. CIRAD and ISSCT; 2000, p. 49–55.
- [5] Croft B, Magarey R, Whittle P. Manual of canegrowing. BSES; 2000, p. 71–88.
- [6] Magarey R, McHardie R, Hession M, Cripps G, Burgess D, Spannagle B, Sutherland P, Di Bella L, Milla R, Millar F, Schembri A, Baxter D, Hetherington M, Turner M, Jakins A, Quinn B, Kalkhoran S, Gibbs L, Ngo C. Incidence and economic effects of ratoon stunting disease on the queensland sugarcane industry : ASSCT peer-reviewed paper. *Proc Aust Soc Sugar Cane Technol* 2021;42:520–6.
- [7] Sugar Research Australia. Ratoon stunting disease. Sugar Research Australia; 2021, URL: [https://sugarsearch.com.au/sugar\\_files/2017/03/RSD-Info-Sheet\\_2021\\_May-2021.pdf](https://sugarsearch.com.au/sugar_files/2017/03/RSD-Info-Sheet_2021_May-2021.pdf).
- [8] Carvalho G, da Silva T, Munhoz A, Monteiro-Vitorello C, Azevedo R, Melotto M, Camargo L. Development of a qPCR for leifsonia xyli subsp. xyli and quantification of the effects of heat treatment of sugarcane cuttings on Lxx. *Crop Prot* 2016;80:51–5. <http://dx.doi.org/10.1016/j.cropro.2015.10.029>.
- [9] Fegan M, Croft BJ, Teakle DS, Hayward AC, Smith GR. Sensitive and specific detection of clavibacter xyli subsp. xyli, causal agent of ratoon stunting disease of sugarcane, with a polymerase chain reaction-based assay. *Plant Pathol* 1998;47(4):495–504. <http://dx.doi.org/10.1046/j.1365-3059.1998.00255.x>.
- [10] Ghai M, Singh V, Martin L, McFarlane S, van Antwerpen T, Rutherford R. A rapid and visual loop-mediated isothermal amplification assay to detect leifsonia xyli subsp. xyli targeting a transposase gene. *Lett Appl Microbiol* 2014;59(6):648–57. <http://dx.doi.org/10.1111/lam.12327>.
- [11] Young AJ, Kawamata A, Ensbey MA, Lambley E, Nock CJ. Efficient diagnosis of ratoon stunting disease of sugarcane by quantitative PCR on pooled leaf sheath biopsies. *Plant Dis* 2016;100(12):2492–8. <http://dx.doi.org/10.1094/PDIS-06-16-0848-RE>, PMID: 30686165.

- [12] Bao D, Zhou J, Bhuiyan SA, Zia A, Ford R, Gao Y. Early detection of sugarcane smut disease in hyperspectral images. In: 2021 36th international conference on image and vision computing New Zealand. 2021, p. 1–6. <http://dx.doi.org/10.1109/IVCNZ54163.2021.9653386>.
- [13] Lu B, Dao PD, Liu J, He Y, Shang J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens* 2020;12(16):2659. <http://dx.doi.org/10.3390/rs12162659>.
- [14] Bao D, Zhou J, Bhuiyan SA, Adhikari P, Tuxworth G, Ford R, Gao Y. Early detection of sugarcane smut and mosaic diseases via hyperspectral imaging and spectral-spatial attention deep neural networks. *J Agric Food Res* 2024;18:101369. <http://dx.doi.org/10.1016/j.jafr.2024.101369>, URL: <https://www.sciencedirect.com/science/article/pii/S266615432400406X>.
- [15] Grisham MP, Johnson RM, Zimba PV. Detecting sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes. *J Virol Methods* 2010;167(2):140–5. <http://dx.doi.org/10.1016/j.jviromet.2010.03.024>.
- [16] Ong P, Jian J, Li X, Zou C, Yin J, Ma G. New approach for sugarcane disease recognition through visible and near-infrared spectroscopy and a modified wavelength selection method using machine learning models. *Spectrochim Acta Part A: Mol Biomol Spectrosc* 2023;302:123037. <http://dx.doi.org/10.1016/j.saa.2023.123037>.
- [17] Vargas LAO, Mendoza GG, Gómez RA, Rivero NA, Espinosa LY. Characterization of diatraea saccharalis in sugarcane (saccharum officinarum) with field spectroradiometry. *Int J Environ Agric Res (JJOEAR)* 2016.
- [18] Abdel-Rahman E, Ahmed F, van den Berg M, Way M. Potential of spectroscopic data sets for sugarcane thrips (Fulmekiola serrata Kobus) damage detection. *Int J Remote Sens* 2010;31:4199–216. <http://dx.doi.org/10.1080/01431160903241981>.
- [19] Soca-Muñoz JL, Rodríguez-Machado E, Aday-Díaz O, Hernández-Santana L, Orozco-Morales R. Spectral signature of brown rust and orange rust in sugarcane. *Rev Fac Ing Univ Antioquia* 2020;9–20. <http://dx.doi.org/10.17533/udea.redin.20191042>.
- [20] Moriya EAS, Imai NN, Tommaselli AMG, Miyoshi GT. Mapping mosaic virus in sugarcane based on hyperspectral images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2017;10(2):740–8. <http://dx.doi.org/10.1109/JSTARS.2016.2635482>.
- [21] Apan A, Held A, Phinn S, Markley J. Detecting sugarcane 'orange rust' disease using EO-1 hyperion hyperspectral imagery. *Int J Remote Sens* 2004;25(2):489–98. <http://dx.doi.org/10.1080/01431160310001618031>.
- [22] Narmilan A, Gonzalez F, Salgadoe ASA, Powell K. Detection of white leaf disease in sugarcane using machine learning techniques over UAV multispectral images. *Drones* 2022;6(9):230. <http://dx.doi.org/10.3390/drones6090230>.
- [23] Simões IO, Rios do Amaral L. UAV-based multispectral data for sugarcane resistance phenotyping of orange and brown RUST. *Smart Agric Technol* 2023;4:100144. <http://dx.doi.org/10.1016/j.atech.2022.100144>.
- [24] Johansen K, Robson A, Samson P, Sallam N, Chandler K, Eaton A, Derby L, Jennings J. Mapping canegrub damage from high spatial resolution satellite imagery. In: Proceedings of the 36th conference of the Australian society of sugar cane technologists. 2014, p. 62–70.
- [25] Johansen K, Sallam N, Robson A, Samson P, Chandler K, Derby L, Eaton A, Jennings J. Using GeoEye-1 imagery for multi-temporal object-based detection of canegrub damage in sugarcane fields in queensland, Australia. *GIScience Remote Sens* 2018;55(2):285–305. <http://dx.doi.org/10.1080/15481603.2017.1417691>.
- [26] Apan A, Held A, Phinn S, Markley J. Formulation and assessment of narrow-band vegetation indices from EO1 hyperion imagery for discriminating sugarcane disease. *Proc Spat Sci Conf* 2003.
- [27] Waters EK, Chen CCM, Rahimi Azghadi M. Sugarcane health monitoring with satellite spectroscopy and machine learning: A review. *Comput Electron Agric* 2025;229:109686. <http://dx.doi.org/10.1016/j.compag.2024.109686>, URL: <https://www.sciencedirect.com/science/article/pii/S0168169924010779>.
- [28] Xue J, Su B. Significant remote sensing vegetation indices: A review of developments and applications. *J Sensors* 2017;2017:1353691. <http://dx.doi.org/10.1155/2017/1353691>.
- [29] Fang H, Liang S. Leaf area index models. Reference module in earth systems and environmental sciences. Elsevier; 2014, <http://dx.doi.org/10.1016/B978-0-12-409548-9.09076-X>.
- [30] European Space Agency. Sentinel-2 overview. 2022, <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/overview>. [Accessed 01 May 2022].
- [31] Gao Bc. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ* 1996;58(3):257–66. [http://dx.doi.org/10.1016/S0034-4257\(96\)00067-3](http://dx.doi.org/10.1016/S0034-4257(96)00067-3).
- [32] Merzlyak MN, Gitelson AA, Chivkunova OB, Rakitin VY. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol Plant* 1999;106(1):135–41. <http://dx.doi.org/10.1034/j.1399-3054.1999.106119.x>.
- [33] Rouse Jr JW, Haas RH, Schell J, Deering D. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Report, Remote Sensing Center Texas A&M University; 1973.
- [34] Fensholt R, Sandholt I. Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sens Environ* 2003;87(1):111–21. <http://dx.doi.org/10.1016/j.rse.2003.07.002>, URL: <https://www.sciencedirect.com/science/article/pii/S0034425703001895>.
- [35] Liu Y, Qian J, Yue H. Comprehensive evaluation of sentinel-2 red edge and shortwave-infrared bands to estimate soil moisture. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2021;14:7448–65. <http://dx.doi.org/10.1109/JSTARS.2021.3098513>.
- [36] Wang C, Chen J, Wu J, Tang Y, Shi P, Black TA, Zhu K. A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems. *Remote Sens Environ* 2017;196:1–12. <http://dx.doi.org/10.1016/j.rse.2017.04.031>, URL: <https://www.sciencedirect.com/science/article/pii/S0034425717301906>.
- [37] Segarra J, Buchailot ML, Araus JL, Kefauver SC. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy* 2020;10(5):641.
- [38] Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci Remote Sens Lett* 2017;14(5):778–82.
- [39] Qiu S, He B, Yin C, Liao Z. Assessments of sentinel-2 vegetation red-edge spectral bands for improving land cover classification. *Int Arch Photogramm Remote Sens Spat Inf Sci* 2017;XLII-2/W7:871–4. <http://dx.doi.org/10.5194/isprs-archives-XLII-2-W7-871-2017>, URL: <https://isprs-archives.copernicus.org/articles/XLII-2-W7/871/2017/>.
- [40] Kaufman YJ, Tanre D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans Geosci Remote Sens* 1992;30(2):261–70. <http://dx.doi.org/10.1109/36.134076>.
- [41] Genc H, Genc L, Turhan H, Smith S, Nation J. Vegetation indices as indicators of damage by the sunn pest (Hemiptera: Scutelleridae) to field grown wheat. *Afr J Biotechnol* 2008;7(2).
- [42] Jordan CF. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 1969;50(4):663–6. <http://dx.doi.org/10.2307/1936256>.
- [43] McFeeters SK. The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int J Remote Sens* 1996;17(7):1425–32. <http://dx.doi.org/10.1080/01431169608948714>.
- [44] Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ* 1979;8(2):127–50. [http://dx.doi.org/10.1016/0034-4257\(79\)90013-0](http://dx.doi.org/10.1016/0034-4257(79)90013-0).
- [45] Gitelson AA, Kaufman YJ, Stark R, Rundquist D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens Environ* 2002;80(1):76–87.
- [46] Datt B. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sens Environ* 1998;66(2):111–21. [http://dx.doi.org/10.1016/S0034-4257\(98\)00046-7](http://dx.doi.org/10.1016/S0034-4257(98)00046-7), URL: <https://www.sciencedirect.com/science/article/pii/S0034425798000467>.
- [47] Jamieson K, Talwalkar A. Non-stochastic best arm identification and hyperparameter optimization. In: Artificial intelligence and statistics. PMLR; 2016, p. 240–8.
- [48] Scikit-learn Developers. HalvingGridSearchCV. 2024, URL: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.HalvingGridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.HalvingGridSearchCV.html). [Accessed 26 September 2024].
- [49] Ting KM. Encyclopedia of machine learning. Boston, MA: Springer US; 2010, [http://dx.doi.org/10.1007/978-0-387-30164-8\\_652](http://dx.doi.org/10.1007/978-0-387-30164-8_652), 781–781.
- [50] Murphy K. Probabilistic machine learning: An introduction. In: Adaptive computation and machine learning series, MIT Press; 2022, URL: <https://books.google.com.au/books?id=wrZNEAAQBAJ>.
- [51] Sugar Research Australia. Variety guide. Sugar Research Australia; 2024, URL: [https://sugarsearch.com.au/sugar\\_files/2024/06/SRA\\_Variety-Guide-2024-25-Herbert.pdf](https://sugarsearch.com.au/sugar_files/2024/06/SRA_Variety-Guide-2024-25-Herbert.pdf).



**Ethan Kane Waters** holds a Bachelor of Engineering (Electrical & Electronics) [Honours] and Bachelor of Science (Data Science) from James Cook University (JCU). Ethan earned the C.N. Barton medal for best Engineering Thesis at JCU and the University Medal for academic excellence. In 2023, Ethan was employed as a Research Engineer at JCU and CSIRO to develop software systems, machine learning algorithms and data processing pipelines with applications in agriculture or marine systems. During this time Ethan served as a Chief Investigator for the Sugar-AI: Smart Sugarcane Health Monitoring Project. Currently, Ethan is employed as a Data Scientist at the Department of Primary Industries



**Dr. Carla C.M. Chen** is a Senior Lecturer in statistics and data sciences at James Cook University. She received her Bachelor of Science in Marine Science in 1999 and her Master of Applied Science in Statistics in 2004, both from James Cook University. She later received her doctoral in statistics from the Queensland University of Technology supervised by the Distinguished Professor Kerrie Mengersen. Carla has strong interests in statistical modelling and methods, particularly in the Bayesian framework. These include, but are not limited to, modelling of complex systems, Bayesian temporal and spatial models, model comparisons, classification, clustering, analysis of big data and Markov Chain Monte Carlo methods. She is especially interested in applying statistics to real problems. To date, she has worked in the field of marine and medical research, genetics, agriculture and fisheries.



**Prof. Mostafa Rahimi Azghadi** is an internationally recognised electronics and computer engineer, specialising in neural-inspired computing and artificial intelligence at James Cook University. Prof Azghadi earned his Bachelor's Degree in Electrical and Computer Engineering in 2006 and completed his Master of Engineering in Computer Architecture with first-class honours in 2009 at the Iranian National University. He secured an Australian International Postgraduate Research Scholarship (IPRS) and an Australian Postgraduate Award (APA) in 2010 to pursue his Ph.D. at The University of Adelaide, earning a Doctoral Research Medal and the University Alumni Medal for doctoral thesis excellence in 2014. His research addresses pressing, real-world challenges in agriculture, aquaculture, medicine, and biosecurity. Ranked among the top 2% of highly-cited researchers in Electrical and Electronic Engineering (EEE) and AI since 2020, Prof Azghadi has made significant contributions to AI-driven innovations and has co-raised nearly \$20M in research funding. He is the founding head of JCU's Centre for AI and Data Science Innovation, a deputy director of the Australian Research Council (ARC) Training Centre in Plant Biosecurity, and the AI leader of the ARC Industrial Transformation Research Hub for Aquaculture.