Journal of Agriculture and Food Research 26 (2026) 102660

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

UAV-based integration of RGB, thermal, and structural features with machine learning for multi-class basal stem rot (BSR) severity detection in oil palm

N.A. Husin [a,b,c,*], N.A.H.M. Baktiar [b], V.U. Tagang [b], S. Khairunniza–Bejo [a,b,c], M.F.M. Yusuf [d]

[a] Smart Farming Technology Research Centre, Malaysia
[b] Department of Biological and Agricultural Engineering, Faculty of Engineering, Malaysia
[c] Institute of Plantation Studies, Universiti Putra Malaysia, Serdang, Selangor, 43400, Malaysia
[d] Felcra Berhad, Lot 4780, Jalan Rejang, Setapak Jaya, Peti Surat 12254, Kuala Lumpur, 50772, Malaysia

ARTICLE INFO

Keywords:
Remote sensing
Machine learning
Ensemble bagged trees
RGB spectral: oil palm

ABSTRACT

Basal Stem Rot (BSR) caused by *Ganoderma boninense* is the foremost threat to global oil palm productivity, yet its early and scalable detection remains profoundly challenging. This study presents an integrated UAV-based framework that combines RGB and thermal imagery with top-view structural palm features - crown area, frond number, and frond angle to classify BSR severity levels (T0–T3) using machine learning. A total of 1278 field-verified oil palm trees were assessed, and the Synthetic Minority Oversampling Technique (SMOTE) was applied to address substantial class imbalance. Vegetation indices (VARI, ExG, GLI), thermal pixel intensities, and canopy structural attributes were extracted and Principal Component Analysis was enabled in Machine Learning before training 30 classification models. Among these, the Ensemble Bagged Trees classifier achieved the most robust and consistent performance, recording 94.20 % accuracy for both validation and testing phases, with high per-class precision up to 98.5 % and recall up to 99.7 %. VARI demonstrated the strongest and most consistent spectral response to disease progression, while ExG and GLI exhibited unstable patterns due to canopy shadowing and radiometric variability. The findings highlight the potential of integrating multisensor UAV data with ensemble learning to develop an accurate, scalable, and cost-efficient BSR severity mapping system, supporting improved surveillance and precision disease management across commercial oil palm plantations.

1. Introduction

Elaeis guineensis, commonly known as oil palm, is a species of palm that is best cultivated with high rainfall, adequate sunlight and humid conditions which explains its cultivation across three continents: Asia, Africa, and South America. Oil palm is considered the most productive oil crop globally due to its higher yield per unit of cultivated land compared to other vegetable oil crops. Data from the Food and Agriculture Organization of the United Nations (FAO) as shown in Fig. 1, oil palm produced about 36 % of vegetable oil with only 8.6 % of land area used compared to sunflower oil that produced only 9 % of oil with almost the same land area requirement [1]. The two adjacent Southeast Asian countries of Indonesia and Malaysia are the vast majority of grown oil palm worldwide which covering about 84 % of global production. Oil palm is the top of three main agricultural commodities in Malaysia besides rubber and cocoa, where the total planted oil palm areas in Malaysia were 5.65 million hectares in 2023 [2]. The production of crude palm oil (CPO) reached up to 18.55 million tons while the revenue for palm oil and oil palm products exports in 2022 was recorded at about RM 130.22 billion (MPOB, 2022). Also, more than half a million employees and about one million people's livelihoods are benefitted from Malaysian palm oil industry [3].

Nevertheless, an existing menace which is a disease called Basal Stem Rot (BSR) may pose a serious threat to the oil palm industry if there are no mitigation strategies implemented. *Ganoderma boninense (G. boninense)* which is a soilborne fungus that acts as the BSR disease pathogen may lessen the oil palm yields by about 50–80 % [4]. BSR detection may take up time due to the fungus development cycles involved which

* Corresponding author. Smart Farming Technology Research Centre, Malaysia.
E-mail addresses: nurazuan@upm.edu.my (N.A. Husin), 201106@student.upm.edu.my (N.A.H.M. Baktiar), 208396@student.upm.edu.my (V.U. Tagang), skbejo@upm.edu.my (S. Khairunniza–Bejo), fairus.yusuf@felcra.com.my (M.F.M. Yusuf).

https://doi.org/10.1016/j.jafr.2026.102660
Received 25 May 2025; Received in revised form 3 December 2025; Accepted 9 January 2026
Available online 10 January 2026

began with the white mycelia formation followed by small white button existence and bracket shaped fruiting body at last. The structural integrity of the palm is compromised as the fungus breaks down lignin and cellulose in the wood, leading to internal decay. Eventually, the tree will collapse and die due to extensive rot and inability to support itself or uptake nutrients. The fungus is highly virulent, making it capable of spreading widely throughout the entire plantation and infecting not only young plants and seedlings but also mature plants. Once the pathogen attacks the basal stem of oil palm trees, water and nutrients uptake from root to the upper palm part especially the foliage will be severely restricted due to the oil palm tissues disruption. The changes are yellowing or drying some leaves, then the apparition of leaf necrosis is largely spread when it reaches a later stage. Healthy fronds require a continuous supply of resources for their growth and maintenance, whereas existing fronds may die off more promptly due to the lack of adequate nutrients and water, which also impedes the development of new fronds. The palm prioritizes its limited resources to essential functions, leading to the shedding of older fronds to conserve energy. The decay of the internal tissues causes a loss of rigidity, making it difficult for the palm to support the weight of the fronds. Consequently, fronds start to droop and bend downward, which leads to more removal of sagging fronds by the plantation workers. As a result, a smaller crown, fewer fronds, and an increased gap between the fronds have occurred due to the disease infection.

In contrast, a reliable early detection of the disease such as laboratory-based approach is complex in procedure, costly, time-consuming and not design for outdoor environments [5]. The approach can be replaced by using Unmanned Aerial Vehicles (UAV) technology based on remote sensing. Remote sensing offers a more suitable, faster, and reliable approach for outdoor monitoring. UAV-based remote sensing enables efficient monitoring of challenging terrains, including areas that are inaccessible or hazardous for manual inspection. UAV-based remote sensing equipped with cameras can capture high-resolution images and data that facilitate the early detection of *Ganoderma* BSR. Traditional ground-based methods of monitoring and detecting BSR are labor-intensive and time-consuming. UAV-based remote sensing can cover large areas rapidly, much faster than manual method and require fewer human resources, reducing labor costs and operational expenses. This rapid assessment capability is highly benefiting in large-scale commercial plantations, allowing for timely interventions and management decisions. Several researchers succeeded in detecting the BSR disease in oil palm trees by using UAV-based remote sensing approaches including hyperspectral [6–8], multispectral [9–11] and RGB [12–14]. Despite significant capital expenditure, remote sensing approaches for *Ganoderma* identification continue to receive significant interest. The primary advantage of remote sensors is the data can be retained for future examination, allowing for systematic expansion of knowledge boundaries. High-resolution imagery from UAVs could provide detailed oil palm maps at the individual or even oil palm leaf level, which is essential for pinpointing affected areas and assessing disease severity. Accurate health mapping enables targeted interventions, reducing the spread of the disease and optimizing resource use.

Furthermore, crop diseases typically result in specific variations in plants' spectral reflection. Due to the presence of various pigments, changes in cell structure, and water content, disease in crops can result in variations in the spectral reflectance of the plant. Remote sensing technology, which analyzes the reflectance of light at different wavelengths, can detect these changes. For instance, healthy leaves contain more chlorophyll, which absorbs blue and red lights. The amount of chlorophyll in a plant may decrease due to disease, which causes a reduction in green light reflectance and an increase in red and near-infrared light reflection. Additionally, a disease may affect the plant's cell structure, which would affect the way light is reflected and scattered. For instance, the appearance of lesions on leaves may lead to a reduction in the amount of energy reflected in the visible and near-infrared spectrum. Therefore, remote sensors that measure the sunlight reflected from terrestrial objects can detect spectral changes between healthy and unhealthy plants. Low to high spatial resolution, ground-based to aircraft or even satellite-based sensors may all detect the amount of reflected solar radiation, but the amount of wavelengths reflectance is varied based on the type of crop and the stage of disease. Different crops and diseases can cause different changes in the spectral reflectance of the plant, which can be detected using remote sensing technology.

In addition, thermal imaging is one of the remote sensing-based approaches in detecting BSR disease since it is done without direct contact with the oil palm trees. Use of thermal bands for *Ganoderma*

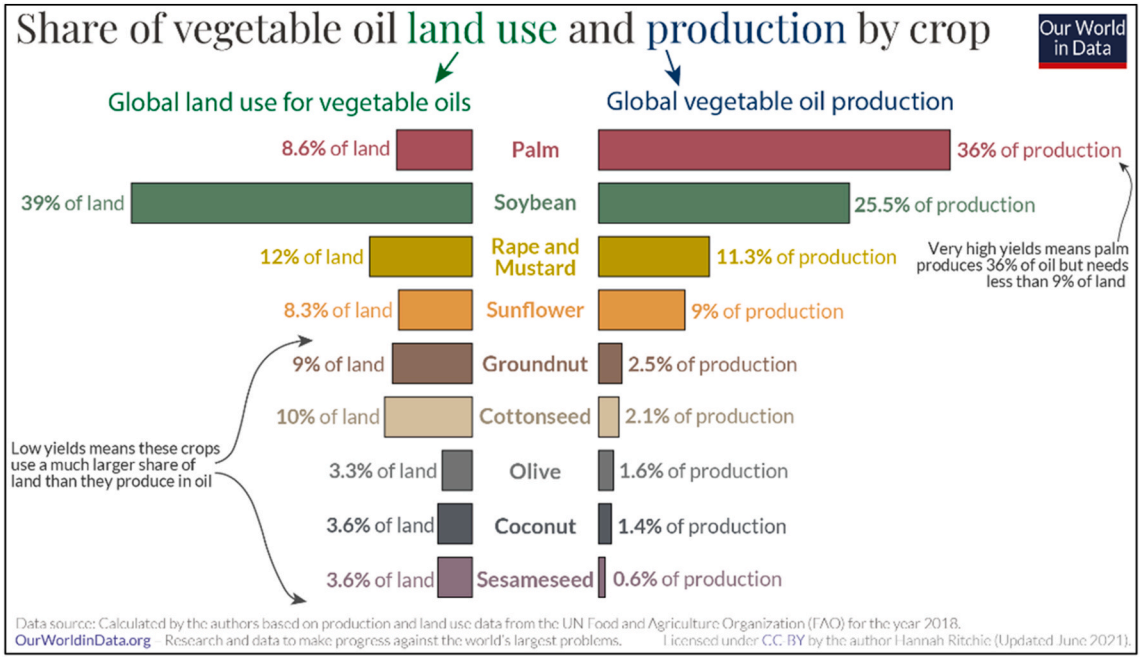

Fig. 1. The breakdown of global vegetable oil production and global land use for vegetable oils in 2018 (Our World Data, 2021).

disease detection in mature oil palm trees were done by researchers limited to only infected and non-infected (two levels) oil palm trees. Hashim et al. [15] used machine learning approach of Random Forest (RF) with the input of temperature feature of Tmax resulted in accuracy of 100 % for infected and 87.10 % non-infected oil palm trees. Meanwhile, Bejo et al. [16] used Support Vector Machines (SVM)-based model approach using average pixel value from thermal images of oil palm trees produced the highest average overall classification accuracy of 89.2 % for training set and 84.4 % for testing. Thermal images of healthy and BSR-infected oil palm seedlings were captured and analyzed to extract key thermal properties, including maximum, minimum, mean, and standard deviation of pixel intensity values. The highest classification accuracy of 80.0 % was achieved using the Support Vector Machine (SVM) with a fine Gaussian kernel, utilizing two principal components, PC1 and PC3 as input features [17]. This research was conducted to study the capability of a thermal imaging device to distinguish BSR infected oil palm trees from healthy oil palm trees based on the temperature of the trees captured in the thermal images. However, there were still limitations to the research such as time of capturing thermal images, the section of the oil palm tree, the feature type, and classification model type which leads to inability to classify the BSR infected trees into their respective severity.

Meanwhile, a Vegetation Index (VI) is a spectral imaging transformation of two or more image bands using mathematical equations into a single value designed to enhance the contribution of vegetation properties of terrestrial photosynthetic activity and canopy structural variations. RGB visible spectral bands is advantageous for its cost-effectiveness, simplicity, ease of use, availability, real-time processing capabilities, aligned with human color perception, and suitability for a broad range of applications. RGB cameras could perform well under a variety of lighting conditions, both natural and artificial, and are ubiquitous which makes them versatile and adaptable for many uses. Vegetation indices using RGB spectral offer a great toolset for disease detection in plants, combining early detection, non-destructive monitoring and high-resolution analysis. By integrating RGB vegetation indices with thermal spectral and UAV oil palm features such as crown area, frond number and frond angle, farmers and researchers can achieve more efficient, data-driven, and environmentally friendly plantation systems. These advantages enhance disease management and contribute to sustainability and productivity of agricultural practices.

BSR infected oil palm trees detection method is currently more convenient and rapid with the aids of machine learning approaches from the remote sensing data. Haw et al. [18] applied deep learning using Convolutional Neural Networks on Terrestrial Laser Scanning (TLS)-derived canopy images to detect Basal Stem Rot (BSR) in oil palm. The fine-tuned DenseNet121 model achieved the best performance with a Macro F1-score of 0.798, demonstrating promise for early-stage disease classification despite challenges like dataset limitations and overfitting. Next, kernel Naïve Bayes (kNB) was used to analyze the TLS data for the four BSR level classification which resulted in 85 % accuracy for four level classification, whereas 100 % accuracy was achieved for healthy level classification [19]. Furthermore, Zheng et al. [20] utilized high-resolution remote sensing imagery in conjunction with a Faster R-CNN framework to enable large-scale detection of individual oil palm trees. Their study demonstrated the robustness of deep convolutional neural networks in managing complex plantation environments characterized by dense canopy structures and heterogeneous backgrounds. Subsequently, Zheng et al. [21] extended this approach by incorporating Unmanned Aerial Vehicle (UAV)-based imagery to evaluate the growing status of individual oil palm trees. Through the integration of a Refined Pyramid Feature module and a class-balanced loss function, their model achieved enhanced precision in distinguishing multiple growth and health categories.

Recent advancements have significantly improved the use of UAV-based imaging and machine learning for diagnosing Ganoderma Basal Stem Rot (BSR) in oil palm. Hashim et al. [22] demonstrated that thermal imagery, combined with an imbalanced data approach, can effectively classify disease severity in oil palms, even under variable field conditions. Kurniawan et al. [23] further applied artificial neural networks to detect Ganoderma BSR, showing that machine learning models can successfully learn discriminative spectral features from UAV data. Complementing these findings, Yang et al. [24] reported high classification performance using UAV-acquired RGB and multispectral imagery coupled with deep learning and vegetation indices such as NGRDI and NDVI. Most recently, Panthakkan et al. [25] evaluated the effectiveness of UAV-based RGB and multispectral vegetation indices in palm tree cultivation, confirming that RGB-based indices can serve as reliable, low-cost alternatives to multispectral systems in precision agriculture applications. Collectively, these studies underscore the growing feasibility of using UAV-derived thermal, visible, and structural data, combined with machine learning algorithms to support early and cost-effective detection of Ganoderma infection in oil palm plantations. Even though there are advantages of NIR (near infra-red) bands compared with visible bands, there is a research gap in exploring the application of ML integrated with inputs from RGB spectral index, thermal band and oil palm features such as crown area, frond number and frond angle for *Ganoderma* BSR disease detection.

2. Materials and methods

2.1. Study area

Fig. 2 shows the study area located at an oil palm plantation in Changkat Lada, Perak, Malaysia. The annual rainfall in this area is between 2000 mm and 2500 mm per year. The size of the study field used for this study is approximately 10 ha and the age of the trees in this field was about 12 years which was planted in 2010. The species of oil palm is *Tenera*, which is a cross between *Pisifera* and *Dura*. Census data were collected from 17 December until December 18, 2022 by tagging and marking the tree's condition and infection level. *Ganoderma* BSR in oil palm manifests through a range of visible foliar and structural symptoms that progress over time. Key indicators include unfolded leaves and unopened spears - young spears fail to open properly, indicating impaired growth at the crown (Fig. 3). Then, drooping of the fronds - mature fronds begin to droop, losing their typical erect posture, which is an early sign of structural weakness. Lastly, drying and yellowing of the leaves - older leaves become chlorotic (yellow) and dry out prematurely, often progressing from the lower canopy upward. These above-ground symptoms typically reflect internal stem decay caused by *G. boninense* and are often observed once the disease is well established, highlighting the importance of early, non-invasive detection methods. A total of 1278 oil palm plants were pre-identified based on apparent symptoms and were divided into four healthiness levels (T0, T1, T2 and T3) [16,26] and its descriptions are tabulated in Table 1.

2.2. Data collection

Aerial images of the oil palm plantation plot area in FELCRA Seberang Perak were captured by using a UAV drone DJI Inspire 2 equipped with Micasense Altum-Pt (33.5 cm/pixel thermal GSD at 122 m) (Fig. 4) for capturing thermal aerial images along with DJI Zenmuse H20 camera (20 MP resolution) to acquire the high resolution RGB aerial images (Fig. 5). The camera sensor was calibrated prior flight using CRP (Calibrated Reflectance Panel) for a high accuracy of light conditions during the flight. CRP also acted as the ground radiometric calibration tool to pre-measure the reflectance values for visible and near-infrared light spectrum. For maximizing flight time, a twin-battery configuration was utilized, extending the maximum flight time to 27 min. The images were captured at an altitude of approximately 75 m above ground with an image overlapping of 80 % under partly overcast weather conditions, which were deemed suitable for flying the drone. During image acquisition, Agisoft Metashape software was utilized to

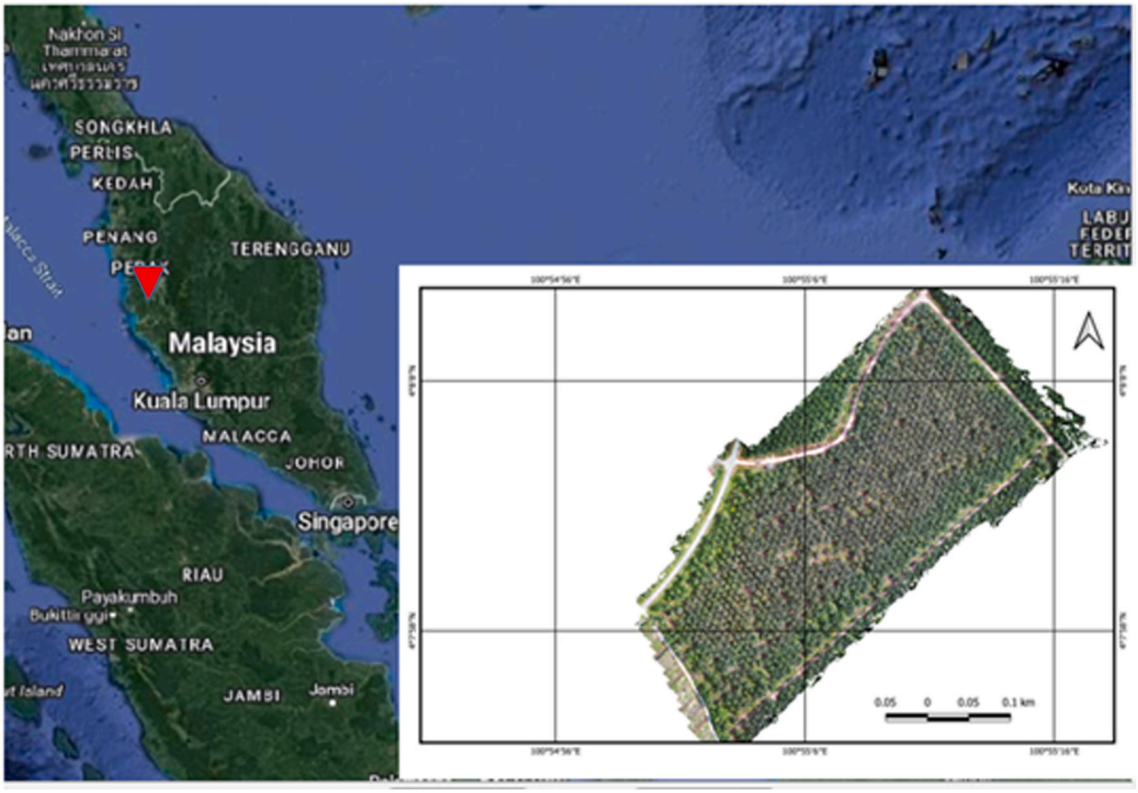

Fig. 2. Study map location of Changkat Lada plantation.

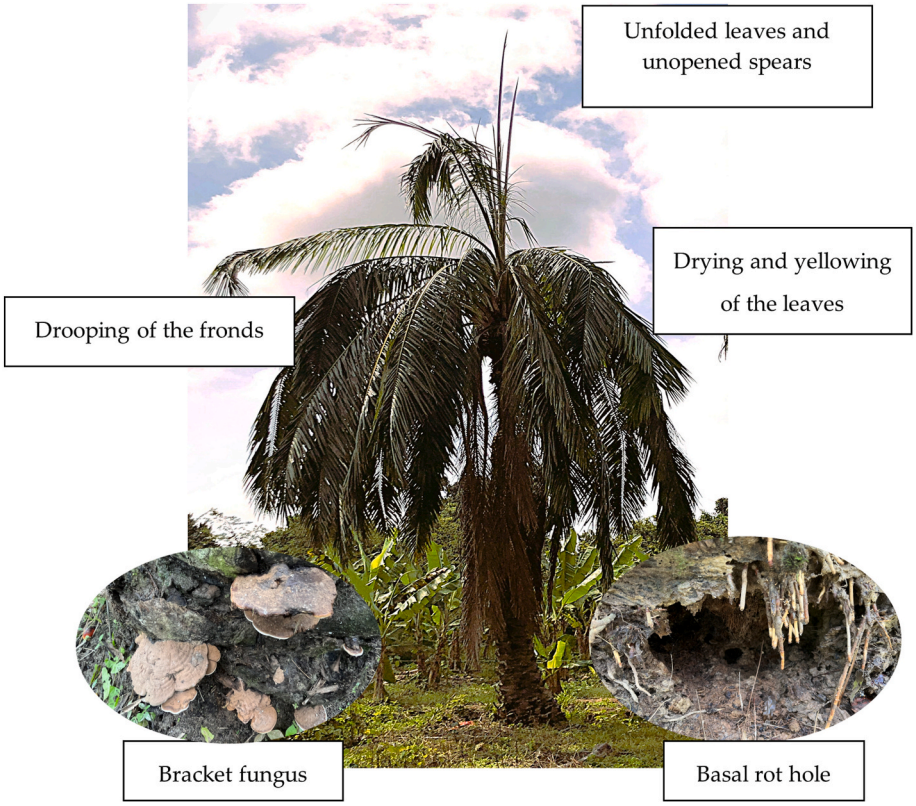

Fig. 3. Symptoms of severe *Ganoderma* BSR infection at foliar – spears, leaves and fronds; at basal – fungus and rot hole.

map the pilot's pathway and enable autopilot functionality for the drone.

Table 1
The categories and the descriptions of *Ganoderma* BSR severity.

Severity level	Descriptions	Number of trees
T0 (Healthy)	Healthy palm, minimal foliage symptoms (0–25 %), no fruiting body	1097
T1 (Mildly Infected)	Early infection, minimal foliage symptoms (0–25 %), fruiting body	128
T2 (Moderately Infected)	Moderate infection, foliage symptoms (25–50 %), fruiting body	35
T3 (Severely Infected)	Severe infection, foliage symptoms (50–75 %), fruiting body	18

2.3. RGB and thermal image processing

The pre-processing of the aerial images was done by employing Agisoft Metashape (Version 1.7.2, Agisoft LLC, St. Petersburg, Russia). At first, 971 images were loaded and aligned by setting the key point at high accuracy of 40 000 for precise camera position estimation, which produced 2 688 005 matching tie point clouds. After calibration, only 967 images were used to build the dense point cloud. High-quality depth map was generated with mild depth filtering setting, which generated 387 825 149-point clouds as a data source for Digital Elevation Model (DEM) with the fixed coordinate system of WGS 84. The DEM was then utilized as the surface along with blending mode as the mosaic to create the orthomosaic map. The "salt-and-pepper" effect was reduced by setting the hole filling option as default to reduce the illumination of multiple small sections in a complex surface. The generated result was a TIF (Tagged Image Format) file comprised of several multi spectral bands. The TIF file image was loaded into Quantum Geographic Information System (QGIS) software (Version 3.28.3, Open-Source Geospatial Foundation) for further processing steps. Every oil palm tree in

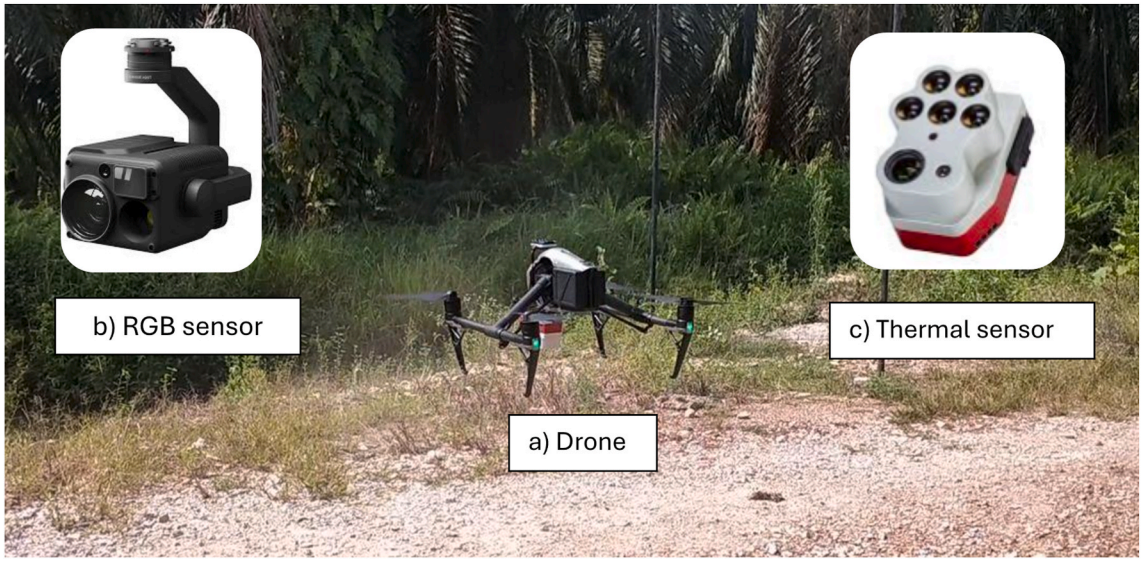

Fig. 4. a) DJI Inspire 2 drone, b) DJI Zenmuse H20 camera and c) Mica Sense Altum-PT sensor.

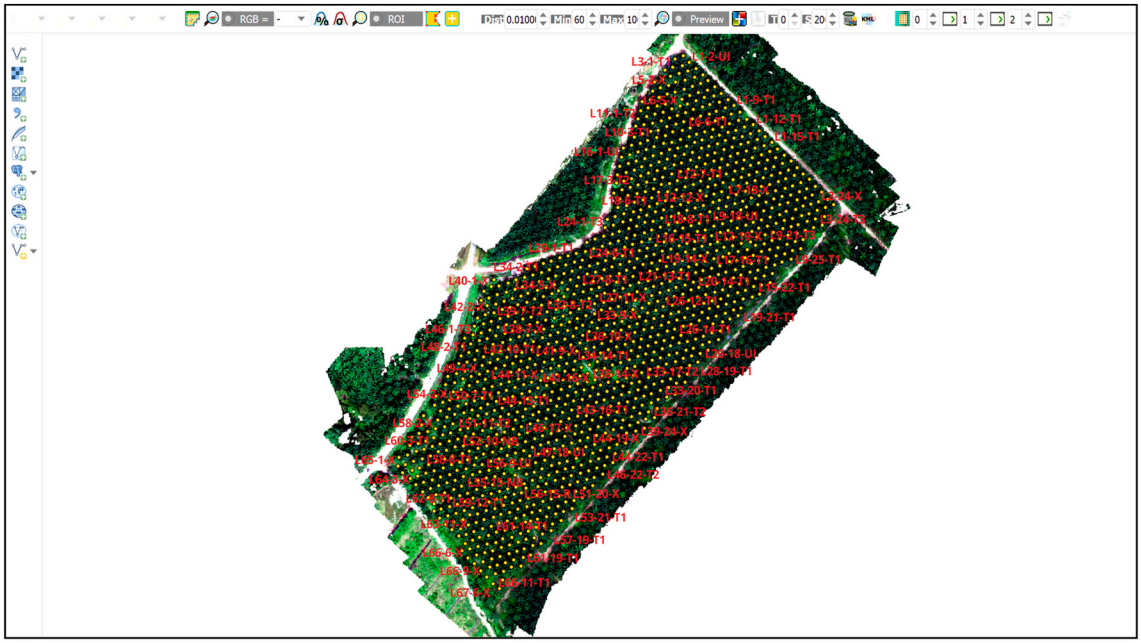

Fig. 5. Oil palm tree identification of exported TIF file image on QGIS software.

the image was labelled with line number and severity level according to the data census reports. The step was done by creating a shapefile layer and choosing the 'point' as the geometry type. Similar steps were made for the thermal band (Band 8 in the camera), where the pixel intensity values were extracted from the image (Fig. 6a).

2.4. RGB image: oil palm features extraction

The TIF image consists of RGB bands used for oil palm features extraction namely crown area, frond number and angle. The crown area was calculated via QGIS software, where Band 1 (Blue), Band 2 (Green) and Band 3 (Red) had been the input layer. The shapefile layer was created with the 'polygon' chosen as geometry type and stream-digitizing was used to draw polygon area on the individual oil palm tree (Fig. 6b) Then an algorithm of zonal statistics was performed to calculate the pixel numbers which represent the coverage of crown area. The statistics results were stored in the attribute table for each created shapefile layer.

Next, the frond number and frond angle were calculated by using the AutoCAD software version 2024 (Autodesk, Inc., San Rafael, USA). Individual oil palm tree was cropped from the TIF file image using the Snipping Tool (Microsoft Corp., USA). The cropped image was attached into the current layer of AutoCAD. Then the 'draw' tool was selected, and the 'polyline' function was used to draw the frond. As the drawing step followed the shape of frond in 2D, thus, the overlapping fronds, which were obscured in the top view, were excluded. The measurement of frond angle was done by using the same cropped image, where the 'dimension' tab was used to measure the space between the fronds. The previous drawn lines were used as the reference to measure the frond angle, where the angle measurement was continued to the adjacent fronds until all frond angles were measured (Fig. 7).

2.5. RGB spectral indices

The calculation of RGB vegetation index was also done by using QGIS software. Three types of indexes used were Visual Atmospheric Resistance Index (VARI), Excess Green Index (EXG) and Green Leaf Index (GLI). The RGB Vegetation Index (RVI) was employed to assess the health and productivity of the vegetation. Explanation and the equation used for the spectral indices is presented in Table 2.

2.6. Statistical analysis

The mean values for pixel intensity values from thermal image and oil palm features (crown area, frond number and frond angle) were analyzed using the JMP software (SAS Institute) for the analysis of variance (ANOVA). The ANOVA test was aimed to analyze the difference between means of different severity categories due to p-values less than significance value in which was set for 0.05 for this study. After determining that there were significant differences between groups using ANOVA, the Tukey-Kramer HSD test was applied to identify which specific groups (severity categories of BSR) were significantly different from each other. This method compares all possible pairs of group means while controlling for Type I error, making it suitable for unequal sample sizes.

2.7. Synthetic Minority Oversampling Technique (SMOTE)

To address the substantial class imbalance present in the original dataset, particularly the limited number of samples representing mild (T1), moderate (T2) and severe BSR infection (T3) - the Synthetic Minority Oversampling Technique (SMOTE) was applied using the WEKA software prior to model training. SMOTE is a well-established oversampling method that synthetically increases the representation of minority classes by generating new samples through interpolation between existing instances. Unlike simple duplication, SMOTE constructs synthetic observations by selecting a minority-class sample, identifying its k nearest minority neighbours (typically $k = 5$), and creating new feature vectors along the line segments joining the sample and its neighbours. This approach preserves the underlying structure within the minority class while reducing the risk of model overfitting that often occurs with random oversampling.

In this study, SMOTE was applied to each minority class (T1–T3) until all classes reached an approximately equal number of samples for training. The balanced dataset was then used as input for the machine learning software. All other preprocessing steps, including feature extraction, principal component transformation, and data normalization, remained identical to those used in the original dataset. Applying SMOTE ensured that the classifier encountered sufficient examples from each severity level, enabling it to learn representative decision boundaries and reducing systematic prediction bias towards the dominant T0 (Healthy) class.

2.8. Machine learning

This study utilized Classification Learner App from the Statistics and Machine Learning Toolbox which was available in the MATLAB software (Version R2024b, The Mathworks Inc., Massachusetts, United States). The Classification Learner application was developed to simplify the process of training and assessing machine learning models for classification tasks. This application enables users to construct and implement classification models without the need for heavy coding. It encompassed a range of algorithms, such as decision trees, support vector machines (SVM), ensemble methods, k-nearest neighbours (KNN), Naive Bayes, neural networks, and logistic regression. A more detailed explanation of the models and kernels can be found in Ciaburro [27].

The data used for the classification had been divided into two sets which were training and testing. Holding out the data sets between training and testing enabled evaluation and comparison of the different model's performance without having any possibilities of overfitting on

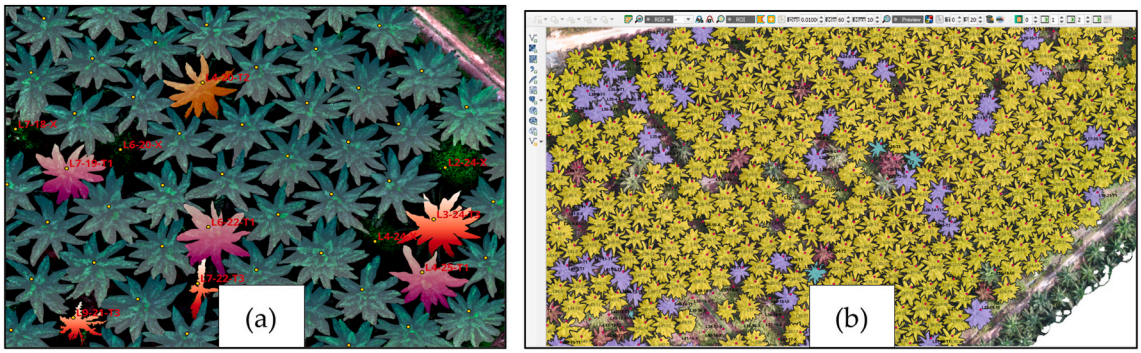

Fig. 6. Manually delineated polygon areas for every oil palm trees in aerial image of a) thermal, b) RGB.

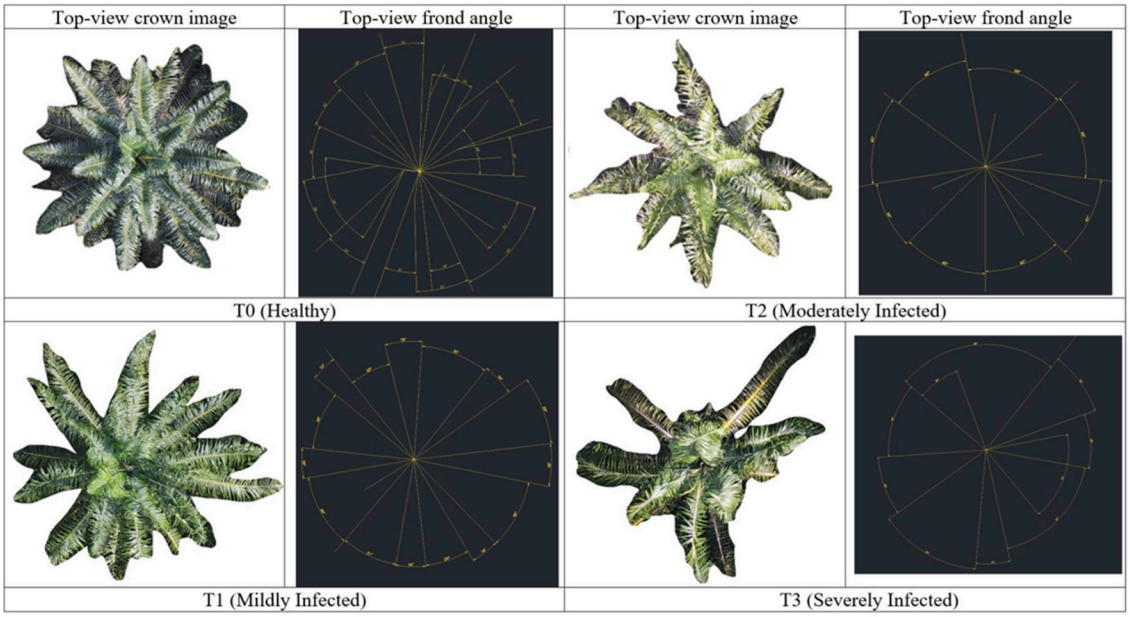

Fig. 7. Line graph of frond number and frond angle at different severity categories.

Table 2
Vegetation used in this study. (G: Green band reflectance, R: Red band reflectance, and B: Blue band reflectance).

Vegetation Index	Description	Formula
ExG	ExG enhances the green component to differentiate vegetation from non-vegetation.	EXG = (2* G) - R-B
GLI	GLI estimates green vegetation cover and chlorophyll content, focusing on plant health and density.	GLI = (2G-R-B)/(2G + R + B)
VARI	VARI is used for minimizing atmospheric effects and highlighting vegetation health.	VARI = (G-R)/(G + R-B)

the training set. The data sets had been split by having training data sets more than testing data sets in order to train the classification model better. The ratio applied was 70:30 (training: testing) as it was commonly used. Gholami et al. [28] also suggested that best results were obtained by having 20–30 % of data for testing and 70–80 % of data for training. All extracted RGB indices, thermal features, and palm structural measurements were used as inputs for model training. Before classification, the feature set was transformed using Principal Component Analysis (PCA) in MATLAB to reduce redundancy and improve model stability. PCA retained 95 % of the variance, ensuring that only the most informative components were used while minimizing noise in the dataset.

3. Model performance evaluation

The performance of the classification models was evaluated using a separate test dataset that was not involved in model training or validation. All analyses were conducted using MATLAB software. The primary evaluation tool was the confusion matrix, which summarizes the number of correct and incorrect classifications across the four disease severity categories (T0–T3). From this matrix, key components such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) were derived. Using the confusion matrix, several per-class performance metrics including precision, recall (sensitivity), specificity and F1-score were calculated to quantify how well the model discriminated between healthy, mildly, moderate, and severe infection levels.

In addition to class-level metrics, overall model performance was assessed using Overall Accuracy, Macro F1, Weighted F1, Micro F1, and Cohen's Kappa. Overall Accuracy measures the proportion of correctly classified samples, while Macro and Weighted F1-scores assess performance balance across classes. The Micro F1-score reflects aggregate predictive performance, and Cohen's Kappa provides an accuracy measure adjusted for chance agreement. To evaluate the model's discrimination capability across threshold settings, Receiver Operating Characteristic (ROC) curves and Precision–Recall (PR) curves were generated for each class. ROC curves provide insight into sensitivity–-specificity trade-offs, while PR curves highlight precision versus recall performance, particularly useful in multi-class problems. All metrics and visual evaluations were derived exclusively from the independent test dataset to ensure an unbiased assessment of model performance. The formulas for all evaluation metrics are provided in Table 3, and the methodological workflow is depicted in Fig. 8.

4. Results and discussion

4.1. Descriptive statistics

4.1.1. Analysis of crown area and thermal

The oil palm crown areas which were indicated by crown pixel numbers is one of the features to detect the BSR-infected oil palm trees in

Table 3
Evaluation metrics developed from the confusion matrix indices.

Evaluation Metrics	Formula
Accuracy	$\frac{TP+TN}{TP+FN+FP+FN}$
Precision (P)	$\frac{TP}{TP+FP}$
Sensitivity/Recall (R)	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
F1 Score	$2\times\frac{P\times R}{P-R}$
Kappa Score	$\frac{OA-EA}{1-EA}$
Micro F1 Score	$\frac{2\times\sum TP}{2\times\sum TP+\sum FP+\sum FN}$

Note: OA = Overall accuracy, EA = Expected accuracy.

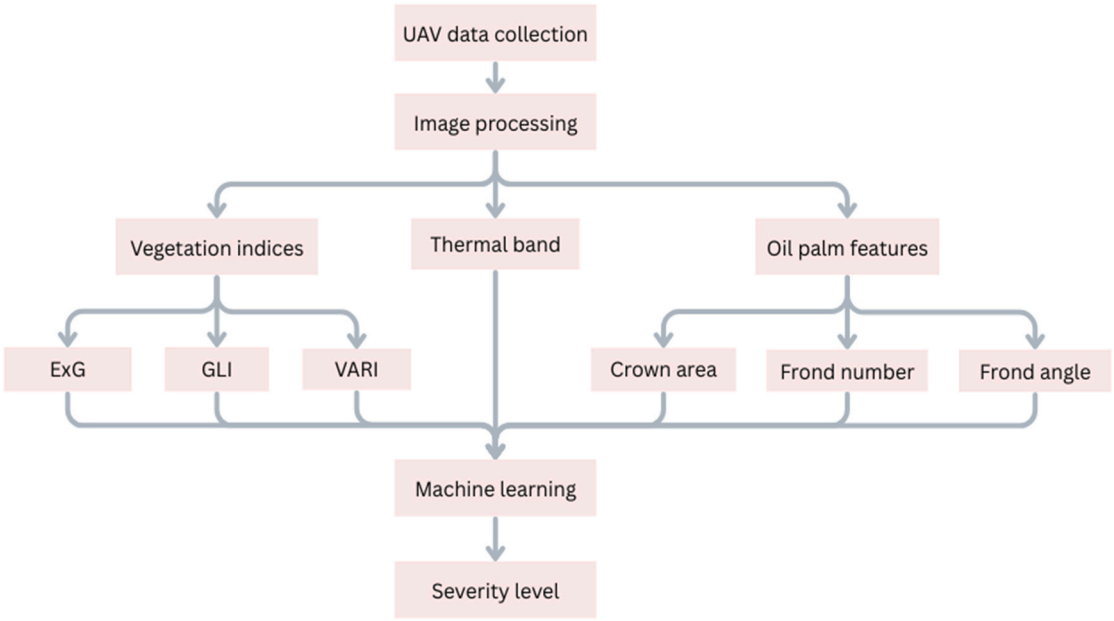

Fig. 8. Flowchart of the steps.

the plantations. The existence of *G. boninense* would probably reduce the crown areas due to internal cell interruption that would affect the crown development and thus, increasing bad foliar symptoms of BSR. Fig. 9 shows the decreasing trend of crown pixel numbers as the severity level increased from T0 to T3. A negative slope and a correlation of R^2 = 0.9634 indicate that the higher the severity level, the smaller the crown area. The oil palm trees with healthy condition (T0) had the largest mean crown area with 368848 crown pixel numbers compared with the non-healthy oil palm trees. As the infection level increased from T1 to T3, the crown areas were progressively smaller. The severely infected trees, T3 only had 131465 mean crown pixel numbers in which approximately less than half of the mean crown pixel numbers of oil palm trees at T0 level. Decreasing crown area of BSR-infected oil palm trees was the effect of foliar symptoms appearance. As *Ganoderma* began to infect the oil palm trees by attacking the root at first, symptoms would appear with leaves becoming dried, yellowish, and chlorotic as the effect of water and nutrients restriction from the root. Finally, as the water stress completely took up the oil palm trees, necrosis happened to the oldest leaves from the lower part and forming skirt-like appearance and spreading the effect upwards through the crown [29].

Fig. 9 also shows the mean thermal spectral of oil palm crown, where all values are increasing from the category of healthy (T0) to mildly infected (T1) followed by moderately infected (T2) and severely infected (T3). The relationship is shown through positive linear regression by the value of determination coefficient or R2 (0.9552). This indicates that the higher the severity level, the higher the crown temperature. As the pixel values increase from level T0 to T3, the severity level of the BSR-infected trees also get higher. A higher temperature was recorded on the canopy of infected oil palm trees compared to the healthy oil palm trees due to *G. boninense* infection. The canopy temperature increased with infection severity category increased from T1 to T3 is well-explained by understanding the bad impact of the pathogen on the internal tissue of trees.

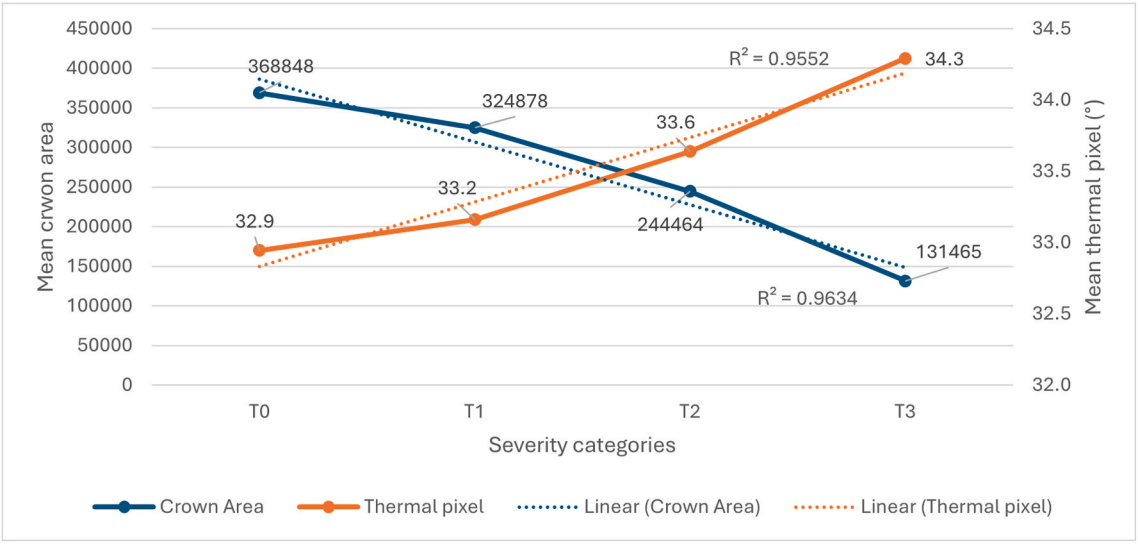

Fig. 9. Line Graph of crown area and thermal band at different severity categories.

Once *G. boninense* attacked the oil palm tree, it slowly damaged the xylem tissue and restricted water uptake from lower part to the upper part of that tree, which induced water stress to the tree [30]. Thus, the minimal water content that cannot reach the upper part of tree canopy amplified its temperature [31].

4.1.2. Analysis of frond number and frond angle

Fig. 10 shows the mean frond numbers for four different severity categories of oil palm trees, where the values are decreasing from healthy (T0) to severely infected (T3). The negative slope and a correlation of $R^2 = 0.9256$ indicate that frond number decreased with the increasing severity category. The mean frond number for healthy oil palm trees was recorded as 23 while the severely infected trees only had 12 mean frond numbers. The effects of parasitic *G. boninense* infection not only decreased the crown area but also lowered the frond numbers of oil palm trees. The declining trend of frond numbers in Fig. 10 was due to ruptured internal cells and causing collapse of oil palm leaves. The BSR is also known as a silent disease with no obvious symptoms when the infection level is around 8 % or less in plant roots [32]. Thus, this explains the almost identical mean frond numbers between oil palm trees in T0 and T1 with 23 and 22, respectively.

It was hypothesized that the oil palm trees with higher severity level of BSR-infection will have larger frond angle. Fig. 10 shows the mean frond angles for four different severity categories of oil palm trees, where the trend is increasing from healthy (T0) to severely infected (T3). The positive linear regression of $R^2 = 0.8379$ indicates that the frond angle expands as the infection level increases. The severely infected (T3) oil palm tree had the largest mean frond angle with 32.6° compared with the other lower severity categories 21.4° (T2), 16.7° (T1) and 16.1° (T0). The difference between healthy and mildly infected oil palm trees is very small coinciding with the issue of early detection of T1 category of infection. Compared with the results of frond numbers, the visual symptoms of *G. boninense* infection could also be observed by analyzing the frond angle. The frond angle of infected oil palm trees became bigger as the neighboring fronds are missing or collapsing. Meanwhile, the healthy trees are not experiencing growth suppression, hence the frond angle will be smaller. As the hypothesis stated, the wider the frond angle, the more severe the infection level in which could be clearly observed as in Fig. 10. It was parallel with the symptoms of BSR disease prevalence including the frond production reduction due to the fungus existence which disturbed the leaves growth, causing more unfolded leaves and unopened spears [33].

In this study, frond angle measurements were performed manually using AutoCAD, a method that, while accurate, is labor-intensive and not scalable for large plantation datasets. Similarly, frond number was counted visually from top-view UAV imagery. These manual approaches, though effective for small sample sizes, pose significant limitations for real-time or large-scale deployment. To address this, future studies should explore the automation of palm structural feature extraction using deep learning-based object detection and instance segmentation models, such as YOLOv8, Mask R-CNN, or Detectron2. These techniques can be trained to recognize individual fronds and their orientations from high-resolution UAV imagery, enabling automated quantification of both frond angle and number with high accuracy and consistency. Incorporating such methods would greatly enhance processing efficiency, reduce subjectivity, and support the development of scalable, intelligent monitoring systems for oil palm health and disease detection.

4.1.3. Analysis of vegetation indices

RGB-based vegetation indices -Green Leaf Index (GLI), Excess Green Index (ExG), and Visible Atmospherically Resistant Index (VARI) were evaluated to examine their relationship with Basal Stem Rot (BSR) severity in oil palm canopies. These indices are commonly used to estimate vegetation greenness, canopy structure, and chlorophyll intensity from visible RGB images, where higher index values typically indicate healthier vegetation. Fig. 11 presents the mean index values across the four severity categories (T0–T3).

Among the three indices, VARI exhibited the most consistent and expected pattern, showing a gradual decline from healthy palms (T0) to severely infected palms (T3). This behavior reflects the reduction of photosynthetically active tissue as infection progresses. VARI's robust performance is likely due to its formulation, which minimizes sensitivity to illumination variability and atmospheric scattering, making it effective in heterogeneous field environments commonly encountered in UAV-based imaging. In this study, VARI emerged as the most reliable

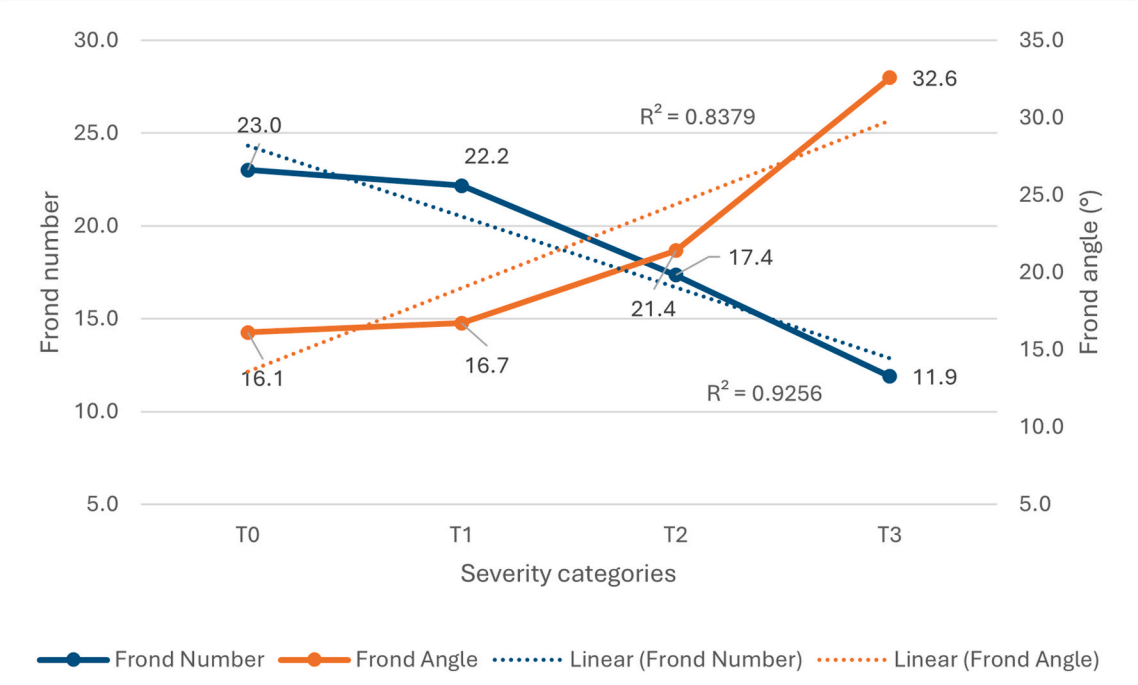

Fig. 10. Line graph of frond number and frond angle at different severity categories.

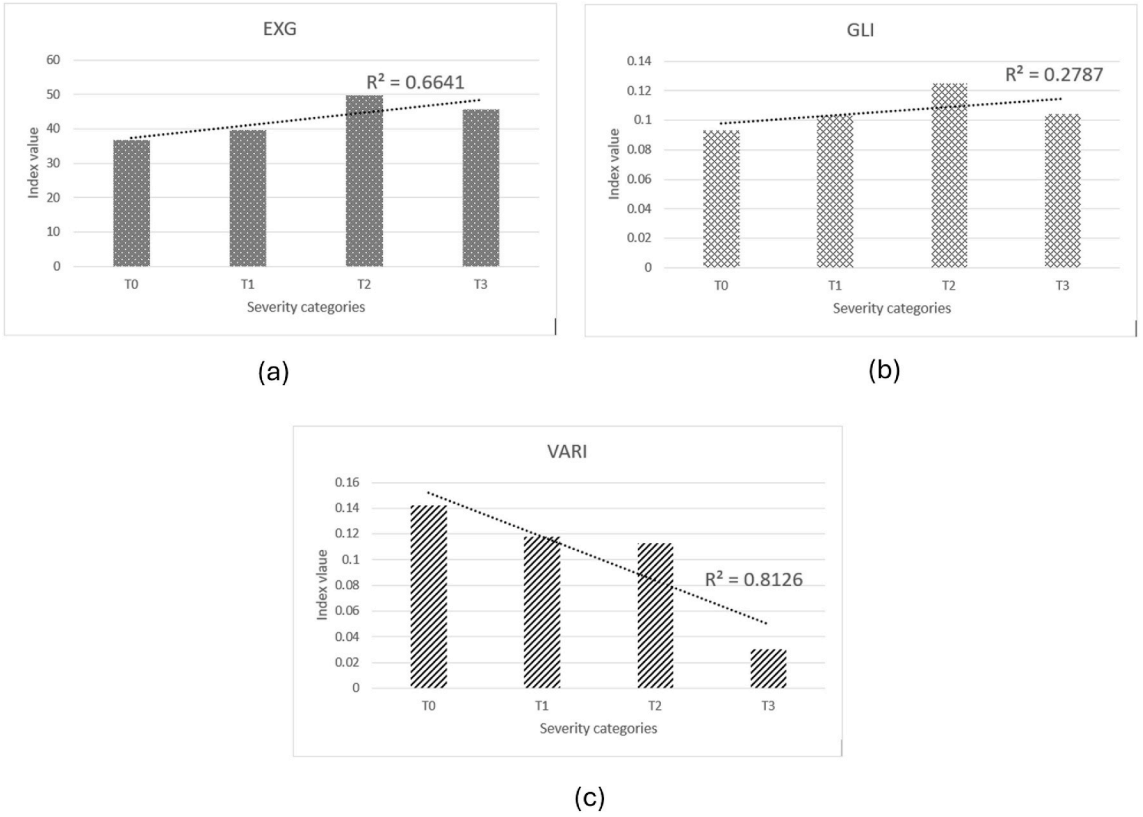

Fig. 11. Mean value for each vegetation index at different severity categories.

RGB indicator of BSR severity.

Although both the ExG and GLI indices demonstrated an inverse and counterintuitive relationship with BSR severity, where mean index values declined as infection became more severe, the patterns observed in Fig. 11 are not strongly apparent or consistent. This is reflected in the relatively low R^2 values for both indices (ExG: 0.6641, GLI: 0.2787), indicating that only a modest proportion of the variation in these indices is explained by disease severity. The weak explanatory power of these linear fits supports the results neither index exhibits a clear, monotonic decline with increasing BSR severity. A closer examination of the bar plots showed that the differences between T1 and T3 are minimal for GLI, and the progression across severity levels is irregular rather than smoothly decreasing. For instance, GLI values increase slightly from T0 to T2, but then drop at T3 disrupting the expected trend of gradual decline. Similarly, ExG shows a noticeable increase from T0 to T2, but then decreases again at T3. These fluctuations illustrate that the progression from healthy to severely infected palms is not reflected consistently in the GLI and ExG measurements.

The observed anomaly can be attributed to several interacting biophysical and radiometric factors inherent to oil palm canopy structure. Unlike short, uniform crops, oil palm crowns exhibit complex vertical layering, overlapping fronds, and substantial internal shadowing. Mild and moderate BSR stages (T1 and T2) often present subtle structural changes such as frond drooping, spear leaf suppression, and partial crown thinning rather than distinct color changes. These structural symptoms increase the proportion of shaded and mixed pixels within the extracted crown region. Because GLI and ExG are highly sensitive to illumination geometry, shadow intensity, and background reflectance (e.g., soil, trunk, and understory vegetation), the increased shadow fraction artificially suppresses the green reflectance component, contributing to the inverted and inconsistent trends observed across severity categories [34]. Collectively, these factors explain why the GLI and ExG patterns appear weak, irregular, and non-monotonic, despite showing downward direction with increasing disease severity. Their sensitivity to canopy structure and radiometric variation limits their reliability as standalone indicators of BSR severity further reinforcing the need for integrating thermal, structural, and machine-learning-derived features to achieve stable multi-class discrimination.

Furthermore, RGB sensors lack near-infrared (NIR) bands, which are essential for detecting early biochemical stress associated with chlorophyll breakdown and water loss. As a result, GLI and ExG primarily respond to structural rather than physiological changes, and their values may not correspond linearly to disease progression. Variability in lighting conditions, viewing angles, radiometric calibration, and canopy background contributions can also destabilize these indices, reducing their reliability for stress detection in tall perennial crops. The empirical formulation of GLI and ExG further contributes to their susceptibility to sensor geometry and radiometric inconsistencies, which may distort reflectance measurements and weaken their correlation with disease severity. Despite their anomalous behavior in regression analysis, GLI and ExG were retained in the machine learning model. These results highlight both the strengths and limitations of RGB-only vegetation indices for BSR detection. VARI proved to be the most robust and biologically interpretable index, showing clear and consistent separation across severity levels. Meanwhile, the complex behavior of GLI and ExG underscores the need for cautious interpretation of RGB-based indices in oil palm canopies, where shadowing, canopy geometry, and mixed-pixel effects are prominent. Their instability also suggests that future studies should incorporate stronger spectral features, particularly from multispectral or hyperspectral sensors to more reliably capture early physiological changes. Nonetheless, when combined with thermal and structural features within a machine learning framework, the RGB indices still provided valuable complementary information, supporting a cost-effective and operational approach to field-scale BSR monitoring.

4.2. Inferential statistics

The one-way ANOVA results (Table 4) showed that all measured parameters - crown area, frond number, frond angle, thermal pixel values, and the three RGB vegetation indices (ExG, GLI, VARI) vary significantly across the four BSR severity categories (T0 - T3), as indicated by their high F-statistics and p-values <0.0001. This demonstrated that each parameter captures meaningful biological differences associated with disease progression. Among these, frond angle exhibits the strongest discriminative power (F = 239.10), followed by crown area (F = 114.23) and frond number (F = 53.12). These three structural features consistently show large between-group differences, confirming that canopy architecture is highly sensitive to BSR-induced deterioration. Thermal pixel values also show significant differentiation (F = 11.39), reflecting the expected canopy heating associated with physiological stress. The vegetation indices demonstrate moderate but statistically significant differences. ExG (F = 16.38), GLI (F = 18.85), and VARI (F = 8.46) all show detectable variation across BSR-severity categories, although their patterns are more complex than those of structural features. The significant F-values indicate that, despite non-linear or inconsistent trends in raw index values, the indices still capture some degree of class separation when analyzed across the full dataset.

The Tukey–Kramer HSD analysis provides deeper insight into how these parameters separate the severity categories (Table 5). Crown area shows the strongest and cleanest separation, with all four severity levels belonging to distinct groups. This confirms progressive crown shrinkage as one of the apparent indicators of BSR severity. Frond number shows the expected decline: T0 and T1 belong to the same group, indicating that mild infection has minimal effect on frond count, whereas T2 and T3 form separate groups, reflecting increased defoliation as the disease advances. Frond angle separates T0 from T3 clearly, with T1 and T2 forming overlapping intermediate groups, suggesting that angular deformation of fronds begins early but intensifies substantially at higher severity. Thermal pixel values follow a similar trend, T0 is distinct, T1 and T2 overlap, and T3 forms its own group, indicating that temperature differences are detectable but become most pronounced only at advanced infection levels. For RGB vegetation indices, the grouping patterns reflect the complexities of visible-band reflectance in oil palm canopies. ExG showed T0, T1 and T2 in a group, while T2 and T3 cluster in different groups, suggesting a non-linear response influenced by shadowing, illumination variability, and canopy structure. GLI, despite significant ANOVA results, separates into overlapping groups. This indicated that while some statistical differences exist, GLI does not monotonically track severity. VARI performs more clearly, with T3 forming its own group, while T0–T2 remain in different group, demonstrating its robustness for detecting severe crown deterioration. The ANOVA and Tukey–Kramer results collectively indicate that structural parameters (crown area, frond number, frond angle) provide the strongest, most consistent separation across disease categories, making them reliable indicators of BSR progression. The results justify the inclusion of all features in the machine learning model, while acknowledging that structural variables remain the most biologically robust predictors.

Table 4
ANOVA of mean parameters in different BSR severity categories.

Parameters	F Ratio	p-value
Crown area	114.23	<0.0001[a]
Frond number	53.12	
Frond angle	239.10	
Thermal pixel	11.39	
EXG	16.38	
GLI	18.85	
VARI	8.46	

[a] Significant at the 5 % level.

Table 5
Homogeneous grouping of treatments (tukey-kramer HSD test).

Parameter	T0	T1	T2	T3
Crown Area	A	B	C	D
Frond Number	A	A	B	C
Frond Angle	A	A, B	B	C
Thermal Pixel	A	A, B	B	C
EXG	A	A	A, B	A
GLI	A	A	A, B	B
VARI	A	A	A	B

Note: Treatments sharing the same letter are not significantly different at p < 0.05.

4.3. Data expansion using SMOTE

After applying SMOTE, the dataset was expanded to a similar number of healthy level (T0), 1097 samples, with approximately equal representation across all four severity levels. SMOTE enabled the model to learn more stable and discriminative boundaries for all severity levels. While synthetic data cannot fully capture the biological variability of real BSR symptoms, especially with very few original T3 samples, The use of SMOTE substantially strengthened model reliability and reduced prediction bias. The total samples for testing data in machine learning became 329 for T0, 328 for T1, 330 for T2 and 329 for T3.

4.4. Classification model development

A total of 30 classification models were developed using PCA-enabled features, and their performances were evaluated based on both validation and testing accuracy (Table 6). Among all the models

Table 6
Classification results of accuracy test using PCA.

Classification Models		Accuracy	
		Validation	Testing
Decision Trees	Fine Tree	90.00	90.60
	Medium Tree	84.30	83.50
	Coarse Tree	73.30	75.30
Discriminant Analysis	Linear Discriminant	87.00	77.40
	Quadratic Discriminant	88.17	84.20
Naïve Bayes	Gaussian Naïve Bayes	88.76	77.30
	Kernel Naïve Bayes	88.56	79.70
Support Vector Machine (SVM)	Linear SVM	88.56	80.70
	Quadratic SVM	86.90	91.30
	Cubic SVM	88.66	94.80
	Fine Gaussian SVM	88.56	91.60
	Medium Gaussian SVM	85.92	93.00
	Coarse Gaussian SVM	88.95	80.70
K-Nearest Neighbour (KNN)	Fine KNN	93.80	95.30
	Medium KNN	91.30	91.90
	Coarse KNN	84.20	84.70
	Cosine KNN	90.20	90.00
	Cubic KNN	90.70	90.90
	Weighted KNN	93.10	93.90
Ensemble	Ensemble Boosted Trees	86.00	85.60
	Ensemble Bagged Trees	**94.20***	**94.20***
	Ensemble Subspace Discriminant	76.00	76.10
	Ensemble Subspace KNN	71.40	70.40
	Ensemble RUSBoosted Trees	84.30	83.50
Neural Network	Narrow Neural Network	90.10	90.30
	Medium Neural Network	92.10	94.00
	Wide Neural Network	94.00	96.70
	Bilayered Neural Network	91.10	92.20
	Trilayered Neural Network	89.90	91.30
Logistic Regression	SVM Kernel	90.30	92.00
	Logistic Regression Kernel	88.60	89.30

assessed, the Ensemble Bagged Trees model achieved the highest and most balanced performance, recording 94.20 % accuracy for both validation and testing, making it the best classifier in this study. The Neural Network models also showed strong performance, with the Wide Neural Network achieving 94.00 % validation accuracy and the highest testing accuracy of 96.70 %, followed closely by the Medium Neural Network (92.10 % validation, 94.00 % testing). These results highlight the suitability of wider and deeper neural architectures for capturing complex nonlinear interactions in RGB, thermal, and structural palm features derived from UAV imagery. However, despite their strong testing performance, the Neural Networks showed slightly less consistency between validation and testing compared to the superior stability of the Ensemble Bagged Trees model.

Several Support Vector Machine (SVM) models performed competitively. The Cubic SVM achieved a high testing accuracy of 94.80 %, while the Medium Gaussian SVM and Quadratic SVM followed closely with testing accuracies of 93.00 % and 91.30 %, respectively. These models demonstrate strong capability in modelling nonlinear boundaries, although performance varied depending on kernel type. K-Nearest Neighbour (KNN) models also performed well, with Fine KNN recording 95.30 % testing accuracy, Weighted KNN achieving 93.90 %, and Medium KNN reaching 91.90 %. However, other KNN variants such as Coarse KNN and Cosine KNN showed moderate performance (84–90 %), reflecting sensitivity to neighbour count and distance metrics.

Traditional models such as Naïve Bayes and Discriminant Analysis produced moderate yet stable performance. Kernel Naïve Bayes achieved 88.56 % validation and 79.70 % testing accuracy, while Quadratic Discriminant Analysis reached 88.17 % validation and 84.20 % testing accuracy indicating that PCA effectively improved feature separability for these parametric models. Logistic Regression and SVM-Kernel Logistic Regression produced consistent but comparatively lower results (testing accuracy 89.30–92.00 %), indicating limitations in capturing the complex, nonlinear relationships present in the multisensor UAV dataset.

The ensemble models exhibited varied performance. Ensemble Boosted Trees achieved 86.00 % validation and 85.60 % testing accuracy, while Ensemble RUSBoosted Trees produced 84.30 % validation and 83.50 % testing accuracy. Ensemble Subspace Discriminant and Ensemble Subspace KNN displayed lower performance (70–76 %). In contrast, Ensemble Bagged Trees outperformed all ensemble and non-ensemble learners, achieving the highest combination of validation and testing accuracies at 94.20 % each. The high and consistent percentage can be attributed to the bagging mechanism, where multiple decision trees are trained on different bootstrap samples and their predictions averaged. This approach reduces overfitting, improves generalization, and captures diverse nonlinear patterns within the PCA-transformed UAV features [35].

The stability of Ensemble Bagged Trees is further enhanced by the use of SMOTE, which balanced the dataset by generating synthetic minority samples for underrepresented classes. Bagging benefits significantly from balanced data, as each tree receives an equitable representation of all classes, enabling the model to learn clearer decision boundaries. Although SMOTE can introduce noise, the averaging effect of bagging naturally reduces overfitting to synthetic samples, resulting in highly robust and reliable performance. While several Neural Network, SVM, and KNN models achieved competitive testing accuracies, Ensemble Bagged Trees remains the best-performing and most stable classifier, providing consistent accuracy across both validation and testing datasets and demonstrating strong suitability for operational UAV-based BSR disease severity classification [36].

4.5. Confusion matrix

The confusion matrix generated using the SMOTE-balanced dataset provides clear evidence of the Ensemble Bagged Trees model's strong classification capability across all four disease severity levels (T0–T3). As shown in Table 7, the diagonal elements of the matrix dominate across all classes, reflecting high correct classification rates and demonstrating that the model successfully learned the distinguishing features associated with each disease level. For T0 (Healthy) palms, 305 out of 329 samples were correctly classified, indicating that the model can reliably differentiate healthy crowns from diseased oil palm trees based on structural signs and vegetation indices extracted from UAV imagery. Misclassifications mainly occurred as T1 (19 samples), which is expected given that early symptoms often manifest subtly and may visually overlap with healthy conditions in top-view aerial imagery. Nonetheless, the misclassification rate for T0 remains low, reaffirming the model's capability to identify non-infected palms accurately. For T1 (Mild infection), 305 of 328 samples were correctly classified. The main misclassifications for T1 were into adjacent classes, particularly T0 (9 samples) and T2 (12 samples) which aligns with the natural progression of symptom severity and the subtle visual similarity shared between mild and moderate infections. Importantly, the confusion matrix reveals that the model rarely misclassified mild infections as severe (only 3 samples), demonstrating good separation between early and advanced disease stages. The T2 (Moderate infection) class also exhibited strong classification stability, with 303 out of 330 samples correctly predicted. Misclassifications predominantly occurred between T1 (25 samples) and T3 (1 sample), which is consistent with the gradual transition of symptom severity in BSR progression.

The T3 (Severe infection) class achieved the strongest performance with 328 out of 329 samples correctly classified, demonstrating near-perfect recognition accuracy. This result is expected as severe BSR infection presents distinct and easily identifiable canopy symptoms such as drastic crown thinning, severe discoloration, and collapsed fronds that are readily captured by UAV imaging. The confusion matrix confirms that after applying SMOTE, the Ensemble Bagged Trees model achieved high and balanced classification performance across all classes. Misclassifications follow biologically reasonable patterns primarily occurring between adjacent disease severity levels indicating that the model learned the continuous nature of the BSR progression rather than producing random or inconsistent errors. These findings underscore the effectiveness of combining SMOTE with Ensemble Bagged Trees for reliable multi-class disease severity detection in oil palm plantations.

4.6. Performance metrics

The per-level performance metrics obtained from the SMOTE-balanced dataset demonstrate that the Ensemble Bagged Trees model can accurately distinguish among all four BSR severity levels (T0–T3). Table 8 presents precision, recall (sensitivity), specificity, and F1-scores for each class, showing consistently strong discrimination across the severity spectrum. For T0 (Healthy) palms, the model achieved a precision of 96.5 %, a recall of 92.7 %, and a high specificity of 98.88 %, indicating that the classifier not only identified healthy palms reliably but also avoided misclassifying diseased palms as healthy. The resulting F1-score of 94.6 % reflects a strong balance between identifying true healthy trees and minimizing false positives, an important requirement for maintaining confidence in plantation baseline health. For T1 (Mild infection), the model recorded a precision of 87.4 %, a recall of 92.7 %, and a specificity of 95.54 %, producing an F1-score of 89.9 %. These values indicate high sensitivity to early-stage symptoms, ensuring that

Table 7
Results of confusion matrix using Ensemble Bagged Trees model True class.

	T0	T1	T2	T3
T0	305	19	4	1
T1	9	304	12	3
T2	1	25	303	1
T3	1	0	0	328

Predicted class.

Table 8
Per-level classification performance.

Level	Metrics			
	Precision	Recall	Specificity	F1 Score
T0	96.5 %	92.7 %	98.88 %	94.6 %
T1	87.4 %	92.7 %	95.54 %	89.9 %
T2	95.0 %	91.8 %	98.38 %	93.4 %
T3	98.5 %	99.7 %	99.49 %	99.1 %

early infections are detected with minimal missed cases. Even though precision is slightly lower than for other classes, the combination of high recall and high specificity confirms that the model can detect early symptoms while maintaining a low rate of false-positive predictions, which is critical for preventing disease escalation.

The T2 (Moderate infection) class achieved a precision of 95.0 %, a recall of 91.8 %, and a specificity of 98.38 %, resulting in an F1-score of 93.4 %. These strong values highlight the model's ability to differentiate moderate infections, which are often visually similar to both early and severe stages. The high specificity shows that the model rarely confuses other classes with T2, while the high recall demonstrates that most moderately infected trees were successfully detected. This is a significant advancement, as moderate infection detection enables timely and targeted management before the disease reaches severe stages. The T3 (Severe infection) category exhibited the highest performance, with precision at 98.5 %, recall at 99.7 %, and specificity at 99.49 %, producing an F1-score of 99.1 %. This near-perfect performance indicates that severe infections produce clear visual and structural cues that the model can detect with exceptional confidence. The extremely high recall means virtually all severely infected palms were correctly identified, while the very high specificity confirms minimal false alarms, both essential for preventing the spread of advanced infection and ensuring accurate field-level decision support.

Furthermore, the performance metrics derived from the SMOTE-balanced dataset demonstrate that the Ensemble Bagged Trees model delivers consistently high performance across all major evaluation indicators, confirming its robustness and reliability for multi-class disease classification in oil palm. As shown in Table 9, the model achieved an overall accuracy of 94.18 %, indicating that more than 94 % of all samples across the four disease severity levels (T0–T3) were correctly classified. The Weighted F1-score of 94.21 % underscores strong performance across all classes while accounting for their relative frequencies in the test set. This metric is particularly important in multi-class classification, as it reflects how well the model performs when both majority and minority class contributions are proportionately represented. The Macro F1-score of 94.24 % also demonstrates that the model performs uniformly well across all classes, regardless of class size.

The Micro F1-score, which aggregates true positives, false negatives, and false positives across all classes to provide measurement of predictive performance, was recorded at 94.22 %, closely matching the overall accuracy. This close alignment indicates a strong consistency between global performance (all predictions combined) and standard accuracy-based assessment. This consistency suggests that the model is not disproportionately influenced by any type of error, such as systematic false negatives or false positives. Another important metric for multi-class evaluation is Cohen's Kappa, which was recorded at 92.30 %.

Table 9
Results of performance metrics using Ensemble Bagged Trees model.

Type of Model	Overall Accuracy (%)	Weighted F1	Macro F1- (%)	Cohen's Kappa (%)	Micro F1 (%)
Ensemble Bagged Trees	94.18	94.21	94.24	92.30	94.22

This value indicated strong agreement between the model's predictions and the ground truth labels after correcting for chance agreement. Kappa score also demonstrated that the performance gains observed are meaningful and not simply a result of high accuracy due to class distribution. Together, these performance metrics showed that the Ensemble Bagged Trees model not only achieves high classification accuracy but also maintains balanced performance across all severity categories. The combination of high F1-scores, strong agreement (Kappa), and stable micro/macro-level performance suggests that the model successfully captures the complexity of RGB indices, thermal spectral and structural canopy features derived from UAV imagery. This robustness makes it suitable for operational deployment in plantation health monitoring systems, where timely and accurate detection of disease severity is critical. Despite evaluating multiple classification models, the Ensemble Bagged Trees algorithm emerged as the most effective, not only in terms of accuracy but also in its suitability for real-world agricultural applications [37]. Its ensemble structure, which combines the predictions of multiple decision trees, improves generalizability by reducing overfitting and enhancing robustness to data variability critical factors in dynamic plantation environments. Furthermore, compared to complex black-box models like deep neural networks, Bagged Trees offer greater interpretability. The ability to trace classification decisions back to specific input features, such as crown area or RGB vegetation indices, provides valuable insights for agronomists and field operators. This interpretability facilitates trust in model outputs and supports integration into plantation decision-making workflows, where transparent and explainable tools are essential for early intervention and disease management.

In addition, the ROC analysis for Model Ensemble Bagged Trees demonstrated good discriminative performance across all four disease severity classes (T0–T3). As shown in Fig. 12a, all class-specific ROC curves are above the diagonal reference line, indicating strong sensitivity-specificity trade-offs. The curves rise sharply towards the upper-left corner with minimal false-positive rate, reflecting that the classifier is able to correctly identify Healthy, Mild, Moderate, and Severe samples with high reliability. The presence of tightly clustered ROC trajectories near the top of the plot suggests that the model maintains consistent performance across classes, with no level exhibiting substantial degradation in predictive ability. The large separation from the chance line and the near-vertical ascent for several classes indicate that the model achieves high true-positive rates even at low false-positive thresholds. These observations are consistent with the high overall accuracy, Macro F1 score, and Cohen's Kappa previously reported, confirming that Model Ensemble Bagged Trees possesses multi-class discriminative ability suitable for UAV-based oil palm disease classification.

The Precision–Recall (PR) curves for Model Ensemble Bagged Trees (Fig. 12b) further confirm the strong discriminative capability of the classifier across all four disease severity classes (T0–T3). All class-specific curves maintain high precision values across a broad range of recall levels, indicating that the model is able to accurately identify true disease cases with very few false positives. The curves show a steep decline only at the very end of the recall axis, which is typical of high-performing multi-class classifiers when approaching maximum sensitivity. The dense clustering of the curves near the upper-right corner with precision values consistently above 0.90 for most recall thresholds reflected a balanced performance between positive predictive value and true positive detection. This behavior suggests that the model is highly reliable even under varying decision thresholds, with minimal trade-offs between sensitivity and precision. These findings are highly aligned with the high percentages of performance metrics reported earlier, reinforcing the robustness of Model Ensemble Bagged Trees for multi-class oil palm disease classification using UAV-based imaging features. Altogether, the ROC and PR analyses revealed that Ensemble Bagged Trees Model not only distinguishes classes with high sensitivity and specificity but also sustains strong predictive reliability, making it

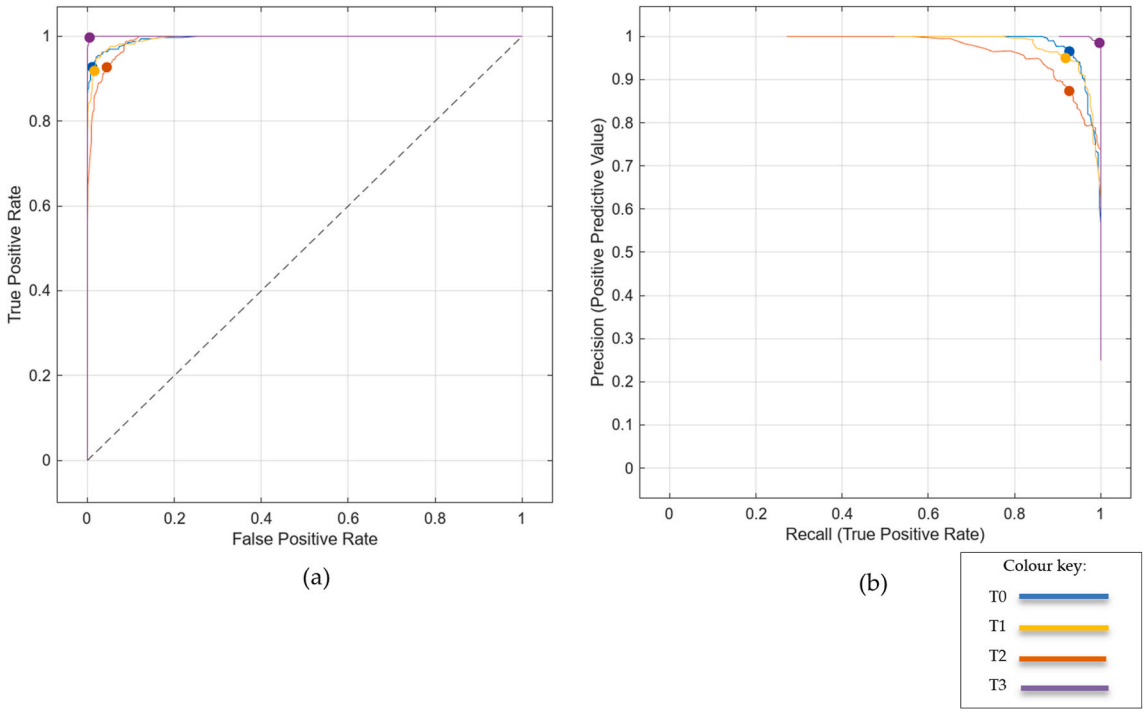

Fig. 12. a) ROC curves and b) Precision–Recall curves.

highly suitable for UAV-based oil palm disease detection.

5. Conclusion

This work demonstrates that pairing ensemble learning with multi-sensor UAV observations enables accurate multi-class assessment of BSR severity in oil palm. The inclusion of SMOTE improved class representation, allowing the Ensemble Bagged Trees model to learn clearer boundaries between severity levels and to perform consistently across all evaluation metrics. The high sensitivity and specificity obtained for each class confirm the model's suitability for operational BSR disease surveillance. Although RGB indices such as GLI and ExG showed unstable correlations with BSR severity due to radiometric and canopy-structural effects, their integration with VARI, thermal features, and structural attributes enhanced overall classification performance. The study underscores the limitations of RGB-only metrics and points toward the value of incorporating multispectral or hyperspectral sensing in future work. In summary, the methodology provides a practical and field framework for aerial BSR monitoring that can support timely intervention and data-driven plantation management.

CRediT authorship contribution statement

N.A. Husin: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **N.A.H.M. Baktiar:** Writing – original draft, Methodology, Formal analysis, Data curation. **V.U. Tagang:** Writing – original draft, Formal analysis, Data curation. **S. Khairunniza–Bejo:** Project administration, Funding acquisition. **M.F. M. Yusuf:** Resources.

Funding

This research was funded by Universiti Putra Malaysia (UPM) under research project code GP-IPM/2021/9697200.

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.

Data availability

Data will be made available on request.

References

[1] Our World Data, Forests and deforestation. https://ourworldindata.org/forests-and-deforestation, 2021. (Accessed 12 December 2022).
[2] Malaysian Palm Oil Board (MPOB), Malaysian Oil Palm Statistics 2021, 41st ed., Bangi: MPOB, 2022.
[3] V. Lai, N.Y.M. Yusoff, A.N. Ahmed, Y.F. Huang, K.B.W. Boo, A. El-Shafie, The benefits and perspectives of the palm oil industry in Malaysia, Environ. Dev. Sustain. (2024) 1–15, https://doi.org/10.1007/s10668-024-03897-1.
[4] D.J. Murphy, K. Goggin, R.R.M. Paterson, Oil palm in the 2020s and beyond: challenges and solutions, CABI agriculture and bioscience 2 (2021) 1–22, https://doi.org/10.1186/s43170-021-00058-3.
[5] N.H. Azuan, S. Khairunniza-Bejo, A.F. Abdullah, M.S.M. Kassim, D. Ahmad, Analysis of changes in oil palm canopy architecture from basal stem rot using terrestrial laser scanner, Plant Dis. 103 (12) (2019) 3218–3225, https://doi.org/10.1094/PDIS-10-18-1721-RE.
[6] C.S. Kwang, S.F.A. Razak, S. Yogarayan, M.A. Zahisham, T.H. Tam, M.A.M. Noor, H. Abidin, *Ganoderma* disease in oil palm trees using hyperspectral imaging and machine learning, J. Hum. Earth Future 6 (1) (2025), https://doi.org/10.28991/HEF-2025-06-01-05.
[7] J. Kurihara, V.C. Koo, C.W. Guey, Y.P. Lee, H. Abidin, Early detection of basal stem rot disease in oil palm tree using unmanned aerial vehicle-based hyperspectral imaging, Remote Sens. 14 (3) (2022) 799, https://doi.org/10.3390/rs14030799.
[8] Y.K. Chan, V.C. Koo, M.Z.A. Zahisham, K.M. Lim, T. Connie, C.S. Lim, H. Abidin, Design and development of a drone based hyperspectral imaging system, in: IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2022, July, pp. 4200–4203, https://doi.org/10.1109/IGARSS46834.2022.9884820.
[9] P. Ahmadi, F.M. Muharam, K. Ahmad, S. Mansor, I.A. Seman, Early detection of *ganoderma* basal stem rot of oil palms using artificial neural network spectral analysis, Plant Dis. 101 (6) (2017) 1009–1016, https://doi.org/10.1094/PDIS-12-16-1699-RE.
[10] C.M. Rafezall, N. Darwin, M.F.M. Ariff, Z. Majid, Detection of palm oil health through multispectral UAV platform, in: 2020 IEEE 10th International Conference

on System Engineering and Technology (ICSET), IEEE, 2020, November, pp. 240–244, https://doi.org/10.1109/ICSET51301.2020.9274721.
[11] M.A. Izzuddin, A. Hamzah, M.N. Nisfariza, A.S. Idris, Analysis of multi spectral imagery from unmanned aerial vehicle (UAV) using object-based image analysis for detection of *ganoderma* disease in oil palm, Journal of Oil Palm Research 32 (3) (2020) 497–508, https://doi.org/10.21894/jopr.2020.0035.
[12] M.A. Kurniawan, J.P. Trinugroho, T. Suparyanto, M. Isnan, D. Sudigyo, B. Pardamean, Detecting *ganoderma* basal stem rot disease on oil palm using artificial neural network method, Commun. Math. Biol. Neurosci. 2023 (2023) 7911, https://doi.org/10.28919/cmbn/7911.
[13] D. Wiratmoko, R.H. Jatmiko, M.A. Yusuf, R. Farrasati, A.E. Prasetyo, Using visible spectral-index as alternative methods for identifying levels of *Ganoderma* Boninese infection, in: IOP Conference Series: Earth and Environmental Science, vol 500, IOP Publishing, 2020, June 012067, https://doi.org/10.1088/1755-1315/500/1/012067, 1.
[14] P. Mimboro, H. Soeparno, B. Soewito, W. Budiharto, Prediction of oil palm conditions using deep learning based on the visible atmospherically resistant index on UAV imagery, in: 2023 International Conference on Informatics Engineering, Science & Technology (INCITEST), IEEE, 2023, October, pp. 1–7, https://doi.org/10.1109/INCITEST57829.2023.10123456.
[15] I.C. Hashim, A.R.M. Shariff, S.K. Bejo, F.M. Muharam, K. Ahmad, Classification of non-infected and infected with basal stem rot disease using thermal images and imbalanced data approach, Agronomy 11 (12) (2021) 2373, https://doi.org/10.3390/agronomy11122373.
[16] S. Bejo, G. Abdol-Lajis, S. Abd-Aziz, I. Abu-Seman, T. Ahamed, Detecting basal stem rot (BSR) disease at oil palm tree using thermal imaging technique. Proceedings of the 14th International Conference on Precision Agriculture, International Society of Precision Agriculture: Monticello, Illinois, U.S.A., 2018.
[17] S.N. Johari, S.K. Bejo, G.A. Lajis, L.D. DaimDai, N.B. Keat, Y.Y. Ci, N. Ithnin, Detecting BSR-Infected oil palm seedlings using thermal imaging technique, Basrah Journal of Agricultural Sciences 34 (2021) 73–80, https://doi.org/10.37077/25200860.2021.34.sp1.8.
[18] Y.H. Haw, Z. Zhao, Y.C. Hum, J.H. Chuah, W. Voon, S. Khairunniza-Bejo, K.W. Lai, Detection of *Ganoderma* boninense diseases of palm oil trees using machine learning, in: 2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE, 2023, pp. 228–232, https://doi.org/10.1109/ISCAIE57739.2023.10165368.
[19] N.A. Husin, S. Khairunniza-Bejo, A.F. Abdullah, M.S. Kassim, D. Ahmad, M.H. Aziz, Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning, Agronomy 10 (11) (2020) 1624, https://doi.org/10.3390/agronomy10111624.
[20] J. Zheng, W. Li, M. Xia, R. Dong, H. Fu, S. Yuan, Large-scale oil palm tree detection from high-resolution remote sensing images using Faster-RCNN. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 1422–1425, https://doi.org/10.1109/IGARSS.2019.8898360. Yokohama, Japan.
[21] J. Zheng, H. Fu, W. Li, W. Wu, L. Yu, S. Yuan, K.D. Kanniah, Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images, ISPRS J. Photogrammetry Remote Sens. 173 (2021) 95–121, https://doi.org/10.1016/j.isprsjprs.2021.01.008.
[22] I.C. Hashim, A.R.M. Shariff, S.K. Bejo, F.M. Muharam, K. Ahmad, Basal stem rot disease classification by machine learning using thermal images and an imbalanced data approach, in: Iot and AI in Agriculture: Self-Sufficiency in Food Production to Achieve Society 5.0 and Sdg's Globally, Springer Nature Singapore, Singapore, 2023, pp. 395–422, https://doi.org/10.1007/978-981-19-8113-5_20.
[23] M.A. Kurniawan, J.P. Trinugroho, T. Suparyanto, M. Isnan, D. Sudigyo, B. Pardamean, Detecting ganoderma basal stem rot disease on oil palm using artificial neural network method, Commun. Math. Biol. Neurosci. 2023 (2023) 40, https://doi.org/10.28919/cmbn/8433.
[24] S.Y. Yang, L.S. Ling, O.S. Yin, Oil palm diseases detection using computer vision techniques, in: American Institute of Physics Conference Series, vol 3153, AIP Publishing LLC, 2024 070002, https://doi.org/10.1063/5.0216581, 1.
[25] A. Panthakkan, S.M. Anzar, K. Sherin, S.A. Mansoori, H. Al-Ahmad, Evaluation of UAV-based RGB and multispectral vegetation indices for precision agriculture in palm tree cultivation, arXiv preprint arXiv:2505.07840 (2025), https://doi.org/10.48550/arXiv.2505.07840.
[26] S. Khairunniza-Bejo, C.N. Vong, Detection of basal stem rot (BSR) infected oil palm tree using laser scanning data, Agriculture and Agricultural Science Procedia 2 (2014) 156–164.
[27] G. Ciaburro, Matlab for Machine Learning, Packt Publishing Ltd, 2017, https://doi.org/10.1016/j.aaspro.2014.11.023.
[28] M. Gholami, M. Ranjbargol, R. Yousefzadeh, Z. Ghorbani, Integrating three smart predictive models using a power-law committee machine for the prediction of compressive strength in masonry made of clay bricks and cement mortar, in: Structures, vol. 55, Elsevier, 2023, September, pp. 951–964, https://doi.org/10.1016/j.istruc.2023.06.012.
[29] M.S.A. Baharim, N.A. Adnan, M.I. Anuar, A.L. Laurence, A.A. Samat, Relationship analysis between *ganoderma* boninense-derived basal stem rot disease severity with multiple leaf physiology parameters in mature oil palm tree responses toward water use efficiency (WUE), J. Plant Pathol. 106 (4) (2024) 1801–1816, https://doi.org/10.1007/s42161-024-01762-5.
[30] N.A. Rebitanim, M.M. Hanafi, A.S. Idris, S.N.A. Abdullah, H. Mohidin, N. Z. Rebitanim, GanoCare® improves oil palm growth and resistance against *ganoderma* basal stem rot disease in nursery and field trials, BioMed Res. Int. 2020 (1) (2020) 3063710, https://doi.org/10.1155/2020/3063710.
[31] M.S.A. Baharim, N.A. Adnan, M.A. Izzuddin, A.L. Laurence, M.K. Karsimen, H. Arof, Modelling water use efficiency (WUE) for estimating the severity of *Ganoderma* boninense-derived basal stem rot disease in oil palm, J. Plant Pathol. (2024) 1–14, https://doi.org/10.1007/s42161-024-01770-5.
[32] S. Supramani, N.A. Rejab, Z. Ilham, W.A.A.Q.I. Wan-Mohtar, S. Ghosh, Basal stem rot of oil palm incited by *Ganoderma* species: a review, Eur. J. Plant Pathol. 164 (1) (2022) 1–20, https://doi.org/10.1007/s10658-022-02546-2.
[33] N.A. Husin, S. Khairunniza–Bejo, A.F. Abdulllah, M.S.M. Kassim, D. Ahmad, Effects of basal stem rot on oil palm inter-frond angles for different severity levels, Journal of Advanced Agricultural Technologies 6 (2) (2019), https://doi.org/10.18178/joaat.6.2.113-117.
[34] L. Gao, X. Wang, B.A. Johnson, Q. Tian, Y. Wang, J. Verrelst, X. Gu, Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: a review, ISPRS J. Photogrammetry Remote Sens. 159 (2020) 364–377, https://doi.org/10.1016/j.isprsjprs.2019.11.019.
[35] X. Yin, Q. Liu, Y. Pan, X. Huang, J. Wu, X. Wang, Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: Comparison of eight single and ensemble models, Natural Resources Research 30 (2021) 1795–1815, https://doi.org/10.1007/s11053-020-09787-0.
[36] S.B. Ramezani, L. Cummins, B. Killen, R. Carley, A. Amirlatifi, S. Rahimi, L. Bian, Scalability, explainability and performance of data-driven algorithms in predicting the remaining useful life: a comprehensive review, IEEE Access 11 (2023) 41741–41769, https://doi.org/10.1109/ACCESS.2023.3267960.
[37] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: A review of machine learning interpretability methods, Entropy 23 (1) (2020) 18. https://doi.org/10.3390/e23010018.