# Classification of Northern Thai Rice Varieties Using Random Forest (RF) and Support Vector Machine (SVM) on Google Earth Engine with Sentinel Imagery: A Case Study in Buak Khang Subdistrict, San Kamphaeng District, Chiang Mai Province

**Boonma, R.,[1] Suwanprasit, C.,[1*] Homhuan, S.[1] and Shahnawaz[2]**
[1]Department of Geography, Faculty of Social Sciences, Chiang Mai University 50200, Thailand
E-mail: raweewan_b@cmu.ac.th, chanida.suwanprasit@cmu.ac.th,* sakda.homhuan@cmu.ac.th
[2]Department of Geoinformatics–Z_GIS, University of Salzburg, 5020 Salzburg, Austria
E-mail: s.shahnawaz@sbg.ac.at
**Corresponding Author*

**DOI: https://doi.org/10.52939/ijg.v20i9.3539**

**Abstract**
*Rice is a crucial agricultural product for Thailand's economy, as the majority of the country's agricultural sector primarily cultivates rice for both domestic consumption and international demand. This research focuses on land use analysis, specifically on rice cultivation areas, and the classification of different rice varieties. The study significantly contributes to the achievement of the Sustainable Development Goals (SDGs) concerning food security. This research utilizes Sentinel-2 Spectral Instrument, Level 2A, and Sentinel-1 polarization VV and VH satellite images from 2023, covering the planting season (June to November). The data processing and analysis are conducted on the cloud platform Google Earth Engine (GEE), employing Support Vector Machine (SVM) and Random Forest (RF) classification methods for both land use analysis and rice variety classification. The analysis results indicate that the RF classification method has higher accuracy than the SVM method. Specifically, for land use analysis, the RF and SVM classification methods achieved accuracy values of 0.91 and 0.87, and kappa values of 0.89 and 0.85, respectively. For rice variety classification, the RF and SVM methods achieved accuracy values of 0.88 and 0.83, and kappa values of 0.73 and 0.61, respectively*.

**Keywords:** Cloud-based Data Processing, Google Earth Engine (GEE), Random Forest (RF), Rice Varieties Classification, Support Vector Machine (SVM)

## 1. Introduction

Rice is considered an important economic crop in Thailand due to the predominant focus on rice cultivation within the country's agricultural sector. This is driven by the need for both domestic consumption and international market demand. In Chiang Mai province, the total rice growing area spans 890.6384 sq.km., with a significant portion located in San Kamphaeng District. This district benefits from a reliable water supply managed by the Royal Irrigation Department, making agriculture a viable occupation for the local population, primarily in rice farming. Accurate mapping of rice cultivation areas and the size of each species' arable land is crucial for achieving the Sustainable Development Goals (SDGs), particularly the goal of zero hunger, which emphasizes food security. Precise data on rice production is essential to measure food security levels [1] and to aid relevant agencies in policy planning for economic development in the region. However, field data collection on farmland is labor-intensive, time-consuming, and costly. In this context, satellite remote sensing technology offers high-frequency, accurate, cost-effective, and dynamic data for effective mapping, tracking, and management [2]. Satellite remote sensing includes optical sensor data and Synthetic Aperture Radar (SAR) data, providing extensive spatial and temporal information [3]. Nonetheless, processing satellite image data over large areas and multiple periods can be challenging.

Recently, the Google Earth Engine (GEE) has become increasingly utilized alongside remote sensing technology due to its highly efficient network services for extracting spatial information from satellite image data. This platform enables access to information from across the earth's surface, allowing for diverse visual resolutions and rapid large-scale data computation through cloud computing.

Previous studies have primarily focused on classifying rice paddies using the Random Forest (RF) algorithm, due to its superior classification performance compared to other algorithms such as Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Decision Trees (DT), and Logistic Regression (LR) [4]. For instance, the Robust Adaptive Spatial and Temporal Fusion Model (RASTFM) model combines Normalized Difference Vegetation Index (NDVI) vegetation index data from MODIS and Sentinel-2 satellite images over multiple periods to obtain high-resolution and temporally resolved data [5]. Other studies include mapping rice cultivation using Sentinel-1 SAR time series, analyzing temporal behavior and backscatter radar polarization of various rice varieties [6], and developing dense time series Sentinel-1 images using GEE [7]. GEE has also been used with Artificial Neural Networks (ANNs) algorithms and Sentinel-1 and Sentinel-2 images to create objective cultivation maps [8]. Rice cultivation analysis was identified using metrics derived from the Gaussian profile of the VV/VH time series with classification methods such as DT and RF algorithms [9]. However, these studies often have limitations due to small area coverage, such as sub-districts and districts, due to low spatial resolution (250 m). Most classification models used in these studies rely on the RF decision tree algorithm and the ANN algorithm. This study aims to evaluate the effectiveness of various land use classification methods, specifically focusing on classifying rice cultivation areas for different rice varieties using machine learning techniques. Among the most effective and widely utilized methods in current research are the RF and SVM algorithms, known for their proficiency in processing satellite imagery to generate accurate land use maps [10]. This research leverages remote sensing data, incorporating optical sensor data and Synthetic Aperture Radar (SAR) data from Sentinel-1 and Sentinel-2 satellite images, processed on the GEE platform.

## 2. Materials and Methods

### *2.1 Study Area*

The research study focused on Buak Khang subdistrict, San Kamphaeng district, Chiang Mai province, located at coordinates 18°43′51.1″N 99°04′33.4″E and 18°40′06.9″N 99°08′26.8″E. The area covers approximately 27.85 sq.km. and features a vast plain with several major rivers flowing through it, including the Mae On, Mae Puka, and Mae Kuang rivers, which are significant water sources. Land use in this region is divided into five main categories: rice field, built-up area, water area, tree cover area, and other areas. The local population primarily engages in agriculture, such as rice farming, tobacco cultivation, and vegetable gardening. Livestock farming, including dairy farming, is also practiced. This study focuses on land use classification and the classification of areas for growing different rice varieties. The total rice cultivation area in Buak Khang Subdistrict is 17.901 sq.km. Field surveys indicated that most farmers in the large-scale farming group prefer to plant rice varieties such as San-Pah-Tawng1, RD-MAEJO2, and CPRICE 888 (Figure 1).

### *2.2 Datasets*

This study utilized satellite imagery data from Sentinel-2, including the Blue, Green, Red, Near Infrared (NIR), and Short-Wave Infrared (SWIR) bands, with spatial resolutions ranging from 10 to 20 meters. Additionally, Sentinel-1 C-Band imagery from descending orbits, featuring VH and VV polarizations, was employed to enhance classification accuracy [11]. The research was conducted from June to November 2023, aligning with the primary rice-growing season in the study area. During June and July, farmers typically begin planting three rice varieties, with harvesting occurring from October to November. Given the optical nature of Sentinel-2 imagery, cloud cover presented a significant challenge during the rainy season. To address this issue and improve classification efficiency, SAR imagery from Sentinel-1, which operates using radar systems, was incorporated into the analysis. In total, eleven Sentinel-1 images were utilized, covering the entire study period. Additionally, five cloud-filtered Sentinel-2 images, each with less than 30% cloud cover, were selected, as detailed in Table 1.

### *2.3 Rice Varieties*

This study involved comprehensive fieldwork to collect detailed information on rice cultivation areas and practices. The data collected encompassed planting dates, rice field sizes, and the specific rice varieties cultivated. Based on the gathered information, it was observed that in Buak Khang Subdistrict in 2023, the majority of farmers planted rice between June and November.

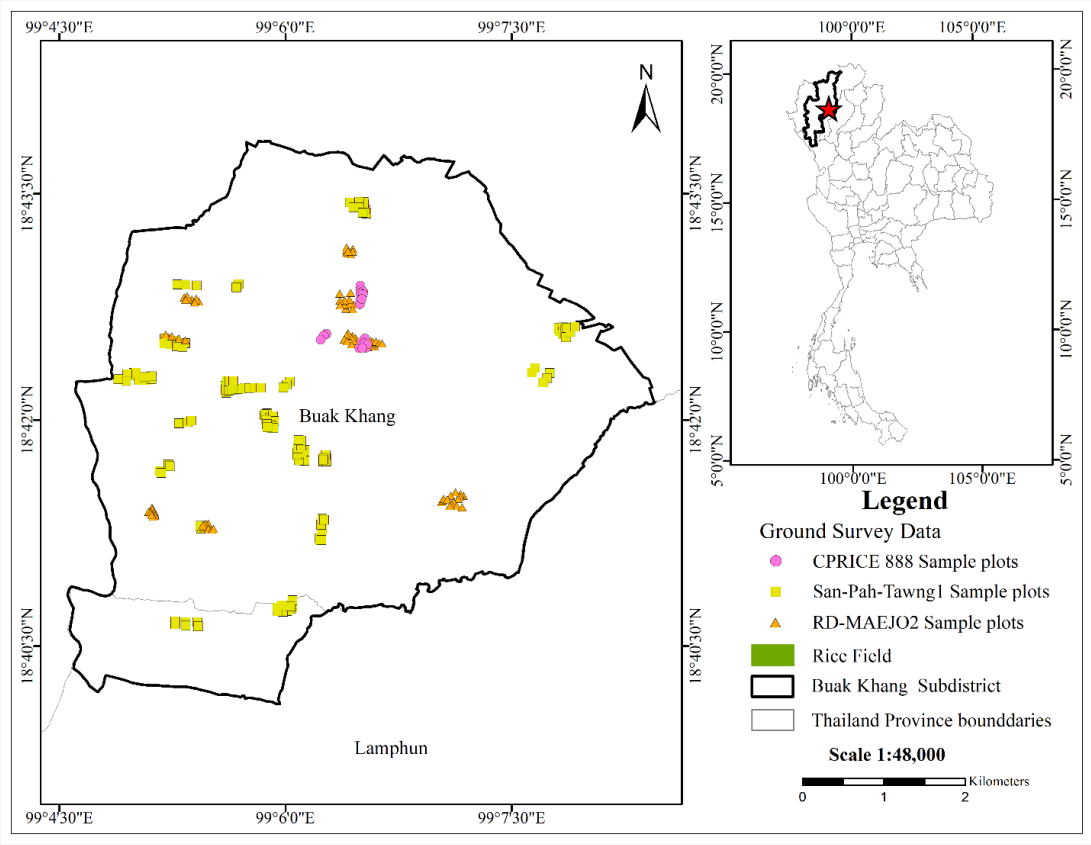



**Figure 1:** Buak Khang Subdistrict, San Kamphaeng District, Chiang Mai Province

**Table 1:** Acquisition dates and number of satellite images used in the study

| Month | Data acquisition dates | |
|---|---|---|
| | Sentinel-1 (Descending) | Sentinel-2 |
| June | 17 | 3 |
| July | 11 | 8 |
| August | 4, 16 | - |
| September | 9, 21 | 21 |
| October | 3, 15, 27 | - |
| November | 8, 20 | 20, 25 |

The three primary rice varieties cultivated were San-Pah-Tawng1, RD-MAEJO2, and CPRICE 888. The distinct characteristics of each of these varieties are detailed as follows: San-Pah-Tawng1 is a photoperiod-insensitive rice variety, making it suitable for year-round cultivation. It has a harvesting period of approximately 130-135 days and yields an average of 40 kg per 100 sq.m, which is 20% higher than the Khao Dawk Mali 10 (KDML10) variety and comparable to the Phrae 1 glutinous rice variety. San-Pah-Tawng1 responds well to nitrogen fertilizer and produces rice that is slightly softer when cooked compared to KDML 10. This variety is resistant to both rice blast disease and bacterial leaf blight, making it particularly suitable for irrigated paddy fields in Northern Thailand, especially in areas frequently affected by these diseases. The physical characteristics of the San-Pah-Tawng1 rice plant include a height of approximately 119 cm, upright culms, green leaves, green leaf sheaths, and an erect flag leaf. The panicles are long and dense with short necks and sturdy straw. Additionally, the leaves age slowly, contributing to the robustness of the plant. This variety's resilience and favorable traits make it an excellent recommendation for farmers in the region [12] and [13]. RD-MAEJO2 is a photoperiod-insensitive sticky rice variety that can be cultivated year-round. It has a harvest period of approximately 135 days in the main growing season and 146 days in the off-season.

This variety exhibits resistance to bacterial leaf blight and brown planthopper. The RD-MAEJO2 rice plant possesses the following physical characteristics: it is a mildly fragrant sticky rice with slender long grains, reaching a height of about 110 cm in the main season and 99 cm in the off-season. The stem is moderately strong, and the leaves are green with an upright leaf tip. The leaves exhibit delayed senescence, with flag leaves measuring 32.16 cm in length and 1.20 cm in width, and an upright flag leaf angle. The panicle length is 29.75 cm [14]. CPRICE 888 is a photoperiod-insensitive sticky rice variety that can be cultivated year-round. This variety is particularly well-suited for irrigated areas in the upper northern region of Thailand. With a harvest period ranging from approximately 110 to 130 days, it offers flexibility and reliability for farmers. The physical characteristics of the CPRICE 888 rice plant are notable. The rice grains are fragrant and sticky, appealing to consumers who prefer aromatic rice. The plant itself reaches a stem height of 91 to 122 cm, which provides a robust structure. The internodes are light yellow, adding to the distinct appearance of the plant. In terms of growth habit, the CPRICE 888 variety forms a spreading clump, which is advantageous for stability and nutrient absorption. The stems are strong, supporting the plant well under various conditions. The flag leaf is upright, measuring approximately 48 cm in length, which aids in efficient photosynthesis [15].

Figure 2 illustrates the growth phases and durations of the three rice varieties—San-Pah-Tawng1, RD-MAEJO2, and CPRICE 888—cultivated in Buak Khang Subdistrict during the 2023 rice-growing season. The timeline at the top spans from June to November, showing the progression of rice growth over 135 days. The growth stages for all three rice varieties—San-Pah-Tawng1, RD-MAEJO2, and CPRICE 888—begin with the vegetative phase, which lasts for 60 days, focusing on the development of leaves and stems. This is followed by the reproductive phase, which also lasts for 90 days, during which the rice plants develop panicles and begin grain formation. The final ripening phase varies slightly among the varieties. Both San-Pah-Tawng1 and RD-MAEJO2 have a ripening phase that lasts 135 days in total, from the start of planting to harvest, while CPRICE 888 has a slightly shorter ripening phase, totaling 130 days from planting to harvest.

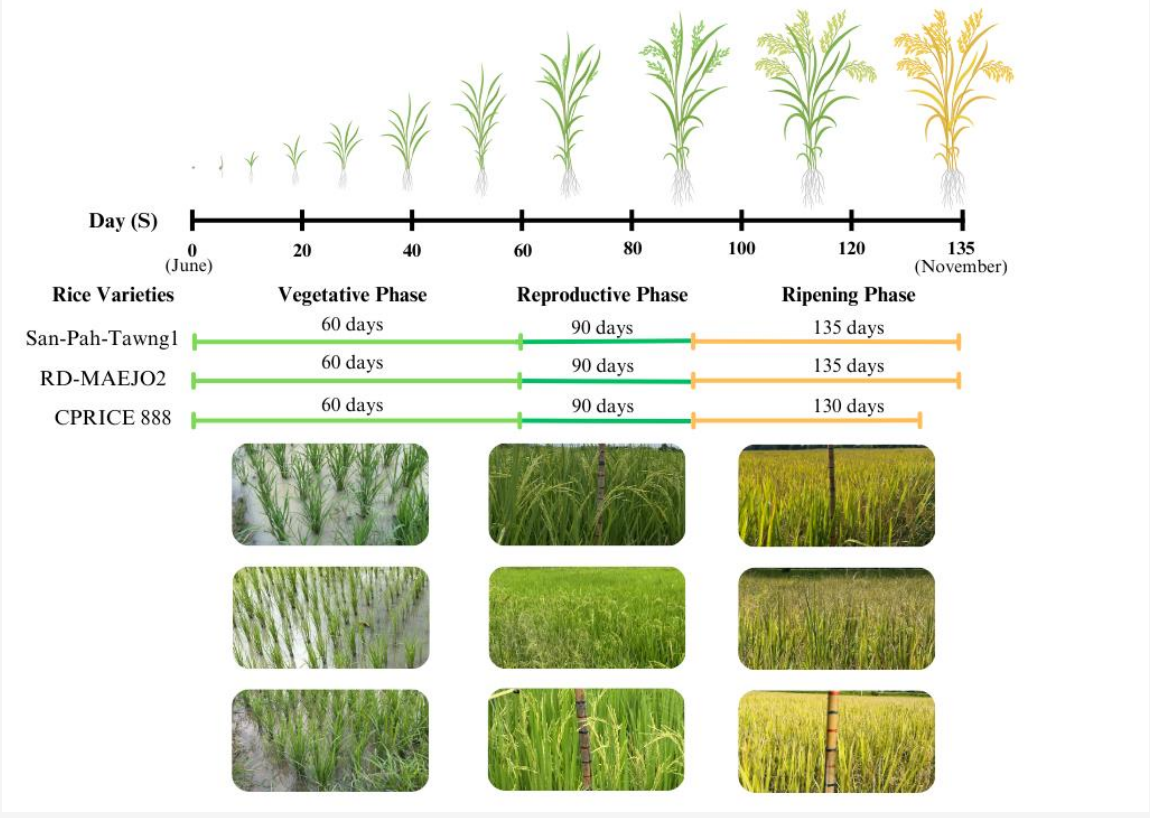



**Figure 2:** Growth phases and duration for rice varieties in Buak Khang Subdistrict (June - November 2023)

*2.4 Machine Learning*

This study employs two machine learning classification methods: Random Forest (RF) and Support Vector Machine (SVM). Both methods are state-of-the-art algorithms for land use classification from satellite imagery [16]. RF and SVM are powerful tools in remote sensing, each with its own strengths. RF is popular for its robustness, ease of use, and ability to handle large datasets with missing values. RF operates by aggregating predictions from multiple decision trees to make a final prediction. This ensemble approach improves the accuracy and robustness of the model. In terms of classification, the final prediction is based on the majority vote of all the trees, as shown in the following equation (Equation 1):

$$\hat{y} = Mode\ \{h1(x), h2(x), \ldots, hB(x)\} \quad \text{Equation 1}$$

Where:

$hB(x)$ = the prediction of the bbb-th tree.

*Mode* = the most frequent class label among the trees' predictions.

On the other hand, SVM performs well in high-dimensional spaces and complex classification tasks but requires careful parameter tuning. SVM is inherently a binary classifier. For multi-class classification, strategies such as One-vs-One (OvO) or One-vs-Rest (OvR) are used. In OvO, each pair of classes has a separate SVM, and the final class prediction is generally determined by majority voting among all the SVMs. In OvR, one SVM is trained for each class against all other classes, and the final class prediction is based on the SVM with the highest confidence score. Therefore, the choice between RF and SVM often depends on the specific requirements of the remote sensing task at hand, including the characteristics of the data and practical outcome needs. It is not uncommon to use both methods complementarity and to compare results to ensure maximum accuracy in remote sensing applications [17]. This study aims to compare the capabilities of RF and SVM in land use classification and rice variety cultivation area classification using both methods.

*2.5 Methodology*

To evaluate the effectiveness of land use classification methods and the classification of rice cultivation areas for different rice varieties, this study utilized satellite imagery data from Sentinel-1 and Sentinel-2. Sentinel-2 MSI Level-2A images, including the RGB, NIR, and SWIR bands, and Sentinel-1 SAR GRD images from descending orbits with VH and VV polarizations, were processed using the GEE platform. The pre-processing of these images involved selecting relevant bands and indices, such as NDVI, NDWI, NDBI, and NDMI, followed by integrating these bands to form a comprehensive dataset. The dataset was then divided into training (70%) and testing (30%) sets for the classification algorithms. Two machine learning algorithms, Random Forest (RF) and Support Vector Machine (SVM), were employed for the classification tasks. Initially, land use maps (LUMs) were generated to identify various land cover types, including rice fields, built-up areas, tree cover, water bodies, and other land uses. Subsequently, the same algorithms were applied to classify the rice varieties based on their unique spectral and temporal characteristics captured in the satellite imagery. Field surveys were conducted to collect ground truth data, which included sample points for the three rice varieties: CPRICE 888, San-Pah-Tawng1, and RD-MAEJO2. This ground truth data was used to assess the accuracy of the classification results. The overall workflow of the methodology is illustrated in Figure 3.

*2.5.1 Data pre-processing*

In this study, the pre-processing of Sentinel-1 and Sentinel-2 data was carried out meticulously to ensure the highest possible level of accuracy and consistency across all subsequent analyses. This section outlines the detailed procedures involved in the pre-processing of these satellite data sets. Sentinel-2 data underwent an extensive pre-processing protocol, starting with the Top-of-Atmosphere (TOA) correction. This correction was crucial for normalizing the radiometric differences caused by varying atmospheric conditions during different acquisition times. TOA correction ensures that the reflectance values are consistent and comparable across different images, which is essential for accurate analysis. Following the TOA correction, the data were geo-referenced to ensure that all images aligned perfectly with the geographical coordinates of the study area. This step involved the application of geometric correction techniques to rectify any spatial distortions inherent in the raw satellite imagery. The precision of geo-referencing is vital for ensuring that the spatial analysis and subsequent indices calculations are accurate and reliable [18]. In this study, the Harmonized Sentinel-2 MSI (Multispectral Instrument), Level-2A product was utilized on the GEE platform for analysis. This dataset provides surface reflectance imagery that has been corrected for atmospheric effects, making it essential for a variety of applications including agriculture, forestry, land use and land cover monitoring, and environmental management.

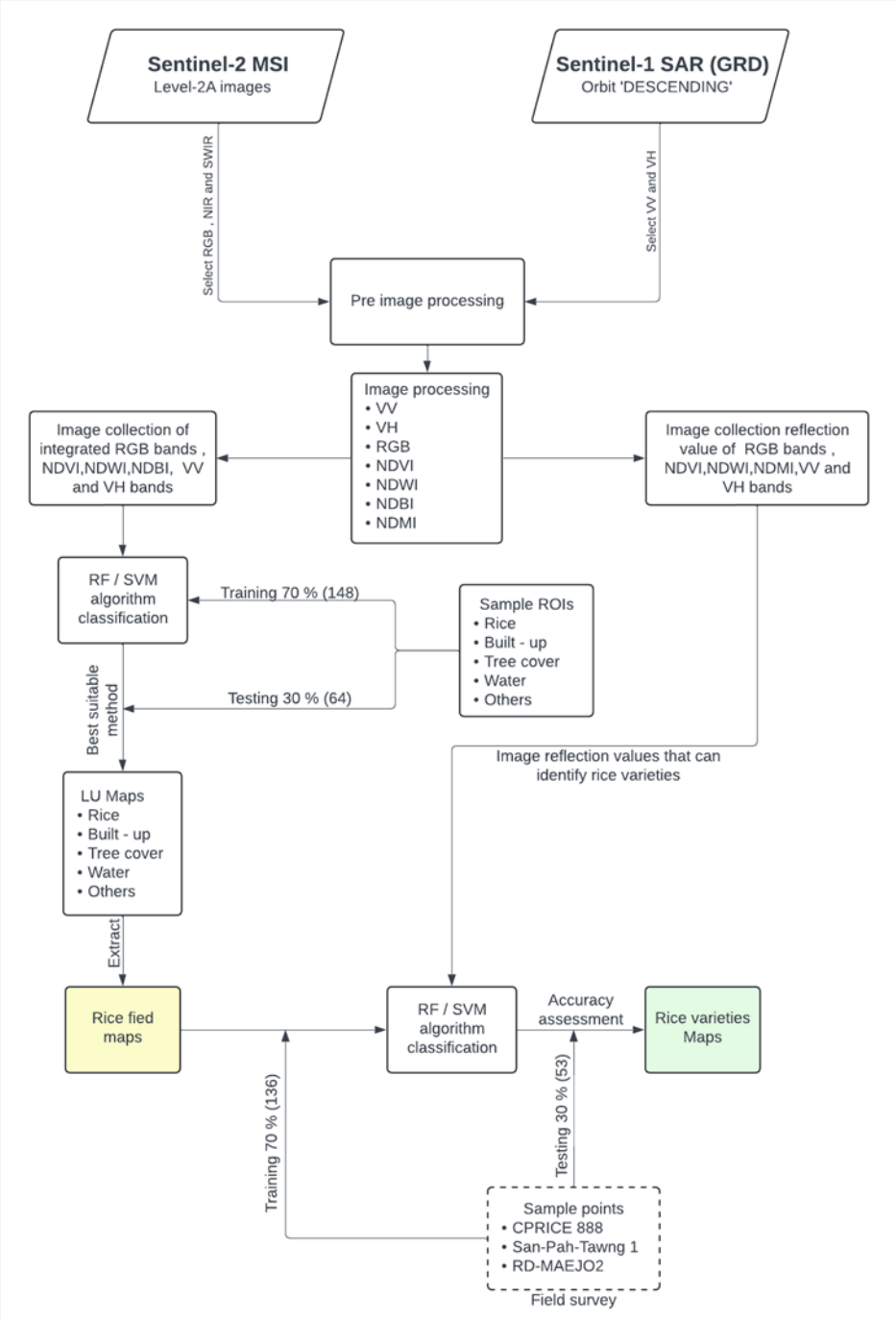



**Figure 3:** The conceptual framework of classification and accuracy assessment

Sentinel-2 offers high-resolution optical imagery, which is invaluable for detailed analysis and monitoring. For Sentinel-1 data, the pre-processing involved several key steps to ensure the data's suitability for further analysis. Sentinel-1 data, being radar-based, provides complementary information to the optical data of Sentinel-2 [19]. The pre-processing of Sentinel-1 data began with the application of radiometric calibration to convert the raw radar backscatter values into calibrated reflectance values. This step ensures that the radar data accurately represents the surface characteristics. Next, the geometric correction was applied to rectify spatial distortions and align the radar images with geographical coordinates accurately.

This geo-referencing process is crucial for ensuring the spatial alignment of Sentinel-1 data with Sentinel-2 data, facilitating integrated analysis. The Sentinel-1 data also underwent speckle filtering to reduce the inherent noise in radar imagery. Speckle noise can obscure important features in the data, so its reduction is essential for improving the clarity and interpretability of the radar images [20].

*2.5.2 Spectral indices*

In this study, Sentinel-2 data underwent TOA correction and geo-referencing to ensure accuracy and spatial consistency. All bands with a spatial resolution of 10 meters and 20 meters, specifically the Blue, Green, Red, Near Infrared (NIR), and Short-Wave Infrared (SWIR) bands [21], were combined into a single composite image using band combination techniques. This composite image was then utilized for further processing steps [22] and [23]. The Normalized Difference Vegetation Index (NDVI) value was determined using the Sentinel-2 imagery. NDVI is a mathematical index derived from the combination of the Red and NIR bands, producing values that range from -1 to +1. This index is crucial for assessing and monitoring vegetation health and density. The formula for NDVI is expressed as follows (Equation 2):

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Equation 2

Where:

*NDVI* = Normalized Difference Vegetation Index
*NIR* = Spectral reflectance acquired from near-infrared region
*RED* = Spectral reflectance acquired from red region

The Normalized Difference Water Index (NDWI) was calculated to improve the detection and analysis of water bodies and moisture content within vegetation. NDWI is derived from the Green and Near Infrared (NIR) bands of the Sentinel-2 data. This index is specifically designed to highlight water features and is essential for monitoring water stress and managing water resources effectively. The NDWI formula is represented as follows (Equation 3):

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$$

Equation 3

Where:

*NDWI* = Normalized Difference Water Index
*NIR* = Spectral reflectance acquired from near-infrared region
*GREEN* = Spectral reflectance acquired from green region

The Normalized Difference Moisture Index (NDMI) was also calculated to further understand the moisture levels in vegetation. NDMI is derived from the Near Infrared (NIR) and Shortwave Infrared (SWIR) bands. It is particularly useful in assessing plant water content and monitoring drought conditions. The NDMI formula is expressed as follows (Equation 4):

$$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1}$$

Equation 4

Where:

*NDMI* = Normalized Difference Moisture Index
*NIR* = Spectral reflectance acquired from near-infrared region
*SWIR1* = Spectral reflectance acquired from short wave infrared region

Lastly, the Normalized Difference Built-up Index (NDBI) was calculated to analyze the built-up and urban areas within the study region. NDBI is derived from the Shortwave Infrared (SWIR) and Near Infrared (NIR) bands. It is specifically designed to highlight urban areas and built-up regions. The NDBI formula is articulated as follows (Equation 5):

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR}$$

Equation 5

Where:

*NDBI* = Normalized Difference Built-up Index
*NIR* = Spectral reflectance acquired from near-infrared region
*SWIR1* = Spectral reflectance acquired from short wave infrared region

*2.5.3 Sentinel-1 polarization VV and VH*

In SAR imaging, polarization refers to the orientation of the electromagnetic waves transmitted and received by the radar. VV polarization means that the radar transmits and receives waves with their electric

field in the vertical direction, while VH polarization means that the radar transmits waves with the electric field in the vertical direction and receives waves with the electric field in the horizontal direction. The choice of polarization mode depends on the specific application and the type of information required. VV polarization is often used for monitoring changes in soil moisture, vegetation, and urban areas, as well as for detecting oil spills on the ocean surface. VH polarization is useful for distinguishing between different types of surfaces, such as water and vegetation, and for detecting changes in surface roughness. In this study, the dual-polarization capability of Sentinel-1, with its VV and VH modes, has been utilized to enhance radar imaging coverage and accuracy. This has aided in the analysis of land use classification and the differentiation of areas suitable for cultivating different rice varieties [24].

*2.5.4 Classification land use process*

In this study, a controlled data classification method was used, employing RF and SVM algorithms with Sentinel-1 and Sentinel-2 data in GEE. This approach demonstrates the power of combining radar and optical remote sensing technologies. By leveraging advanced preprocessing, feature extraction, and machine learning techniques, this process provides a robust framework for generating accurate and reliable land use maps essential for environmental monitoring, urban planning, and resource management. In the land use classification process, data from Sentinel-2 satellite imagery and SAR images from Sentinel-1, which had undergone preliminary corrections, were collectively analyzed to generate a composite image. The dataset includes true-color composite images from Sentinel-2, NDVI, NDBI, and NDWI indices analyzed from Sentinel-2 [25], and VV and VH bands from Sentinel-1 SAR images. Subsequently, the pixel units were defined for training purposes in conjunction with machine learning techniques, specifically SVM and RF. These techniques involve assigning pixel values based on the maximum likelihood and implementing supervised classification methods. The analysis was conducted to compare the results obtained from each machine learning technique, evaluating their efficacy in classifying different land use types. The land use types were categorized into five classes including rice field, built-up area, water area, tree cover area, and other areas. The classification procedures were executed utilizing the GEE platform.

*2.5.5 Classification rice varieties process*

In the study of rice cultivation areas for different rice varieties, a total of 189 sample points were identified to classify each rice variety. The sample group selection was done using the Stratified Random Sampling method, based on surveys conducted with farmers in the large-scale rice farming project. The criteria for selecting the rice field samples were as follows: 1) rice cultivation area of 8,000 sq.m or more, 2) similar planting period (June - November 2023), and 3) cultivation of rice varieties CPRICE 888, San-Pah-Tawng1, and RD-MAEJO2. The survey found that most farmers in the area preferred to plant the San-Pah-Tawng1 variety, resulting in more sample plots for this variety compared to others. Sample points were determined based on the number of selected sample plots, with each sample plot having 3-9 points, depending on the plot size and the convenience of data collection for the researchers. The number of sample points for each rice variety was as follows: CPRICE 888 had 30 points, San-Pah-Tawng1 had 111 points, and RD-MAEJO2 had 48 points. These points were chosen to comprehensively cover the study area and were analyzed alongside Sentinel-2 satellite imagery, which was filtered to have less than 30% cloud cover during the planting period (June to November). The analysis focused on calculating vegetation indices such as NDVI, NDWI, and NDMI, as well as VV and VH backscatter coefficients from Sentinel-1 satellite images [26]. These analysis results were used to identify variables that showed distinct reflectance values for each rice variety at the same time. The identified variables were then analyzed using machine learning techniques, specifically SVM and RF. These techniques involved assigning pixel values based on maximum probability and employing supervised classification methods. The analysis compared the results of each machine learning technique to evaluate their performance in classifying the cultivation areas of each rice variety [27].

*2.5.6 Accuracy assessment*

The accuracy assessment is based on a dataset of samples, defined using the same method employed for identifying the training dataset. This assessment involves the computation of the error matrix and the derived accuracy measures for each classification as well as the final aggregated classification. The measures derived from the error matrix include overall accuracy, producer’s accuracy, and user’s accuracy [28]. To evaluate the performance of the classification, validation data is used. This validation dataset comprises a separate set of labeled points that were not used during the training phase. Several metrics are calculated to assess the model's performance comprehensively. These metrics include overall accuracy, which provides a general measure of the proportion of correctly classified instances; the Kappa coefficient, which accounts for

agreement occurring by chance. By analyzing these metrics, the effectiveness and reliability of the classification model can be thoroughly evaluated.

## 3. Results and Discussion

### *3.1 Analysis of Land Use Classification*

The study aimed to classify land use using advanced machine learning techniques, focusing on the SVM and RF algorithms. The classification process involved a comprehensive dataset consisting of 212 polygons designated as training samples. These samples were meticulously categorized into five distinct land use types within the specified study area: water area, built-up area, other areas, tree cover area, and rice field. Specifically, the water area category comprised 24 polygons, the built-up area included 51 polygons, the other areas category had 14 polygons, the tree cover area contained 32 polygons, and the rice field category consisted of 91 polygons. The strategic division of the total sample area ensured that 70% of the data was allocated for training purposes, while the remaining 30% was reserved for accuracy assessment. This division facilitated a robust evaluation of the classification performance of the two algorithms. The analysis revealed a significant difference in the performance of the RF and SVM algorithms (Figure 4). The RF algorithm demonstrated a superior ability to classify land use accurately, achieving an impressive accuracy rate of 0.91. In comparison, the SVM algorithm achieved a slightly lower accuracy rate of 0.87. These findings are detailed in Table 2, highlighting the efficacy of the RF algorithm in land use classification tasks within the context of this study. Overall, the study underscores the potential of the RF algorithm as a more effective tool for land use classification compared to the SVM algorithm. The higher accuracy achieved by the RF algorithm suggests its greater suitability for similar applications in the future. The detailed categorization of the polygons and the rigorous division of the dataset for training and accuracy assessment reinforce the reliability of these findings, providing a solid foundation for future research in this domain. The table presents the results of land use classification across various categories, comparing two machine learning RF and SVM. The outcomes were evaluated against land use data previously collected by relevant agencies to ensure accuracy. The analysis indicates that each algorithm exhibits different strengths in classifying land use types. Generally, the RF algorithm outperforms the SVM algorithm in both training and overall accuracy verification, as well as in the Kappa coefficient, which indicates more reliable classification results.

### *3.2 Analysis of Rice Varieties Classification*

The classification results of different rice varieties were analyzed by selecting and evaluating the reflectance values of NDVI, NDWI, NDMI indices, and VV, VH backscatter coefficients from Sentinel-2 and Sentinel-1 satellite images. These values demonstrated significant differences in the reflectance of each rice variety. Field surveys conducted from June to August indicated the onset of rice planting in the study area. During these three months, most rice in the area was in the vegetative phase, with some in the reproductive phase. Reflectance values during this period were not suitable for variety classification due to asynchronous planting, which could lead to inaccurate results. In September, most rice fields entered the reproductive phase, with some in the ripening phase. Reflectance values during this month were the most suitable for variety classification. From October to November, all rice fields were in the ripening phase, and harvesting began, making reflectance values during this period unsuitable for variety classification. The study found that the normal RGB reflectance values from Sentinel-2 were more effective for classification than other reflectance values (Figure 5). Consequently, these selected variables were used to analyze rice cultivation area classification for each variety using machine learning techniques, specifically the SVM and RF algorithms. A total of 189 training samples were designated for the three rice varieties, with 70% (136 samples) allocated for training and 30% (53 samples) for accuracy assessment.

**Table 2:** Land use data in the study area were compared with data obtained from RF and SVM classification analysis

| Algorithm | Water area (sq.km.) | Built-up area (sq.km.) | Other area (sq.km.) | Tree cover area (sq.km.) | Rice field (sq.km.) | Training overall accuracy | Validation overall accuracy | Kappa |
|---|---|---|---|---|---|---|---|---|
| Actual | 0.876 | 4.401 | 1.196 | 3.477 | 17.901 | - | - | - |
| RF | 0.849 | 4.209 | 1.107 | 3.349 | 17.940 | 0.98 | 0.91 | 0.89 |
| SVM | 0.777 | 4.512 | 1.158 | 3.397 | 17.204 | 0.90 | 0.87 | 0.85 |

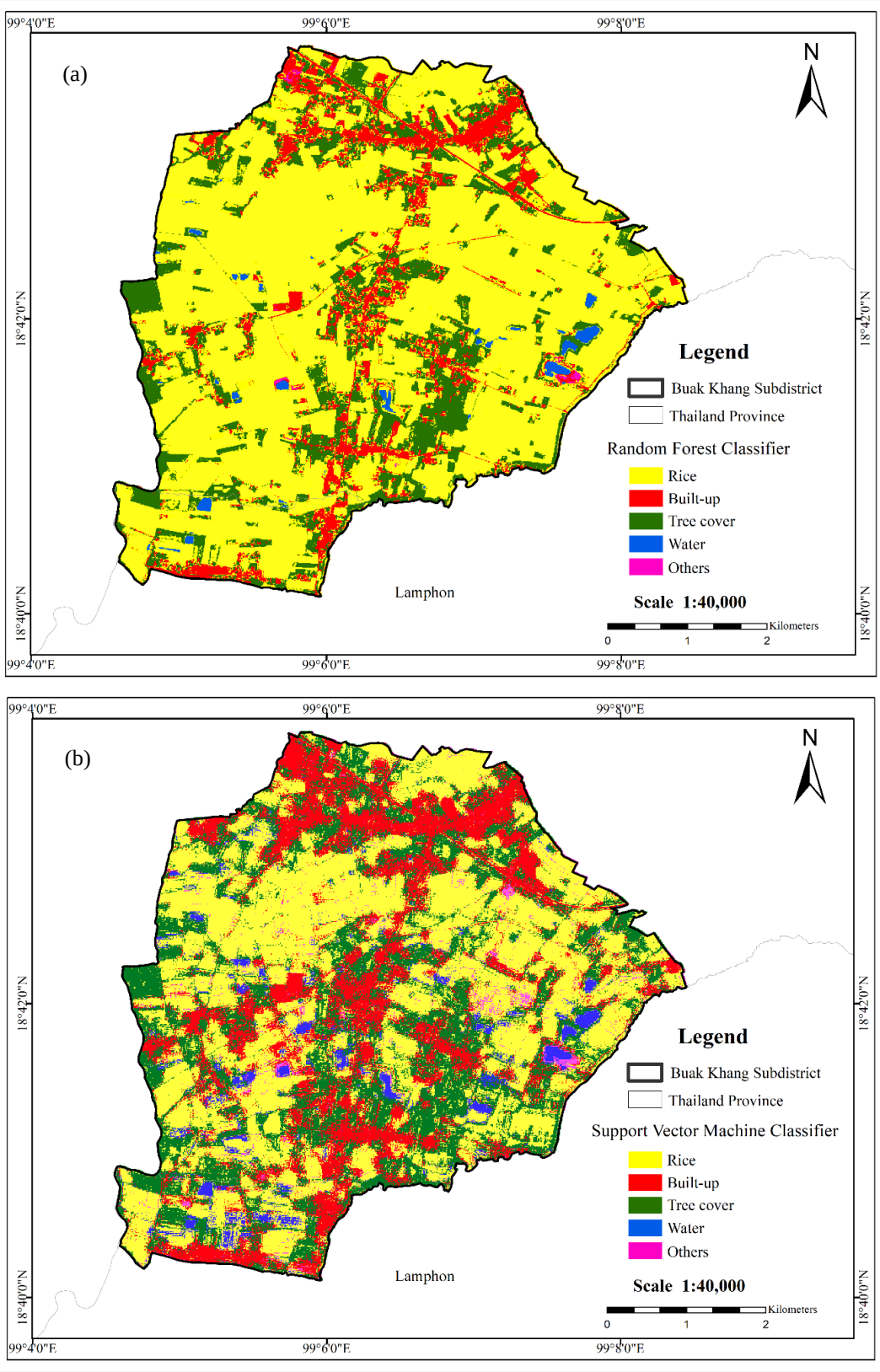



**Figure 4:** Land use classification of different machine learning techniques (a) RF (b) SVM

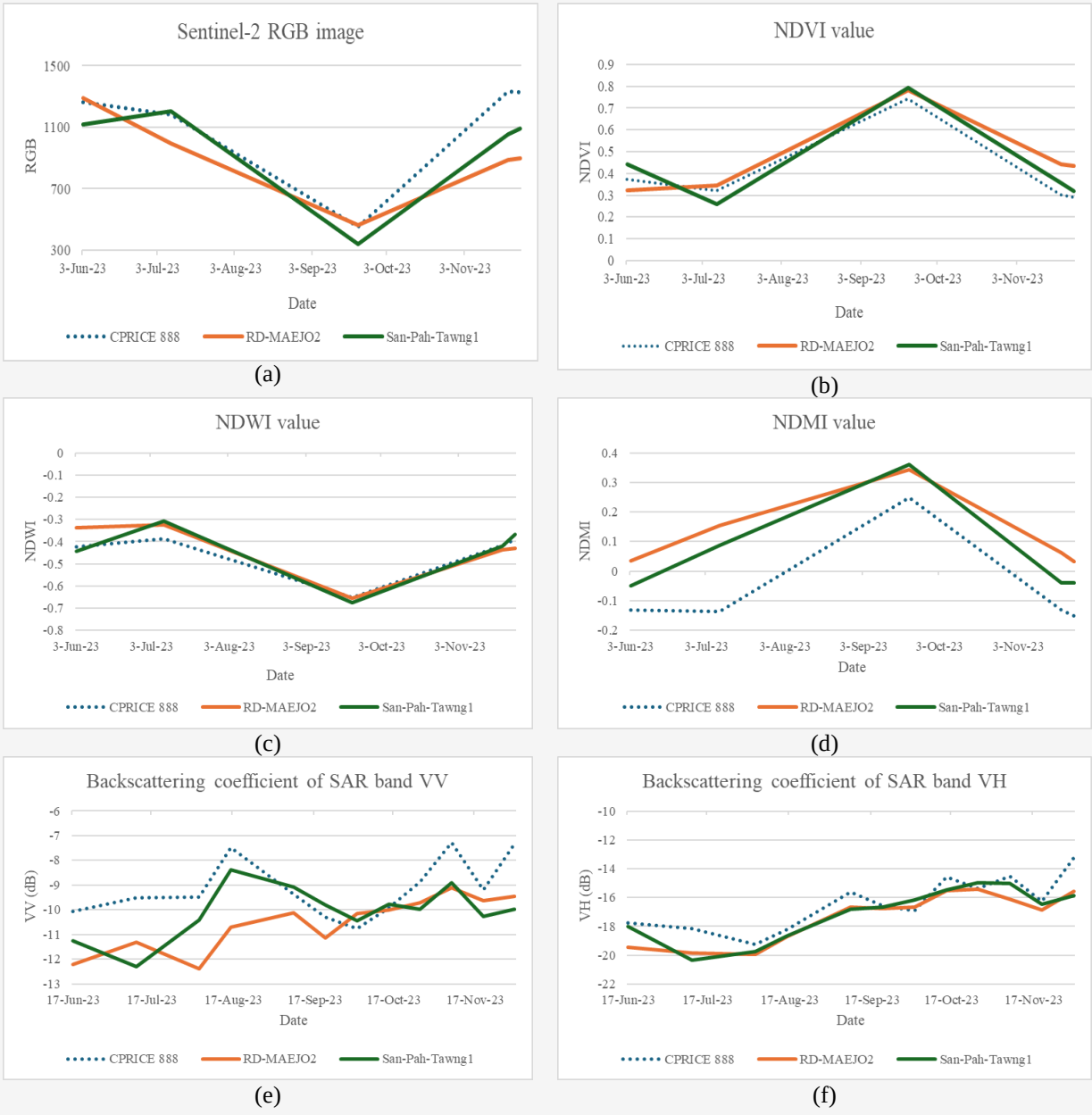
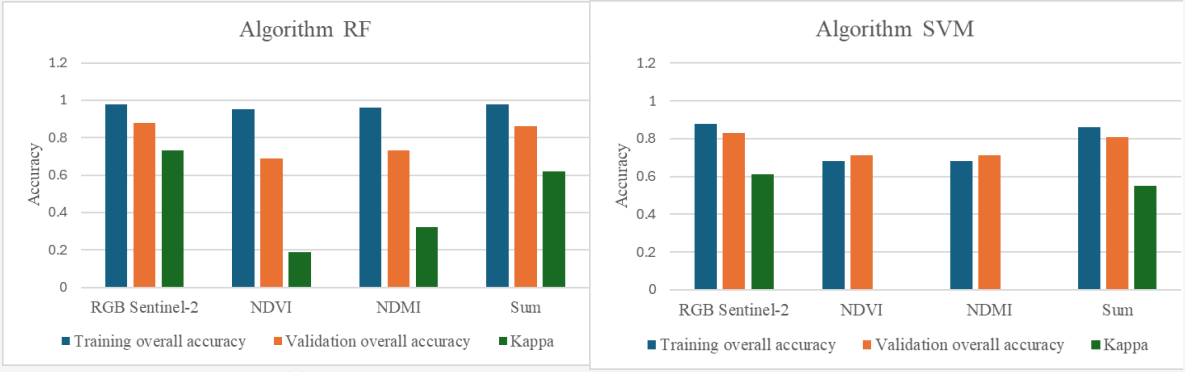



**Figure 5:** The average values of various spectral indices and backscatter coefficients derived from Sentinel-2 and Sentinel-1 satellite imagery for different rice field varieties. These indices include (a) RGB, (b) NDVI, (c) NDWI, (d) NDMI from Sentinel-2, and (e) VV, (f) VH from Sentinel-1

The study found that the RF algorithm had superior classification performance compared to the SVM algorithm, with accuracy rates of 0.88 and 0.83, respectively (Table 3). The classification results indicated that most of the study area was predominantly cultivated with the San-Pah-Tawng1 rice variety, followed by RD-MAEJO2 and CPRICE 888, respectively. Figure 7 show that the San-Pah-Tawng1 variety was classified more accurately than the other varieties due to the larger number of sample points, as most farmers in the area planted this variety extensively. In contrast, the CPRICE 888 variety had fewer sample points due to its limited cultivation area, which could lead to less accurate classification results. The rice varieties examined are CPRICE 888, San-Pah-Tawng1, and RD-MAEJO2. From the results graph, the reflectance values of the three rice varieties for each variable show clear diversity and differences. It is evident that the NDVI and NDMI index values for all three varieties peak in September, indicating increased vegetation density during this period.

The NDWI index values also show a significant decrease in reflectance during this month, corresponding to the growth stage of rice, which is at the end of the reproductive phase and entering the ripening phase. This decrease in NDWI index values is because farmers typically stop irrigating the rice fields in preparation for the harvest season. For the SAR Band VV and VH backscatter coefficients, radar waves interact less with the vegetation structure, resulting in fluctuating reflectance values, which are not suitable for rice variety classification. Conversely, the RGB image from Sentinel-2 show distinct differences for each variety in July and September. However, since most farmers in the area begin planting in July, classification results may be highly inaccurate during this month, making it unsuitable for variety classification. Therefore, the most appropriate reflectance values for classifying rice cultivation areas by variety are the RGB Sentinel-2, NDVI, NDMI, and NDWI values in September. Sentinel-2 satellite imagery for September includes two dates: the 9th and the 21st. However, considering the graph, on September 9th, the reflectance values for each index are quite similar, except for the NDMI index, where the reflectance of the CPRICE 888 variety differs from other varieties. On September 21st, the RGB Sentinel-2, NDVI, and NDMI reflectance values show clear differences that can distinguish the varieties. When these selected reflectance values were used for classifying rice cultivation areas by variety, it was found that the best reflectance values for rice variety classification were the RGB Sentinel-2 values, as detailed in Figure 6. The Figure 6 presents the performance metrics of the RF and SVM algorithms in classifying rice varieties using different indices including RGB Sentinel-2, NDVI, and NDMI. The metrics include training accuracy, validation accuracy, and the kappa coefficient. Training accuracy indicates the proportion of correctly classified instances during the model training phase, reflecting how well the model has learned from the training data. Validation Accuracy measures the proportion of correctly classified instances using a separate validation dataset, providing an assessment of the model's performance on unseen data. Kappa is a statistical measure that accounts for agreement occurring by chance, offering a more robust evaluation of classification accuracy beyond simple percentage agreements. The results indicate that the RF algorithm consistently outperforms the SVM algorithm across all indices. Specifically, for the RGB Sentinel-2 index, the RF algorithm achieved the highest validation accuracy (0.88) and kappa (0.73), demonstrating its superior capability in distinguishing rice varieties. In contrast, the NDVI and NDMI indices exhibited lower performance, with the SVM algorithm showing no reliability (kappa of 0.00) for these indices. Table 3 presents the classification results of rice varieties using the RF and SVM algorithms.




(a) (b)

**Figure 6:** Performance metrics of (a) RF and (b) SVM for rice variety classification using different indices

**Table 3:** Classification results of rice varieties using RF and SVM algorithms from Sentinel-**2** RGB satellite imagery data

| **Algorithms** | **CPRICE 888 (sq.km.)** | **San-Pah-Tawng1 (sq.km.)** | **RD-MAEJO2 (sq.km.)** | **Training overall accuracy** | **Validation overall accuracy** | **Kappa** |
|---|---|---|---|---|---|---|
| RF | 1.974 | 10.635 | 5.323 | 0.98 | 0.88 | 0.73 |
| SVM | 2.281 | 8.458 | 5.468 | 0.88 | 0.83 | 0.61 |

The Table 3 details the area (in sq.km.) classified for each rice variety—CPRICE 888, San-Pah-Tawng1, and RD-MAEJO2—along with the training accuracy, validation accuracy, and kappa coefficient for each algorithm. The RF algorithm classified 1.974 sq.km. for CPRICE 888, 10.635 sq.km. for San-Pah-Tawng1, and 5.323 sq.km. for RD-MAEJO2, achieving a training accuracy of 0.98, validation accuracy of 0.88, and a kappa coefficient of 0.73. In comparison, the SVM algorithm classified 2.281 sq.km. for CPRICE 888, 8.458 sq.km. for San-Pah-Tawng1, and 5.468 sq.km. for RD-MAEJO2, with a training accuracy of 0.88, validation accuracy of 0.83, and a kappa coefficient of 0.61. These results indicate that the RF algorithm outperforms the SVM algorithm in terms of both validation accuracy and kappa coefficient, suggesting that RF provides more reliable and accurate classification results for the rice varieties in this study. The larger area classified for San-Pah-Tawng1 by RF aligns with the higher number of sample points for this variety, contributing to its higher classification accuracy.

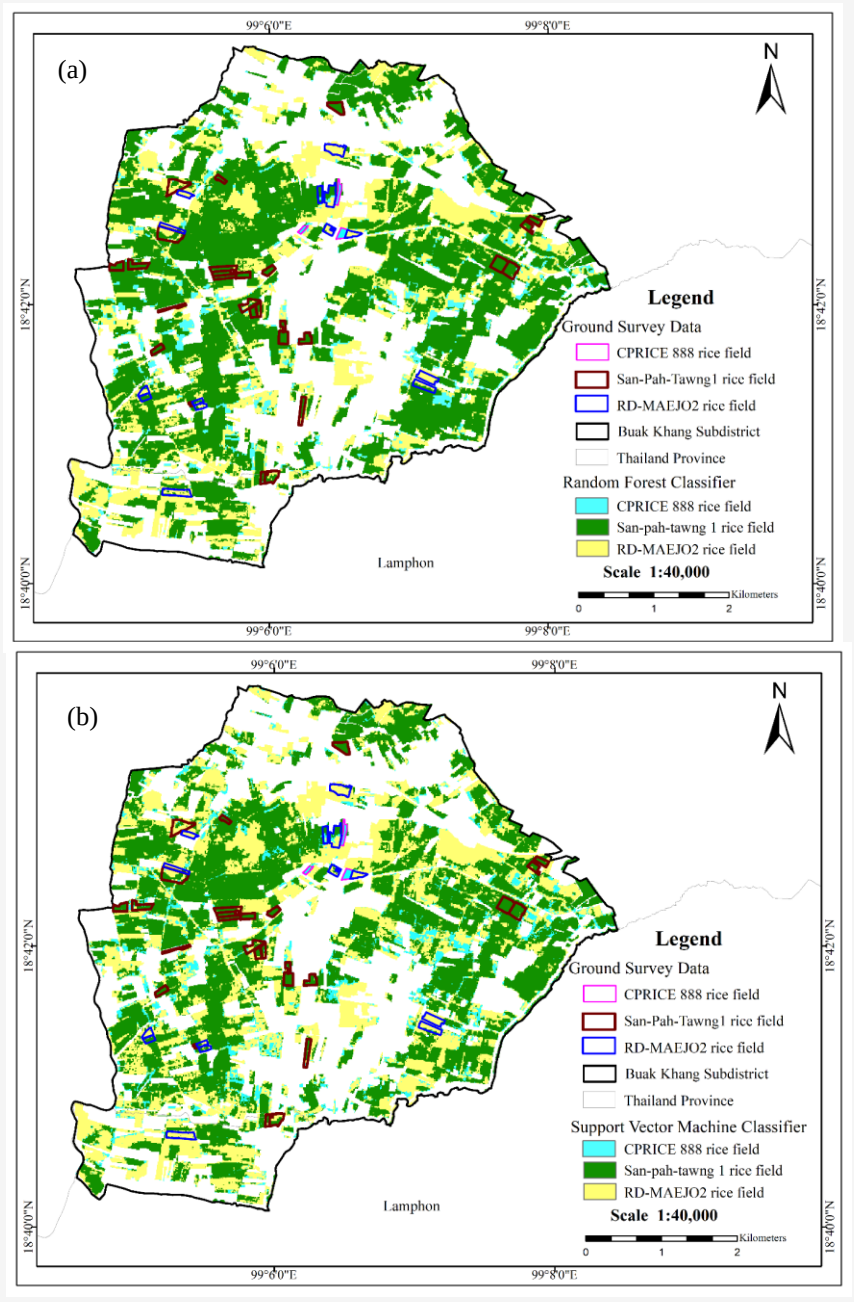



**Figure 7:** Rice variety classification map for Buak Khang subdistrict using (a) RF (b) SVM algorithm

*3.3 Discussion*

In this study, satellite images from both Sentinel-2 and Sentinel-1 were used to improve the accuracy of land use classification and rice cultivation area identification. Previous studies typically used either Sentinel-2 or Sentinel-1 images for analysis, which often resulted in less accurate outcomes compared to using both types of data together [8]. Prior research primarily focused on land use classification using indices derived from optical images, such as NDVI, NDWI, and NDBI [19]. However, this study combines both optical and SAR images, using NDVI, NDWI, and NDBI indices along with VH and VV backscatter coefficients from SAR images for rice variety classification. Numerous studies have employed NDVI, NDWI, and NDMI indices from optical images in conjunction with VH and VV backscatter coefficients from SAR images [24]. In some studies [27], it has been found that the reflectance values of different rice varieties depend on the inherent characteristics of the varieties, with differences observed across the VNIR to SWIR spectral ranges. Our study analyzed all relevant reflectance variables to differentiate the characteristics of different rice varieties, selecting only the variables that effectively distinguish between them. The machine learning algorithms used in this study were RF (Random Forest) and SVM (Support Vector Machine). The results indicated that the RF model outperformed the SVM model in both land use classification and rice variety classification. This finding aligns with previous studies that have demonstrated the superior performance of RF in similar contexts [4]. Although some studies have reported that the SVM algorithm performs better in rice variety classification than RF [10], the difference in accuracy was minimal. This difference may be due to variations in the variables, study area, or time period used in those studies. In this study, the use of a large sample size with complex relationships made the RF classification algorithm more effective than the SVM algorithm. RF's flexibility and efficiency, particularly in scenarios involving large datasets with complex relationships and noise, contributed to its superior classification performance.

## 4.Conclusion

Mapping high-resolution rice cultivation areas accurately and promptly is essential for advancing global food security. This study emphasizes the value of high-resolution and freely accessible Sentinel-2 and Sentinel-1 satellite imagery in monitoring and analyzing different rice varieties. The study employed RF and SVM models, with the RF model demonstrating superior classification performance over the SVM model. The overall land use classification accuracy was 0.91, while the accuracy for classifying individual rice varieties was 0.88, showcasing the effectiveness of these data sources. However, several limitations were identified, including significant cloud cover during the rainy season, which affected the accuracy of Sentinel-2 imagery, and insufficient field data collection, which limited the scope of the analysis. Additionally, the indices used for rice variety classification showed relatively minor differences, which could reduce the classification accuracy.

Future research should address these limitations by incorporating additional satellite data from multiple platforms, using a broader range of wavelengths and indices for classification, and improving cloud and shadow masking techniques [29]. Moreover, increasing the density and distribution of field reference data would enhance the accuracy and representativeness of classification results. Despite these challenges, the combined use of Sentinel-2 and Sentinel-1 imagery, integrated with advanced machine learning techniques, provides a robust and precise approach for land use and rice variety classification. This method holds significant potential for improving agricultural planning and management, thereby supporting food security efforts. Overcoming current limitations will enable future studies to further enhance the reliability and applicability of these methods, ensuring that the results are accurate and valuable for stakeholders.

**Acknowledgements**

This research endeavor has been generously supported by the ERASMUS+ KA107: Student and Staff Mobility project. We wish to convey our profound gratitude to the European Commission and the various participating institutions for their invaluable financial support and the myriad collaborative opportunities facilitated through this esteemed program. Our deepest appreciation is directed towards the University of Salzburg for its exceptional role in hosting and facilitating the exchange, providing an environment conducive to academic and professional growth. Furthermore, we extend our heartfelt thanks to Chiang Mai University for their unwavering and sustained support throughout the duration of this project. The extensive experiences and substantial knowledge acquired during this mobility have played a pivotal role in the successful completion and high quality of this academic document.

**References**


[1] Suryono, H., Kuswanto, H. and Iriawan, N., (2022). Rice Phenology Classification Based on Random Forest Algorithm. *Procedia Computer Science*, Vol. 197, 668 - 676. https://doi.org/10.1016/j.procs.2021.12.201.

[2] Choudhary, K., Shi, W., Dong, Y. and Paringer, R., (2022). Random Forest for Rice Yield Mapping and Prediction Using Sentinel-2 Data with Google Earth Engine. *Advances in Space Research*, Vol.70(8), 2443 2457. https://doi.org/10.1016/j.asr.2022.06.073.

[4] Singha, C. and Swain, K. C., (2023). Rice Crop Growth Monitoring with Sentinel 1 SAR Data Using Machine Learning Models in Google Earth Engine Cloud. *Remote Sensing Applications: Society and Environment*, Vol. 32. https://doi.org/10.1016/j.rsase.2023.101029

[5] Cai, Y., Lin, H. and Zhang, M., (2019). Mapping Paddy Rice by the Object-based Random Forest Method Using Time Series Sentinel-1/Sentinel-2 Data. *Advances in Space Research*, Vol.64(11), 2233-2244. https://doi.org/10.1016/j.asr.2019.08.042.

[6] Phan, H., Le Toan, T. and Bouvet, A., (2021). Understanding Dense Time Series of Sentinel-1 Backscatter from Rice Fields: Case Study in a Province of the Mekong Delta, Vietnam. *Remote Sensing*, Vol.13(5). https://doi.org/10.3390/rs13050921.

[7] Mandal, D., Kumar, V., Bhattacharya, A., Rao, Y. S., Siqueira, P. and Bera, S., (2018). Sen4Rice: A Processing Chain for Differentiating Early and Late Transplanted Rice Using Time-Series Sentinel-1 SAR Data with Google Earth Engine. *IEEE Geoscience and Remote Sensing Letters*, Vol.15(12), 1947-1951. https://doi.org/10.1109/lgrs.2018.2865816.

[8] Amani, M., Kakooei, M., Moghimi, A., Ghorbanian, A., Ranjgar, B., Mahdavi, S., Davidson, A., Fisette, T., Rollin, P., Brisco, B. and Mohammadzadeh, A., (2020). Application of Google Earth Engine Cloud Computing Platform, Sentinel Imagery, and Neural Networks for Crop Mapping in Canada. *Remote Sensing*, Vol.12(21). https://doi.org/10.3390/rs12213561.

[9] Bazzi, H., Baghdadi, N., El Hajj, M., Zribi, M., Minh, D. H. T., Ndikumana, E., Courault, D. and Belhouchette, H., (2019). Mapping Paddy Rice Using Sentinel-1 SAR Time Series in Camargue, France. *Remote Sensing*, Vol. 11(7). https://doi.org/10.3390/rs11070887.

[10] Jamali, A., (2020). Sentinel-1 Image Classification Using Machine Learning Algorithms Based on the Support Vector Machine and Random Forest. *International Journal of Geoinformatics*, Vol. 16(2). 15–22. https://journals.sfu.ca/ijg/index.php/journal/article/view/1809/923.

[11] Eisfelder, C., Boemke, B., Gessner, U., Sogno, P., Alemu, G., Hailu, R., Mesmer, C. and Huth, J., (2024). Cropland and Crop Type Classification with Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine for Agricultural Monitoring in Ethiopia. *Remote Sensing*, Vol. 16(5). https://doi.org/10.3390/rs16050866.

[12] Pindhaya, P., Sorachai, C., Chotinisakorn, P., Techa, P. and Suchaibunsiri, S., (2006). Sanpatong 1 Seed Production and Distribution (in Thai). *Agricultural Knowledge Base*. Available: https://agkb.lib.ku.ac.th/doa/search_detail/result/156074. [Accessed Mar. 14, 2024].

[13] Department of Rice, Ministry of Agriculture. *Technical Document on San Pa Tong 1 Rice Variety (in Thai). Retrieved from Department of Rice, Upper Northern Thailand Rice Research Center, Rice Research and Development Office*. [Online]. Available: https://newwebs2.ricethailand.go.th/upload/doc. [Accessed Mar. 14, 2024].

[14] Songchantuk, S., (2017). *Breeder and Foundation Seeds Production of Rice Variety: RD-Maejo 2(in Thai).* Genetics, Faculty of Science, Maejo University.

[15] Name Change of Plant Varieties in the Plant Variety Registration Certificate According to the Plant Variety Act B. E. 2518 (in Thai). *Department of Agriculture*. [Online]. Available: https://www.doa.go.th/pvp/wpcontent/uploads/2020/11/AnnoDOA_Public182.pdf. [Accessed Mar. 14, 2024].

[16] Erdanaev, E., Kappas, M., and Wyss, D. (2022). The Identification of Irrigated Crop Types Using Support Vector Machine, Random Forest and Maximum Likelihood Classification Methods with Sentinel-2 Data in 2018: Tashkent Province, Uzbekistan. *International Journal of Geoinformatics*, Vol. 18(2), 37–53. https://doi.org/10.52939/ijg.v18i2.2151.

[17] Belgiu, M. and Dragut, L., (2016). Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[18] Eisfelder, C., Boemke, B., Gessner, U., Sogno, P., Alemu, G., Hailu, R., Mesmer, C. and Huth, J., (2024). Cropland and Crop Type Classification with Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine for Agricultural Monitoring in Ethiopia. *Remote Sensing*, Vol. 16(5). https://doi.org/10.3390/rs16050866.

[19] Mema, P. and Charoenpanyanet, A., (2021). Model Creation for Assessing Productivity of Dry Season Rice with Backscatter Coefficient from Sentinel-1. *The Journal of King Mongkut's University of Technology North Bangkok*, Vol. 31(3). https://doi.org/10.14416/j.kmutnb.2021.05.018.

[20] Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N. and Reiche, J., (2021). Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine. *Remote Sensing*, Vol. 13(10). https://doi.org/10.3390/rs13101954.

[21] Aji, A., Husna, V. and Purnama, S., (2024). Multi-Temporal Data for Land Use Change Analysis Using a Machine Learning Approach (Google Earth Engine). *International Journal of Geoinformatics*, Vol. 20(4), 19–28. https://doi.org/10.52939/ijg.v20i4.3145.

[22] Sameh, S., Zarzoura, F. and El-Mewafi, M., (2022). Automated Mapping of Urban Heat Island to Predict Land Surface Temperature and Land Use/Cover Change Using Machine Learning Algorithms: Mansoura City. *International Journal of Geoinformatics*, Vol. 18(6), 47–67. https://doi.org/10.52939/ijg.v18i6.2461.

[23] Soriano-González, J., Angelats, E., Martínez-Eixarch, M. and Alcaraz, C., (2022). Monitoring Rice Crop and Yield Estimation with Sentinel-2 Data. *Field Crops Research*, Vol. 281. https://doi.org/10.1016/j.fcr.2022.108507.

[24] Dineshkumar, C. and Satishkumar, J., (2019). Rice Crop Monitoring Using Sentinel-1 C-Band Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: Workshop on "Earth Observations for Agricultural Monitoring"*. 73–77. https://doi.org/10.5194/isprs-archives-XLII-3-W6-73-2019.

[25] Miranda, E., Mutiara, A. B., Ernastuti, E. and Wibowo, W. C., (2020). Land Cover Classification through Ontology Approach from Sentinel-2 Satellite Imagery. *International Journal of Geoinformatics*, Vol. 16, 61-72. https://journals.sfu.ca/ijg/index.php/journal/article/view/1779/893.

[26] Rauf, U., Qureshi, W. S., Jabbar, H., Zeb, A., Mirza, A., Alanazi, E., Khan, U. S. and Rashid, N., (2022). A New Method for Pixel Classification for Rice Variety Identification Using Spectral and Time Series Data from Sentinel-2 Satellite Imagery. *Computers and Electronics in Agriculture*, Vol. 193. https://doi.org/10.1016/j.compag.2022.106731.

[27] Sarapin, P., (2018). Plication of Hyperspectral Remote Sensing in Classifying the Rice Varieties and Bacterial Leaf Blight Disease. *Inter-Department of Environmental Science*, Chulalongkorn University: Bangkok.

[28] Belcore, E., Piras, M. and Wozniak, E., (2020). Specific Alpine Environment Land Cover Classification Methodology: Google Earth Engine Processing for Sentinel-2 Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2020*, 663-670. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-663-2020.

[29] Clauss, K., Ottinger, M. and Kuenzer, C., (2017). Mapping Rice Areas with Sentinel-1 Time-series and Superpixel Segmentation. *International Journal of Remote Sensing*, Vol. 39(5), 1399 - 1420. https://doi.org/10.1080/01431161.2017.1404162.