

[Open Peer Review on Qeios](#)

Do SHapley Additive exPlanations Enhance Classification Accuracy? An Investigation of Game Theory for Mapping Cityscape Blue-Green Infrastructures Using Sentinel-2

Mehedi Hasan¹, Malay Pramanik¹, Iftexharul Alam¹, Atul Kumar²

¹ Asian Institute of Technology

² Chaudhary Charan Singh University

Funding: This research funded by Research Initiation Grant from Asian Institute of Technology, Thailand (Project number: SERD-2023-029).

Potential competing interests: No potential competing interests to declare.

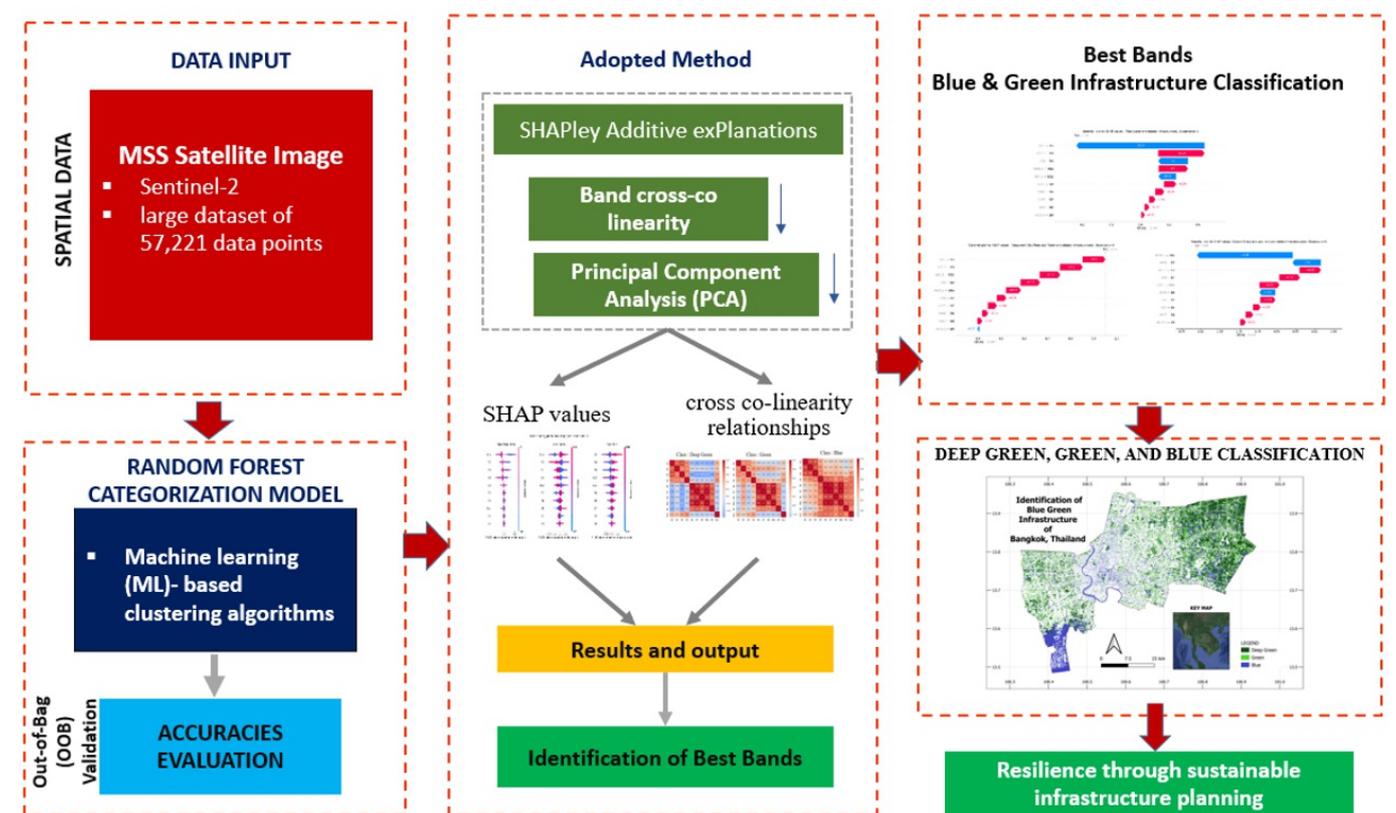
Abstract

Urban feature identification and classification play a pivotal role in urban planning, environmental management, and resource allocation, especially in blue and green infrastructure (BGI). Despite a number of studies demonstrating band combinations for improving classification accuracy using machine learning and artificial intelligence, there remains a gap in understanding the consideration of band combinations² across diverse urban settings and regions, particularly in addressing the variability and complexity of city-level BGI characteristics and planning. Therefore, in this research, we present a novel framework for BGI classification and accuracy enhancement using Game Theory SHapley Additive exPlanations (SHAP) values integrating with machine learning techniques and Sentinel-2 satellite imagery for scalability of the methodology. In order to examine blue and green infrastructures at the city level, this study examines Game Theory SHAP values integrating with the random forest (RF) model used in the study. Using a large dataset of 57,221 data points, the performance of the model is methodically assessed, with a focus on improving interpretability by utilizing SHAP values. Beyond conventional accuracy assessments, this research investigates the subtle spectral characteristics influencing the recognition of infrastructure across distinct classes, including Deep Green, Green, and

Blue. Notably, spectral bands such as B8A, B7, B6, and B8 demonstrate remarkable precision in categorizing the Deep Green class, while the importance of bands like B8A, B5, B4, and B3 is shown in distinguishing green infrastructures. Additionally, the combination of bands B5, B4, B3, and B12B5 demonstrates importance as a discriminator for accurately identifying blue infrastructures. We enhanced infrastructure classification through granular analysis utilizing Shapley values, offering insights into the importance of specific spectral bands for classification outcomes. These findings carry significant implications for urban planners, environmentalists, and policymakers, offering valuable insights into optimizing urban feature classification accuracy and refining models for precise infrastructure mapping.

Keywords: Cityscape classification; feature identification; game theory; LULC enhancement; SHAPley value.

Graphical Abstract



Highlights

- Investigated game theory SHAP for classifying city-level blue-green infrastructure (BGI)
- Shapley identified accurate spectral features influencing classification across all classes
- Identified optimal band combinations for improved classification of BGI using PCA

- Blue infrastructure accurately represented by bands B5, B4, B3, and B12
- Green infrastructure class accurately classified in B8A, B7, B6, and B8 combinations

1. Introduction

Satellite image-based classification for land use land cover (LULC) analysis has been utilized since the advent of satellite remote sensing technology. The launch of high-resolution satellite sensors with improved spectral, spatial, and temporal resolutions has enabled more detailed and accurate LULC mapping (Luo et al., 2021; Pandey et al., 2021). However, the field has evolved significantly over time with advancements in satellite technology, sensor capabilities, and image processing techniques (Balha et al., 2021). These initial classifiers were typically parametric and grounded in the statistical assumption that land surface reflectance follows a normal distribution. As a result, they gained popularity for their simplicity and the convenience of being prepackaged with remote sensing software. In contrast, modern methods like Object-Based Image Analysis (OBIA) and Machine Learning (ML) algorithms (including Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), Deep Learning (DL), and Convolutional Neural Networks (CNN)) as well as Ensemble Methods enable rapid and highly precise image classification with enhanced accuracy and robustness (Hosseiny et al., 2022). These intricate algorithms are based on a wide range of different architectures, such as neural networks; e.g., deep learning (Bengio, 2009), random forests for decision trees (Breiman, 2001), and support vector machines for hyperplanes (Cortes and Vapnik, 1995). Because statistical classifiers depend on the underlying statistical features of the data, there was a transition in LULC classification from statistical classification to machine learning (Hosseiny et al., 2022).

The transition from statistical classification to ML in LULC classification largely occurred because statistical classifiers depend heavily on the basic statistical properties of the data. The exponential growth of geospatial big data, of which remote sensing (RS) big data forms a substantial portion, presents considerable challenges to traditional geographic information systems (GIS) and remote sensing methodologies and platforms. The availability of satellite data and the viability of multitemporal analysis enabled the substitution of statistical classifiers with more sophisticated algorithms. For instance, the maximum likelihood approach presumes a normally distributed set of data, the Mahalanobis distance approach assumes equal covariances across all classes, and the parallelepiped approach builds patterns using the averages and standard deviations of individual classes (Hosseiny et al., 2022). The restrictions resulting from these data assumptions lower the classifiers' relative accuracy (Richards, 2005).

While multivariable large datasets can be clustered using ML-based clustering methods, the interpretation of the cluster findings may be unclear from these techniques (Murdoch et al., 2019). Concerns over how to effectively evaluate ML model findings are growing as the use of these models in data analysis evolves (Rai, 2020). It becomes especially difficult to figure out how a particular feature drives cluster prediction when several factors are strong predictors of a cluster (Brandsaeter and Glad, 2022). Algorithms for ML-based clustering usually use unsupervised learning, which means that there is no preset solution. Because of this, the interpretation of the findings is mostly based on the analysts' perception of the data and research subject matter. Therefore, users might misunderstand model results due to lack of expertise with

how ML models operate or insufficient knowledge of the study setting, which could have an impact on the findings of the research (Wang and Biljecki, 2022).

In order to address this problem, explanatory techniques like SHAP have become more popular recently. These techniques improve the interpretability and transparency of machine learning models (Khadem et al., 2022). By estimating each feature's contribution to the final prediction and averaging it across all potential feature coalitions, the SHAP technique explains the output of a machine learning model (Lundberg and Lee, 2017). The SHAP technique is useful for highlighting the inner workings of intricate machine learning models (Ekanayake et al., 2022). The SHAP tool is increasingly being accepted for use in evaluating cluster data, as evidenced by the tool's use in studies carried out in the recent few years in a variety of fields, including engineering (Meddage et al., 2022), finance (Mokhtari et al., 2019), and health (Zheng et al., 2020).

Urban green space typically refers to green infrastructure comprising vegetated areas such as urban parks, roadside greenery, and green spaces within workplaces (WHO, 2017). Urban development consumes green spaces and exerts adverse effects on urban environments (Benedict and McMahon, 2012). The World Health Organization has recognized green spaces as innovative means to enhance the quality of urban environments by bolstering local resilience and promoting sustainable lifestyles (WHO, 2017). Understanding the extent and characteristics of green spaces is crucial for sustainable urban planning and management. Bangkok, the capital city of Thailand, is known for its dense urban fabric and high population density. The city faces various environmental challenges, including water and air pollution, heat island effects, and loss of biodiversity. In this context, green spaces play a vital role in mitigating these challenges by providing essential ecosystem services and enhancing the overall quality of urban life. Moreover, information on green space distribution and its impact on the urban ecosystem helps in developing sustainable urban planning policies.

As one of the most widely used machine learning tools, the RF is one of the many algorithms that are now being utilised in LULC mapping, and many more are moving from the computer/data science domain to the field of remote sensing (Avci et al., 2023; Nguyen et al., 2018). These algorithms may not create the same variable importance hierarchies since they are frequently based on different underlying architectures (Avci et al., 2023; Nguyen et al., 2018). Nevertheless, generalizable consistencies in variable ranking can identify consistent subsets of variables that are important for identifying the relevant LULC classes (Lundberg and Lee, 2017; Shapley, 1953). In order to improve the RF algorithm's interpretability, the present study employs Shapley additive explanations, which offer insights into the significance of a particular set of features for LULC classes. Through identifying the variables or bands that are consistently present among classifiers, this study aims to enable well-informed conclusions.

Incorporating a comprehensive array of geospatial data, including both spectral and auxiliary information, across complex urban morphologies, this approach enhances identification accuracy, considering the significance of variables in LULC classification, particularly for blue and greenspace classification, through the application of widely recognized algorithms such as Random Forest integrating with SHAP. To deepen our understanding of the outcomes, we employed Shapley Additive Explanations and Principal Component Analysis (PCA). By leveraging Shapley additive explanations, we investigated the interpretability of the Random Forest model, offering insights into feature importance at the level of

individual LULC classes. It can help to understand urban landscape characteristics and optimise spectral band combinations; this research can inform a scalable methodology for decision-making in urban planning and management, thereby advancing sustainable blue-green infrastructure planning. Also, the mapping outcomes could involve promoting green infrastructure development, integrating green spaces into urban design, and protecting existing green areas.

2. Data and methods

2.1. Study area

Thailand's capital city, Bangkok, is situated in the nation's centre and possesses a number of unique topographical features. BMR is situated on the Chao Phraya River delta in the Gulf of Thailand, along the country's central coastline. It is positioned on the eastern bank of the river, approximately 40 km from the Gulf. The Chao Phraya River plays a vital role in Bangkok's geography (*figure 1*). It flows through the city, dividing it into the western and eastern banks. The river serves as a major transportation route and has influenced the city's development and urban planning. BMR is characterized by a relatively flat topography. Its elevation ranges from just a few meters above sea level to around 2 meters in some areas (Goeyinsup, 2022). The city's low-lying terrain makes it susceptible to flooding, especially during the monsoon season (Limthongsakul et al., 2017; Marks et al., 2020). BMR has an extensive network of canals, known as khlongs, which were historically used for transportation and irrigation (Davivongs et al., 2012). Many of these canals have been converted into roads or filled in over time, but some still exist and contribute to the city's unique charm. The khlongs are interconnected and connected to the Chao Phraya River, providing additional water transport options (Davivongs et al., 2012). While BMR is not directly on the coast, it is located near the Gulf of Thailand. The city's proximity to the coast has influenced its climate and made it an important economic and transportation hub for the country (Davivongs et al., 2012).

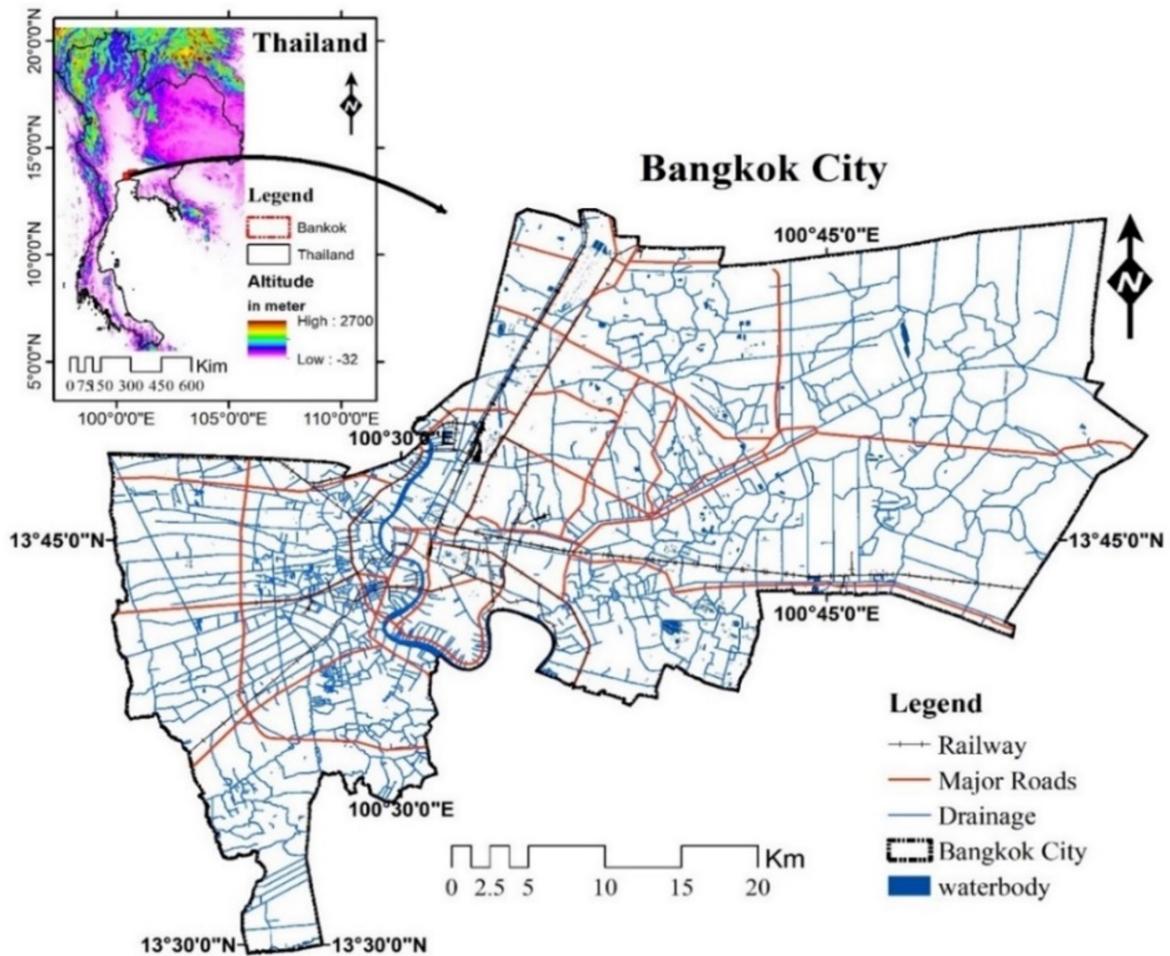


Figure 1. Showing the geographical location of Bangkok city, Thailand

BMR is experiencing subsidence, which is the sinking of the land surface. Factors such as rapid urbanization, groundwater extraction, and the weight of infrastructure contribute to the sinking (Goeyinsup, 2022). As a result, some areas of the city are gradually sinking below sea level, increasing vulnerability to flooding. Bangkok has witnessed rapid urban expansion, with the city growing in all directions (Goeyinsup, 2022). As a result, high-rise buildings and skyscrapers dominate the city's skyline, especially in the central business district. The urban landscape is a mix of modern high-rises and traditional architecture. Despite being a highly urbanized city, Bangkok has several green spaces and parks that provide recreational areas and serve as lungs for the city. Lumpini Park and Chatuchak Park are among the notable green spaces that offer respite from bustling city life (Petcharat et al., 2020). These geographical attributes contribute to Bangkok's unique landscape and shape its urban development, transportation systems, and environmental challenges.

2.2. Dataset used

In this research, we conducted a comprehensive seasonal LULC assessment of the Bangkok Metropolitan region. To achieve this, we acquired satellite images from the European Sentinel Agency

(<https://sentinels.copernicus.eu/web/sentinel/copernicus/sentinel-2>). These Sentinel-2 datasets were obtained for the winter seasons of 2022, allowing us to observe and analyze LULC changes, particularly for mapping blue and green landscape features in the Bangkok Metropolitan Region (BMR). Table 1, presented below, highlights the essential details of the multispectral satellite datasets used in the study. It includes information on the spectral bands, acquisition dates, spatial resolutions, and wavelengths used in the current research. The availability of these datasets enhances the accuracy and depth of our analysis, facilitating a more precise assessment of identifying blue and green spaces in BMR.

Table 1. Shows the details of the multispectral satellite datasets used in the current study

Datasets	Sentinel-2
Season	Winter
Image collection date	2021-11-01 to 2022-02-28
Cloud cover	< 1%
Spatial Resolution	10 m

2.3. Dataset composition

ESA, provided by Sentinel-2, is an important dataset in the field of Earth observation. It features a multispectral instrument (MSI) with an array of high-resolution spectral bands (Figure 2). These bands are strategically designed to capture a wide range of information about land cover and land use.

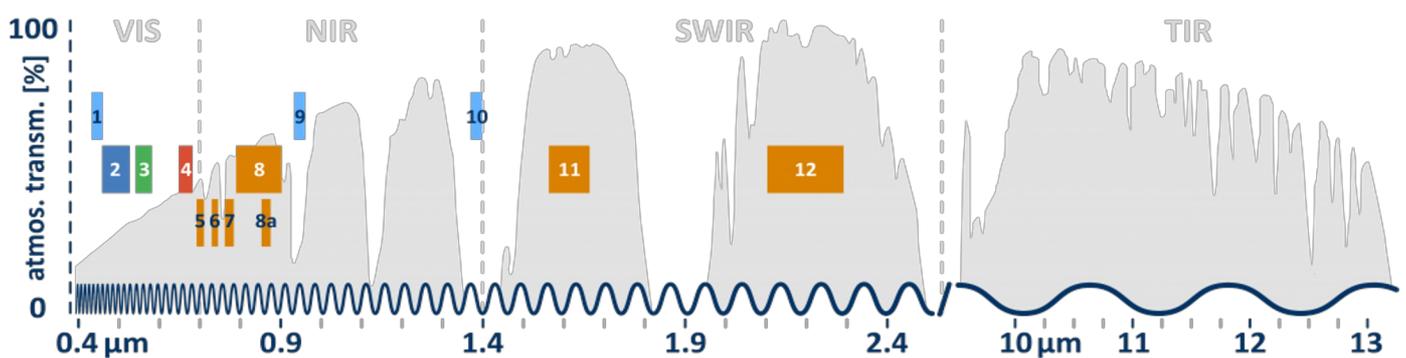
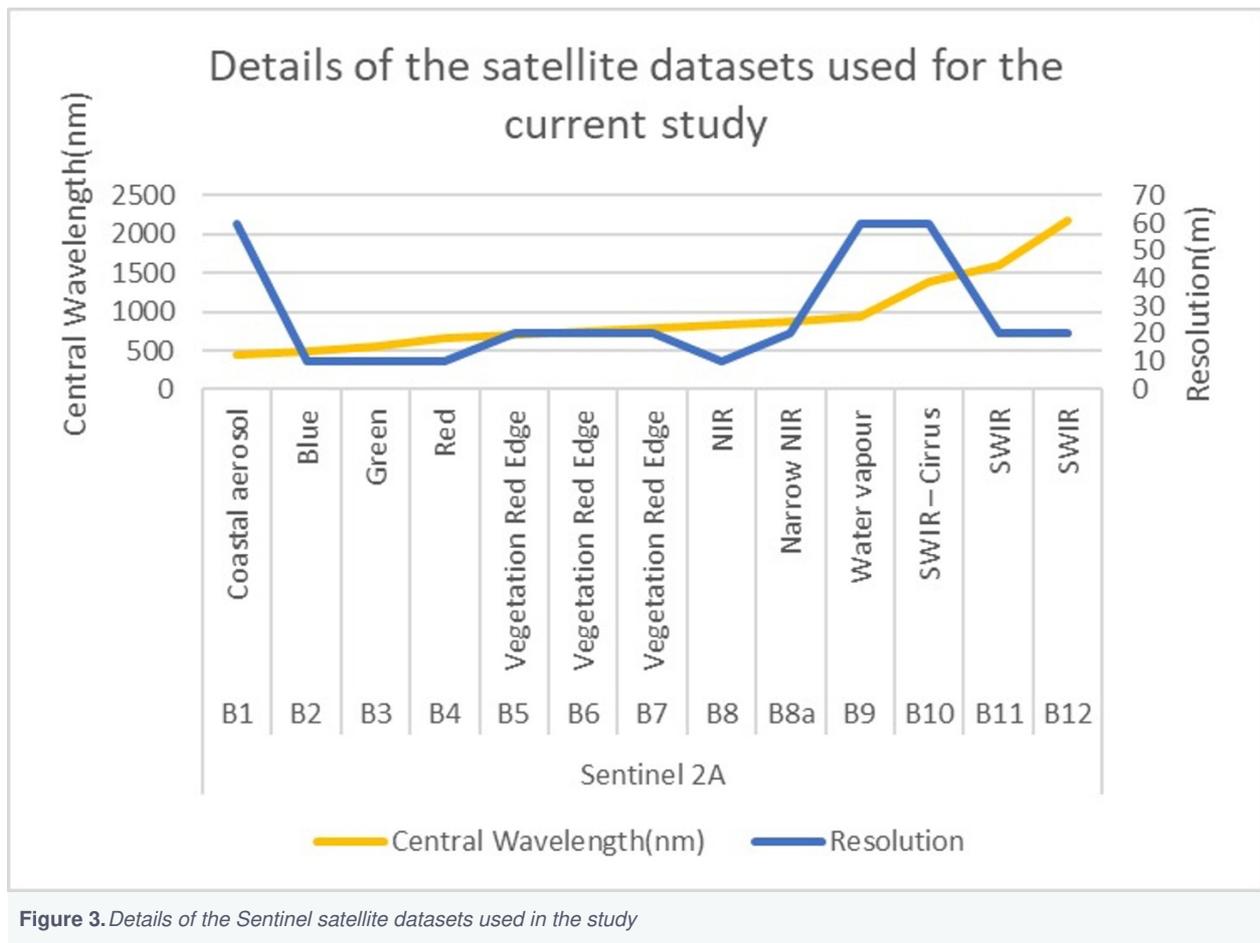


Figure 2. Sentinel-2 band characteristics (Source: Freie Universität Berlin)



The Sentinel-2 datasets typically include bands in the visible, near-infrared, and shortwave-infrared regions of the electromagnetic spectrum. Researchers can leverage these bands to monitor vegetation health, detect changes in land use, and assess the impact of natural disasters, among various other applications. Figure 3 provides a comprehensive overview of the satellite datasets employed in the present study, including band descriptions and corresponding wavelengths.

2.4. Machine learning classification using Google Earth Engine

In geospatial analysis using Google Earth Engine (GEE), various machine learning models, as well as the Random Forest (RF) classification technique, proved to be effective tools. This is an ensemble learning approach in machine learning called the RF classifier (Breiman, 2001). Multiple tree predictors are combined by RF, and the process begins with the careful collection and preparation of geospatial data, processing and classification, including satellite images and environmental variables. Once the data is in place, the random forest model is configured. The number of decision trees to be generated is given while configuring the forest model. The six input parameters that are typically used by the available RF classifiers in GEE are the number of trees used in classification, the number of variables used at each node, the random seed variable for decision tree construction, the minimum leaf population, the bagged fraction of the input variables for each decision tree, and the Out-of-Bag (OoB) mode. Up until overfitting-free convergence, the total classification accuracy increases with the number of trees (Rimal et al., 2017). Overfitting can be prevented if the RF

classifier's parameter values are improved. One way to select optimal parameters is to use the OoB outputs.

The core of the model consists of these decision trees, which have been trained to categorise different forms of land cover using the available data. Random Forest stands out due to its ensemble learning strategy. It builds an ensemble of decision trees, commonly referred to as K , rather than just one decision tree during training. The unpredictability introduced throughout this procedure is what gives it its distinctiveness. A selection of predictor variables is randomly selected at each node of the decision tree, reducing the possibility that any one variable would dominate the classification.

Importantly, the Gini index (Dangeti, 2017) plays a pivotal role in the decision-making process within each decision tree. It assesses data purity and helps in deciding how to optimally split the data. High Gini index values indicate the significance of specific predictor variables in influencing classification outcomes. A high Gini index value in Random Forest (RF) indicates the significance of the influencing factors. In this study, the Gini index $i(t)$ is measured as follows;

$$I_G(a_{X(x_i)}) = 1 - \sum_{j=1}^m f(a_{X(x_i)}, j)^2 \quad (\text{Equation 1})$$

Where, $f(a_{X(x_i)}, j)$ represents the proportion of the sample value X_i belonging to leaf j in node t . The decision tree's splitting criteria are determined based on the lowest Gini index value (IG).

Breiman (2001) proposed an advancement in decision trees with RF, which combines multiple single Random Trees (RT) using a rule-based approach. This approach aims to achieve higher accuracy compared to other machine learning algorithms, leveraging the combination of predictions.

In RF regression, typically, K regression trees are built, and the results are averaged. The predictor equation is as follows;

In RF regression, typically, K regression trees are built, and the results are averaged. After k number of trees $\{T(a)\}_1^k$ are grown, the predictor equation shown is the following;

$$\hat{f}_R^K(a) = \frac{1}{K} \sum_{k=1}^K T(a) \quad (\text{Equation 2})$$

In RF regression, the generalization error can be expressed as (Chen et al., 2018; Masetic and Subasi, 2016),

$$mg(x, y) = av_k I(h_k(x) = y) - \max_{j \neq y} av_k I(h_k(x) = j) \quad (\text{Equation 3})$$

Following the training phase as stated above, the ensemble of decision trees comes together to deliver more robust and accurate classifications. For classification tasks, the majority vote mechanism is often employed to assign the final class label, while in regression, individual tree predictions are averaged. To ensure the model's reliability, it undergoes rigorous evaluation, typically employing metrics such as accuracy and the Kappa coefficient. This validation phase ensures that the Random Forest model generalizes well to unseen data and produces dependable results.

Once validated, the Random Forest model becomes a powerful tool for various geospatial applications. It can be utilized to

generate geospatial maps, classify land cover, detect land use changes, and analyze a wide array of geospatial phenomena. This versatility makes Random Forest in Google Earth Engine an invaluable asset for researchers and analysts in fields like environmental monitoring and land management. It's particularly well-suited for handling complex geospatial datasets and delivering accurate, actionable insights (Rimal et al., 2017).

2.4.1. Training, validation, and testing

In the initial phase of data preparation for this study, a labeled dataset was essential for training the Random Forest model. The dataset was meticulously divided into features, representing independent variables, and labels, signifying the target variable. A total of 57,221 data points were thoughtfully selected and categorized for the purpose of classification. Subsequently, the bootstrapped sampling technique was employed during each iteration of tree building, involving the random selection of a subset of data with replacement. This ensured diversity among the trees in the forest. To prevent overfitting, feature randomization was implemented, where only a subset of features was considered at each split in the decision tree. The randomization steps were facilitated by the "random Column" function in Python. For training and testing, 70% of the data was utilized for model training, with the remaining 30% reserved for testing. The actual construction of multiple decision trees independently, each using a distinct subset of the training data, was efficiently managed by the "Random Forest Classifier" function in Python, culminating in an ensemble of predictions that constitutes the final output of the Random Forest model.

In the testing phase, a separate dataset constituting 30%, distinct from the training data, is reserved for the final evaluation of the model. This dataset remains untouched during both the training and validation phases, ensuring an unbiased assessment of the model's performance.

Validation was performed through Out-of-Bag (OOB) validation, utilizing a distinct validation set as out-of-bag samples for internal validation purposes. Additionally, interpretability was enhanced using SHAP values, a technique applied through the SHAP library. By employing SHAP values, the analysis delves into feature importance and individual predictions, shedding light on the contribution of each feature to the model's output. This approach enhances transparency and interpretability, providing valuable insights into the model's decision-making process.

2.4.2. Band Cross Co-linearity and PCA

The band cross co-linearity analysis used the product moment correlation (r) to assess the cross-co-linearity of Sentinel-2 band values at a predetermined training point. The correlation coefficients between different bands had to be specifically calculated at the designated training points. The degree of linear correlation between bands was measured in the study using cross-co-linearity and principal component analysis. A heat map was created in order to display the calculated correlation coefficients. This heat map was an effective instrument for determining the correlation values, which revealed whether or not there were stronger linear relationships between the bands. This all-inclusive method provided a visual aid to improve comprehension of the inter-band relationships at the designated training point, in addition to facilitating a quantitative assessment.

2.5. Principal Component Analysis (PCA)

The first step in working with a data matrix that has n samples and p variables is to centre the data according to each variable's mean. In this process, the spatial correlations and variances along the variables are preserved, and the data cloud is centered at the origin of our primary components. The first principal component (Y_1) is then derived as a linear combination of the variables X_1, X_2, \dots, X_p .

$$Y_1 = X_1 a_{11} + X_2 a_{12} + \dots + X_p a_{1p} \quad (\text{Equation 4})$$

or, in matrix notation,

$$Y_1 = X a_1^T \quad (\text{Equation 5})$$

The computation of the first principal component is designed to capture the maximum variance within the dataset. While it is possible to maximize the variance of Y_1 by selecting large values for the weights $a_{11}, a_{12}, \dots, a_{1p}$, measures are taken to prevent this. The weights are calculated with the aim of striking a balance, ensuring meaningful representation without disproportionately favoring large weight values.

$$a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1 \varphi_1^S \quad (\text{Equation 6})$$

The calculation of the second principal component follows a similar process, with the additional condition that it must be uncorrelated (perpendicular) to the first principal component. The objective remains to capture the next highest variance in the dataset while maintaining independence from the first principal component.

$$Y_2 = X_1 a_{21} + X_2 a_{22} + \dots + X_p a_{2p} \quad (\text{Equation 7})$$

This procedure is repeated until p principal components—the same number of variables as initially entered—have been calculated. At this point, all of the original data has been clarified or taken into consideration because the total of the variances of all the major components equals the sum of the variances of all the variables. When taken as a whole, these transformations of the original variables into principal components offer a comprehensive summary of the dataset.

$$Y = XA \quad (\text{Equation 8})$$

The computation of these transformations or weights is computationally intensive, requiring a computer for all but the smallest matrices. The rows of matrix A are referred to as the eigenvectors of matrix S_x , which represent the original data's variance-covariance matrix. An eigenvector is made up of weights (a_{ij}), also known as loadings. The diagonal members of matrix S_y , which represents the primary components' variance-covariance matrix, are also referred to as eigenvalues. These eigenvalues represent the variance explained by each principal component and, more critically, are required to reject monotonically from the first to the last. Typically, these eigenvalues are represented on a scree plot, demonstrating the decreasing rate at which variation is explained by additional principal components.

The positions of each observation in this new coordinate system of primary components are known as scores. These scores are calculated using linear combinations of the original variables and weights (a_{ij}). Specifically, the score for the r th sample on the k th principal component is calculated as follows:

$$Y_{rk} = X_{r1}a_{1k} + X_{r2}a_{2k} + \dots + X_{rp}a_{pk} \quad (\text{Equation 9})$$

Understanding the correlations between the original variables and the principal components might be useful when interpreting principal components. The relationship between variable X_i and primary component Y_j is represented by.:

$$r_{ij} = \sqrt{a_{ij}^2 V_{ar}(Y_j) / s_{ij}} \quad (\text{Equation 10})$$

2.6. SHAPley Additive exPlanations (SHAP)

The Shapley value is a concept in cooperative game theory that is used to divide the payoff among a group of players who work together to accomplish a shared objective in an equitable manner. When used with machine learning models, the Shapley value turns into an effective tool for figuring out how each attribute affects the model's output prediction. In a machine learning model, the Shapley value for a specific feature (let's call it feature i) must be calculated by taking into account all potential feature orderings or permutations. SHAP explains how input data layers or features contribute to final predictions using Shapley values (Kavzoglu and Bilucan, 2023; Shapley, 1997). In a coalition game, each agent (player) contributes to the outcome, as per game theory principles. The Shapley value determines an agent's efficacy and allows for fair payment based on their contribution to the game (Shapley, 1997).

The Shapley value represents an agent's marginal contribution among all conceivable combinations. The average marginal contribution of feature i for each of these permutations is then the Shapley value for feature i . The term "marginal contribution" describes how adding a particular feature alters the model's forecast. The Shapley value offers a thorough and equitable assessment of each feature's relevance by averaging across all potential feature orderings and accounting for the interactions between features (Shapley, 1997). The computation entails taking into account each potential feature coalition and assessing the effects of including feature i in each coalition. By taking into account every potential combination of characteristics, the Shapley value makes sure that every feature contributes fairly to the model's final forecast (Kavzoglu and Bilucan, 2023; Shapley, 1997). This method is very useful for feature significance analysis and interpretability in machine learning. It assists in answering inquiries such as, "How much does a particular feature contribute to the model's predictions, on average, when considering all possible interactions with other features?" The use of the Shapley value in machine learning not only improves our comprehension of the significance of each individual feature but also sheds light on the relationships and dependencies between features. Building confidence in the predictions produced by sophisticated machine learning models, optimizing them, and troubleshooting them may all benefit from this comprehensive knowledge.

The Shapley value ϕ_i for feature i is calculated using the following formula:

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (\text{Equation 11})$$

Where:

- N is the set of all features.

- S is a subset of features excluding feature i .
- $|S|$ is the number of features in subset S .
- $f(S)$ is the model's prediction when considering only the features in S .
- $f(S \cup \{i\})$ is the model's prediction when adding feature i to subset S .

This formula considers all possible combinations of features, calculates the marginal contribution of feature i to the model prediction, and takes the average over all permutations.

For Random Forest models, SHAP values are often approximated due to the complexity of the model. The SHAP library uses a kernel-shapley regression to approximate Shapley values for each instance. The formula for SHAP values is as follows:

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \cdot \phi_i^S \quad (\text{Equation 12})$$

Where:

- $\phi_i(f)$ is the SHAP value for feature i .
- ϕ_i^S is the Shapley value for feature i in the subset S .

Understanding the complex links between input data and model predictions in machine learning is made easier with the help of the SHapley Additive exPlanations library. Its ability to effectively approximate Shapley values, a notion with strong roots in cooperative game theory, is one of its most notable qualities. These values promote interpretability in intricate models such as Random Forest by providing a detailed knowledge of the contribution of each component to the overall forecast (Hosseiny et al., 2022). Shapley values are combinatorial in nature, which directly contributes to their implementation difficulty. Shapley values are calculated by taking into account every conceivable combination of characteristics, which soon becomes unfeasible as the number of features increases. Understanding this computing issue, the SHAP library uses a clever approach to overcome these barriers by combining sampling and regression approaches. An essential component of the SHAP library's methodology is sampling. By selectively sampling subsets, SHAP creates a representative subset of the power set of features as opposed to exhaustively analyzing every conceivable subset of features (Kavzoglu and Bilucan, 2023). The library is able to capture the essence of feature interactions thanks to this creative sampling method, all without giving in to the excessive computing demands of analyzing every possible combination. Regression methods are employed to supplement the sampling approach. For every sampled subset of characteristics, the library uses regression models to approximate the machine learning model's output. Through the careful selection of a sample of data for training, SHAP is able to capture the complex link between characteristics and predictions. The SHAP library uses the tree structure to speed up the Shapley value computation for models like Random Forest, which is complicated by nature because they are ensemble models (Kavzoglu and Bilucan, 2023; Shapley, 1997). It makes use of the additive properties of tree ensembles to streamline the process of calculating Shapley values for individual trees and then logically merging these values (Temenos et al., 2023).

2.7. Experiments on band compositions

The section focused on the examination of Sentinel-2 image band values, particularly analysing how they differed based on different classifications. Based on existing data points, data variations for deep green, green, and blue bands were plotted in this process. Determining the cross-correlations between these bands revealed important information about their interactions. Then, using these correlation values as a guide, a carefully designed set of tests was organized. Finding the best band combinations to improve Random Forest classification performance was the main objective of this analysis. Following the utilization of the chosen combinations in the categorization procedure, the accuracy level was precisely evaluated. Shapley values were calculated in order to provide further support for and interpretation of the categorization results. These values were useful in attributing the contributions of different bands to the overall classification findings and offered an in-depth knowledge of each band's usefulness in the process. Additionally, QGIS visual inspections were used to assess the findings' reliability. In order to validate and comprehend the classification findings, this phase entailed a thorough analysis of the categorized results within the geographic information system. The objective of this multifaceted method was to assure a thorough and reliable examination of the Sentinel-2 image categorization findings by integrating quantitative analysis, correlation evaluations, Shapley value computation, and visual inspections.

3. Results and discussion

In this study, our research identified blue and green infrastructure using a dataset including 57,221 data points, and a random forest classification model was carefully developed in Python. The intricacies of the model's accuracy are systematically presented and quantified in Table 2, providing a comprehensive insight into its performance. After the model evaluation, the calculation of SHAP values for every classification was done as a deeper level of analysis. This detailed analysis provided significant information about the contributing factors affecting the classification results. The subsequent exploration of these SHAP values culminated in the computation of mean SHAP values, an aggregation that provides a consolidated perspective across all three classification categories.

Table 2. Shows the classification report for random forest classifier

Class	Precision	Recall	f1-score	Support
Deep Green	0.97	0.96	0.97	7166
Green	0.96	0.96	0.96	4975
Blue	0.99	0.98	0.99	5026
accuracy			0.97	17167
macro avg	0.97	0.97	0.97	17167
weighted avg	0.97	0.97	0.97	17167

The graphical representation of mean SHAP values in Figure 4 presents interesting information about how specific

spectral bands play an important role in the classification process. Bands B8A, B5, B7, B3, B6, and B12 show as essential contributors, each with an identifiable weight in influencing classification outcomes. Of particular significance is band B8A's prominence, which has been identified as the most important among the factors considered, emphasising its significant role in the overall classification framework. This thorough research not only improves our understanding of the model's fundamental workings, but it also provides insight into the spectral properties that play an important role in accurately distinguishing between blue and green infrastructures.

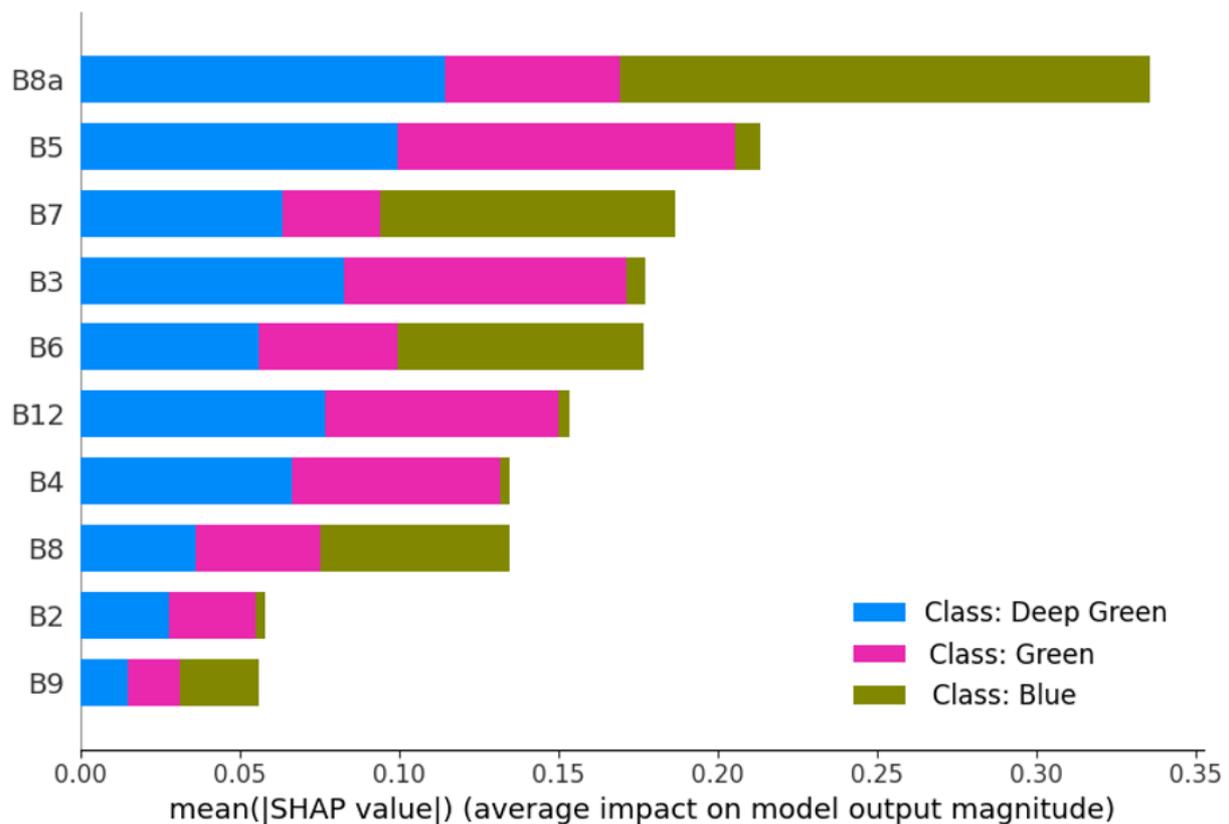


Figure 4. Shows the output for mean SHAP value random forest classifier

Figure 4 illustrates how taking a deeper understanding of the SHAP values within each class reveals unique spectral characteristics that are essential to the identification method. The SHAP values highlight the fundamental significance of bands B8A, B7, B6, and B8 within the Deep Green class. These particular spectral bands show evident as significant contributors, defining the subtle qualities that set Deep Green infrastructure apart in the sample. In the green class, bands B8A, B5, B4, and B3 play a crucial role, as indicated by the SHAP values. The key determinants that characterise and define the green infrastructure class are all of these spectral bands considered altogether. In this particular context, the small variations that the SHAP values provide offer important new understandings of the different spectral fingerprints associated with green infrastructure.

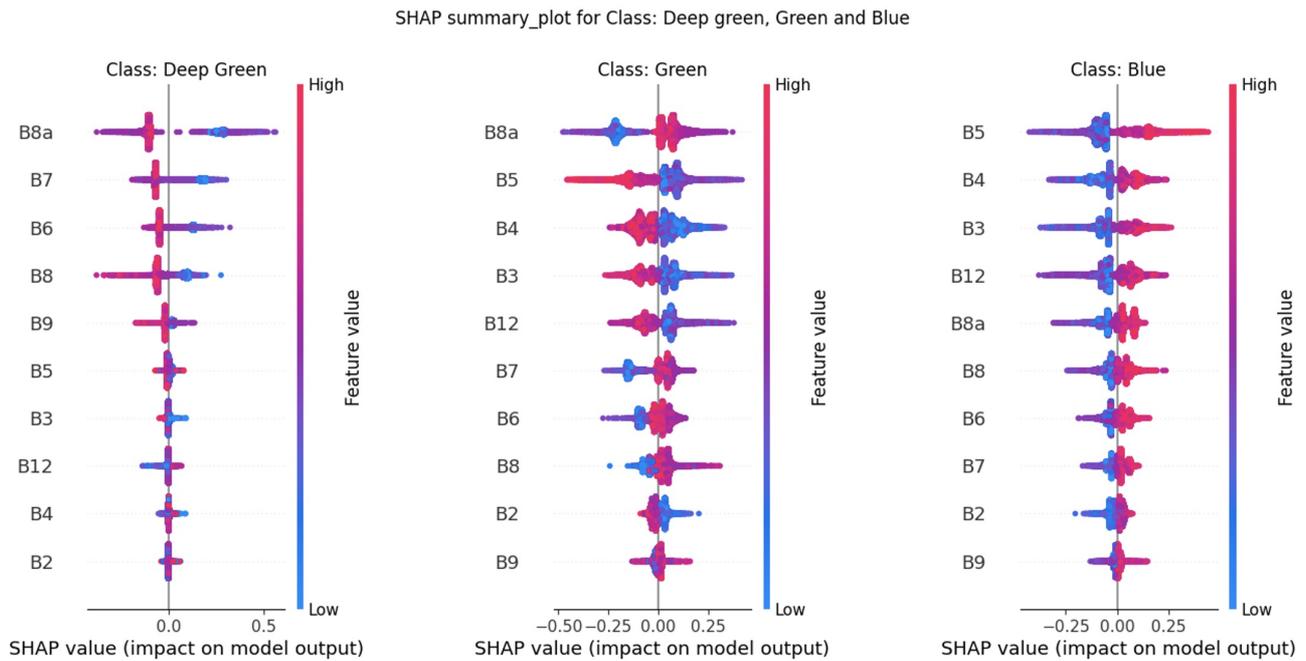


Figure 5. Bee swarm plot for bands for deep green, green, and blue categories. The gradient colour of the dots represents the band's value transitioning from low (blue) to high (red). The input variables are arranged on the y-axis according to their relevance, with the most influential variables at the top. The SHAP value on the x-axis represents the chance of being predicted as the target's textural class. The higher the SHAP value, the greater the likelihood.

In the blue class, a specific set that comprises important bands resulted. The SHAP values in this case clearly show that band B5 is the most important factor in the classifying method. In addition, bands B4, B3, and B12 are important in identifying characteristics distinctive to the blue class. This thorough analysis highlights how crucial each spectral band is to capturing the unique characteristics of blue infrastructures in the dataset.

In addition to improving our comprehension of the spectral characteristics influencing categorization results, the detailed examination of SHAP values within each class allows for a focused and knowledgeable appraisal of the key bands or factors in each infrastructure category—Deep Green, Green, and Blue. This thorough spectral categorization offers an accurate understanding for improving the model and enhancing our understanding of the complex relationship between infrastructure classification and spectral bands.

3.1. Class: Blue (water) and related infrastructure

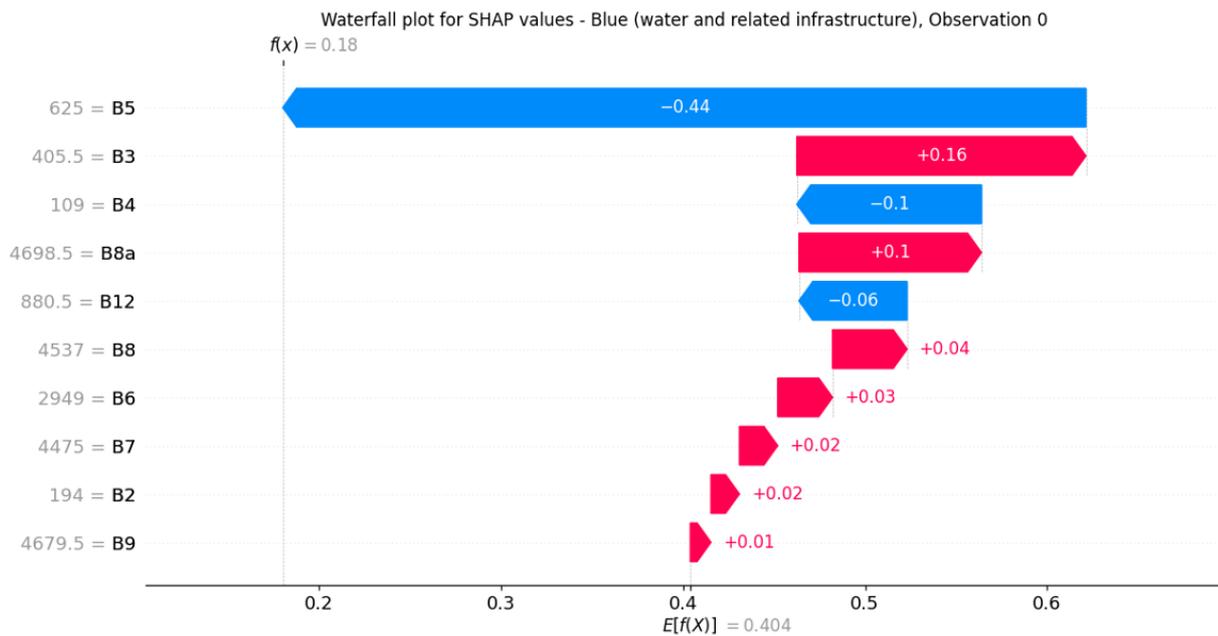


Figure 6. Waterfall plot for SHAP values for the Class - Blue (water) and related infrastructures

In Figure 6, Bands B5, B4, and B2 are negatively correlated to the water features with the values -0.44, -0.1, and -0.06, respectively, which are in the INR and visible ranges. The remaining bands exhibit positive correlation in the SHAP values. From the figure, it is also visible that water primarily reflects electromagnetic radiation in the visible range (Band B2 – B4). However, since water reflects very little in the NIR range (B5), the SHAP negative values are higher in B5.

3.2. Class: Deep green (Big Trees and Forest) and related infrastructure

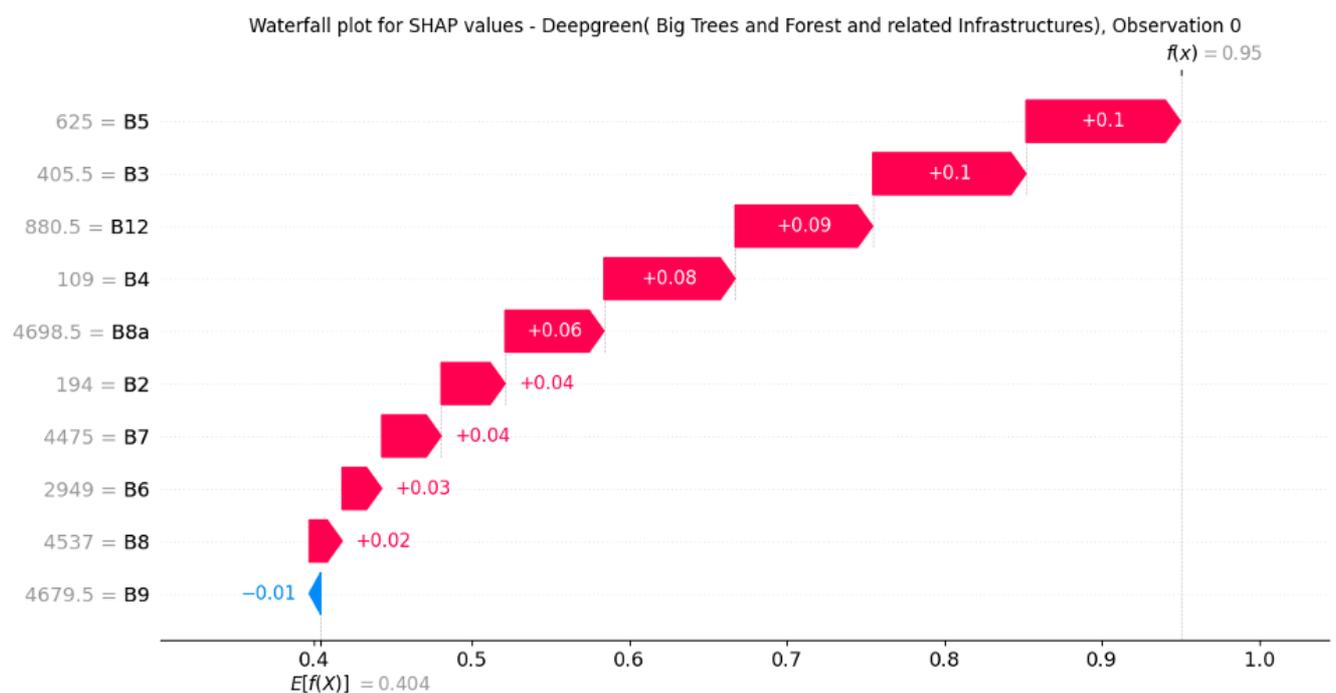


Figure 7. Waterfall plot for SHAP values for the Class - Deep green (Big Trees and Forest) and related infrastructures

In Figure 7, all bands except B9 are found positively correlated for trees and forest class identification. In vegetation, a significant reflectance increment happens in the red-edge bands (B5, B6, B7) and NIR bands (B8, B8a), compared with the red band (B4). The Red Edge Band is sensitive to changes in the chlorophyll content of plants, which can help distinguish between different types of vegetation. The Near-Infrared (NIR) Bands (B6 – B8) are highly reflective for vegetation. They are sensitive to the structure of plant leaves and canopies. That is why they are showing positive correlation in the SHAP values. The Shortwave-Infrared (SWIR) Bands (B8 – B12) are sensitive to the water content of plants. That is why B12 and B11 are also showing positive SHAP values in the figure.

3.3. Class: Green (Grass land and Harb) and related infrastructure

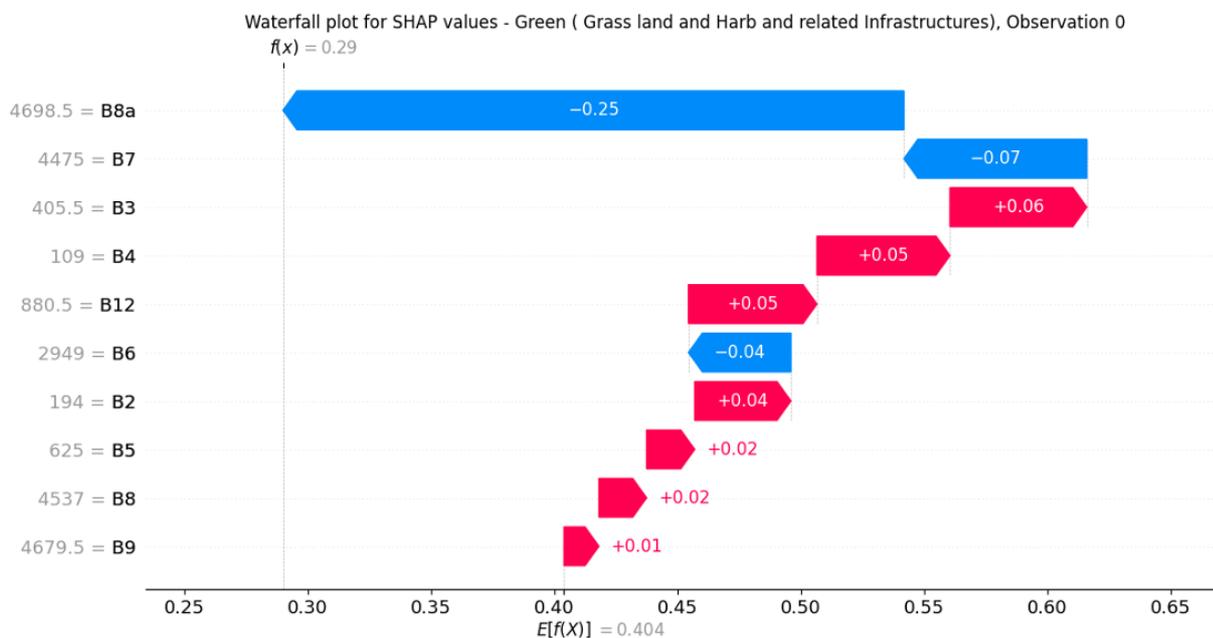


Figure 8. Waterfall plot for SHAP values for the Class – Green (Grass land and Harb) and related infrastructures

Since the visible range bands are sensitive to chlorophyll absorption in plant leaves, both forest and grass cover have similar spectral profiles. The Green Band (B3) is useful for soil and vegetation discrimination. However, in Figure 8, Bands B8A and B7 show stronger negative correlation. The Red and Blue Edge Bands (B4 and B3) are sensitive to chlorophyll content and can help distinguish between different types of vegetation. The positive response of the B5 band value here distinguishes between large forest and grass cover. Near-Infrared (NIR) bands are highly reflective for vegetation. They are sensitive to the structure of plant leaves and canopies. The Red Band (B4) is absorbed by chlorophyll, which results in darker plants. The difference in the B4 value is visible between Figure 6 and Figure 7, which differentiates two different types of vegetation: forest cover and grassland.

3.4. Cross-Co Linearity and PCA

The heat map produced (Figure 9) during the cross-collinearity analysis reveals several interesting patterns, especially in the Deep Green and Green classes. There is a notable similarity between the cross-co-linearity patterns of these two classes; however, the Deep Green class shows a stronger association. When examining Band B3's impact on the cross-collinearity dynamics between the Deep Green and Green classes, a significant difference becomes apparent.

In the case of the Deep Green class, Band B3 shows a negative correlation, introducing a unique characteristic to this class. However, in the Green class, the dynamics change, with Band B3 showcasing a negative correlation with Bands B6, B7, B8, B8A, and B9. This contrast highlights the important role of Band B3 as a differentiator between Deep Green and Green infrastructure. The complicated nature of the spectral properties contributing to the classification outcomes is highlighted by the nuanced interplay of correlations, providing insightful information for future refinement and interpretation.

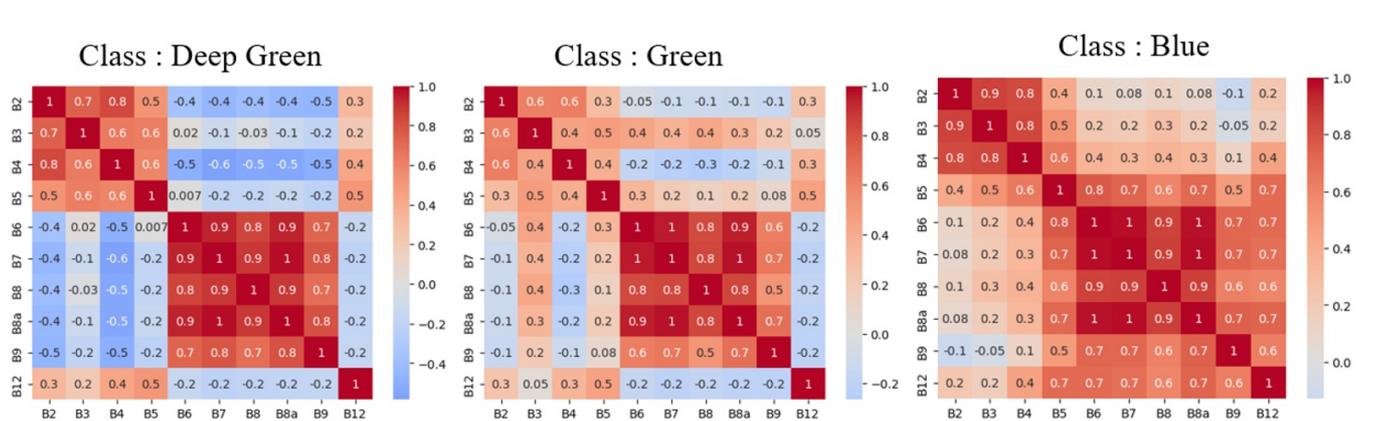
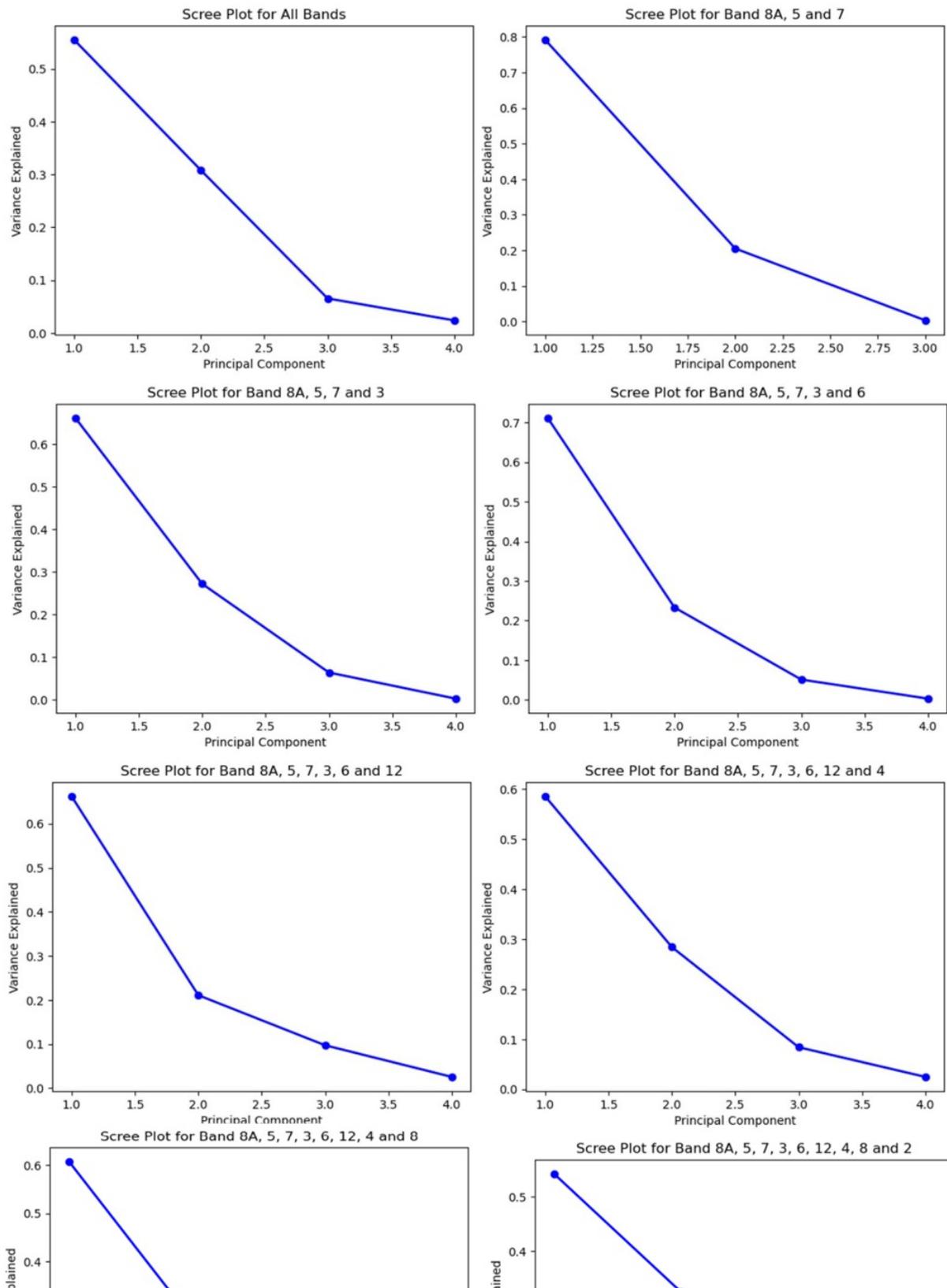


Figure 9. Heat map for cross-co-linearity analysis.

For the Blue class, a distinct set of cross-co-linearity relationships emerges. Specifically, Bands B2 and B3 display a negative correlation with Band 9. This finding implies a unique spectral interplay within the Blue class, where these particular bands contribute to the classification process by exhibiting noticeable correlations. Understanding these cross-co-linearity tests provides an accurate understanding of the spectral relationships associating the distinctions within the Blue class.

Figure 10 shows the Scree plot produced using Principal Component Analysis (PCA), which offers a thorough evaluation of variance for each of the bands: B2, B3, B4, B5, B6, B7, B8, B8A, B9, and B12. Major band combinations were carefully examined to determine their effect on the variance explained in order to further explore the analysis. These key combinations include (B8A, B5, and B7), (B8A, B5, B7, and B3), (B8A, B5, B7, B3, and B6), (B8A, B5, B7, B3, B6, and B12), (B8A, B5, B7, B3, B6, B12, and B4), (B8A, B5, B7, B3, B6, B12, B4, and B8), and (B8A, B5, B7, B3, B6, B12, B4, B8, and B2).

Within these combinations, the variance explained was carefully evaluated. Remarkably, the combinations (B8A, B5, and B7) and (B8A, B5, B7, B3, and B6) showed as particularly influential in capturing a substantial variance. The particular analysis of these combinations revealed that they not only contributed significantly to explaining variance but also held promise for enhancing the overall classification results.



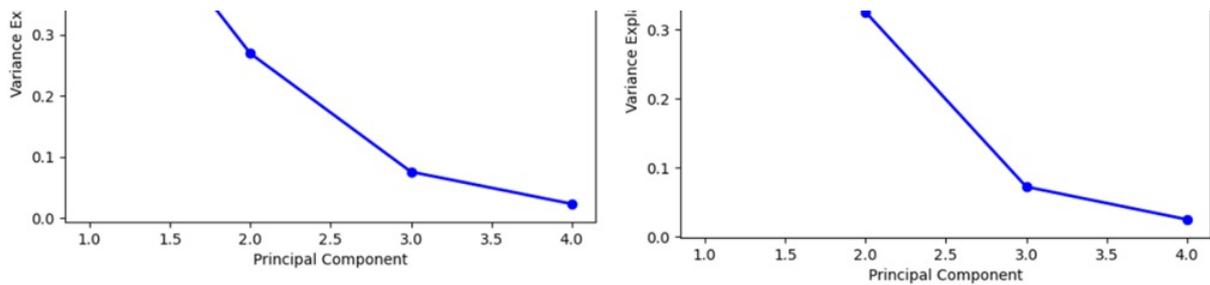


Figure 10. Shows the scree plot of principal component analysis

Figure 11 visually summarizes the comparative classification outcomes, indicating that the combination (B8A, B5, B7, B3, and B6) yielded better results. This finding highlights the efficacy of this specific band combination in enhancing the discriminatory power of the classification model. Consequently, the exploration of PCA-driven band combinations not only helps in understanding the underlying variance but also provides practical insights for optimizing the model's performance through optimal band selection (Table 3). This detailed approach towards band combinations within the PCA framework contributes to refining the model's capabilities and, consequently, the accuracy of infrastructure classification.

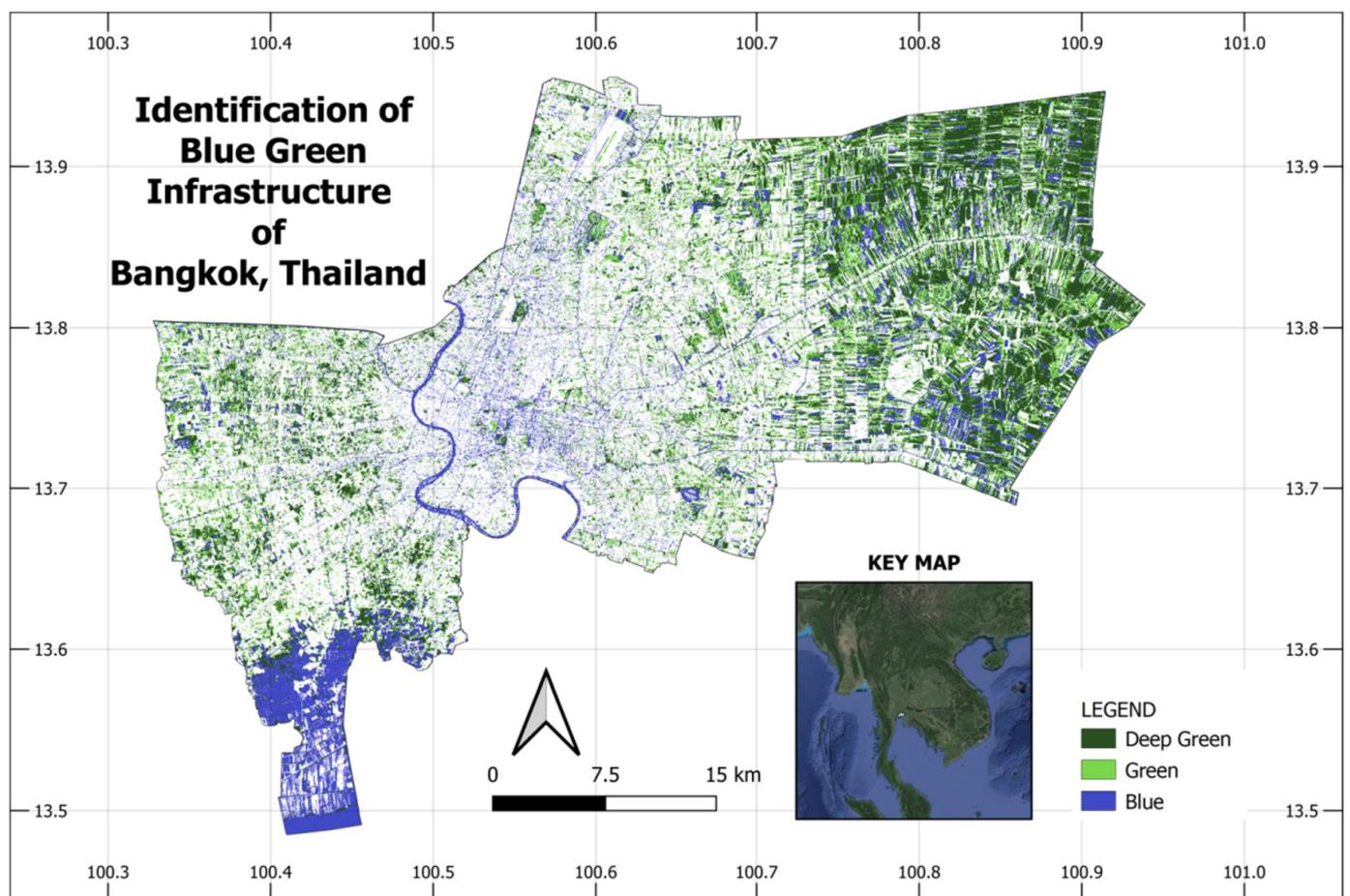


Figure 11. Final classification output for blue and green infrastructures in the Bangkok metropolitan region

In this study, we found that water features show a strong negative correlation in SHAP values. The gradual decrease in reflectance from band 1 to band 12 was specific to water bodies, which makes it distinct from other classes (Marwal and Silva, 2023). Water reflects electromagnetic radiation (EMR) in the visible range. However, water has a lower reflectance than vegetation and soils. Vegetation can reflect up to 50% of incident radiation, soils 30-40%, and water no more than 10% (Du et al., 2016). Water reflects little in the NIR range. Due to the fact that water bodies have a tendency to absorb more NIR light and produce lower reflectance values, the NIR band is frequently used in water body detection. The SWIR bands are particularly useful for water body mapping. The Modified Normalized Difference Water Index (MNDWI), a popular method for water body mapping, is calculated from the Green and SWIR bands (Du et al., 2016; Jiang et al., 2020).

In our research, we found that b4 and b12 showed a strong negative correlation and b8 and b8a a positive correlation. In vegetation, there is a significant reflectance increment in the red-edge bands (B5, B6, B7) and NIR bands (B8, B8a), compared with the red band (B4). The chlorophyll inside is sensitive to visible light: blue, green, and red. A good portion of incoming radiation is reflected in the green band. That's why leaves appear to be green in color. The Red Edge Band is sensitive to changes in the chlorophyll content of plants, which can help distinguish between different types of vegetation (Schuster et al., 2012). Wang et al., 2018 reported that the three most important bands of Sentinel-2 MSI data for vegetation class identification were band 4 (Red), band 12 (SWIR2), and band 8a (NIR2).

Several studies have demonstrated relationships between spectral properties in optical data and vegetation proxies, such as grass cover percentage and grassland biomass (Gao, 2006; Wang et al., 2019). Several bands like green, red, red edge, NIR, and SWIR are very helpful in identifying this land cover. The Green Band (B3) is useful for soil and vegetation discrimination (Schuster et al., 2012). Red wavelengths are absorbed by chlorophyll, which results in darker plants. The Red Edge wavelength is sensitive to chlorophyll content and can help distinguish between different types of vegetation (Schuster et al., 2012). Near-Infrared (NIR) Bands (B8, B8A): These bands are particularly useful for vegetation mapping as they are highly reflective for vegetation. Shortwave-Infrared (SWIR) Bands (B11, B12): These bands can help differentiate between different types of vegetation and are used in the agriculture band combination (B11, B8, B2).

The suggested classification scheme serves as the basis for assessing blue and green infrastructure for specific urban planning and climate mitigation aims. However, it may also be beneficial for evaluating the performance of other urban ecological and ecosystem services. The study aims to develop a method based on a complete set of sentinel datasheets and can be used by researchers and practitioners to classify and describe existing blue green infrastructure conditions, report and compare observations, and predict future scenarios with consistency and efficiency (Labib and Harris, 2018). New opportunities for expanding the present research to a larger audience include implementing the current methodology and testing it in combination with comparable approaches to enhance classification accuracy for future urban planning initiatives.

4. Conclusion

In conclusion, using a dataset of 57,221 data points, this study successfully created a robust random forest classification model for analysing blue and green infrastructures. The model's accuracy was systematically evaluated, and its interpretability was enhanced through SHAPley Additive exPlanations values, offering insights into the importance of specific spectral bands. Within individual infrastructure classes, i.e., Deep Green, Green, and Blue, the study uncovered distinctive spectral signatures influencing the identification process. Particularly, bands such as B8A, B7, B6, and B8 played significant roles in the Deep Green class, while B8A, B5, B4, and B3 were pivotal for the Green class. In the Blue class, B5 stood out as the most crucial contributor, emphasizing the importance of individual spectral bands in characterizing infrastructure classes. The findings underscore the utility of machine learning models in infrastructure classification and highlight the significance of specific spectral features in distinguishing between different land cover types. The study contributes valuable insights for optimizing classification results and refining models for accurate blue and green infrastructure mapping.

References

- Avci, C., Budak, M., Yağmur, N., Balçık, F., 2023. Comparison between random forest and support vector machine algorithms for LULC classification. *International Journal of Engineering and Geosciences* 8, 1–10. <https://doi.org/10.26833/ijeg.987605>
- Benedict, M.A., McMahon, E.T., 2012. *Green infrastructure: linking landscapes and communities*. Island press.
- Bengio, Y., 2009. Learning Deep Architectures for AI. *FNT in Machine Learning* 2, 1–127. <https://doi.org/10.1561/2200000006>
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cortes, C., Vapnik, V., 1995. Support-Vector Networks. *Machine Learning* 20, 273–297. <https://doi.org/10.1023/A:1022627411411>
- Davivongs, V., Yokohari, M., Hara, Y., 2012. Neglected Canals: Deterioration of Indigenous Irrigation System by Urbanization in the West Peri-Urban Area of Bangkok Metropolitan Region. *Water* 4, 12–27. <https://doi.org/10.3390/w4010012>
- Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band. *Remote Sensing* 8, 354. <https://doi.org/10.3390/rs8040354>
- Ekanayake, B., Ahmadian Fard Fini, A., Wong, J.K.W., Smith, P., 2022. A deep learning-based approach to facilitate the as-built state recognition of indoor construction works. *CI*. <https://doi.org/10.1108/CI-05-2022-0121>
- Gao, J., 2006. Quantification of grassland properties: how it can benefit from geoinformatic technologies? *International Journal of Remote Sensing* 27, 1351–1365. <https://doi.org/10.1080/01431160500474357>
- Goeyinsup, K., 2022. Impacts of future sea level rise on fluvial and coastal flood risk in Bangkok, Thailand. <https://doi.org/10.26021/12899>
- Hosseiny, B., Abdi, A.M., Jamali, S., 2022. Urban land use and land cover classification with interpretable machine learning – A case study using Sentinel-2 and auxiliary data. *Remote Sensing Applications: Society and Environment*

- 28, 100843. <https://doi.org/10.1016/j.rsase.2022.100843>
- Jiang, W., Ni, Y., Pang, Z., He, G., Fu, J., Lu, J., Yang, K., Long, T., Lei, T., 2020. A NEW INDEX FOR IDENTIFYING WATER BODY FROM SENTINEL-2 SATELLITE REMOTE SENSING IMAGERY. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* V-3–2020, 33–38. <https://doi.org/10.5194/isprs-annals-V-3-2020-33-2020>
 - Kavzoglu, T., Bilucan, F., 2023. Effects of auxiliary and ancillary data on LULC classification in a heterogeneous environment using optimized random forest algorithm. *Earth Sci Inform* 16, 415–435. <https://doi.org/10.1007/s12145-022-00874-9>
 - Khadem, H., Nemat, H., Elliott, J., Benaissa, M., 2022. Interpretable Machine Learning for Inpatient COVID-19 Mortality Risk Assessments: Diabetes Mellitus Exclusive Interplay. *Sensors* 22, 8757. <https://doi.org/10.3390/s22228757>
 - Labib, S.M., Harris, A., 2018. The potentials of Sentinel-2 and LandSat-8 data in green infrastructure extraction, using object-based image analysis (OBIA) method. *European Journal of Remote Sensing* 51, 231–240. <https://doi.org/10.1080/22797254.2017.1419441>
 - Limthongsakul, S., Nitivattananon, V., Arifwidodo, S.D., 2017. Localized flooding and autonomous adaptation in peri-urban Bangkok. *Environment and Urbanization* 29, 51–68. <https://doi.org/10.1177/0956247816683854>
 - Luo, X., Tong, X., Pan, H., 2021. Integrating Multiresolution and Multitemporal Sentinel-2 Imagery for Land-Cover Mapping in the Xiongan New Area, China. *IEEE Trans. Geosci. Remote Sensing* 59, 1029–1040. <https://doi.org/10.1109/TGRS.2020.2999558>
 - Marks, D., Connell, J., Ferrara, F., 2020. Contested notions of disaster justice during the 2011 Bangkok floods: Unequal risk, unrest and claims to the city. *Asia Pacific Viewpoint* 61, 19–36. <https://doi.org/10.1111/apv.12250>
 - Marwal, A., Silva, E.A., 2023. Exploring residential built-up form typologies in Delhi: a grid-based clustering approach towards sustainable urbanisation. *npj Urban Sustain* 3, 40. <https://doi.org/10.1038/s42949-023-00112-1>
 - Meddage, D.P.P., Ekanayake, I.U., Weerasuriya, A.U., Lewangamage, C.S., Tse, K.T., Miyanawala, T.P., Ramanayaka, C.D.E., 2022. Explainable Machine Learning (XML) to predict external wind pressure of a low-rise building in urban-like settings. *Journal of Wind Engineering and Industrial Aerodynamics* 226, 105027. <https://doi.org/10.1016/j.jweia.2022.105027>
 - Mokhtari, R., Ayatollahi, S., Fatemi, M., 2019. Experimental investigation of the influence of fluid-fluid interactions on oil recovery during low salinity water flooding. *Journal of Petroleum Science and Engineering* 182, 106194. <https://doi.org/10.1016/j.petrol.2019.106194>
 - Pandey, P.C., Koutsias, N., Petropoulos, G.P., Srivastava, P.K., Ben Dor, E., 2021. Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers—a review of the state of the art. *Geocarto International* 36, 957–988. <https://doi.org/10.1080/10106049.2019.1629647>
 - Petcharat, A., Lee, Y., Chang, J.B., 2020. Choice Experiments for Estimating the Non-Market Value of Ecosystem Services in the Bang Kachao Green Area, Thailand. *Sustainability* 12, 7637. <https://doi.org/10.3390/su12187637>
 - Rai, A., 2020. Explainable AI: from black box to glass box. *J. of the Acad. Mark. Sci.* 48, 137–141. <https://doi.org/10.1007/s11747-019-00710-5>
 - Richards, J., 2005. Is there a best classifier? in: Bruzzone, L. (Ed.),. Presented at the Remote Sensing, Bruges, Belgium, p. 59820A. <https://doi.org/10.1117/12.637223>

- Rimal, B., Zhang, L., Keshtkar, H., Wang, N., Lin, Y., 2017. Monitoring and Modeling of Spatiotemporal Urban Expansion and Land-Use/Land-Cover Change Using Integrated Markov Chain Cellular Automata Model. *IJGI* 6, 288. <https://doi.org/10.3390/ijgi6090288>
- Schuster, C., Förster, M., Kleinschmit, B., 2012. Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing* 33, 5583–5599. <https://doi.org/10.1080/01431161.2012.666812>
- Shapley, L., 1997. 7. A Value for n-Person Games. *Contributions to the Theory of Games II (1953)* 307-317., in: Kuhn, H.W. (Ed.), *Classics in Game Theory*. Princeton University Press, pp. 69–79. <https://doi.org/10.1515/9781400829156-012>
- Temenos, A., Temenos, N., Kaselimi, M., Doulamis, A., Doulamis, N., 2023. Interpretable Deep Learning Framework for Land Use and Land Cover Classification in Remote Sensing Using SHAP. *IEEE Geosci. Remote Sensing Lett.* 20, 1–5. <https://doi.org/10.1109/LGRS.2023.3251652>
- Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., Yao, Y., Zhang, X., 2018. Assessment of Sentinel-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover. *Remote Sensing* 10, 1927. <https://doi.org/10.3390/rs10121927>
- Wang, J., Biljecki, F., 2022. Unsupervised machine learning in urban studies: A systematic review of applications. *Cities* 129, 103925. <https://doi.org/10.1016/j.cities.2022.103925>
- Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing* 154, 189–201. <https://doi.org/10.1016/j.isprsjprs.2019.06.007>
- WHO, 2017. Urban green spaces: a brief for action. World Health Organization. Abgerufen von. Available from <https://apps.who.int/iris/handle/10665/344116>
- Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H.S., Zhang, L., 2020. Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers. <https://doi.org/10.48550/ARXIV.2012.15840>