A COMPARATIVE STUDY OF SUPERVISED MACHINE LEARNING CLASSIFIERS FOR ARCGIS

MIRVJEN ULQINAKU, ANA KTONA

Department of Computer Sciences, Faculty of Natural Sciences,

University of Tirana

e-mail: mirvjen.ulqinaku@fshn.edu.al

Abstract

Classification of ortho-imagery is a fundamental task in remote sensing applications, aiding in land cover analysis, urban planning, and environmental monitoring. In this article, we evaluate the performance of various classification algorithms in ArcGIS for five distinct land cover classes, using both Visible and InfraRed bands. The evaluated algorithms include Maximum Likelihood, K-Nearest Neighbors, Random Trees, and Support Vector Machine. In addition, this article will further focus on the performance of the Random Trees algorithm, particularly examining its utilization of entropy. Ortho-imagery datasets were acquired and preprocessed to ensure consistency and accuracy across all tests. Each classification algorithm was trained and tested using the same dataset and performance metrics such as overall accuracy, kappa coefficient, precision, recall and F1-score. Results reveal notable variations in classification accuracy among the algorithms tested, on both bands. For the visible band, Random Trees algorithm emerged with the best performance, followed by the Support Vector Machine algorithm. In contrast, for the infrared band, the Support Vector Machine algorithm exhibited the highest accuracy among the classifiers studied, closely followed by Maximum Likelihood algorithm. These findings provide valuable insights in remote sensing applications, aiding in the selection of appropriate classification algorithms for ortho-imagery analysis across multiple spectral bands.

Key words: *Raster classification, Maximum Likelihood, K-Nearest Neighbors, Random Trees, Support Vector Machine, Entropy.*

Përmbledhje

Klasifikimi i ortoimazherisë është një detyrë parësore e aplikacioneve të detektimit në distancë, e cila ndihmon analizën e kategorizimit të terrenit, planifikimin urban dhe monitorimin e ambientit. Në këtë artikull, vlerësohet performanca e algoritmeve të ndryshme klasifikues në ArcGIS për pesë klasa të ndryshme të kategorizimit të terrenit, duke përdorur dy banda, atë të dukshmen dhe atë Infra-të-kuqe. Algoritmet e vlerësuara përfshijnë Maximum Likelihood, K-Nearest Neighbors, Random Trees dhe Support Vector Machine. Përveç kësaj, ky artikull do të përqendrohet më tej në performancën e algoritmit Random Trees, duke ekzaminuar veçanërisht përdorimin e entropisë. Për të siguruar një konsistencë dhe saktësi në të gjitha testet, setet e të dhënave për ortoimazherinë u përftuan dhe u para-procesuan. Çdo algoritëm klasifikimi është trajnuar dhe testuar duke përdorur të njëjtin set të dhënash dhe metrika performance si overall accuracy, kappa coefficient, precision, recall dhe F1-score. Rezultatet treguan ndryshime të dukshme në saktësinë e klasifikimit ndërmjet algoritmeve të testuara, kjo për të dy bandat. Për bandën e dukshme, algoritmi Random Trees doli me performancën më të mirë, ndjekur nga algoritmi Support Vector Machine. Nga ana tjetër, për bandën infra-të-kuqe, algoritmi Support Vector Machine tregoi saktësinë më të lartë midis klasifikuesve të studiuar, i ndjekur afër nga algoritmi i Maximum Likelihood. Këto gjetje ofrojnë njohuri të vlefshme në aplikacionet e detektimit në distancë, duke ndihmuar në zgjedhjen e algoritmeve të përshtatshme të klasifikimit për analiza ortoimazherie përmes shumë bandave spektrale.

Fjalë kyçe: *Klasifikim Raster, Maximum Likelihood, K-Nearest Neighbors, Random Trees, Support Vector Machine, Entropi.*

1. Introduction

Remote sensing technologies have revolutionized the way we perceive and analyze our environment, offering insights into land cover dynamics, urban planning, and environmental monitoring. Central to these applications is the classification of ortho-imagery, a fundamental task that involves categorizing pixels into distinct land cover classes. By leveraging the spectral information captured by various bands, classification algorithms play a crucial role in extracting meaningful information from imagery datasets.

This research evaluates the performance of several classification algorithms within ArcGIS Pro, focusing on ortho-imagery datasets of Tirana, the capital city of Albania. With its diverse landscape encompassing urban areas, water bodies, agricultural lands, and natural vegetation, Tirana serves as an excellent study area for assessing the effectiveness of classification techniques.

This study explores the performance of four classification algorithms: Maximum Likelihood, K-Nearest Neighbors, Random Trees, and Support Vector Machine. These algorithms are evaluated using both Visible and Infrared bands of Sentinel-2 and World Imagery satellites, ensuring a comprehensive analysis across different spectral ranges.

The research methodology involves acquiring and preprocessing ortho-imagery datasets to ensure consistency and accuracy across all tests. Each classification algorithm is trained and tested using the same dataset, and performance metrics such as overall accuracy, kappa coefficient, precision, recall, and F1-score are employed to assess their effectiveness.

The study explores further the performance of the Random Trees algorithm, specifically examining its use of entropy as a criterion for decision-making. By understanding how different algorithms take advantage of spectral and spatial patterns for classification, this article aims to provide valuable information for remote sensing applications and aid in the selection of appropriate algorithms for ortho-imagery analysis.

Through this research, we seek to contribute to the advancement of remote sensing techniques, facilitating more accurate and efficient land cover analysis, urban planning, and environmental monitoring efforts in diverse geographical contexts.

2. Data

2.1 Area

The focus of this article is on Tirana, the capital city of Albania. Situated in the central part of Albania, Tirana is bordered to the north by the cities of Vore, Kamez, Kruje, and Klos; to the west by Shijak, Durres, and Kavaje; to the south by Rrogozhine, Peqin, and Elbasan; and to the east by Librazhd and Bulqize.

In terms of administrative and territorial divisions, the Tirana County includes the municipalities of Tirana, Kamez, Kavaje, Rrogozhine, and Vore. This article specifically focuses on the municipality of Tirana, which consists of a total of 27 Administrative Units: 1–11, District 12–14, Baldushk, Bërzhitë, Dajt, Farkë, Kashar, Krrabë, Ndroq, Petrelë, Pezë, Shëngjergj, Vaqarr, Zall–Bastar, and Zall–Herr.

Situated at an average altitude of 110 meters above sea level, Tirana is delimited by its flat area to the north-northwest, by mountainous regions to the north-east and east, where the highest point of 1,828 meters (Mali me Gropa) is located, and by hilly terrain to the south, southwest, and west.

In terms of area, Tirana covers 1,120 km^2, with Shëngjergj being the largest Administrative Unit (208.5 km^2) and Nr. 10 the smallest (0.77 km^2).

2.2 Satellite Imagery

Aerial and satellite imagery, which constitutes a fundamental data source for Geographic Information Systems, refers to data captured from a distance, commonly acquired through satellite or aerial platforms. These images serve as a rich source of spatial data, offering advantages such as extensive coverage, extended spectral bands, and the potential for geometric accuracy through advanced correction techniques.

The acquired imagery for processing originates from a satellite source, with the following specifications:

- Sentinel-2 L2A: Sentinel-2 supplies data to Copernicus services and is equipped with a variety of technologies including multi-spectral imaging instruments for land, ocean, and atmospheric surveillance. It gives high-resolution imagery essential for land monitoring, emergency response, and security operations.

The resolution chosen for the images from this satellite is 10 m; Sensor MultiSpectral Instrument, with 13 bands[1], where for this study B4 (red, 665 nm), B3 (green, 560 nm) and B2 (blue, 490 nm) have been chosen as processing templates for Visual Band, and B8 (near-infra red, 842 nm), B4 and B3 for infrared band. For better comparison, both images from this satellite are dated 03.10.2023.

The coordinate system for all acquired images is ETRS 1989 Albania 2010.

[1] https://docs.sentinel-hub.com/api/latest/data/sentinel-2-l2a/

2.3 Models

Vector

The vector model is one of the primary models for representing and manipulating geographic data. In this model, geographic data is portrayed as geometric entities known as vectors, composed of points, lines and polygons. These vectors can be precisely described and modified to align with various geographic application requirements. The vector model is used in numerous functionalities including spatial analysis, and data visualization, making it a crucial option for a multitude of GIS applications.

The entities used in this article are polygons, which represent the administrative units of Tirana, as shown in Figure 1. In total, there are 27 polygons that include the entirety of the Tirana municipality, which will also be the focus of this study.

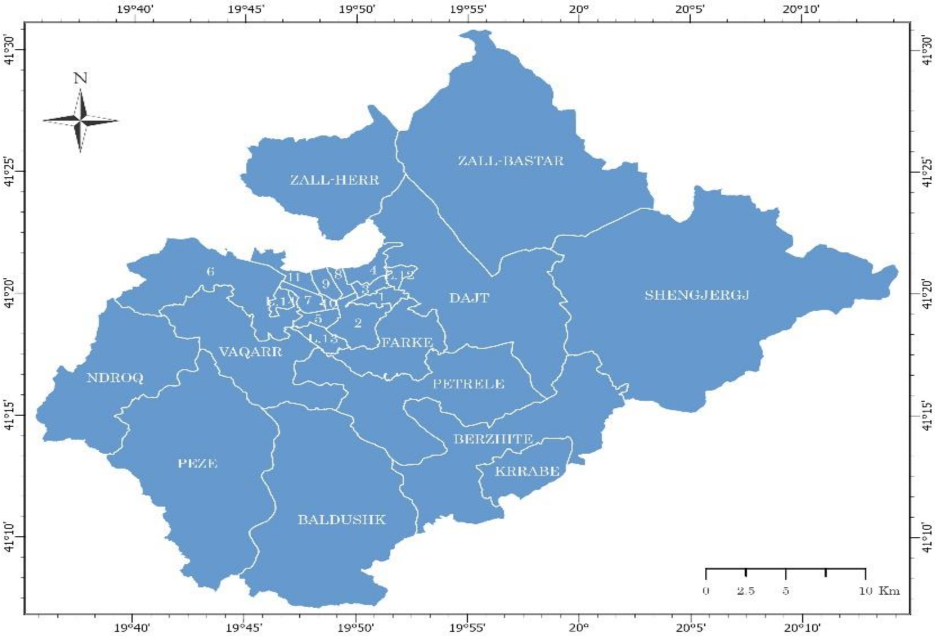


Figure 1: Vector layer of the 27 Administrative Units of Tirana.

Raster

Orthophoto technology, manned or unmanned, utilizes sophisticated digital processing to rectify aerial photographs like satellite or drone imagery, removing distortions caused by terrain relief, camera tilt, and lens distortion (Bolstad, 2016). By aligning the images to a uniform scale, orthophotos offer a seamless, geometrically accurate depiction of an area's topography, allowing for precise measurements and analysis.

Orthophotos rely on raster model, which consists of a grid of cells or pixels, each containing specific information about the terrain, land cover, and other features. Raster data forms the foundation for generating orthophotos, as it captures the details of an area's landscape with low, medium, or high resolution. Through raster analysis, orthophotos can be enhanced and manipulated to extract valuable insights about meteorology, land use land cover (LULC), agriculture, wildlife, infrastructure, disaster management, and even epidemic emergencies (Shahbazi et al., 2014).

The rasterization of Tirana's zones presents a mosaic-like structure, resulting from the amalgamation of raster data from all its administrative units, where each administrative unit is clipped based on the vector layer mentioned in the previous point. Two final orthophotos will be generated using the same method: one captured in the visual band and the other in the infrared band. This dual approach allows for comprehensive analysis across different spectral ranges.

The study area for Tirana, representing the final image upon which the study will be based, will be smaller in both cases compared to the final raster, given that the raster model forms a rectangular grid of rows and columns (Bugya & Farkas, 2018).

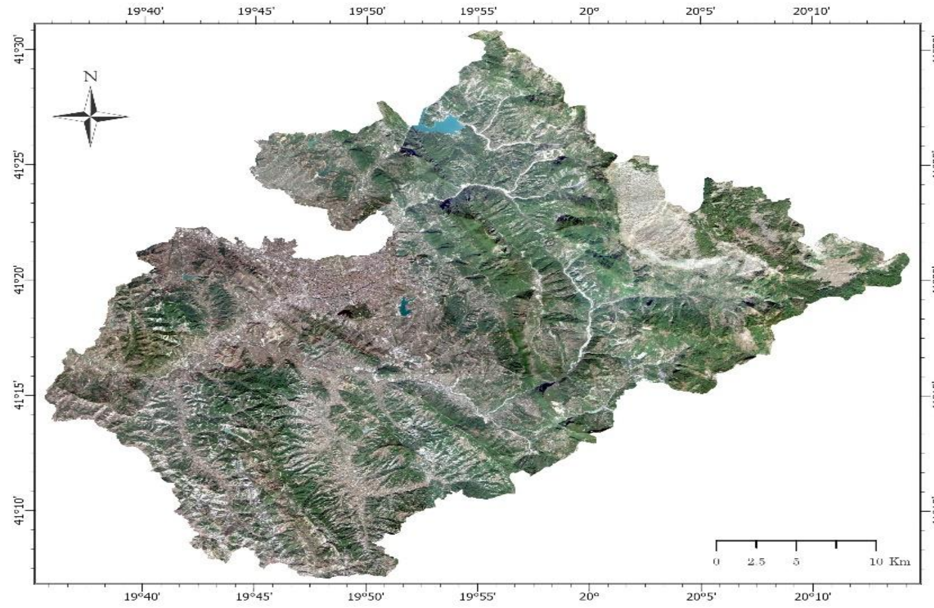


Figure 2: Visual Band raster of the study area.

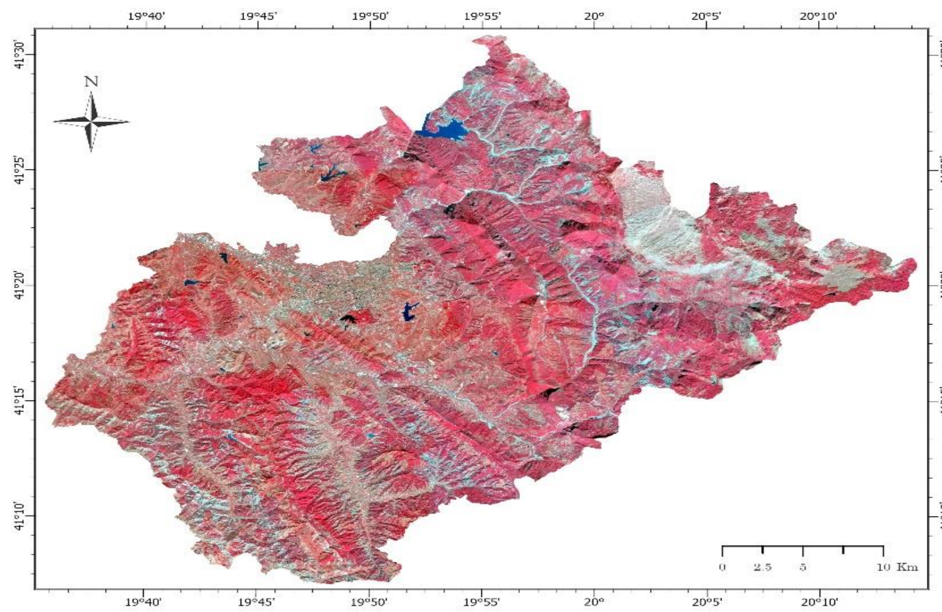


Figure 3: Infrared band raster of the study area.

2.4 Classes

One of the essential aspects of machine learning classifiers is the class, which is the output category of the classification. In this study, five major classes have been chosen for the study area, each representing distinct categories within a given dataset. With the aim of achieving a consistent environment for the comparison of classifiers for the area of Tirana, the same classes have been selected for both visual and infrared bands:

- Water: water bodies, which in the study area and resolution of the image consist of lakes and rivers.
- Developed: areas built by human activities, like residential, commercial, and industrial.
- Green: natural areas or areas covered by vegetation, which mainly include forests, grasslands, woodlands, and parks.
- Barren: areas which lack significant vegetation and contain very limited biodiversity, such as bare rocks, gravel, or soil.
- Agriculture: areas designated for farming or crops' cultivation.

2.5 Datasets

Classifiers, a particular type of machine learning algorithm, depend on datasets for training, testing, and performance evaluation. The best results are achieved through the use of balanced training and testing datasets (Wei & Dunbrack, 2013), which serve as the basis for teaching classifiers to recognize and classify data into categories. Selecting the right dataset is essential to ensure that classifiers develop effective and valuable models for their purpose.

For the training set of both the visual and infrared bands, efforts have been made to approximately capture the same surface area, aiming for a better comparison. The approximations of surface area for each class are detailed in Table 1.

The test set is composed of the whole map of Tirana, giving a total area of 1,120.9 km^2, or 11,209,000 pixels for classification.

Class	No. of pixels	Area (km^2)
Water	9,700	0.97
Developed	33,000	3.30
Green	54,000	5.40
Barren	54,000	5.40
Agriculture	57,000	5.70
TOTAL	207,700	20.77

Table 1: Number of pixels and area for each class included in the training set for the entire area of Tirana.

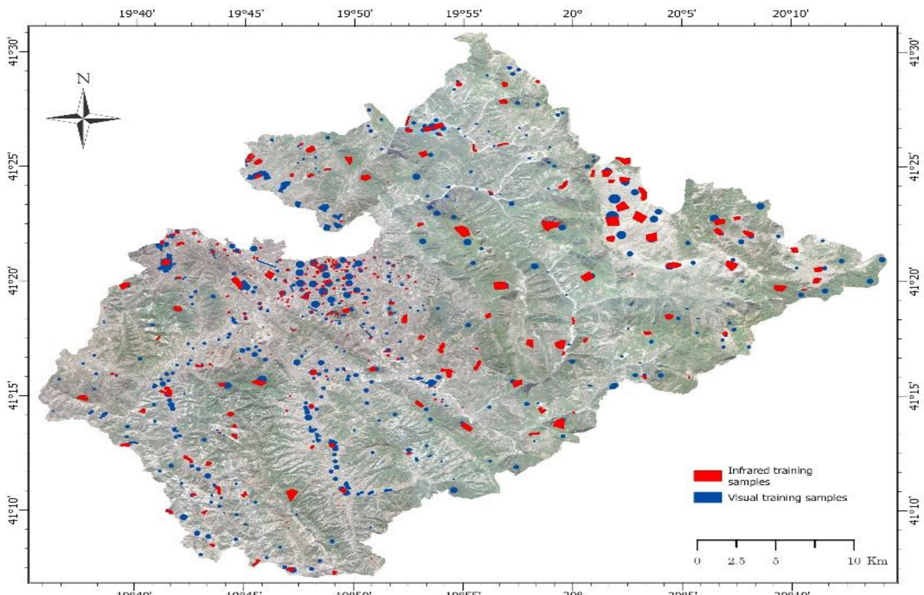


Figure 4: A map showing the distribution of training samples for both Visual and Infrared bands.

3. Classifiers

The objective of image classification is to categorize pixels into LULC classes using either spectral or spatial patterns. Spectral pattern recognition groups pixels based on spectral reflectance and classifies them individually, while spatial pattern recognition considers spatial relationships with surrounding pixels, like texture, proximity, and context (Lillesand et al., 2015).

Spectrally oriented procedures, like supervised classification, involve specifying groups of pixels with spectral attributes for LULC classes, a process called training. Then, other untrained pixels are compared to these values, thus creating the classification process. The output of the classification includes thematic maps, statistical tables, and digital files for GIS integration.

Whereas unsupervised classification begins with classifying image data into spectral clusters without prior training. Then, LULC classes are identified by comparing these clusters to ground reference data.

The concept of using unsupervised training areas to aid in identifying spectral classes for supervised classification, selected to represent diverse land cover types across the map, gives rise to a third type of procedure: hybrid classification.

The above approaches can be combined in hybrid modes, such as object-based image analysis, where both spectral and spatial pattern recognition are used. In this study, our focus lies on supervised classification, where we elaborate, use, and compare four classifiers available in ArcGIS Pro.

3.1 Maximum Likelihood

Maximum Likelihood Classifier (MLC) is a parametric method used in machine learning to classify data based on the likelihood of each class, given observed data. It involves collecting data, modeling its probability distribution for each category, calculating likelihoods, and assigning data to the category with the highest likelihood. It is presented by Equation 1, based on the Bayesian estimation:

$$P(H_i|D) = \frac{P(H_i) \cdot P(D|H_i)}{\sum_i P(H_i) \cdot P(D|H_i)} \quad (1)$$

where: $P(H_i|D)$ the probability of a true hypothesis H with given data D, also called posterior probability; $P(H_i)$ the probability of a true hypothesis before seeing the data, also called prior probability; $P(D|H_i)$ the probability of seeing the data D with a true hypothesis.

This classifier uses ellipses to represent the probability distribution of various classes, especially when using multivariate data. These ellipsoids are defined by the mean vectors and variance-covariance matrices of individual classes (Kavzoglu & Reis, 2008). Concentric ellipses centered on each class's mean vector are used to assess the likelihood probabilities of pixels for classification, and when new data points arrive, their likelihood of belonging to each class is evaluated based on how well they fit into the ellipses.

One disadvantage of MLC classification is the significant computational load needed to classify each pixel, especially when dealing with numerous spectral channels, so differentiating many spectral classes. But with the recent advancements in computational power, this drawback is now less of a concern for most applications (Lillesand et al., 2015). Another challenge that MLC faces is the distinction between probability distributions. When distributions are not clearly distinguished from each other, they tend to overlap, thus making it unclear which distribution corresponds to which class (Susaki & Shibasaki, 2000).

Despite its limitations, MLC is extensively used in classification tasks, particularly in scenarios where classes exhibit a Gaussian probability distribution, which renders it an optimal choice for classification (Tso & Mather, 2001).

Aside these drawbacks, MLC is widely used in classification, mostly when there is a Gaussian probability distribution of classes, which makes it an optimal classifier.

3.2 K-Nearest Neighbors

K-Nearest Neighbor (KNN) is a machine-learning method for regression and classification. It finds the k closest points (neighbors) to a given instance, based on a distance metric (Mitchell, 1997). The resulting equation using the standard Euclidian distance is:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (2)$$

where: x and y are instances and i is the instance's attribute.

KNN is a non-parametric classification method that operates on the principle of proximity. It finds the distance between a target instance and its nearest neighbors based on the specified value of k. The classification decision is then made based on the majority vote of these k neighboring instances (Avand et al., 2019).

KNN operates effectively with both discrete and continuous data types. It computes distances, such as Hamming distance, between instances for discrete data, while employing metrics like Euclidean, Manhattan, or Minkowski distances for continuous data. These distances are utilized to assign class labels or numerical values, facilitating the classification process. However, it's crucial to preprocess the data adequately to ensure that the algorithm can effectively handle the diverse nature of the data types involved.

Regardless of its popularity and simplicity in implementation, KNN method can display some issues. A main problem with KNN is the selection of k; if it is excessively large, there is a risk of including irrelevant data as nearest neighbors and ignoring minorities; if k is too small, it may fail to yield accurate results (Agrawal, 2014). Another prevalent issue encountered in practical applications of KNN is the uneven distribution of data across classes, leading to a bias towards the majority class, irrespective of distance measurements (Sun & Huang, 2010).

3.3 Random Trees

Random Trees (RT) is a classification method that uses the Random Forest (RF) algorithm (Breiman, 2001). RF is an ensemble learning method that constructs multiple decision trees during training and aggregates their outputs to make predictions. Each decision tree is built based on a subset of the training data and a random subset of features selected at each node. This randomness helps ensure diversity among the individual trees. The decision tree in itself, is a recursive supervised learning technique extensively applied in both classification and regression tasks. Unlike single-step methods, the decision tree adopts a hierarchical strategy to systematically explore, classify, and render decisions (Ulqinaku & Ktona, 2023). The goal of decision tree is to maximize Information Gain at each split (Quinlan, 1986), and it achieves this by using the concept of Entropy, more specifically Shannon's entropy, expressed by the equation:

$$H_i = -\sum_{i=1}^{n} p_i \cdot log(p_i) \qquad (3)$$

where $log \equiv log_2$, n is the number of system's different states and p_i is the probability of event i. The connection between Information Gain and Entropy is shown in the equation:

$$IG(TA, SA) = H(TA) - \sum_{v \in SA} \frac{TA_v}{TA} \cdot H(TA_v) \qquad (4)$$

where, H is the entropy, TA represents the target attribute, while SA is the supporting attribute, and TA_v denotes a subset of TA where SA attribute takes on the value v. Entropy, a measure of randomness or disorder in a dataset, is commonly used as a criterion to quantify the randomness of the data before and after a split. Information gain, on the other hand, measures the reduction in entropy achieved by splitting the data on a specific attribute.

As in other models, each tree in the RT model is trained on a sample of training data. For classification tasks, predictions are made by majority vote of the trees, while for regression tasks, the average output of the trees is used. RT builds multiple decision trees, each trained on a random subset of the training data, a process called bagging, and considering only a random subset of features for splitting. The use of bagging methods, compared to employing single classifiers like decision trees, resulted in higher accuracy while also reducing classification variance, but it had minimal impact on classification bias (Belgiu & Drăguţ, 2016).

According to (Belgiu & Drăguţ, 2016), RT classifiers can be susceptible to several disadvantages such as: struggling with imbalanced training data, leading them to favor the most representative classes and potentially causing biased classifications; sensitivity to spatial autocorrelation within the training classes, which implies that the spatial distribution of training samples may influence the classifier's performance; the frequent need for large number of training samples (Kulkarni & Sinha, 2013), which may not always be feasible to obtain, particularly in situations where collecting extensive training data is difficult or resource-intensive.

RT is applied in banking to identify loyal customers, detect fraud, and assess profitability; in medicine for diagnosing diseases through patient record analysis; in the stock market for predicting profits or losses from stock behavior; and in e-commerce for customer segmentation based on purchase patterns (Shaik & Srinivasan, 2019).

3.4 Support Vector Machine

Support Vector Machine (SVM) represents an advanced form of supervised learning, based on the principles of statistical learning theory (Vapnik, 1995). In classification tasks, SVM functions by identifying a hyperplane within the input space. The fundamental strategy in SVM is to find a hyperplane that achieves the best possible separation between classes. According to Kavzoglu & Colkesen (2009), there are multiple hyperplanes that can be used to separate two classes. Yet, among them, only one hyperplane has the ability to maximize the margin between these classes. This special hyperplane, known as the optimum hyperplane, is crucial in discerning the boundary with the greatest distance from the data points of each class. The data points that significantly influence the width of this margin are referred to as support vectors.

If we consider a binary classification scenario with two classes, the hyperplane is mathematically represented by:

$$w \cdot x_i + b = 0 \qquad (5)$$

where w represents the weight vector, which perpendicular to the hyperplane and determines its orientation; x_i is a feature vector containing i features; and b is the bias, indicating the distance of the hyperplane from the origin. If the left part of the equation gives a negative result, it indicates that the data points belong to one class, and if the result is positive, the data points belong to a different class. In real-world scenarios, linearly separating classes may not always be feasible. In such cases, SVM uses a Kernel function, which maps the data points into a higher-dimensional space where linear separation becomes possible (Oommen et al., 2008). In this case, the equation of the hyperplane becomes:

$$w \cdot \Phi(x_i) + b = 0 \qquad (6)$$

where $\Phi(x_i)$ is the kernel function.

Despite its ability to handle complex classification tasks through kernel functions, SVM also exhibits certain issues. The main one lies in their excessive computational cost, particularly when dealing with large datasets. This is attributed to the quadratic growth involved in determining the optimal hyperplane, demanding substantial computational resources and memory.

According to Mountrakis et al. (2011), SVM displays other limitation such as: the selection of kernel functions, where choosing a small kernel width parameter may result in overfitting, while large values may lead to over-smoothing; inadequacy for handling noisy data commonly encountered in remotely sensed data due to factors like measurement errors and atmospheric distortions; and lastly, the effectiveness of an SVM classifier can significantly diminish even with just a few mislabeled examples.

Even with their limitations, SVM have been employed to address numerous real-world issues such as (Cervantes et al, 2020): image classification, text categorization, bioinformatics, hand-written character recognition, face detection, and other complex classification problems.

4. Methodology

4.1 Workflow

After identifying the areas to be used for the study, they were located and obtained from imagery providers, then integrated into ArcGIS Pro. The images were processed both in terms of geographical extent and raster layer parameters. Additionally, a geographic database was established.

Subsequently, a classification scheme was developed, outlining the categories to be addressed, along with training samples for each category. The next step involved classification using the four classifiers provided by ArcGIS Pro. Following classification, the attributes of each resulting layer were processed to obtain meaningful data such as the number of pixels for each category and the respective area.

Following that, a layer of accuracy points was created, where 200 such points were taken for each classified layer, based on manually performed ground truthing. Using the accuracy points as a basis, a confusion matrix was generated to compare the outputs of the classifiers, based on accuracy and performance metrics mentioned in section 4.3.

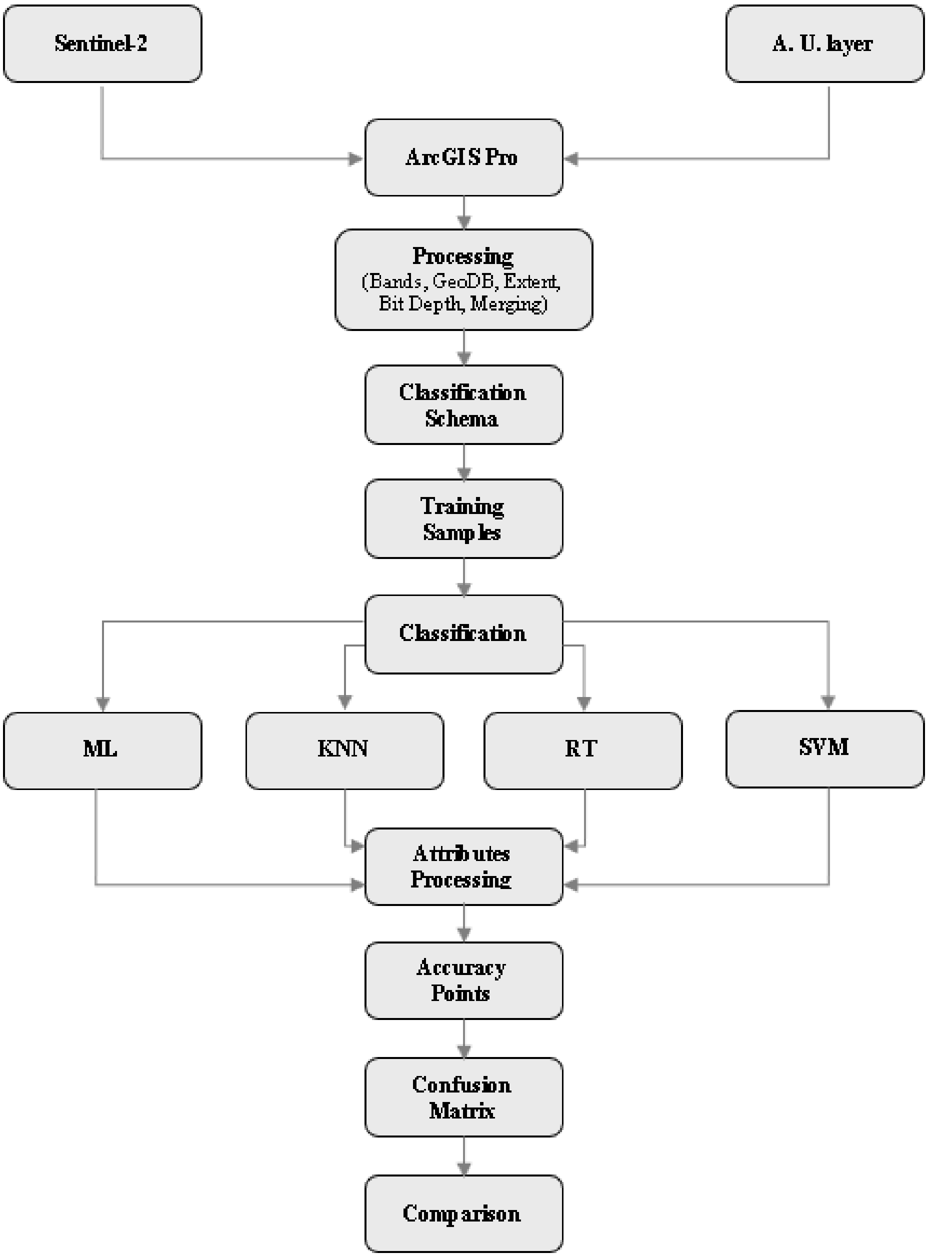


Figure 5: Methodology workflow of the study.

4.2 Classifiers' parameters

To enhance the classification output of each classifier, it is essential to tailor the parameters according to the desired outcomes. It is important to note that the parameters for each classifier were adjusted using a trial-and-error approach until satisfactory results were obtained. The parameters configured for each classification are as follows:

- For ML classifier, no input parameter configuration is required as it does not possess any.
- For KNN classification, two parameters were set: K-Nearest Neighbor: 5, and Maximum Number of Samples per Class: 100.
- For Random Trees, three parameters were set: Maximum Number of Trees: 50, Maximum Tree Depth: 30, and Maximum Number of Samples per Class: 100.
- In the case of SVM classifier, only one parameter can be set, which is Maximum Number of Samples per Class: 100.

4.3 Accuracy and performance metrics

As already mentioned in section 4.1, for each classified layer, another layer of accuracy points was created, as shown in Figure 6. It was created using "Accuracy assessment points" tool, which generates a set of random points and assigns them a class based on already classified data.

The sampling strategy for these points was "Stratified random", where points were created randomly within each class, with the number of points in each class being proportional to its relative area. This ensures that the distribution of samples is representative and minimizes biases.

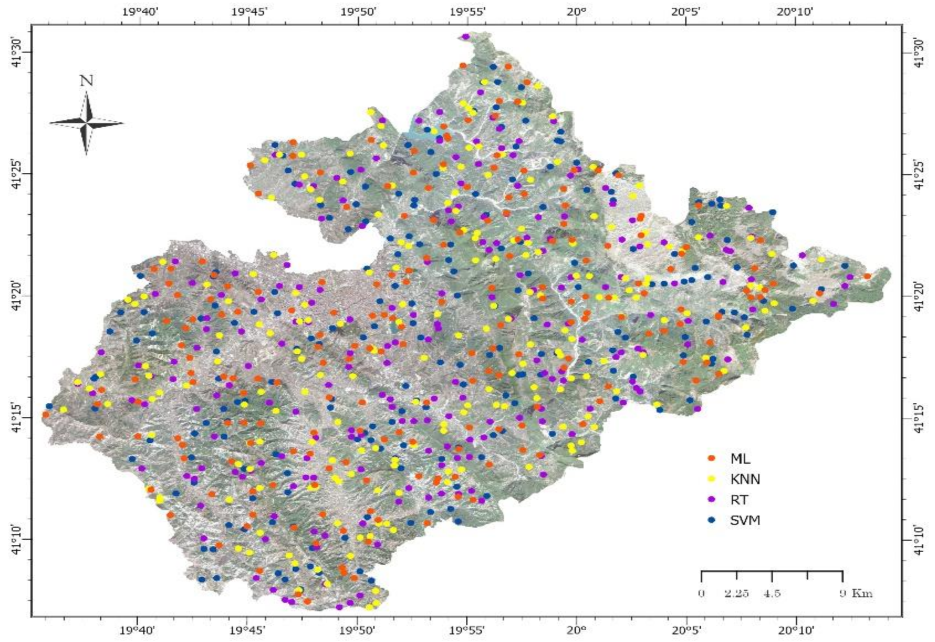


Figure 6: Distribution of accuracy points for each classification.

The confusion matrix created has four components: TP – true positive, TN – true negative, FP – false positive, and FN – false negative. The performance metrics, derived from the confusion matrix, and used in this study were:

- Overall accuracy: represents the ratio of correctly classified instances – TP and TN – to the total number of instances, as shown in the equation 7:

$$OA = \frac{TP + TN}{Total\ no.of\ instances} \qquad (7)$$

- Kappa: assesses the agreement between predicted classifications and true values, and is expressed by equation 8:

$$k = \frac{OA - RA}{1 - RA} \qquad (8)$$

where RA is the random accuracy.

- Precision: derived from the confusion matrix, precision measures the ratio of true positive classified instances among all positive classified instances. Its equation 9 is as follows:

$$Precision = \frac{TP}{(TP + FP)} \quad (9)$$

- Recall: measures the proportion of true positive classified instances, among all actual positive instances, and its equation 10 is:

$$Recall = \frac{TP}{(TP + FN)} \quad (10)$$

- F1-Score: provides a balanced measure between precision and recall, and is expressed by the equation 11:

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (11)$$

Results

The classification of Tirana's zones is conducted using the four methods outlined in Section 3, with the outcomes presented in the classified maps below.

Figure 7 shows that the ML classifier performs well in the visual band overall, but it tends to incorrectly classify water bodies, leading to numerous false positives. In the infrared band (Figure 8), this misclassification improves significantly; however, there is a notable issue with false positives in areas classified as Developed instead of Barren lands.

Similar to the ML classifier, the KNN method also encountered challenges with water bodies. However, rather than producing false positives, it resulted in false negatives. This means that many water bodies were incorrectly classified as a different class, indicating significant difficulties for this classifier in accurately identifying water (Figure 9 and Figure 10).

Among the four classifiers, Random Trees (RT) yielded the most accurate results in classifying the five classes. Like the other classifiers, it encountered challenges, particularly with identifying water bodies (Figure 11). While Figure 12 demonstrates some improvement in the infrared band, these issues persisted to some extent.

The SVM classifier continues to struggle with identifying water bodies, resulting in false positive outcomes, although it shows slight improvement in the infrared band. Additionally, it misclassifies barren land as agriculture in the visual band (see Figure 13) and as developed in the infrared band (see Figure 14).

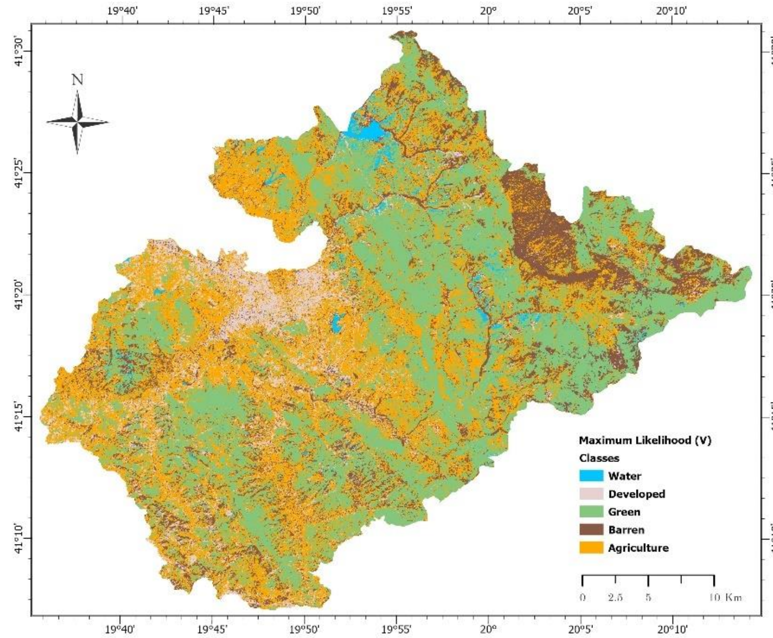


Figure 7: ML classification output in visual band

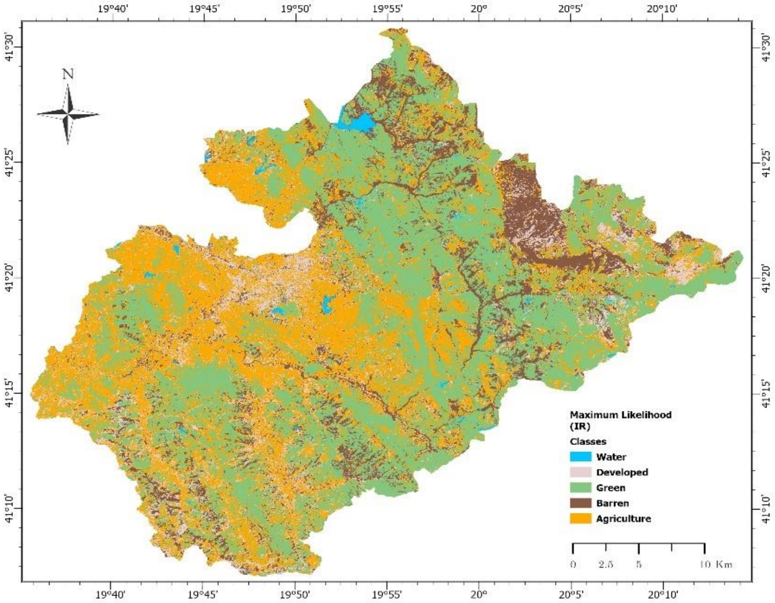


Figure 8: ML classification output in infrared band

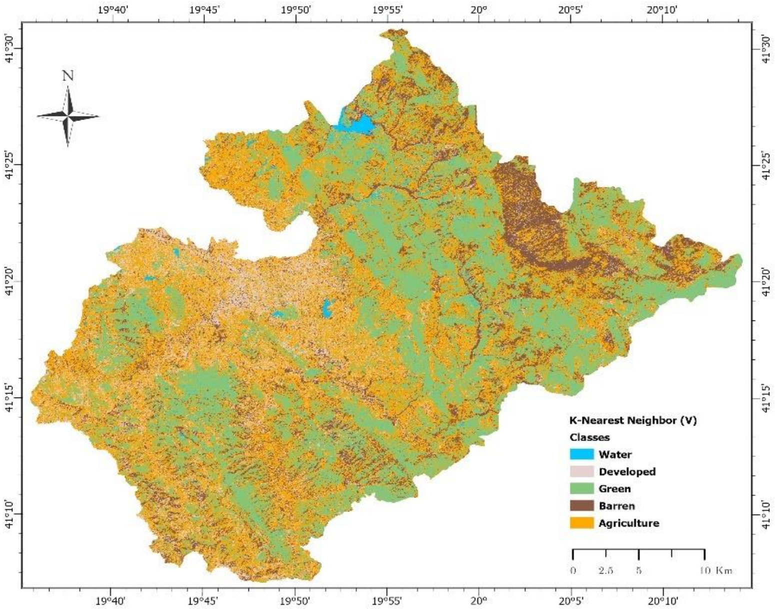


Figure 9: KNN classification output in visual band

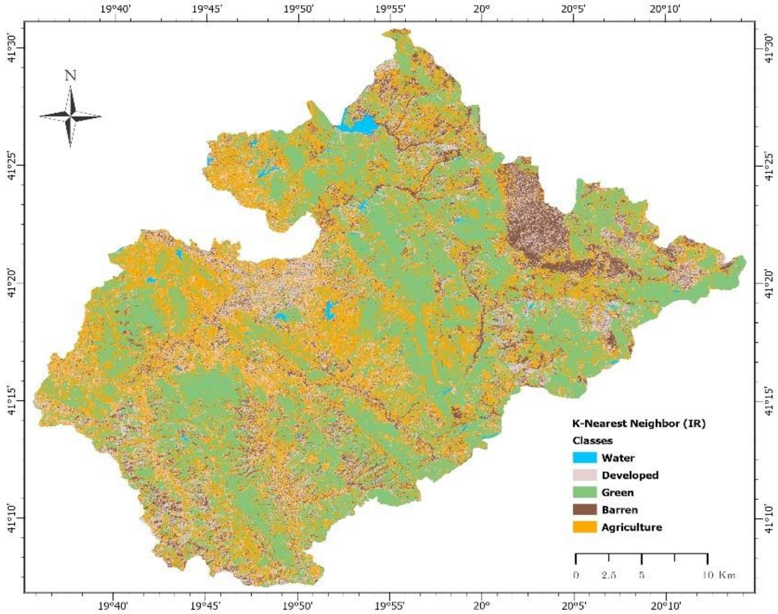


Figure 10: KNN classification output in infrared band

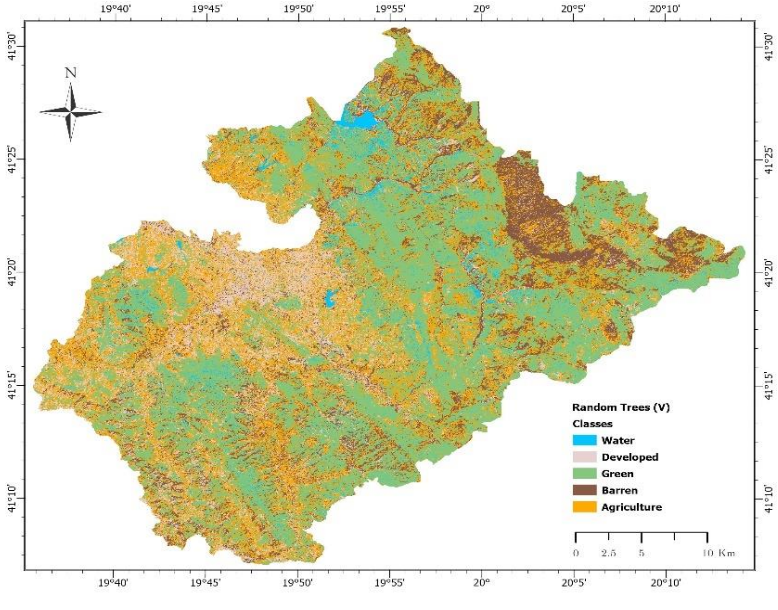


Figure 11: RT classification output in visual band

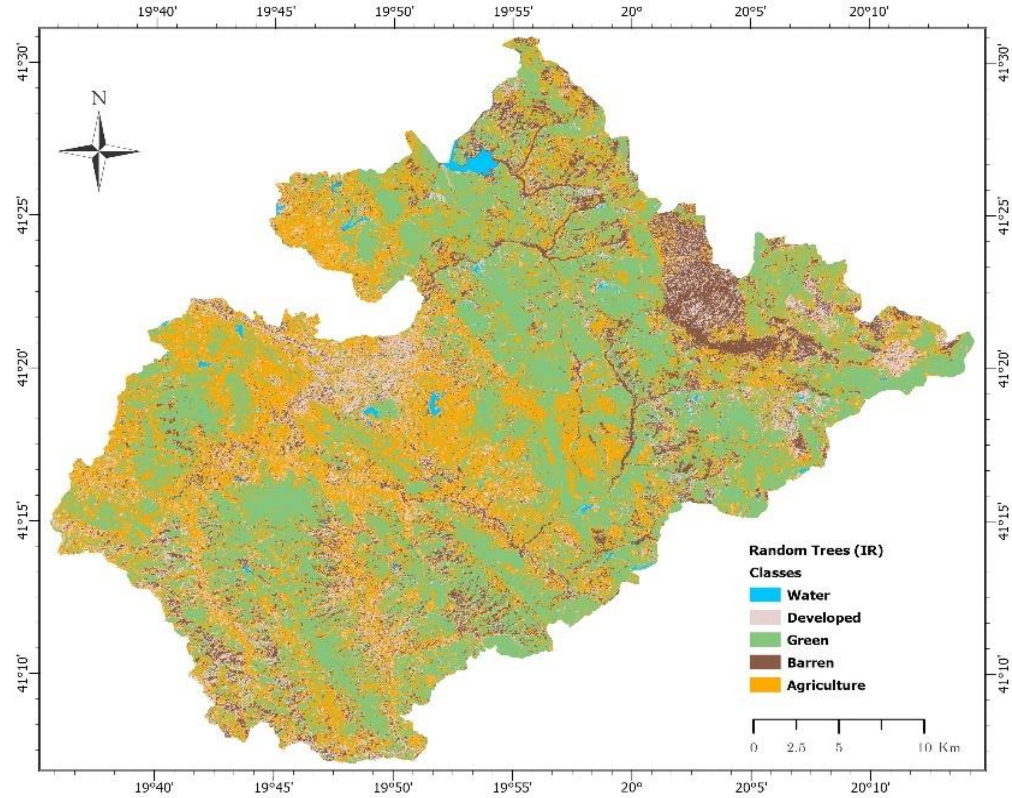
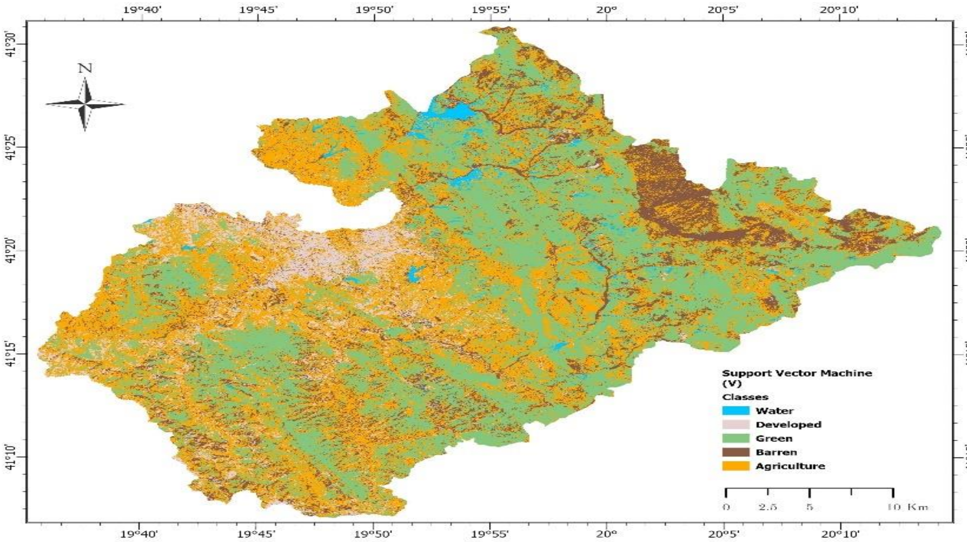


Figure 12: RT classification output in infrared ban

Figure 13: SVM classification output in visual band



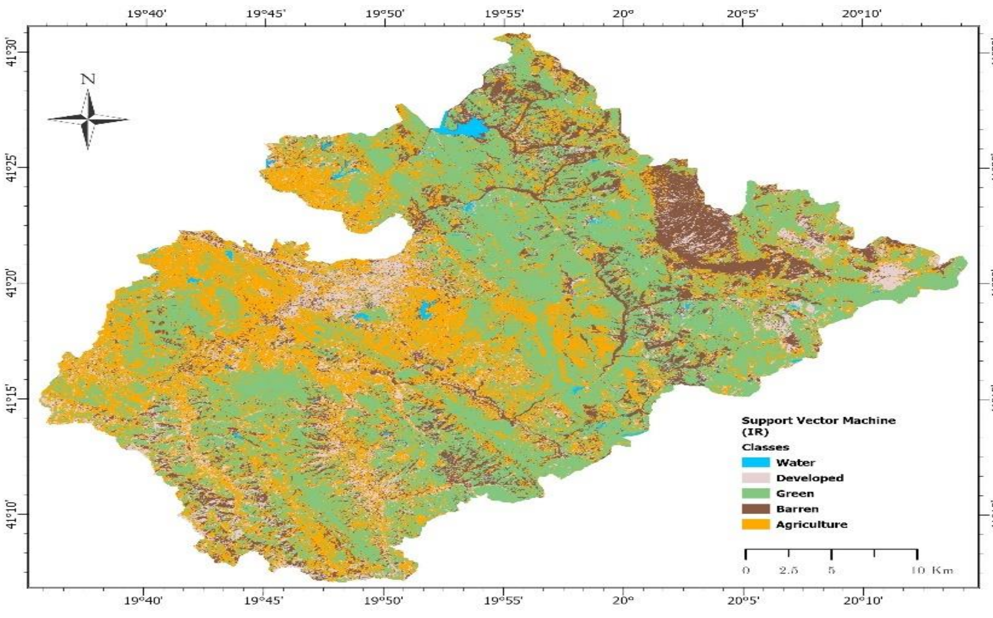


Figure 14: SVM classification output in infrared band

Table 2 and Table 3 present the area covered by each class for every classification method, both in the visual and infrared bands. As mentioned earlier, classifiers faced a significant challenge in detecting water in the visual band, leading to inflated numbers in Table 2, showing more area covered by water bodies than actually exists. In contrast, the numbers in Table 3 for the infrared band are more accurate. This is because water absorbs most of the infrared radiation, resulting in a dark or black appearance, which is easier to train classifiers to recognize in various classifications.

Class	ML	KNN	RT	SVM
Water	16.45	11.99	27.51	19.66
Developed	100.59	120.02	142.81	92.80
Green	420.29	349.94	452.58	415.13
Barren	179.22	178.85	170.13	161.81
Agriculture	404.37	460.1	327.88	431.52

Table 2: Area in km^2 covered by each class, for every classification method, in the visual band.

Class	ML	KNN	RT	SVM
Water	8.73	8.93	8.14	9.34
Developed	119.32	205.31	167.00	125.41
Green	445.72	431.24	445.96	464.49
Barren	154.67	124.05	132.69	169.34
Agriculture	392.44	351.35	367.08	352.30

Table 3: Area in km^2 covered by each class, for every classification method, in the infrared band.

Table 4 presents the overall accuracy (OA) and kappa coefficient for each classification method applied to both visual and infrared bands. In the visual band, Random Forest (RT) classification yielded the highest accuracy at 90.24%. This classifier accurately detected greenery and developed land but struggled with overdetecting water and misclassifying barren land as agricultural land. Instead, the SVM classifier (with an overall accuracy of

87.44%) effectively detected greenery and agricultural land but had issues with misclassifying developed land as agricultural land and, like the RT classifier, overdetecting water bodies. In this band, ML and KNN did not quite meet expectations, giving less satisfactory outcomes.

In the infrared band, SVM performed the best, closely trailed by ML. Whereas KNN and RT produced inferior results in this band. SVM achieved an overall accuracy of 88.94%, demonstrating excellent water detection and an improvement over the visual band. Most classes were well detected, except for developed land, which was occasionally misclassified as either agricultural or barren land. The performance of the RT classifier in the infrared band was reduced, resulting in an overall accuracy of 85.17%. While it effectively detected water and green areas, it sometimes misclassified barren land as developed land.

	Visual				**Infrared**			
	ML	**KNN**	**RT**	**SVM**	**ML**	**KNN**	**RT**	**SVM**
OA %	80.70	79.71	90.24	87.44	87.08	80.86	85.17	88.94
Kappa	0.73	0.72	0.86	0.82	0.82	0.74	0.79	0.85

Table 4: OA and Kappa for each classifier, in visual and infrared band.

The results discussed above are also presented in two charts in Figures 15 and 16, providing a better visualization, understanding, and comparison.

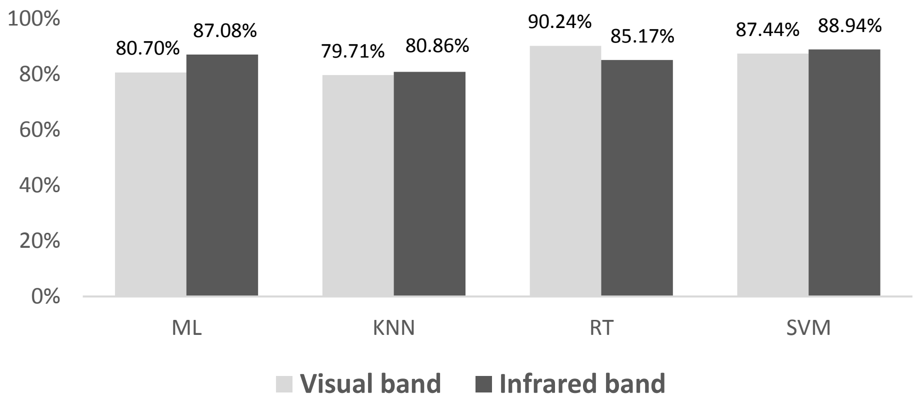


Figure 15: OA for each classifier, in visual and infrared band.

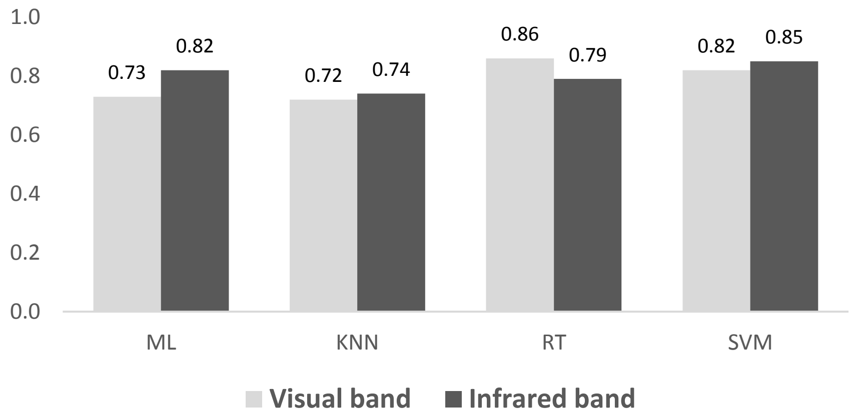


Figure 16: Kappa for each classifier, in visual and infrared band.

Tables 5, 6, 7, and 8 present the precision, recall, and F1-score for each class in both the visual and infrared bands. In all visual band classifications, there were significant issues with detecting water. Shadows cast by nearby mountainous regions or very dark green areas were often misclassified as water. However, this issue showed considerable improvement in the infrared band, achieving an F1-score of at least 0.89, as indicated in the tables below.

The developed class also faced classification issues in the visual band, and did not significantly improve in the infrared band, resulting in some developed areas being misclassified as agricultural land.

Regarding the greenery class, detection was relatively good in both bands, with slightly better results in the infrared band. This improvement may be attributed to the high reflectance of infrared radiation by leaves, which results in a brighter red for healthier vegetation.

Barren land did not perform well in the visual band, with some areas being misclassified as agricultural land. Although there was an improvement in the infrared band, misclassifications still occurred, with barren land often being mistaken for developed land. This issue was particularly noticeable in the largest area of barren land in Tirana, Mali me Gropa.

Agricultural land generally produced better results in the visual band. However, in both bands, there was a tendency to overfit this class, leading to the creation of agricultural areas instead of developed land. This issue was most prominent in the main urban area of Tirana, where many pixels belonging to the developed class were misclassified as agricultural land.

Class	Visual			Infrared		
	Prec.	Recall	F1	Prec.	Recall	F1
Water	0.40	0.80	0.53	0.90	1.00	0.95
Developed	0.50	0.64	0.56	0.62	0.76	0.68
Green	0.95	0.87	0.90	0.96	0.89	0.92
Barren	0.97	0.62	0.76	0.93	0.76	0.84
Agriculture	0.72	0.93	0.81	0.81	0.92	0.86

Table 5: Precision, Recall and F1-Score for ML classification, in visual and infrared band.

Class	Visual			Infrared		
	Prec.	Recall	F1	Prec.	Recall	F1
Water	0.30	0.60	0.40	0.90	1.00	0.95
Developed	0.48	0.67	0.56	0.54	0.87	0.67
Green	1.00	0.81	0.89	0.96	0.86	0.91
Barren	0.81	0.70	0.75	0.82	0.50	0.62
Agriculture	0.78	0.88	0.83	0.76	0.87	0.81

Table 6: Precision, Recall and F1-Score for KNN classification, in visual and infrared band.

Class	Visual			Infrared		
	Prec.	Recall	F1	Prec.	Recall	F1
Water	0.70	1.00	0.82	0.90	1.00	0.95
Developed	0.72	0.90	0.80	0.80	0.80	0.80
Green	0.96	0.95	0.96	0.97	0.89	0.93
Barren	0.97	0.72	0.83	0.83	0.65	0.73
Agriculture	0.90	0.95	0.92	0.72	0.92	0.81

Table 7: Precision, Recall and F1-Score for RT classification, in visual and infrared band.

Class	Visual			Infrared		
	Prec.	Recall	F1	Prec.	Recall	F1
Water	0.70	1.00	0.82	0.80	1.00	0.89
Developed	0.76	0.72	0.74	0.73	0.70	0.71
Green	0.92	0.93	0.93	0.99	0.93	0.96
Barren	0.93	0.69	0.79	0.90	0.79	0.84
Agriculture	0.86	0.94	0.90	0.83	0.95	0.88

Table 8: Precision, Recall and F1-Score for SVM classification, in visual and infrared band.

Conclusions

This study evaluated the performance of four classification methods – Maximum Likelihood (ML), K-Nearest Neighbors (KNN), Random Trees (RT), and Support Vector Machine (SVM) – in classifying land cover types in Tirana County using both visual and infrared bands. The findings reveal several key insights into the efficacy and limitations of each classifier, particularly in terms of overall accuracy, kappa coefficient, and class-specific performance metrics such as precision, recall, and F1-score.

The RT classifier demonstrated the highest overall accuracy of 90.24% and a kappa coefficient of 0.86 in the visual band. This indicates a superior performance in detecting various land cover classes accurately.

In the infrared band, the SVM classifier achieved the highest overall accuracy of 88.94% and a kappa coefficient of 0.85, showing a better performance in this spectral band.

Both the ML and KNN classifiers performed less satisfactorily in comparison to RT and SVM across both bands. This may be due to the fact that many classification methods operate under the assumption that the prior probability of occurrence for each class is the same. This assumption is made because accurately estimating this prior probability for each class is challenging due to factors such as variability in data, terrain, and environmental conditions.

Water bodies posed a significant challenge for all classifiers, particularly in the visual band. Shadows and dark areas were often misclassified as water, leading to high rates of false positives. The infrared band substantially improved the classification of water bodies across all classifiers, with F1-scores reaching up to 0.95. This improvement is attributed to water's strong absorption of infrared radiation, which enhances its distinguishability from other land cover types.

Developed land was frequently misclassified as agricultural land in both the visual and infrared bands. This issue was more pronounced in urban areas of Tirana, affecting the classifiers' precision for this class. While the infrared band showed some improvement, misclassification of developed land persisted, indicating the complexity of distinguishing developed areas from other similar classes such as barren lands.

Green areas were generally well detected in both bands, with slight improvements in the infrared band due to the high reflectance of infrared radiation by the vegetation. This resulted in high precision for the greenery class.

Agricultural land exhibited a tendency for overfitting in both bands, leading to some misclassifications of developed land as agricultural. This was particularly evident in urban areas, affecting the classifiers' overall performance.

The classification of barren land was less accurate in the visual band, with some areas being misclassified as agricultural land. Although the infrared band improved this classification, misclassifications still occurred, often confusing barren land with developed land.

One notable issue that may have impacted the classification results is the disparity in the quantity and area of training samples for the Developed and Water classes compared to the other three classes. The smaller training sets for these classes could lead to less robust models, as the classifiers may not have been exposed to a sufficient variety of examples to accurately learn the distinguishing features of these two land cover types. This imbalance can result in misclassifications and reduced overall performance. Specifically, the limited data for Developed land might contribute to its misclassification as agricultural land, while the smaller sample size for Water areas could affect the distinguishing of water bodies from shadows and dark areas, particularly in the visual band.

Choosing the appropriate spectral band is crucial for accurate land cover classification, as no single band is universally adequate for all classes in general. Different land cover types exhibit unique spectral signatures in various bands, making some bands more suitable for certain classes than others. For example, the infrared band enhances the detection of water bodies due to their strong absorption of infrared radiation, while the visual band might better capture certain urban features. Therefore, understanding the strengths and limitations of each band is essential to optimize the classifications performance across diverse land cover types.

A limitation in this study is the absence of a reference or base map against which to compare the classified maps. This lack of an external benchmark prevents achieving absolute certainty in the classification accuracy. Therefore, the evaluation of classifiers performance relies exclusively on internal metrics and visual comparison, which may overlook undetected errors or misclassifications, thereby affecting the overall reliability of the results.

Supervised classification techniques play an essential role in accurately identifying and mapping land cover types, as demonstrated in this study. By using labeled training data in GIS software, these methods can learn to distinguish between different land cover classes with great precision. The information gained from evaluating various classifiers highlights the necessity of selecting appropriate methods and spectral bands, customized to the specific characteristics of the land cover being studied. The use of supervised classification techniques stands as an indispensable tool in accurately mapping land cover types, offering valuable information necessary for a wide range of applications across diverse fields.

References

Agrawal, R. (2014). K-nearest neighbor for uncertain data. International Journal of Computer Applications, 105(11), 13-16.

Avand, M., Janizadeh, S., Naghibi, S. A., Pourghasemi, H. R., Khosrobeigi Bozchaloei, S., & Blaschke, T. (2019). A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. Water, 11(10), 2076.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.

Bolstad P. (2016). GIS Fundamentals - A First Text on Geographic Information Systems (Fifth Edition). Eider Press.

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

Bugya, T., & Farkas, G. (2018). An Alternative Raster Display Model. In GISTAM (pp. 262-268).

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189-215.

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. International Journal of Applied Earth Observation and Geoinformation, 11(5), 352-359.

Kavzoglu, T., & Reis, S. (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. GIScience & Remote Sensing, 45(3), 330-342.

Kulkarni, V. Y., & Sinha, P. K. (2013). Random forest classifiers: a survey and future research directions. Int. J. Adv. Comput, 36(1), 1144-1153.

Lillesand Thomas M., Kiefer Ralph W., Chipman Jonathan W. (2015). Remote Sensing and Image Interpretation (Seventh Edition). Wiley.

Mitchell T. M. (1997). Machine Learning. McGraw-Hill.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS journal of photogrammetry and remote sensing, 66(3), 247-259.

Oommen, T., Misra, D., Twarakavi, N. K., Prakash, A., Sahoo, B., & Bandopadhyay, S. (2008). An objective analysis of support vector machine-based classification for remote sensing. Mathematical geosciences, 40, 409-424.

Quinlan J. R. (1986): Induction of Decision Trees. Kluwer Academic Publishers, Boston. Machine Learning 1: 81-106.

Shahbazi, M., Théau, J., & Ménard, P. (2014). Recent applications of unmanned aerial imagery in natural resource management. GIScience & Remote Sensing, 51(4), 339-365.

Shaik, A. B., & Srinivasan, S. (2019). A brief survey on random forest ensembles in classification model. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2 (pp. 253-260). Springer Singapore.

Sun, S., & Huang, R. (2010). An adaptive k-nearest neighbor algorithm. In 2010 seventh international conference on fuzzy systems and knowledge discovery (Vol. 1, pp. 91-94). IEEE.

Susaki, J., & Shibasaki, R. (2000). Maximum likelihood method modified in estimating a prior probability and in improving misclassification errors. International Archives of Photogrammetry and Remote Sensing, 33(B7/4; PART 7), 1499-1504.

Tso B., & Mather P. M. (2001). Classification Methods for Remotely Sensed Data. Taylor and Francis.

Ulqinaku, M., & Ktona, A. (2023). Entropy-based decision trees – A Review. Journal of Natural Sciences, No. 34.

Vapnik V. N. (1995). The Nature of Statistical Learning Theory. Springer.

Wei, Q., & Dunbrack Jr, R. L. (2013). The role of balanced training and testing data sets for binary classifiers in bioinformatics. PloS one, 8(7), e67863.