Evaluating Quantum Machine Learning for GIS-Based Land Suitability Analysis: A Case Study of Reforestation

More, N.,[1*] Patil, S.,[1] Trimbakkar, V.,[1] Shrivastav, S.[1] and Shah, P.[1] Pasi, A.[1] and Nikam, V.[2]
[1]K. J. Somaiya School of Engineering, Somaiya Vidyavihar University, Department of Information Technology, Mumbai 400077, India, E-mail: nilkamalpmore@gmail.com,[*] ORCID ID: 0000-0003-09649490[*]
suchitrapatil@somaiya.edu, varad.trimbakkar@somaiya.edu, satyam.ns@somaiya.edu
purav.shah1@somaiya.edu, abhijeet.p@somaiya.edu
[2]Veermata Jijabai Technological Institute, Matunga, Mumbai, 400019, India, E-mail: vbnikam@it.vjti.ac.in
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v22i4.4943

Abstract
Quantum Machine Learning (QML) has emerged as an active area of research, yet its practical readiness for real-world geospatial applications remains uncertain due to constraints related to data availability, preprocessing requirements, and current quantum hardware limitations. This study evaluates the feasibility and behaviour of quantum kernel-based learning models within a realistic GIS workflow, using land suitability analysis for reforestation in the Mumbai metropolitan region as a representative case study although demonstrated using the Mumbai metropolitan region as a case study, the proposed GIS-driven preprocessing pipeline and quantum kernel evaluation framework are transferable to other regions with comparable remote-sensing data availability and environmental indicators. A geospatial dataset is constructed from satellite-derived environmental variables, including Normalized Difference Vegetation Index (NDVI), annual rainfall, terrain slope, soil pH, and soil organic carbon, with suitability labels assigned using ecologically motivated rule-based thresholds that serve as heuristic ground truth. The dataset is processed through a standardized pipeline involving feature scaling, Principal Component Analysis (PCA), stratified train–test splitting, and Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. Quantum Support Vector Machines (QSVMs) employing Pauli and ZZ feature maps are implemented using Qiskit and evaluated under noiseless simulation conditions, with performance compared against a classical Support Vector Machine baseline using identical preprocessing and evaluation metrics. While classical models achieve superior computational efficiency and higher predictive performance, QSVMs exhibit non-trivial classification behaviour on small, structured datasets when appropriate preprocessing and feature-to-qubit alignment are applied. Rather than claiming quantum advantage, this work provides an empirical assessment of the current capabilities and limitations of QML in applied geospatial analysis, positioning quantum kernel methods as complementary tools for exploratory environmental modelling under present technological constraints.

Keywords: Environmental Modelling, GIS, Quantum Kernel, Quantum Machine Learning, Reforestation

1. Introduction

Effective land-use planning has become increasingly important in the context of accelerating deforestation, climate change, and urban expansion. Reforestation initiatives play a critical role in ecological restoration by mitigating soil erosion, supporting biodiversity, and enhancing carbon sequestration, particularly in environmentally sensitive regions such as coastal and peri-urban ecosystems. Identifying land parcels suitable for reforestation is a complex task, as it requires the integration of heterogeneous geospatial variables including vegetation indices, climatic factors, terrain characteristics, and soil properties derived from remote sensing and geographic information systems (GIS) data [1] and [2].

Classical machine learning methods have been widely applied to land suitability and environmental modeling tasks and have demonstrated strong predictive performance across numerous ecological applications. Techniques such as Support Vector Machines and Random Forests are commonly used to model nonlinear relationships in remote-sensing data. However, real-world geospatial datasets frequently exhibit characteristics such as limited sample sizes, correlated features, and class imbalance, which can complicate model training

and interpretation [1][2][3] and [4]. These challenges motivate continued exploration of alternative computational approaches that may offer different representational capabilities under constrained data conditions.

Quantum Machine Learning (QML) has emerged as a promising paradigm that embeds classical data into high-dimensional Hilbert spaces using quantum feature maps and kernel-based methods. Quantum Support Vector Machines (QSVMs) have attracted interest due to their ability to represent complex feature interactions through entanglement and nonlinear quantum kernels. Recent studies have shown that QSVMs can exhibit behavior comparable to classical baselines in selected remote-sensing and Earth observation tasks when evaluated under simulated, noise-free conditions [5] and [6]. At the same time, other investigations emphasize that the practical applicability of quantum kernel methods remains strongly influenced by preprocessing choices, circuit design, dataset size, and current hardware limitations [7] and [8].

Despite growing interest in QML, there remains a lack of systematic evaluation of QSVM feasibility for GIS-based reforestation suitability analysis under realistic preprocessing constraints. In particular, the combined effects of dimensionality reduction, feature-to-qubit alignment, and class balancing on QSVM performance in applied geospatial settings are not yet well understood. Addressing this gap is essential for developing a grounded understanding of where QML methods currently stand relative to classical techniques in environmental analytics.

This study addresses this need by evaluating the feasibility of integrating QSVMs into a GIS-driven land suitability analysis pipeline using reforestation planning in the Mumbai metropolitan region as a representative case study evaluated under noiseless simulation conditions on classical computing infrastructure. The objectives are to (i) construct a satellite-derived geospatial dataset using environmental indicators, (ii) evaluate Pauli and ZZ-based QSVMs under identical preprocessing conditions, and (iii) compare their behavior against a classical Support Vector Machine baseline. Rather than claiming predictive superiority, this work provides an empirical assessment of the capabilities, limitations, and trade-offs associated with quantum kernel methods in applied geospatial modeling, positioning QSVMs as complementary analytical tools within the current landscape of environmental machine learning.

2. Literature Review

Machine learning techniques have been widely applied in environmental modelling and remote-sensing applications, with numerous studies reporting strong performance of classical models for reforestation assessment, land-cover mapping, and ecological suitability analysis [1][2][3][6] and [9]. Ensemble-based and margin-based classifiers, including Random Forests and Support Vector Machines, have demonstrated robustness and high predictive accuracy, particularly when sufficient training data and well-separated feature spaces are available [1] and [3].

More recently, quantum machine learning approaches have been explored for remote-sensing and Earth-observation tasks, primarily under simulated or constrained computational environments. Existing studies indicate that quantum kernel methods and quantum annealing–based classifiers can achieve performance parity with classical baselines, although their practical deployment remains limited by current hardware capabilities [5][6] and [10]. Quantum support vector machine algorithms have been found to deliver similar results for image classification and feature-based remote sensing problems as the conventional ones when noise-free simulators are used in quantum computing [5] and [6]. It can be inferred from other research reports that quantum kernels may have a potential advantage for specific high-dimensional or correlated datasets [7] and [8].

Based on these findings, this research applies Quantum Support Vector Machines to a GIS-derived land suitability dataset for reforestation and compares them with classical Support Vector Machines under identical preprocessing conditions. The rationale for this is to evaluate the usability of Quantum Kernels in suitability modelling under realistic parameters such as sample size, class balancing, and dimensionality reduction. Despite the progress reported in prior studies, several recurring challenges can be identified in the application of machine learning and quantum machine learning to remote-sensing based environmental analysis. A central gap in much of the existing literature is the limited evaluation of quantum algorithms within realistic, end-to-end application workflows that reflect practical data and computational constraints. Many studies focus on proof-of-concept demonstrations or highly controlled experimental settings, often using small, region-specific datasets, noiseless simulation environments, and heuristic or proxy-based labels, which makes it difficult to assess how quantum methods behave when embedded within applied geospatial pipelines. In addition, the interaction

between preprocessing choices such as feature scaling, dimensionality reduction, and class balancing and the behavior of quantum kernel methods is frequently underexplored in applied contexts. The present study does not claim to resolve these challenges outright. Instead, it seeks to evaluate the feasibility and behavior of quantum kernel methods when applied within a realistic GIS-driven workflow, operating on satellite-derived data and executed on conventional computing infrastructure. By comparing QSVMs with classical SVMs under identical preprocessing and evaluation conditions, and by explicitly discussing limitations, this work aims to provide a grounded reference point for understanding the current practical role of quantum machine learning in geospatial analysis.

3. Methodology

3.1 Overall Workflow

The system combines a GIS data construction pipeline with a classical and quantum machine learning model. The entire process has four main steps:

(1) retrieving environmental characteristics for the Mumbai area via Google Earth Engine (GEE),

(2) rule-based sample classification into suitable and unsuitable groups, 6792525650

(3) preprocessing steps involving standardization, Min-Max scaling to [-1, 1], PCA dimensional reduction, stratified splitting, and SMOTE class balancing, and

(4) Training and testing classical SVM and QSVM models on the same preprocessing environment.

All quantum experiments were executed using Qiskit simulators in Google Colab, enabling evaluation on standard workstation hardware without requiring access to physical quantum devices.

3.2 Dataset Construction with GEE

Figure 1 presents the study area map of the Mumbai Metropolitan Region. The blue lines represent the major river networks within the region, while the red rectangular boundary delineates the defined study area used for analysis. The selected map scale provides an optimal balance between regional context and local detail, ensuring that key geographic features such as drainage patterns and land cover distribution are clearly visible without overcrowding the visualization. This scale is therefore appropriate for supporting the GIS-based land suitability assessment conducted in this study. The environmental variables for the Mumbai region (72.75°E–73.05°E, 18.85°N–19.35°N) were extracted from widely used open-access satellite datasets. A high-level overview of the complete data construction pipeline is illustrated in Figure 2.

The extracted variables include:

- Normalized Difference Vegetation Index (NDVI) from Sentinel-2 surface reflectance (median composite),
- Annual rainfall from CHIRPS daily precipitation data,
- Terrain slope derived from the SRTM digital elevation model, and
- Soil pH and soil organic carbon (SOC) from SoilGrids products.

The ruleset applied to each variable for suitability classification is summarized in Table 1. All layers were resampled to a spatial resolution of 30 m and combined into a multiband raster. A total of 500 uniformly distributed sampling points were generated across the region, and corresponding feature values were extracted to construct the initial dataset. To ensure transparency and reproducibility, the characteristics of the satellite-derived variables used to construct the dataset including their sources, spatial resolution, and temporal coverage are summarized in Table 1.

3.2.1 Study area

The study area selected for this work is the Mumbai metropolitan region, located along the western coast of India (approximately 18.85°N–19.35°N latitude and 72.75°E–73.05°E longitude). This region represents a complex urban–peri-urban environment characterized by heterogeneous land cover, coastal climate conditions, and increasing pressure on natural ecosystems due to urban expansion. The area is therefore well suited as a representative case for evaluating GIS-based land suitability analysis for reforestation using remote-sensing–derived environmental indicators.

3.2.2 Dataset description

The geospatial dataset used in this study was constructed by integrating multiple openly available remote-sensing and environmental data products commonly employed in GIS-based land suitability and reforestation analyses. These datasets differ in spatial resolution, temporal characteristics, and thematic focus, requiring harmonization prior to feature extraction. To provide a clear overview of the inputs used in this study, Table 1 summarizes the satellite platforms, data products, spatial resolution, temporal coverage, and the role of each variable within the suitability modelling framework.

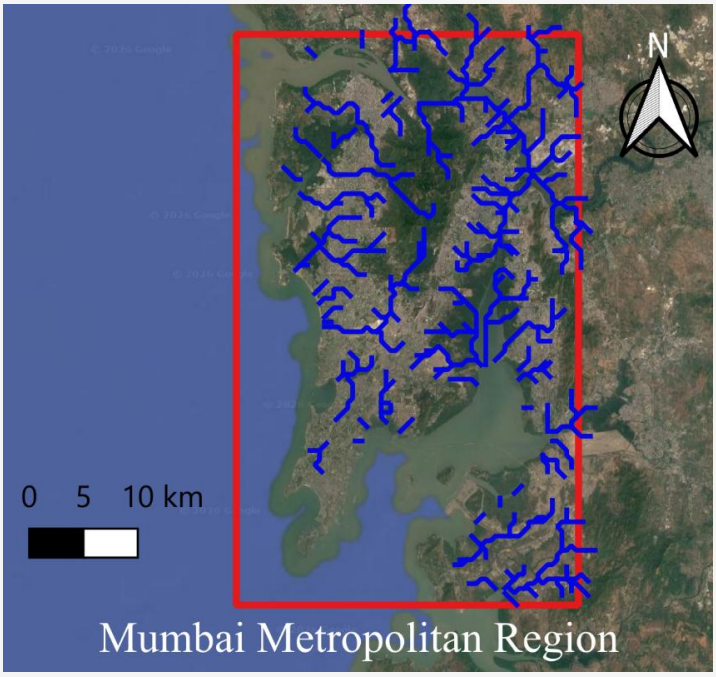


Figure 1: Mumbai Metropolitan Region illustrating the spatial extent of the analysis, including the defined boundary and major river networks relevant to the GIS-based land suitability assessment

Table 1: Summary of satellite-derived variables used for dataset construction

Variable	Data Source	Satellite / Product	Spatial Resolution	Temporal Coverage	Description / Role
NDVI	Copernicus Open Access Hub	Sentinel-2 Surface Reflectance	10–20 m (resampled to 30 m)	2021–2022 (median composite)	Proxy for vegetation health and existing biomass
Rainfall	Climate Hazards Center	CHIRPS Daily Precipitation	~5 km (resampled to 30 m)	2021–2022 (annual aggregate)	Indicator of moisture availability for reforestation
Slope	NASA / USGS	SRTM Digital Elevation Model	30 m	Static	Terrain accessibility and erosion risk
Soil pH	ISRIC	SoilGrids	~250 m (resampled to 30 m)	Static	Indicator of soil chemical suitability
Soil Organic Carbon (SOC)	ISRIC	SoilGrids	~250 m (resampled to 30 m)	Static	Proxy for soil fertility and carbon content

The selected environmental variables represent commonly used indicators in GIS-based land suitability and reforestation studies. Sentinel-2 NDVI captures surface vegetation conditions, while CHIRPS rainfall data provides a climatic proxy for long-term moisture availability. Terrain slope derived from SRTM informs accessibility and erosion risk, whereas soil pH and soil organic carbon reflect subsurface conditions relevant to vegetation growth. Together, these variables form a compact yet representative feature set for evaluating machine learning and quantum kernel methods under realistic geospatial data constraints.

3.3 Rule-based labelling

Each sampled location was labelled as suitable (1) or unsuitable (0) for reforestation using ecologically motivated threshold rules based on Normalized Difference Vegetation Index (NDVI), terrain slope, and annual rainfall. These variables are widely adopted indicators in land suitability and reforestation studies and are commonly used as proxy measures of vegetation health, terrain

accessibility, and moisture availability [1][3] and [4].

To examine the robustness of model behaviour under varying ecological assumptions, three labelling variants were defined by adjusting threshold combinations, as summarized in Table 2. These rule-based labels are treated as heuristic ground truth, serving as a controlled reference for evaluating and comparing machine learning models rather than asserting definitive ecological classification. This design choice enables a feasibility-oriented assessment of model performance under realistic data construction constraints. Soil pH and Soil Organic Carbon (SOC) were intentionally excluded from the labelling rules to avoid circular learning, while being retained as predictor variables during training. This allows both classical and quantum models to independently infer the relevance of these soil properties during classification, consistent with practices adopted in prior remote-sensing and environmental modelling studies [1][3] and [11]. While NDVI, terrain slope, and rainfall are used both in heuristic labelling and as predictive features, this does not invalidate the comparative evaluation performed in this study. Both classical SVM and QSVM models are trained and evaluated on identical feature sets and label definitions, ensuring that any potential bias introduced by heuristic labelling affects all models uniformly. As a result, the reported performance differences reflect relative model behaviour rather than absolute predictive accuracy with respect to ecological ground truth. The objective of this work is therefore not to claim real-world suitability prediction, but to assess the feasibility, expressivity, and generalization characteristics of quantum and classical models under controlled and consistent data construction assumptions.

3.4 Preprocessing Pipeline

Prior to model training, the dataset was processed through a structured preprocessing pipeline designed to ensure compatibility with both classical and quantum classifiers following standard practices in geo-environmental ML modelling [1] and [3] and QML preprocessing workflows [5] and [6]. All preprocessing transformations were fitted exclusively on the training subset to prevent information leakage.

3.4.1 Standardization

Feature standardization was applied to center each variable to zero mean and unit variance, following standard practice in geospatial machine learning. This step mitigates scale disparities between heterogeneous variables (e.g., rainfall measured in millimetres versus NDVI expressed as a ratio) and ensures numerical stability during subsequent processing. Standardization parameters were computed using the training data only and applied unchanged to the test data to prevent information leakage [1] and [3].

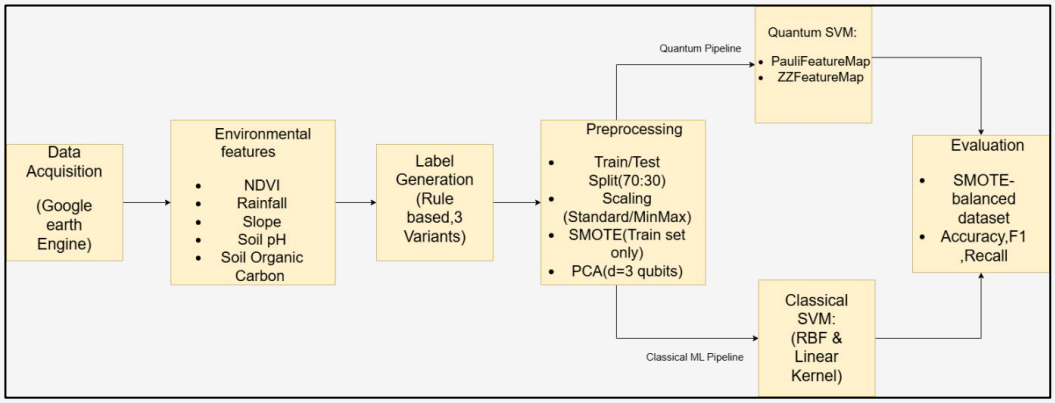


Figure 2: End-to-end workflow for the quantum-enhanced land classification pipeline, showing GIS-based data acquisition, rule-based labelling, pre-processing (scaling, SMOTE, PCA), and evaluation of both quantum (QSVM) and classical SVM models

Table 2: Rule-based labelling

Variant	NDVI	Slope	Rainfall (mm/yr)	Purpose
Primary	> 0.40	< 20°	> 1000	Baseline regional criteria
Conservative	> 0.45	< 15°	> 1200	Focuses on long-term survival
Lenient	> 0.35	< 25°	> 800	Includes marginal areas

3.4.2 Min–Max scaling

Following standardization, features were rescaled to the range [−1, 1] using min–max normalization, consistent with preprocessing requirements commonly adopted in quantum machine learning workflows [5] and [6]. This transformation aligns feature values with the rotation-angle constraints of quantum feature maps and reduces the risk of numerical instability during quantum state preparation. As with standardization, scaling parameters were learned exclusively from the training subset and subsequently applied to the test data.

3.4.3 Dimensionality reduction with PCA

PCA was applied to reduce the dimensionality of the standardized feature space consistent with approaches used in other QML–EO studies [5] and [6]. The transformation is defined in Equation 1:

$$Z = XW_d$$

Equation 1

Where:

X is the standardized feature matrix

W_d is the matrix containing the top principal components

Z is the transformed feature matrix in the reduced-dimensional space

The explained variance for each principal component is summarized in Table 3, indicating that the first three components capture over 70% of the total variance. The cumulative explained variance curve is shown in Figure 3, motivating the selection of $d = 3$ for the quantum models. This results in a three-qubit configuration for most experiments.

3.4.4 Train–Test split and SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data to address class imbalance by generating synthetic minority-class samples in feature space. Following recommended practice in remote-sensing ML workflows [1] and [3]. The transformation is defined in Equation 2:

$$x_{new} = x_i + \lambda\left(x_j - x_i\right), \lambda \in [0,1]$$

Equation 2

Where new synthetic samples are generated along the line between a minority sample x_i and one of its k-nearest neighbors x_j.

It is acknowledged that, when applied to datasets with a limited number of minority-class samples, SMOTE may introduce synthetic patterns that do not fully reflect real-world variability. In this study, SMOTE is therefore used strictly as an exploratory preprocessing step to examine model behaviour under balanced conditions, rather than as a mechanism for improving ecological realism or predictive validity.

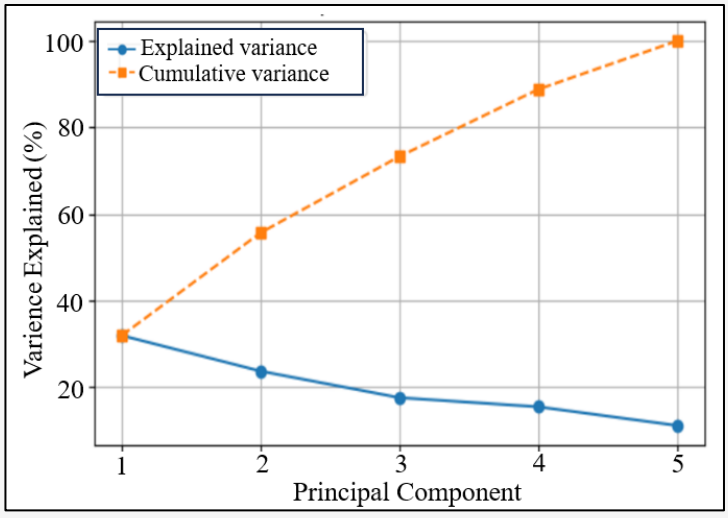


Figure 3: Explained variance and cumulative variance captured by PCA components on the training set. The first three principal components together explain more than 70% of the dataset variance, motivating the choice d=3 for the quantum feature maps

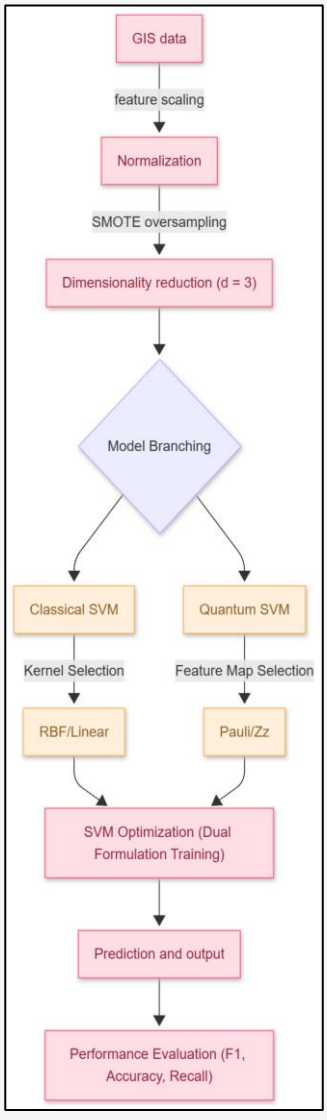


Figure 4: Overall model workflow including preprocessing, branching, and evaluation

The effects of SMOTE balancing are illustrated in detail in the Results section. A high-level overview of preprocessing, model branching, and evaluation is shown in Figure 4. The post-SMOTE embedding reveals more balanced and overlapping clusters, which favours margin-based classifiers such as SVMs and QSVMs.

3.5 Quantum Machine Learning Model

The quantum classification model is based on the QSVM framework implemented in Qiskit. Instead of analytic kernels (e.g., RBF or linear), QSVMs compute similarities using state fidelity in a quantum Hilbert space. Let $x \in R^d$ denote a PCA-reduced feature vector. A quantum feature map is defined in Equation 3:

$$x \mapsto |\phi(x)\rangle$$

Equation 3

Embeds x into a d-qubit quantum state. Two feature maps were investigated.

3.5.1 ZZFeatureMap

The ZZFeatureMap introduces nonlinear feature interactions through single-qubit Z rotations followed by pairwise entangling ZZ operations between qubits. The number of qubits is set equal to the number of retained PCA components, directly linking the feature dimension to the quantum circuit design. This alignment ensures that each reduced feature is encoded into a corresponding qubit. The circuit depth, controlled by the parameter *reps*, governs the degree of entanglement and expressive power of the embedding. A deeper circuit means a more complex interaction of features but results in increased simulation times and a higher danger of overfitting. The encoding process for a feature vector $x = (x_j, x_j, x_j, \ldots, x_d)$, the encoding consists of a single-qubit rotation stage followed by pairwise ZZ entanglement.

Single-qubit encoding is defined in Equation 4:

$$U_Z(x) = \prod_{i=1}^{d} e^{ix_i Z_i}$$

Equation 4

Pairwise ZZ-entanglement is defined in Equation 5:

$$U_{ZZ}(x) = \prod_{i<1}^{d} e^{i\cdot\eta\cdot x_i\cdot x_i\cdot(Z_i\otimes Z_j)}$$

Equation 5

Full ZZFeatureMap is defined in Equation 6:

$$|\phi_{ZZ}(X)\rangle = U_{ZZ}(X)U_Z(x)\,|0\rangle^{\otimes d}$$

Equation 6

Where η is the entanglement strength. It entangles every qubit pair which gives rich nonlinearity but heavier computation.

In this study, a quantum feature embedding with $d = 3$ qubits (matching the PCA-reduced feature dimension) corresponding to the PCA-reduced feature dimension and a circuit depth of *reps* = 16 is **employed to explore the upper bounds of** expressivity achievable through entanglement-rich quantum kernels.

This configuration intentionally prioritizes representational capacity over computational efficiency, allowing an examination of how highly expressive quantum embeddings behave under small-sample geospatial datasets. While such depth introduces substantial computational overhead and potential overparameterization, it serves as a stress-test configuration rather than an optimized deployment setting. Similar expressive embeddings have been explored in prior quantum kernel studies to analyse nonlinear feature interactions under controlled, noise-free simulation conditions [3].

3.5.2 PauliFeatureMap

The PauliFeatureMap encodes classical features using sequences of parameterized Pauli rotations applied independently to each qubit. For a given input feature x_i, the encoding is a rotation about a Pauli operator. The complete embedding of a PCA-reduced feature vector into a *d-qubit* quantum state is realized by the application of the Pauli rotations across all qubits.This rotation-based structure yields a smooth and stable embedding of the input data into the quantum Hilbert space. Compared to the feature maps reliant on entanglement, the PauliFeatureMap relies more on local, qubit-wise transformations than on pairwise interactions. This property contributes to the low circuit complexity while retaining the expressiveness-good embedding for a moderate-size dataset. In the current work, PauliFeatureMap in Equation 7 provided the best quantum performance; it outperformed the ZZ-based configuration tested here under identical conditions of preprocessing [6].

$$U_{Pauli}(x_i) = e^{i \cdot x_i \cdot P_i}$$

Equation 7

The Full embedding is defined in Equation 8:

$$|\phi_{Pauli}(x)\rangle = \left(\prod_{i=1}^{d} U_{Pauli}(x)\right)|0\rangle^{\otimes d}$$

Equation 8

The kernel is computed using Equation 9:

$$K_{Pauli}\left(x_i, x_j\right) = |\langle \phi_{Pauli}(x_i) \,|\, \phi_{Pauli}(x_j)\rangle|^2$$

Equation 9

Where $|\phi_{Pauli}(x)\rangle$ denotes the state prepared by the Pauli encoding circuit.

3.5.3 Quantum Kernel and QSVM Objective

Pairwise similarities between samples x_i and x_j are computed using state fidelity as presented in Equation 10:

$$K(x_i, x_j) = \left|\langle \phi(x_i) \,|\, \phi(x_j)\rangle\right|^2$$

Equation 10

The QSVM follows the standard SVM dual objective as shown in Equation 11:

$$\max(a)\left[\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)\right]$$

Equation 11

Which leads to: $0 \le \alpha_i \le C$, and $\sum_{i=1}^{n} \alpha_i y_i = 0$

Where:

$y \in \{-1, +1\}$ are labels

C is penalty parameter

$K(x_i, x_j)$ is the quantum fidelity kernel

The decision function is defined in Equation 12:

$$f(x) = sign\left(\sum_{i=1}^{n} a_i y_i K(x_i, x) + b\right)$$

Equation 12

This hybrid classical–quantum formulation is also used in QMSVM and QSVM literature [5] and [6]. All quantum simulations were executed using Qiskit's StatevectorSampler, consistent with other feasibility-focused QML studies [5][6] and [10]. The F1-score is computed using Equation 13:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Equation 13

3.6 Classical SVM Baseline

Classical SVMs with linear and RBF kernels were implemented using scikit-learn. These models were trained on the same PCA-reduced, scaled, and SMOTE-balanced data as the QSVMs, enabling a fair comparison. The linear kernel is defined in Equation 14:

$$K_{linear}(x_i, x_j) = x_i, x_j$$

Equation 14

The Gaussian RBF kernel is defined in Equation 15:

$$K_{RBF}(x_i, x_j) = \exp(-\gamma \,||\, x_i - x_j \,||^2)$$

Equation 15

4. Results

The performance of classical and quantum models was evaluated on the GIS-derived dataset after preprocessing, dimensionality reduction, and SMOTE balancing. The final dataset contained 332 labelled samples characterized by five

environmental features. This preprocessing strategy follows established practices in remote-sensing machine learning and is consistent with prior quantum machine learning feasibility studies in Earth-observation applications [1][3][5] and [6]. The effect of SMOTE on the class distribution (before as shown in Figure 5(a) and after as shown in Figure 5(b)) is illustrated in Figure 5. The original dataset is moderately imbalanced, with the minority (suitable) class underrepresented, whereas SMOTE produces an approximately balanced training distribution. Figures 7(a) and 7(b) show the before and after comparisons on a scatter plot to provide a visual representation of the effects of SMOTE balancing. Table 4 summarizes the performance of classical and quantum models on the test set. This outcome is consistent with QML performance reports in NISQ-era studies, where deeper embeddings can be more expressive yet significantly more computationally expensive [2][6] and [10]. Accuracy, macro F1-score, weighted F1-score, recall for the minority class, and approximate runtime are reported. with macro and weighted variants averaging over classes with equal and support-based weights, respectively.

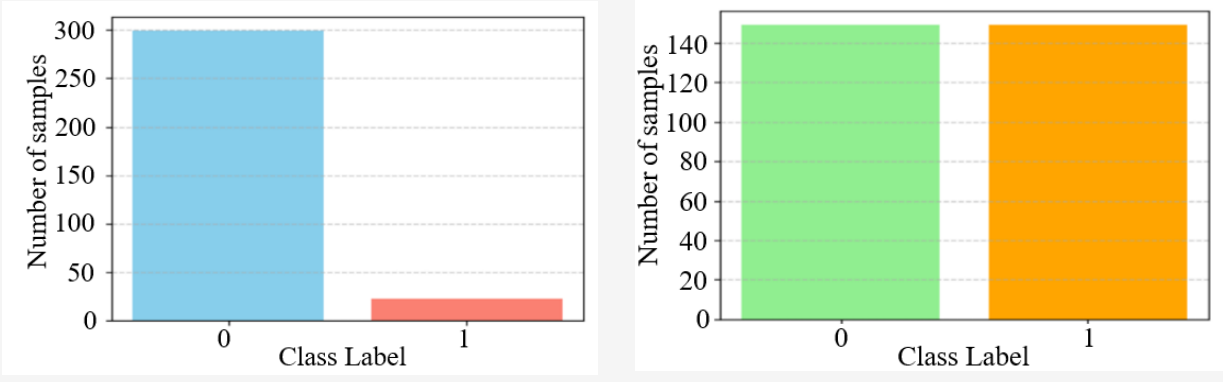


Figure 5: Class distribution in the training data:
(a) before SMOTE balancing, and (b) after SMOTE balancing

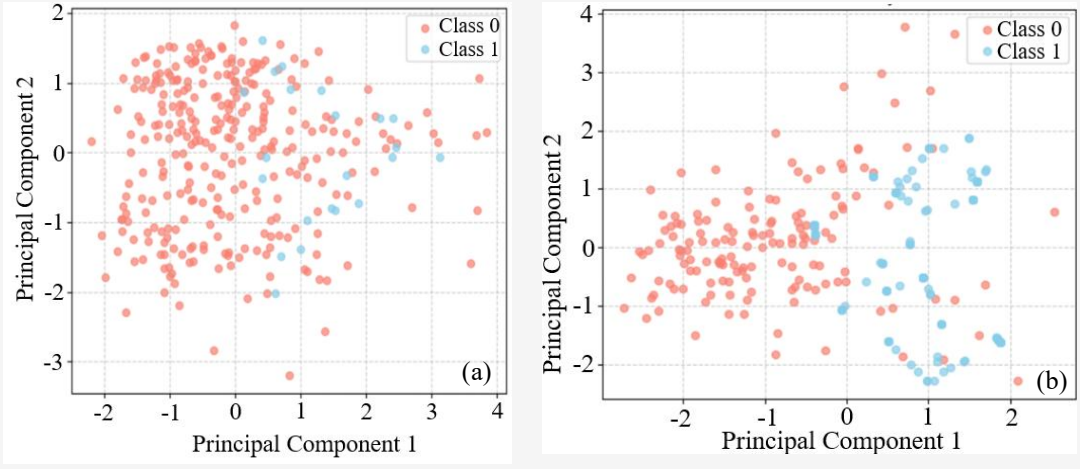


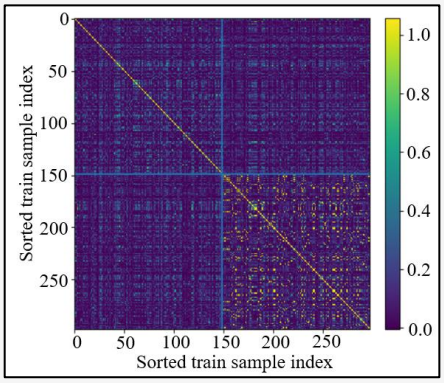
Figure 6: Training data projected onto the first two principal components as a 2D projection:
(a) before SMOTE balancing, and (b) after SMOTE balancing

Table 3: PCA components and explained variance (training set)

PCA Component	Explained Variance (%)	Cumulative Variance (%)
PC1	31.97	31.97
PC2	23.74	55.71
PC3	17.61	73.32
PC4	15.50	82.82
PC5	11.18	100.00

Table 4: Comparative performance of quantum and classical SVM models

Model	Feature Map / Kernel	Accuracy (%)	Recall	Macro F1	Weighted F1	Runtime
Classical SVM	Linear	98.09	1.00	0.98	0.97	~0.04 s (train) / ~0.02 s (predict)
Classical SVM	RBF	98.09	1.00	0.98	0.97	~0.04 s (train) / ~0.02 s (predict)
QSVM	PauliFeatureMap (3 qubits)	89.47	1.00	0.88	0.90	~17 min (train) / ~24 min (predict)
QSVM	ZZFeatureMap (3 qubits)	77.51	0.85	0.75	0.79	~8 min (train) / ~11 min (predict)



(a)

```
QSVM Accuracy (PauliFeatureMap, 3 qubits): 0.9091
F1-score (macro): 0.8968
F1-score (weighted): 0.9123

Confusion Matrix:
[[131  19]
 [  0  59]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.87      0.93       150
           1       0.76      1.00      0.86        59

    accuracy                           0.91       209
   macro avg       0.88      0.94      0.90       209
weighted avg       0.93      0.91      0.91       209
```

(b)

Figure 7: PauliFeatureMap QSVM:
(a) quantum kernel similarity structure, and (b) classification report on the test set

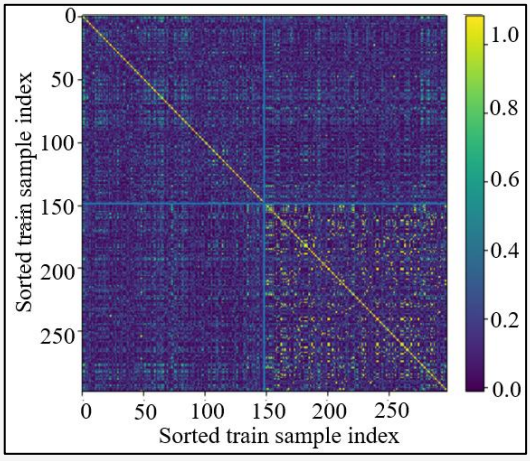
4.1 PauliFeatureMap QSVM

The PauliFeatureMap-based QSVM exhibited the strongest performance among the quantum models, achieving an accuracy of 89.47%, a macro F1-score of 0.88, and a weighted F1-score of 0.90. Notably, the classifier achieved a recall of 1.00 for the minority class, indicating that no suitable sample was misclassified, a desirable characteristic in ecological modelling where overlooking viable restoration regions is costly. This behaviour aligns with trends observed in recent QML literature, where rotation-based or Pauli-based encodings often achieve stronger stability than entanglement-heavy circuits when data is moderately sized and well-scaled [2][5] and [6]. The PauliFeatureMap encodes each principal component into parameterized rotations across the {X, Y, Z} Pauli basis. Since the encoded feature space remains smooth and continuous, the resulting kernel matrix tends to preserve geometric relationships between samples.

Figure 7 illustrates both the kernel similarity matrix (Figure 7(a)) and confusion matrix (Figure 7(b)). The heatmap shows higher intra-class similarity and a relatively smooth block structure, consistent with well-separated but compact decision regions. The confusion matrix confirms that misclassifications are limited to the majority class, preserving the minority (suitable) regions. These findings are consistent with reports that rotation-based feature maps generalize well under PCA-reduced and normalized feature spaces [5][6] and [7].

4.2 ZZFeatureMap QSVM

The QSVM model using the ZZFeatureMap yielded lower performance relative to the Pauli implementation, with an accuracy of 77.51% and a macro F1-score of 0.75. However, minority-class recall remained relatively strong at 0.85 as shown in Figure 7(b), suggesting that the entanglement structure may enhance the sensitivity of the feature map to subtle patterns in the geospatial data. The ZZFeatureMap introduces nonlinear interactions between features through parameterized entangling operations and gives us Equation 7. Figure 8 shows the kernel similarity matrix (Figure 8(a)) and confusion matrix (Figure 8(b)) for the ZZ-based QSVM. The heatmap demonstrates more irregular clustering and a noisier similarity structure compared to the Pauli encoding, indicating that the feature map may be over-expressive for the given sample size. This behaviour helps explain the lower precision and overall accuracy, even though minority patterns were somewhat preserved. Runtime was shorter than the Pauli configuration because of gate structure differences, despite a deeper circuit topology. Hyperparameters were tuned using cross-validation, and the cross-validated performance is summarized in Figure 9.



(a)

```
QSVM Accuracy (ZZFeatureMap, 3 qubits): 0.7703
F1-score (macro): 0.7457
F1-score (weighted): 0.7802

Confusion Matrix:
[[113  37]
 [ 11  48]]

Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.75      0.82       150
           1       0.56      0.81      0.67        59

    accuracy                           0.77       209
   macro avg       0.74      0.78      0.75       209
weighted avg       0.81      0.77      0.78       209
```

(b)

Figure 8: ZZFeatureMap QSVM:
(a) quantum kernel similarity structure; (b) classification report on the test set

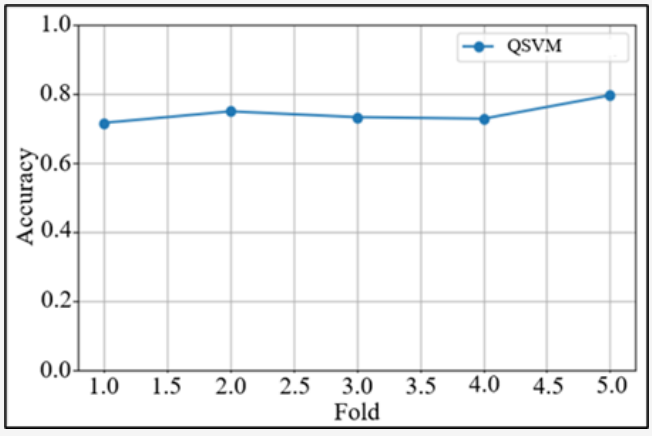
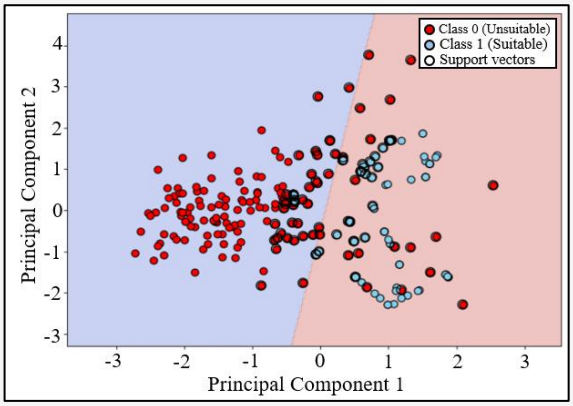


Figure 9: Cross-validation scores for classical SVM configurations on the pre-processed dataset



(a)

```
Classical SVM Accuracy: 0.9809
F1-score (macro): 0.9768
F1-score (weighted): 0.9810

Confusion Matrix:
[[146   4]
 [  0  59]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99       150
           1       0.94      1.00      0.97        59

    accuracy                           0.98       209
   macro avg       0.97      0.99      0.98       209
weighted avg       0.98      0.98      0.98       209
```

(b)

Figure 10: SVM performance assessment: (a) classical SVM, and (b) classification report

4.3 Classical SVM Baseline

Both linear and RBF classical SVM models achieved an accuracy of 98.09%, with strong macro and weighted F1-scores, demonstrating that the dataset becomes nearly linearly separable after preprocessing (standardization, scaling to $[-1, 1]$, SMOTE, and PCA). Compared with the quantum models, the classical SVM exhibited significantly faster training and inference times (milliseconds versus several minutes). Figure 10 visualizes the classical SVM performance, the confusion matrix (Figure 10(b)) indicates near-perfect predictions, while the 2D decision boundary plot in PCA space (Figure 10(a)) confirms well-separated class clusters, reinforcing why linear and RBF SVMs perform similarly. A qualitative sensitivity analysis with respect to circuit depth was conducted by examining representative shallow and deep embedding configurations reported in prior QSVM studies. Existing literature indicates that increasing circuit depth beyond moderate values often yields diminishing performance gains for small datasets while substantially increasing kernel evaluation cost [4][5] and [10]. In the present setting, the choice of a deeper embedding (reps = 16) is therefore not intended as an optimal configuration but as a demonstrative upper bound on expressivity, highlighting the trade-off between representational capacity and overfitting risk in small-sample geospatial applications.

5. Discussion

These results are interpreted in the context of algorithmic feasibility rather than predictive superiority, in line with the exploratory objectives of this study. The experimental results demonstrate that QSVMs can achieve non-trivial classification behaviour relative to classical Support Vector Machines when applied to carefully pre-processed remote-sensing data, consistent with prior studies evaluating quantum kernel methods in Earth observation tasks [4][5] and [6]. Nevertheless, in today's laboratory setups, QSVMs' accuracy and speed are not superior to those of traditional classifiers, which aligns with findings reported in recent comparative analyses of classical and quantum machine learning approaches [1] and [3]. Rather than emphasizing performance metrics in isolation, the discussion focuses on interpreting model behaviour, kernel structure, and computational trade-offs to contextualize the role of quantum kernel methods within applied geospatial analysis. It is important to note that the reported quantum runtimes correspond to statevector simulation overhead on classical hardware and are therefore not directly comparable to classical CPU training times or execution on physical quantum devices, as also highlighted in existing QML feasibility studies [5] and [10].

Uncertainties in NDVI, rainfall, and other environmental predictors may influence model sensitivity and decision boundary stability, a challenge commonly observed in remote-sensing-based environmental modelling [1] and [3]. The PauliFeatureMap demonstrated strong performance, combining expressiveness with robustness. This suggests that smooth function classes with rotational encodings can effectively capture underlying dependencies between NDVI, rainfall, topography, soil pH, and SOC, consistent with observations in prior quantum kernel studies [5] and [6]. The corresponding kernel similarity matrix exhibits a structured block pattern indicative of well-separated but compact classes. In contrast, the ZZFeatureMap achieved lower overall accuracy while maintaining relatively high minority-class recall, indicating that its entanglement-based structure enhances sensitivity to subtle feature interactions. However, the irregular kernel similarity patterns suggest potential overparameterization for the given dataset size.

The use of deep and entanglement-heavy quantum feature maps in conjunction with limited training samples introduces a clear risk of overfitting. The observed behaviour of the ZZFeatureMap supports this, as excessive expressivity can distort kernel similarity structures when data availability is limited, resulting in reduced generalization performance despite high minority-class recall. This aligns with observations in prior QML studies, where increasing circuit depth improves representational capacity but amplifies sensitivity to noise, sample sparsity, and kernel concentration effects [5][6] and [10]. These findings underscore the importance of carefully balancing circuit depth and dataset size when designing quantum kernels for practical geospatial applications.

The strong performance of classical SVMs on PCA-transformed data indicates that the dataset becomes nearly linearly separable after preprocessing, a phenomenon frequently reported in geospatial machine learning literature [1] and [3]. In such scenarios, the use of quantum kernels does not provide a clear advantage. Additionally, the quantum kernel evaluation times, obtained via statevector simulation on classical hardware, are several orders of magnitude larger than classical SVM training times, further emphasizing current simulation and hardware constraints. Overall, these findings support the notion of quantum parity reported in prior research, where quantum kernel

methods demonstrate comparable performance to classical models under controlled conditions but do not yet surpass them in practical applications [5][6][7][10] and [12]. The similarity matrices, confusion matrices, and PCA-based decision boundary visualizations indicate that both quantum and classical models learn comparable geometric structures in reduced feature space, despite significant differences in computational complexity.

6. Conclusion

This study evaluated the feasibility of applying Quantum Support Vector Machines to GIS-based land suitability analysis for reforestation using remotely sensed environmental data. A structured geospatial pipeline was developed in which satellite-derived indicators were integrated with classical and quantum machine learning models under identical preprocessing conditions, enabling a controlled comparison between QSVMs employing Pauli and ZZ feature maps and a classical Support Vector Machine baseline. The experimental results indicate that classical SVMs achieve higher predictive accuracy and substantially lower computational cost than QSVMs under the current experimental setting. While QSVMs based on Pauli and ZZ feature maps do not outperform classical models, they exhibit non-trivial classification behavior on small, structured datasets and reveal meaningful variations in kernel similarity and decision boundaries. These observations suggest that quantum kernel methods can capture complex feature interactions, albeit with increased sensitivity to dataset size, circuit depth, and preprocessing choices.

Importantly, the elevated computational cost observed in QSVM experiments reflects the overhead of exstatevector-based quantum simulation on classical hardware rather than execution on physical quantum devices. As such, runtime comparisons should be interpreted as indicative of current simulation and hardware constraints rather than inherent algorithmic inefficiency. The findings reinforce the view that, at present, QSVMs are best understood as exploratory analytical tools rather than replacements for established classical classifiers in applied geospatial modeling. Overall, this study demonstrates that while classical machine learning methods remain more efficient and accurate under current practical conditions, quantum kernel methods can exhibit meaningful and interpretable behaviour when embedded within realistic GIS-based workflows. By explicitly addressing limitations and focusing on feasibility rather than advantage, this work provides a cautious but informative reference for future investigations as quantum computing technologies and geospatial datasets continue to evolve.

6.1 Limitations

Nonetheless, considering these promising outcomes, several aspects of this research that could be improved will now be mentioned. First, all quantum computations were performed on noiseless simulators, rather than actual quantum hardware. Although this makes it possible to specifically assess the role of algorithmics, it does not include the effect of decoherence, gate errors, or qubit connectivity, which are relevant to actual hardware. Second, the number of examples in this study was relatively small and confined to one geographical area, lessening the ability to generalize successfully to different, or more varied, environmental conditions. Third, PCA was necessary to reduce dimensionality to match the number of available qubits, but this may have suppressed high spatial resolution relevant to reforestation, if this data indeed varies in this manner. Lastly, the rule-governed assignment scheme, though biologically inspired, involves certain threshold assumptions, potentially glossing over complexities in actual environmental dynamics. Future research might remove these present limitations, adding larger, more extensive regional data sets, validating these assignments against data sets of greater spatial resolution, and testing the QSVM algorithm against actual, more widely available quantum hardware. Consequently, absolute performance metrics reported in this study should be interpreted as indicative of model behaviour under heuristic labelling rather than definitive measures of ecological suitability. Similarly, the use of SMOTE for class balancing may introduce artificial feature relationships in the minority class, particularly under limited sample sizes, and the resulting performance metrics should therefore be interpreted as indicative of relative model behaviour rather than absolute predictive capability.

Acknowledgements

The authors would like to appreciate the contributions of open-access satellite data and geospatial platforms, which enabled performing this study. In addition, they would appreciate the contributions made by the authors of Qiskit and Google Earth Engine, which are free platforms for quantum machine learning and earth observations analysis. The authors would also appreciate the academic environment and computational support provided by Mr. Bohir Soham.

References

[1] Buthelezi, M. N. M., Lottering, R. T., Peerbhay, K. Y. and Mutanga, O., (2024). A Machine Learning Approach to Mapping Suitable Areas for Forest Vegetation in the eThekwini Municipality. *Remote Sensing Applications: Society and Environment,* Vol. 35. https://doi.org/10.1016/j.rsase.2024.101208.

[2] Şenol, H. İ., Yiğit, A. Y. and Ulvi, A., (2025). GIS-Based Multi-Criteria Analysis for Urban Afforestation Planning in Semi-Arid Cities. *Forests,* Vol. 16(7). https://doi.org/10.3390/f16071064.

[3] Adugna, T., Xu, W. and Fan, J., (2022). Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sensing, Vol. 14(3). https://doi.org/10.3390/rs14030574.

[4] Manaouch, M., Sadiki, M., Pham, Q. B., Zouagui, A., Batchi, M. and Al Karkouri, J., (2023). Predicting Potential Reforestation Areas by Quercus ilex (L.) Species Using Machine Learning Algorithms: Case of Upper Ziz, Southeastern Morocco. *Environmental Monitoring and Assessment,* Vol. 195. https://doi.org/10.1007/s10661-023-11680-1.

[5] Delilbasic, A., Le Saux, B., Riedel, M., Michielsen, K. and Cavallaro, G., (2024). A Single-Step Multiclass SVM Based on Quantum Annealing for Remote Sensing Data Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* Vol. 17; 1434–1445. https://doi.org/10.1109/JSTARS.2023.3336926.

[6] Miroszewski, A., Mielczarek, J., Czelusta, G., Szczepanek, F., Grabowski, B. and Le Saux, B., (2023). Detecting Clouds in Multispectral Satellite Images Using Quantum-Kernel Support Vector Machines, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 16; 7601–7613. https://doi.org/10.1109/JSTARS.2023.3304122.

[7] Zollner, J. M., Walther, P. and Werner, M., (2024). *Satellite Image Representations for Quantum Classifiers.* Datenbank Spektrum, Vol. 24, 33–41. https://doi.org/10.1007/s13222-024-00464-7.

[8] Zaidenberg, D. A., Sebastianelli, A., Spiller, D., Le Saux, B. and Ullo, S. L., (2021). Advantages and Bottlenecks of Quantum Machine Learning for Remote Sensing. *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium;* 5680–5683. https://doi.org/10.1109/IGARSS47720.2021.9553133.

[9] Buthelezi, M. N. M., Lottering, R. T., Peerbhay, K. Y. and Mutanga, O., (2024). The Use of Suitability Models and Remote Sensing to Map Forest Suitability: A Systematic Review. *Southern Forests: A Journal of Forest Science,* Vol. 86(4); 262–277. https://doi.org/10.2989/20702620.2024.2373749.

[10] Qniverse. (2024). *Quantum Support Vector Machine Explained with Qiskit, Qniverse Blog.* [Online]. Available: https://www.qniverse.com/qsvm-qiskit-tutorial [Accessed: Aug. 8, 2025].

[11]Alvarez, O. and Montaño, T., (2018). Functional Relationships of a Geospatial System for Reforestation of a Territory Using Geographic Information Systems. *OALib,* Vol. 6. https://doi.org/10.4236/oalib.1105193

[12] European Commission Joint Research Centre. (2023). Proceedings of the 2023 Conference on Big Data from Space (BiDS'23): From Foresight to Impact, Vienna, Austria; 6–9 November. https://doi.org/10.2760/46796.