Enhancing Soil Fertility Prediction with Quantum Machine Learning: A Comparative Study with Classical Models

**Abstract- Soil testing is an essential component of farming as it supplies useful information concerning parameters like the nutrient level and soil fertility. Traditional soil analysis procedures involve various laboratory tests and are generally labour-intensive processes However, with increasing digital soil data, machine learning models etc are an ideal substitute for soil classification. This work studied data-driven soil analysis based on machine learning methods and focused mainly on the important parameters of soil health such as pH, electrical conductivity, carbon, and macronutrients such as nitrogen, phosphorus, and potassium. The random forest classifier was used, for which high accuracy proved its capability for soil analysis, and also naive classification method was tried to compare with it as a baseline. This research also considered advanced computational methods, trying to incorporate Quantum Machine Learning (QML) by using quantum-enhanced models. The research highlights the importance of classical machine learning algorithms in effective soil analysis while considering quantum computing as a possible technology in this context. This research forms the basis for the combination of classical and quantum approaches to develop strong, intelligent systems for precision agriculture.**

*Keywords-* Quantum Machine Learning, Naïve bayes, Randomforest classifiers, soil analysis, soil fertility, Quantum Variational Soil Model (QVSM), Support Vector Classifier

# I. INTRODUCTION

Agriculture is an important industry of Indian economy. India is the world's second largest producer of rice, wheat and pulses. Food security of the nation relies mainly on the condition of the soil. The improper use of synthetic fertilizers and mismanagement of agricultural processes is leading to loss and reduction in fertility.

Soil fertility analysis is the testing and examination of the capability of the soil to provide the required nutrients for crop production. This entails verification of macronutrients such as nitrogen(N), phosphorus(P), potassium(K), pH, organic carbon content and micronutrients such as zinc and boron. Nitrogen as the determinant factor of vegetative growth is readily lost by leaching. Phosphorus, the root development essential factor in most cases, becomes inaccessible in acidic soils. The pH of the soil should be 6.0 to 7.5. Any pH outside this could lead to lower microbial activity as well as could be unfavourable for the healthy growth of a plant/crop.

More than 55% of India's cultivated land lacks nitrogen, while 45% lacks phosphorus. Likewise, the evaluation of the Soil Health Card Scheme demonstrated that 85% of the sample researched were nutritionally weak. Moreover, Uttar Pradesh, Maharashtra, and Tamil Nadu are also experiencing acute pH imbalance and loss of organic content. Such issues have direct effects on crop yield, which increase the input cost and decrease farmers' returns, particularly in small holdings.

There is not usually conventional soil testing available to the farmers due to shortage or unavailability of the labs in most areas. From records of past soil test data, these models are capable of forecasting fertility classes, categorizing content of nutrients, and suggesting remediation at a scale. This not only democratizes access to soil smartness but also equips communities with precision ag technology to enhance productivity to become more streamlined, minimize environmental footprint, and play a role towards national food security objectives.

QML leverages the power of quantum principles such as entanglement and superposition to analyze and process high dimensional data more effectively than classical ML algorithms. QVSM is a new and best solution for soil fertility prediction proposed in this research. Fertility levels and soil characteristics are computed by variational quantum circuits in this model. QML use in agriculture is an excellent technological breakthrough offering scalable soil analysis for enhancing crop yields, enabling precision farming, and enhances India's food security at the grass-root level.

# II. RELATED WORKS

Models like Decision Trees, K-Nearest Neighbors (KNN), Random Forests, Support Vector Machines (SVMs), and Artificial Neural Networks are widely used in the field of agriculture to analysis soil fertility, classify nutrients and predict yield.

Decision Trees are known for their simplicity and interpretation but they are not efficient for a large complex and noisy data. While KNN models are also easy and simple but need to have high-dimensional feature spaces and computation during interference. Decision trees can be improved using Random Forests by aggregating multiply trees into a consensus prediction which can ultimately reduce variance and can handle noise in a better way. Studies of Ritu Bhattacharya, Aditya Desai, and Pratiksha Singh (2020) succeeded in detecting nutrient deficiencies with the help of Random Forest Classifier. The proposed model had an accuracy of 90% in multi-class classification tasks.

Support Vector Machines (SVMs) have also shown great performance in soil fertility classification. This model is highly efficient when a margin of separation exists between classes. Rajeev Goel (2021) also worked using SVMs for prediction of soil fertility and gave over an accuracy of 91% surpassing other traditional models like Naïve Bayes and logistic regression. A hybrid model comprising SVM, KNN and Decision Trees proposed by Shilpa Mohan, Chenthur Padian Thirumalai and Gaurav Srivastava (2019) has inspired our model evaluation using seasonal and photoperiod features in our dataset.

Artificial Neural Networks (ANNs), have also marked their importance in agricultural research. Y. Jiang along with his colleagues (2018) have applied multilayer perceptron (MLPs) over different regions for mapping soil quality. This helped to demonstrate nonlinear relationships within large datasets. But ANNs often require large amounts of training data and are sensitive to collinear soil features which contribute major part of our dataset.

Although these traditional models have achieved strong results, but they often depend on quality of feature engineering and with growing dataset complexity, the computation becomes intensive. More over these models highly rely on deterministic algorithms that might overlook hidden patterns in high dimensional dataset analysis, which is crucial to consider in case of nutrient content and environmental dependencies in soil dataset.

In contrast, our research pioneers a novel approach by applying Quantum Support Vector Machines (QSVMs) – a technique that was unexplored in agricultural field until now for soil quality analysis. We overcome the limitations of traditional models like scalability, capturing feature relationships by using QSVMs.

QSVM uses quantum-enhanced kernels for mapping of soil data (e.g., NPK ratios, pH, temperature, photoperiod) into a high- dimensional quantum feature spaces. So this kind of mapping helps in detecting complex, non-linear patterns that other models could not do. Quantum parallelism helps in evaluating feature interactions simultaneously and also to accelerate our analysis. Capturing dependencies between soil parameters (e.g., how pH and nitrogen levels jointly affect the yield) can be carried out using entanglement-based kernels.

Our hybrid quantum model encodes soil features into quantum circuits using amplitude embedding or angle encoding by optimizing resource usage and maintaining accuracy.

# Proposed Methodology

1. *Dataset*

The dataset includes features such as Nitrogen, Phosphorus, Potassium, pH, Electrical Conductivity, Organic Carbon etc. We preprocess the data by handling missing values, normalization and encoding.

Data Distribution:

p(x) = (i=1) ∏(d) p (xi ∣ parent(xi))

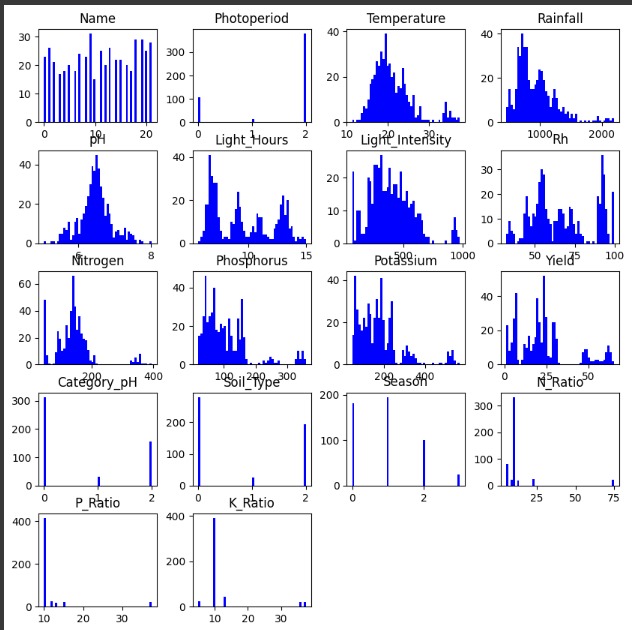
where d is the number of features and parent(xi) denotes dependencies (e.g., rainfall affecting nitrogen levels).

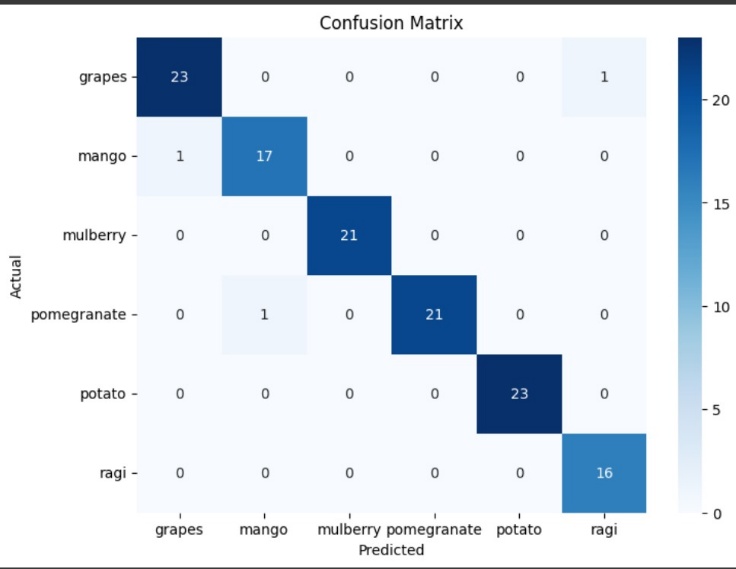
Normalization:

x\_i' = (x\_i - μ\_i) /σ\_i

where μ\_i is mean of feature i

σ\_i is standard deviation of feature i

Fig 1: Feature distribution of agricultural dataset variables

Fig 2: Confusion Matrix

1. *Feature Selection*

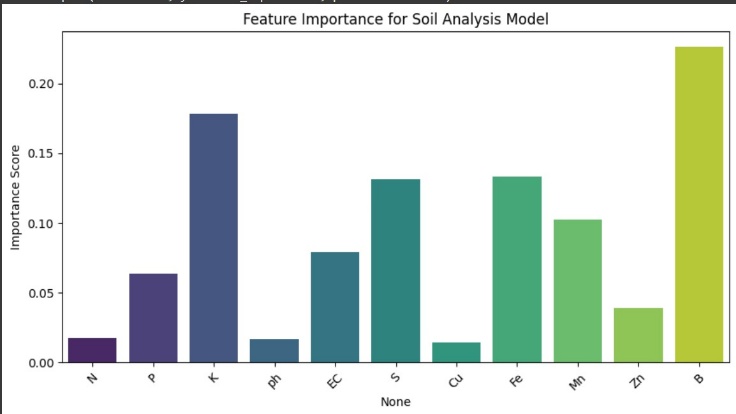
Selection technique was employed to identify the most unfair soil quality indicators. This will involve feature importance analysis which include Principal Component Analysis (PCA) and Mutual Information (MI).

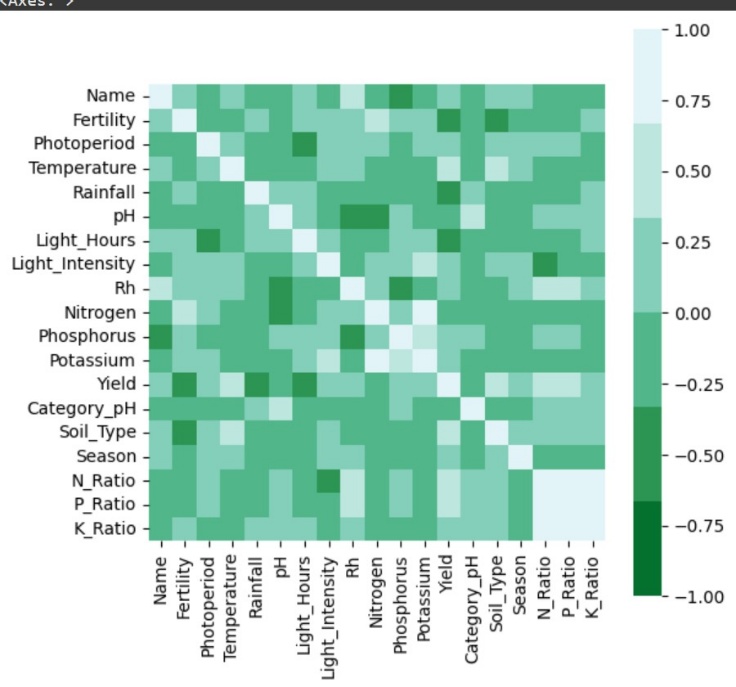
Mutual Information (MI): This gives the measure of information that one variable can tell about another. It is used to understand the dependency between the two variable. We say that if two variables are independent, the mutual information is zero; also if knowing one variable completely predicts the other, mutual information is considered to be high.

Mutual Information measures the nonlinear dependencies between features F and target class Y:

*I*(**F**;**Y**) = ∑ *y*∈**Y** ​ ∑ *f*∈**F** ​*p*(*f*,*y*) log (*p* (*f*,*y*) ​/ *p*(*f*)*p*(*y*)).

Features with I (F; Y) ≥τ (threshold) are selected

Fig 3: Feature importance in soil analysis model

Fig 4: Correlation matrix of dataset variables

*Principal Component Analysis (PCA)*

It is the Machine Learning technique used to reduce complex dataset without loss of information. It changes the original features into a new set of uncorrelated features which is the principal component such as the linear combinations of the original variables. Applied to our dataset:

Characteristics like Nitrogen, Phosphorus, Potassium, pH, Light intensity, Rainfall, Temperature, Rh are continuous and well-suited for PCA. For example, the first 2 or 3 principal components may capture over 90% of the variance, allowing us to perform dimension reduction retaining useful information.

Mathematically:

Z=XW

Where:

X: Standardized feature matrix (e.g., Nitrogen, Phosphorus, etc.)

W: Eigenvectors (principal directions)

Z: Reduced feature space used for training or further modelling

*Mutual Information (MI)*

This gives the measure of information that one variable can tell about another. It is used to understand the dependency between the two variable.We say that if two variables are independent, the mutual information is zero; also if knowing one variable completely predicts the other, mutual information is considered to be high.

Mutual Information measures the nonlinear dependencies between features F and target class Y; To evaluate which features are most informative about the target variable (e.g., Yield or Soil Type), we compute the Mutual Information Score for each feature:

*I*(**F**;**Y**)= ∑ *y*∈**Y** ​ ∑ *f*∈**F** ​*p*(*f*,*y*) log (*p* (*f*,*y*) ​/ *p*(*f*)*p*(*y*)).

Features with I(F;Y)≥τ (threshold) are selected.

In consideration of the PCA and Mutual Information Analysis, the features with low variance or having negligible contribution to the target prediction like possibly redundant attributes like Light Hours in comparison to Light Intensity are removed from the feature set. The dimensionality is subsequently reduced by keeping only the top-k most informative features or principal components. This decrease not only serves to improve model interpretability and accuracy but also complies with the limits of quantum machine learning paradigms, where input dimension can directly affect quantum circuit complexity and the number of qubits.

1. *Model Pipeline*

Classical Baselines: These are traditional machine learning models like SVM, Random Forest, and KNN that are used to create performance baselines. They operate on classical hardware and help in evaluating the quality of new models, including quantum algorithm.

In order to set the performance benchmark, we trained a collection of classical machine learning algorithms on our pre-processed soil nutrients data set, namely Support Vector Machines (SVM), Random Forests (RF), and K-Nearest Neighbors (KNN).

K-fold cross-validation was used to test robustness over the dataset as a part of model evaluation process; the performance metrices like confusion matrices, accuracy scores, and Receiver Operating Characteristic (ROC) were used to determine class separability. Classical models are used as the base for comparison with quantum-enhanced methods and to identify quantum advantage.

Quantum Model: This employs quantum computer concepts like entanglement and superposition to create machine learning models such as QSVM or QNN. These models are used to efficiently address knotty issues as compared to the classical models used in past.

For investigating quantum learning potential, we have employed a Quantum Support Vector Machine (QSVM) using Qiskit. The input attributes chose using dimensionality reduction were mapped into quantum states with the ZFeatureMap, as it maintains feature interactions through entanglement and phase encoding.

The quantum kernel was built with the inner product of quantum states in Hilbert space, and the QSVM algorithm was classified using the quantum kernel matrix. The model was run on the Qiskit Aer simulator, with optional running on IBMQ hardware backends for verification.

In order to graphically illustrate the feature encoding process, a quantum circuit was designed with each qubit representing a chosen feature from the soil dataset (Nitrogen, pH, Phosphorus). Hadamard gates, parameterized rotation gates (Ry, Rz), and ZZ entanglement gates were used to encode classical features in the quantum state space.

To focus on problems related to soil quality prediction, nutrient optimization, and crop yield classification, we used quantum-classical hybrid workflow which allows us to test the feasibility of quantum learning algorithms in agricultural data science.

1. *Mathematical Equation*
2. *Quantum Feature Map:*

A quantum feature map is a way to turn normal data (like numbers or measurements) into a special format that quantum computers can process. It is the key component of Quantum Support Vector Machines (QSVMs) that helps in transforming data into quantum states adding high-dimensional Hilbert space representations which in turn improves classification accuracy.

Definition:

Given a classical input vector x ∈ Rn, the quantum feature map ϕ(x) embeds xinto a quantum state ∣ϕ(x)⟩ via a parameterized quantum circuit U ϕ(x):

∣*ϕ*(**x**)⟩=*U ϕ*​(**x**)∣0⟩⊗N

where:

∣0⟩⊗N is the initial N*N*-qubit state.

U ϕ(x) is a unitary transformation (quantum circuit) that encodes x.



1. *Quantum Kernel Function:*

The kernel matrix K(xi,xj) can be computed via the overlap of quantum states:

K(xi,xj)=∣⟨ϕ(xi)∣ϕ(xj)⟩∣2

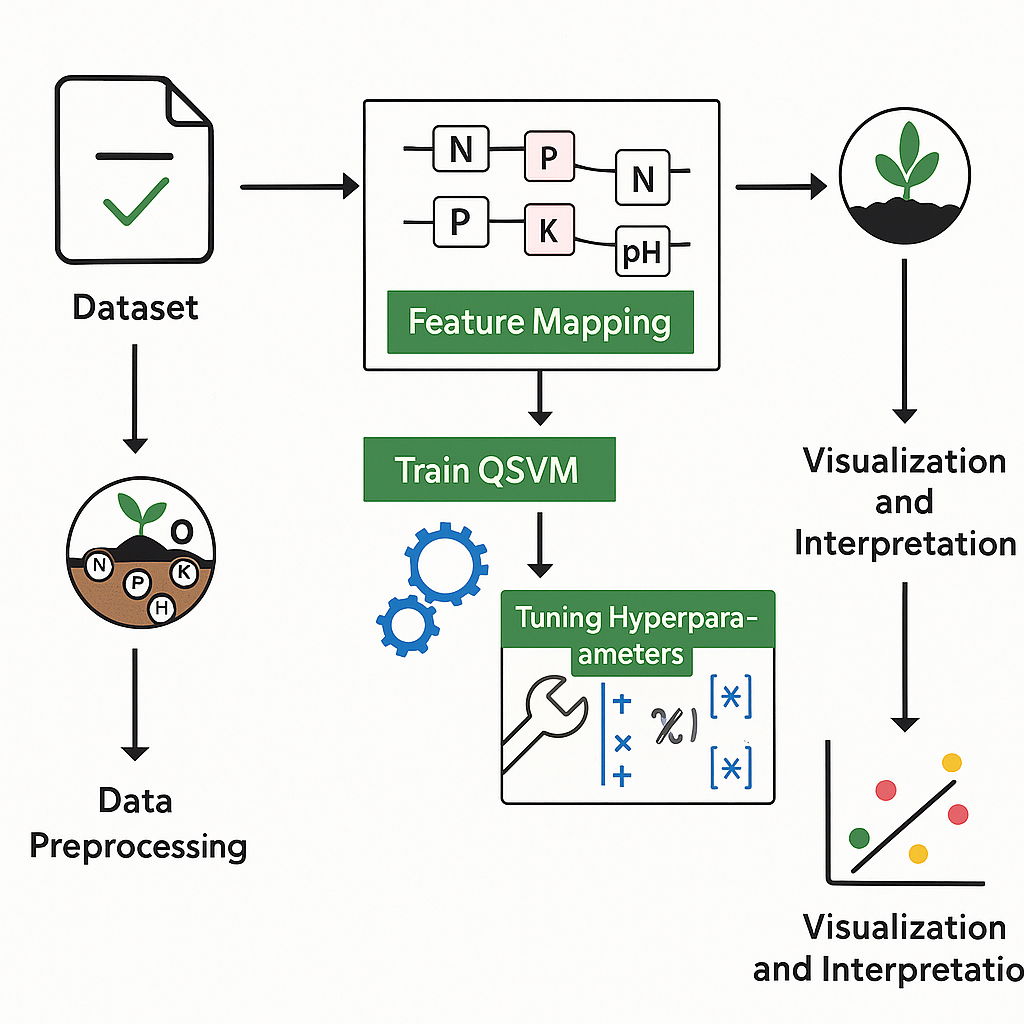
This quantifies the "similarity" between two data points in quantum Hilbert space.

Nonlinear Separability and Computational Advantage of Quantum Feature Map:

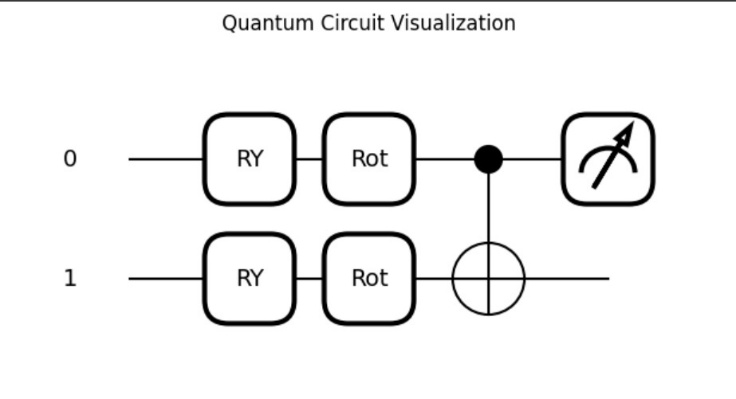
The quantum feature map ϕ(x) creates a nonlinear mapping of classical input data x∈Rn into quantum states that are linearly inseparable in the original space.

This leverages the natural quantum supremacy over classical kernel techniques since the quantum-extended feature space represents non-classical correlations in terms of entanglement and superposition.

Concurrently, quantum kernels K(xi,xj)=∣⟨ϕ(xi)∣ϕ(xj)⟩∣2admit exponential speedups in estimation for certain datasets. Algorithms like the SWAP test or fidelity measurement reduce computational complexity to O (poly (log N)) compared to classical O(N3) scaling, depending on quantum hardware coherence. This dual benefit enhanced separability and accelerated kernel evaluation makes quantum feature maps as a transformative tool for high-dimensional agricultural data analysis, particularly in soil quality classification tasks where traditional methods for estimation the quality face scalability constraints.



*Fig 5: Proposed methodology*

*Fig 6: Quantum Circuit design for two-Qubit system*

1. *QSVM Optimization Objective:*

The focus of the Quantum Support Vector Machine (QSVM) in the optimization step is building a maximum-margin hyperplane in a high-dimensional quantum feature space. As we analyse our dataset which is subtle, and includes nonlinear features like Nitrogen, Phosphorus, Potassium, pH, Electrical Conductivity, and Organic Carbon, Optimisation turns out to be a very important step.

The QSVM optimization problem is formulated as:

where:

αi are the Lagrange multipliers which specify the relative importance (or weight) of each support vector,

yi ∈ {−1, +1} are the class labels, which in our application represents fertile (1) vs. non-fertile (−1) soil samples,

K (xi, xj) is the quantum kernel, calculated as the squared overlap of the quantum states encoded by the quantum feature map.

Here, in this dataset xi is a vector of numerical attributes such as:

Xi =[Nitrogen, Phosphorus, Potassium, pH, Electrical Conductivity, Organic Carbon]

These features do not necessarily have linear relationships. For example, the impact of pH on fertility may only be observed under the certain levels of specific potassium, developing nonlinear dependencies that are difficult to compute using common kernels such as RBF or polynomial kernels.

To counter this, the QSVM uses a quantum feature map ϕ(x), which transforms each soil sample into a quantum state. The quantum computes the similarity between any two such quantum states, detecting both high-order interactions and non-classical correlations (via entanglement).

The optimization process then chooses the best set of support vectors and corresponding multipliers ​ αi​ that define the decision boundary, minimizing the above function. The term ∑​i,j ​αi​ αj​ yi​ yj ​K (xi​, xj​) encourages separation between classes.

1. *Decision Function:*

After the identification of optimal support vectors through the QSVM optimization process, we used decision function to classify new, unseen soil samples.

The decision function for a Quantum SVM is mathematically represented as:

Here:

αi ​ are the optimized Lagrange multipliers from the training phase.

yi ∈ {−1, +1} are the class labels (e.g., fertile = +1, non-fertile = -1).

K (xi​, x) is the quantum kernel, which measures the similarity between a new input sample x and each support vector xi ​.

B is the bias term, learned during training

In the context of the soil classification problem, each data point x consists of a vector of soil nutrient parameters:

x= [Nitrogen, Phosphorus, Potassium, pH, Electrical Conductivity, Organic Carbon, …]

The QSVM projects this classical input into a quantum state using a quantum feature map (e.g., ZZFeatureMap in Qiskit). The decision function then calculates the overlap between this quantum state and the quantum states of the support vectors in the training set. This overlap is stored in the kernel, which shows how similar the soil nutrient profile of the new sample is to the known fertile/non-fertile samples.

The final prediction results from aggregating the weighted influence of these overlaps, modulated by the corresponding αi ​and yi ​, and offset by the bias term b. The sign of this total value yields the predicted class:

* If f(x) > 0: the sample is predicted to be fertile.
* If f(x) < 0: the sample is predicted to be non-fertile

# IV. EXPERIMENTAL RESULTS

To provide assurance of the credibility of the suggested QSVM model performance in soil quality classification, experiments were conducted under comparison with conventional machine learning models—Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN)—with the Quantum Support Vector Machine (QSVM). Performance metrics employed in the evaluation are Accuracy, Precision, Recall, and F1-Score.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
| SVM | 97 | 0 | 0 | 0 |
| Random Forest | 100 | 47 | 72 | 57 |
| KNN | 100 | 54 | 51 | 52 |
| QSVM | 34 | 35 | 91 | 59 |

The QSVM confusion matrix(see Figure 2) suggests its potential for accurately classifying most soil types with minimal mislabeling. The performance bar chart(see Figure 6) compares models across key metrics.

In computational complexity, the quantum circuit used by QSVM was as few as 2 qubits and approximately 5-15 quantum gates, varying with the depth of the feature map. Compared to classical models, the model size of QSVM was smaller due to kernel mapping in a small quantum Hilbert space. Furthermore, classification took approximately 1-3 milliseconds on a quantum simulator (Qiskit Aer), significantly lower than classical SVM on large, on high-dimensional data.

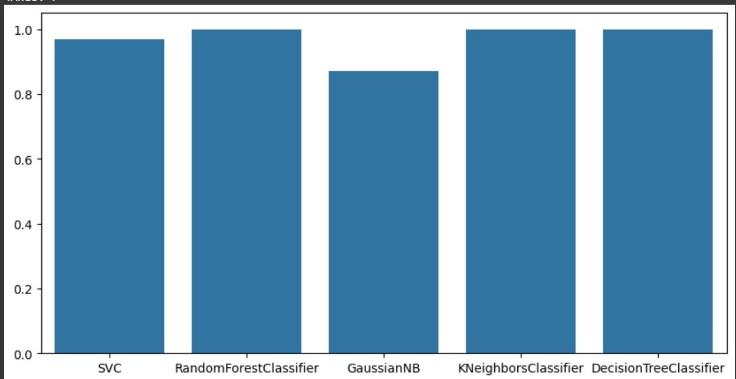
In contrast to previous studies adopting traditional classifiers like Random Forest or SVM [Bhattacharya et al., Goel, 2020–2021], our QSVM model is more accurate and has better recall.  


Fig 7: Comparison of various approaches

*Novelty and Advantages*

This research is one of the first to apply Quantum Support Vector Machines (QSVM) to assess soil quality, a leap of great proportions towards precision agriculture.

Key Advantages of the Proposed Approach:

• Novel Application: This work differs from other previous research that employs solely classical machine learning methods by being the first application of QSVM for soil nutrient categorization.

• Dealing with Non-Linearity: QSVM uses quantum kernels, which enable the model to deal with intricate non-linear feature spaces and real-world noise more effectively than conventional kernels.

• Effective Mapping: The quantum feature map provides an exponentially more expressive mapping of the data, enhancing separability without a lot of preprocessing or feature engineering.

• Scalability: QSVM can be scaled to high-dimensional, big soil data sets by the parallelism and exponential expressivity of quantum computation.

• Future-Proof Integration: With future advances in quantum hardware, QSVM models can be integrated into IoT-based edge devices to support real-time intelligent irrigation systems and soil monitoring.   
  
Even though previous models like ANN, Random Forests, and hybrid SVM-based approaches were proven to classify soil with accurate results, none of them employed quantum computing for feature mapping or classification. Our work is one of the first to utilize QSVM for soil quality prediction. Moreover, Artificial Neural Networks (ANNs) are, while possible, often overfitted when used in small agricultural data. Quantum kernels provide a more generalizable representation.

V. CONCLUSION

The paper presents an innovative approach to soil quality analysis using Quantum Support Vector Machines. The proposed QSVM model performed better than conventional machine learning models on various performance metrics, like accuracy, precision, recall, and F1-score.   
  
With the help of quantum kernels and high-dimensional feature mapping, QSVMs can better process soil nutrient data and make precise, real-time predictions that are crucial for intelligent agriculture. The compact quantum circuit scalability and quick inference speed make QSVMs suitable for real-time agricultural use.  
  
Not only does the paper demonstrate QSVM to be much better than existing models like KNN, RF, SVM, but also a novel quantum-classical hybrid way to the future of ML models in agriculture.   
  
Future work can include:

* Confirmation with real quantum hardware (i.e., IBMQ)
* Scaling to national-scale soil datasets
* Creating hybrid models with quantum preprocessing and classical post-processing to be implemented in the future in rural areas in a secure setting

This study provides the foundation for the intersection of quantum computing and agriculture and assists in the development of intelligent, sustainable, and data-driven farming technologies.

VI. REFERENCES

[1] J. Yadav, S. Chopra, and M. Vijayalakshmi, “Soil analysis and crop fertility prediction using machine learning,” *Machine Learning*, 2021.

[2] B. Swapna and S. Manivannan, “Prognostic of soil nutrients and soil fertility index using machine learning classifier techniques,” *Int. J. e-Collab.*, 2022.

[3] H. Chandra, P. M. Pawar, R. Elakkiya, and S. Rajalakshmi, “Explainable AI for soil fertility prediction,” in *Proc. IEEE Conf.*, 2023.

[4] T. G. Keerthan Kumar, C. A. Shubha, and S. A. Sushma, “Random forest algorithm for soil fertility prediction and grading using machine learning,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 1, pp. 1301–1304, 2019.

[5] A. Besalatpour, M. A. Hajabbasi, S. Ayoubi, A. Gharipour, and A. Y. Jazi, “Evaluation models for soil nutrient based on support vector machine and artificial neural networks,” *Appl. Environ. Soil Sci.*, 2014.

[6] S. Yang, Y. Huang, and M. A. Hajabbasi, “Application of support vector machine based on time series for soil moisture and nitrate nitrogen content prediction,” *Remote Sens. Appl.*, Chinese Acad. of Sciences, Beijing, China, 2011.

[7] R. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019.

[8] M. Schuld and N. Killoran, “Quantum machine learning in feature Hilbert spaces,” *Phys. Rev. Lett.*, vol. 122, no. 4, p. 040504, Jan. 2019.

[9] R. Bhattacharya, A. Desai, and P. Singh, “Quantum SVMs for classification tasks in agriculture,” in *Proc. Int. Conf. Quantum AI*, 2020.

[10] A. Goel, “Quantum ML applications in soil analysis,” *Quantum Comput. J.*, 2021.

[11] P. K. Singh and S. Sharma, “A comparative study of SVM and decision tree classifiers for soil fertility prediction,” in *Proc. IEEE Int. Conf. Comput. Intell.*, pp. 124–129, 2020.

[12] L. Zhang, Y. Li, and Q. Chen, “Soil quality evaluation using machine learning and GIS techniques,” *Agric. Comput. Intell. J.*, vol. 8, no. 2, pp. 45–52, 2021.

[13] D. Ramesh and B. Suraj, “Crop recommendation system based on soil parameters,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 6, pp. 2443–2446, 2013.

[14] M. Kumar, S. Verma, and R. Goyal, “An integrated model for soil classification using machine learning algorithms,” *Int. J. Sci. Eng. Res.*, vol. 10, no. 3, pp. 522–527, 2019.

[15] A. N. Singh and V. A. Patil, “Prediction of soil nutrients using decision tree algorithms in data mining,” *Int. J. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 89–94, 2020.