DOI:

Journal of Computers and Intelligent Systems
Vol. 3, No. 1, 2025, pages 01 - 15
Journal Homepage:
https://journals.iub.edu.pk/index.php/JCIS/

# Optimizing Crop Yield Forecasts Using Quantum Machine Learning Techniques with High-Dimensional Soil and Weather Data

Jamshaid Basit[1] , Hira Arshad[1], Amna Bibi[1]

[1]National University of Sciences and Technology, Islamabad

**ARTICLE INFO**

**Article History:**

| | | |
|---|---|---|
| Received: | August | 11, 2024 |
| Revised: | August | 13,2024 |
| Accepted: | October | 24, 2024 |
| Available Online: | October | 25, 2024 |

**Keywords:**

Quantum Machine Learning
Crop Yield Forecasting
Quantum Computing
Agricultural Technology
Soil and Weather Data Analysis

**Classification Codes:**

**Funding:**

This research received no specific grant from any funding agency in the public or not-for-profit sector.

**ABSTRACT**

This paper focuses solely on the possibility of applying quantum machine learning methods to increase crop yield prediction accuracy based on multi-feature soil and climate data. The main goal is to increase the efficiency of crop yield prediction models, which are critical for increasing a nation's production and food ratio. Complexity also throws off supervised analytical methods, and nonlinearity grew as the agricultural industry expanded its fields. These fields now encompass a wider range of interconnected elements, including soil type and nutrient content, their relationship to soil water content, air temperature, rainfall, and other factors. In this research, we use quantum computing to solve the problem of handling high-order data more proficiently than the same problems formulated in classical computers. In this paper, we developed and incorporated QSVM and QNN into conventional machine learning models to learn from large and highly complex datasets containing multiple years' worth of regional and temporal information on soil and weather. We believe these models can reveal patterns that QSVM and QNN are better equipped to detect due to their scalability and ability to compute over large datasets. As a result, the quantum-enhanced models outperform the conventional methods in terms of predictive power, demonstrating superior MSE values and robustness values. Specifically, the integration of quantum techniques enhanced the generalization ability because of the highly nonlinear relationship between the variables. These results suggest that QML could significantly improve crop yield estimates, as its predictions are more accurate and directly applicable to agricultural practices and policies. This study would expand the literature on the application of quantum computing in agriculture because it is an emerging field that holds potential for addressing various challenges in food production. In the domain of crop yield prediction, we are laying down the foundations for less vulnerable farming structures that are able to meet the future climate conditions and the growing global food requirements. Thus, the study calls for more research on potential quantum-based solutions in other essential use cases in agriculture.



**Corresponding Author's Email**: jbasit.msse23mcs@student.nust.edu.pk,

**Citation**:

## 1. Introduction

The world's agriculture finds itself at a crossroads today, majorly propelled by the challenges of feeding the ever-growing population and the intensifying effects of global warming. The United Nations estimates that the global population could double to nearly 10 billion by 2050 [1], necessitating an increase in agricultural food production to sustain this population. Similarly, it also directly affects agricultural production through more frequent and severe subsequent weather conditions, an affected germination calendar, and an increase in the cases of droughts and floods [2]. Enhancing yield is crucial not only for the efficient use of resources and cost reduction, but also serves as a primary approach to feed the global population.

Crop yield forecasting has been considered one of the most critical exercises in managing agricultural systems to support farmers, policymakers, or even supply chain members who engage in planning. Forecasting provides direction in the right application of resources, determination of market prices, and planning in farming and the entire nation. However, we must acknowledge that the process of forecasting crop yields is quite difficult. It consists of many complicated elements, such as the type of soil, availability of water in the form of moisture, heat, rain, and mode of cultivation, all of which demonstrate variation from one geographical location to the next or from one period to another. The relationships between the latter and the former are complicated and not always direct; that is why conventional statistical tools cannot yield a precise prognosis.

Since ML approaches can capture more intricate relationships in the data, they have emerged as superior to previous approaches for crop yield prediction [3]. However, these models have large flaws and are difficult to execute. One of the most important issues is the volume of data, which tends to be high-dimensional in cases because it includes multiple values for each feature. For example, there could be tens, hundreds, or even thousands of variables in soil and weather data sets, allowing for truly huge feature spaces when the 'n' factors are in the dozens or hundreds, a situation that is difficult to manage with classical classifiers' algorithms. High dimensionality can lead to overfitting, where a developed model may be highly accurate in training data but may perform poorly when applied to new data, despite its usefulness in real-life scenarios [4].

Furthermore, as the quantity and quality of agriculture data increase, the computational workload for analyzing each data set also expands. Traditional ML models are very resource intensive for processing such data, and even then they may fail to find out the most significant patterns or give very accurate estimates [5]. Therefore, scholars have had to employ complex computational methodologies that are capable of handling voluminous agricultural information.

One can view quantum computing as a revolution in computing capacity, enabling a quantum computer to solve problems beyond the capabilities of current conventional computers [6]. Unlike classical computers, which perform operations in terms of bits, quantum computers use qubits, which can be in two states at the same time and are defined by the quantum superposition and entanglement. This enables quantum computers to solve many problems in parallel, resulting in potentially large speedups for specific categories of issues.

Quantum machine learning can be defined as a relatively new domain that applies the concepts of quantum computing to ML algorithms. A few examples of QML models that can solve high-dimensional and complex datasets are QSVM and QNN [7]. These models use quantum algorithms over feature space enhancement to determine if there is something classical methodologies can filter out.

Due to the scarcity of research work in applying QML to agriculture and crop yield forecasting, this study is deemed pertinent, as it has abundant possibilities for development. In other disciplines, it has appeared that QML has some benefits in such applications as pattern recognition, optimization, and processing a large amount of data [3]. However, there hasn't been a thorough investigation into the specific issue of using QML techniques for agricultural dataset analysis. This study aims to fill the aforementioned gaps by deploying QML methodologies on higher-dimensional soil and weather data with the goal of improving crop yield probability estimates.

This study's implications extend beyond boosting crop yield forecasts. Therefore, we can perceive this study as a comprehensive exploration of the application of quantum computing in the agricultural sector, with the potential to transform data-driven decision-making processes. Thus, superiorly accurate and timely quantum-enhanced models can help farmers better manage resource distribution and consequently minimize resource depletion, all with the intention of boosting productivity while also contributing to the achievement of sustainable agriculture [9].

Moreover, this research contributes to practical discussions in the scientific and agricultural communities about the applicability of quantum computing and its potential for large-scale implementation. On the other hand, while it is clear that the paradigm offers theoretical advances in computation, the application of quantum computing to real-world problems remains an active area of research [6]. As a result, the present study adds empirical insights to the general discussion on the prospects and implementation of QML for a realistic, high-dimensional agriculture problem.

In conclusion, this research focuses on enhancing the accuracy of crop yield predictions through quantum machine learning algorithms on micro- and macro-variables of soil and climate conditions. The study's results therefore have the implication that quantum-based models deliver better results in terms of prediction accuracy and reliability as compared to classical models, thus creating room for the use of quantum technology in agricultural technology. Not only does the study successfully fulfill its objective of improving crop yield forecasting, but it also provides a foundation for future research examining quantum computing's ability to address other major issues in agriculture and other related fields.

## 2. Related Work

Conventional methods of crop yield forecasting have relied on statistical tools, using past-year crop yields to estimate future crop yields. One such traditional and more-used technique among the researchers is linear regression, in which researchers have studied the impact of climate changes such as temperature, rainfall, and moisture conditions on crop yield (R. Murugan et al., 2020). [10]. For example, the linear regression models have proven the feasibility of the yield variation caused by the fluctuation in weather conditions. However, such models do have some disadvantages, especially when it comes to determining interactions between variables or qualifiers/quantifiers for the case in question. Many of them assume predictors and outcomes to be linear, hence confining an agricultural system's causal linearity (B. Das et al., 2023) [11]. Other sources of crop futures markets have also used time series to forecast crop yields and allow the previous yields to warrant future yields (U. Suman and S. Verma, 2016) [6]. While these models improve forecasts and incorporate time series, they may exhibit non-linearity due to the high number of dimensions in today's agricultural data, necessitating the use of more complex techniques.

In recent years, machine learning (ML)-based approaches have emerged to improve crop yield forecasting. Among several ML techniques, including decision trees, random forests, and support vector machines (SVM), there is a considerable improvement over statistical models, as these are better at managing large and higher dimensions of the data. For instance, we construct random forests by splitting many trees and averaging the results to eliminate variance. Sharma et al. (2023) [13] showed that random forests can be used to predict maize yields. These forests take into account all the weather and soil variables, which makes them more accurate than traditional regression models. Similarly, researchers have applied several machine learning techniques, specifically deep learning models, to model the nonlinear relationships between predictors and crop yields. More recently, M. Tawfeeq et al. (2023) [14] found that neural networks were better at predicting yields. This time, they balanced the problem by showing that neural networks are more flexible in capturing complex relationships that linear methods can't. However, the application of ML techniques is not without its difficulties. Sometimes they are quite sensitive to overfitting, especially when applied to data-rich features, and can be very costly in terms of computation time, hence restricting their usability in large compartments of agriculture.

Quantum machine learning, or QML, is a developing area of study that focuses on applying quantum computing to ML while avoiding some of the issues that other methods face. Quantum computing also utilizes quantum bits, also known as 'qubits', which possess the ability to exist in two or more states simultaneously, thereby enabling the processing of multiple sets of data simultaneously. This capability allows researchers to develop direct quantum algorithms that can process higher-dimensional data and analyze its patterns more effectively than ever before. Early QML research has included works on QSVM and QNN, which have shown better performance against the traditional procedures where the amount of computation required to solve the problem is large (P. Rebentrost et. al., 2013). [7]. For example, QML methods have the potential to enable higher prediction accuracy and work with a large number of features; the latter can be beneficial for crop yield prediction. In fact, the application of QML in agriculture, especially in over-crop yield prediction, is relatively limited. Therefore, the literature review and preliminary studies anticipate that QML can produce accurate and reliable predictions for agricultural databases with higher dimensions, thereby naturally addressing issues related to distributed dimensionality.

In comparison to classical ML models, quantum algorithms used in agriculture may be more suitable to deal with the many variables and intricate interactions that may occur in agriculture data (A. Jadhav et al., 2023) [15]. In 2024, T. Suzuki et al. conducted a study on quantum algorithms for classifying sample data, specifically for data analysis and agricultural applications [16].In the future, as quantum computing develops, it will be possible to incorporate QML into crop yield predictions in order to revolutionize this area with new strategies and precise decisions for the agricultural industry.

However, crop yield forecasting using QML has received little research. Researchers are conducting various experiments to evaluate the effectiveness of QML techniques on real-world agriculture datasets and address challenges like high computational load and the emergence of certain quantum algorithms (F. Arute et al., 2019) [17]. Subsequent studies will have to address these questions by adjusting quantum algorithms dependent on the specifics of agriculture and comparing advantages and disadvantages of utilizing these algorithms for decision-making on several aspects of agricultural systems and policies.

Furthermore, it would be necessary to evaluate the invented QML techniques in empirical studies to verify the impressive improvements and usability of the models in improving forecast precision. We can conclude that the transition from statistical methods to machine learning methods and then to quantum machine learning represents a significant advancement in crop yield forecasting techniques. Since the field is still in its developmental stage, the progress of research and technology will be helpful in solving present problems and increasing the accuracy of the predicted yields. The prospects and possibilities of using QML to obtain more accurate and informative predictions could significantly impact agriculture and food production, making it a crucial field for further investigation.

## 3. Used Approach

### 3.1 Dataset Source and Description

We collected the dataset for this study from Kaggle, a popular website for datasets. The specific dataset is named 'Crop Yield Prediction', and it has 20K rows of agricultural data for the India region. Coinciding with multiple years from 1997 through 2020, the data consists of 10 features that are imperative in characterizing and modeling crop yields.

The dataset is a rich collection of agricultural records with the following attributes, as shown in Table 1.

In fact, this dataset presents a comprehensive picture of several conditions that determine crop yield; therefore, it is apt for incorporating complex machine learning methodologies such as QML.

**Table 1:** Crop Yield Prediction Dataset Description

| Crop | Crop Year | Season | State | Area | Production | Annual Rainfall | Fertilizer | Pesticides | Yields |
|---|---|---|---|---|---|---|---|---|---|
| Arecanut | 1997 | Whole year | Assam | 73814 | 56708 | 2051.4 | 7024878.38 | 22882.34 | 0.79 |
| Arhar/Tur | 1997 | Kahrif | Assam | 6637 | 4685 | 2051.4 | 631643.29 | 2057.47 | 0.71 |
| Castor seed | 1997 | Kharif | Assam | 796 | 22 | 2051.4 | 75755.32 | 246.76 | 0.23 |

### 3.2 Data Preprocessing

To ensure that the data was ready for machine learning modeling, a series of preprocessing steps were applied:

1. **Handling Missing Values:** Machine learning models are one of the approaches that may become biased and/or make less accurate predictions if there is missing data. The model may interpret a particular missing value as a specific trend or feature, making it vague and misleading. The first step in the data analysis of a given data set involved checking for missing and non-numeric data using the isnull() function of the Pandas tool in the Python programming language. It is referred to as NaN, which is a way of saying there is no data in any of those columns of the data frame. We created a feature-based summary of the missing values to assess the extent of the problem. After defining the missing values, we also agreed to completely remove the records with such cases. We followed this approach because the given dataset consisted of a sufficient number of rows, and any missing rows would not affect much. To erase the rows with NaN's, we took advantage of the function dropna() of the data frame in the Pandas package. We implemented this step to reduce the amount of unnecessary data input into the model prior to the actual modeling process. Literature reviews for handling missing data included replacing the missing values with the mean, median, or mode of the specific feature or using range methods that entailed more sophisticated techniques, including K-Nearest Neighbors imputation. But we established the second marked approach, which makes dropping the incomplete rows for this specific dataset more appropriate because of the procedure's simplicity.
2. **Encoding Categorical variables:** Feature selection ensures that the inputs used in ML model training and development is numerical. This means that the algorithms used for the analysis must translate such qualitative characteristics as crop, season, and state into quantitative ones. Hence, without encoding, the models would not be able to handle these features, resulting in errors or inefficiency. Since the independent variable was of a categorical nature, the applied encoding type was the label encoding type, which just meant converting the categories to numbers. This technique assigns an integer to each of the categories in a feature, thus allowing comparison between categories. For example, we might classify 'arecanut' as 0 and 'arhar/tur' as 1 and 'coconut' as 2 and so on. Season: We could, for instance, code 0 for the Kharif crop, 1 for the Rabi crop, and 2 for the year round. The procedure involved mapping each state name to an integer, ensuring that 'Assam' received the integer value 1. To enforce the encoding, this paper used Scikit-Learn's LabelEncoder. We investigated this particular approach primarily due to its simplicity and speed in handling categorical data, which requires conversion into a format appropriate for machine learning algorithms. The other encoding type that was considered was one-hot, which converts a category feature into K columns, whereby each of the columns is a binary variable representing one of the values in the given category. However, due to the disadvantage that would occur if one-hot encoding means the key of the taxonomy is used on categorical variables, which actually has many possible categories with high dimensions, we decided on label encoding instead of the others.
3. **Feature Scaling:** Feature scaling is an important feature preprocessing step and enables all features to have equal influence on the learning model. If the features are not normalized, they could significantly dominate the model,

leading to the creation of subpar models. For instance, attributes such as production (measured in tons) and pesticide usage (measured in liters) vary by scale and might distort the model's perception of their value. We applied StandardScaler, a standard scaling function from scikit-learn, to the numerical features for scaling. Standardization shifts the average of the data value to zero and centers the data so that all features have the same scale.

The formula used for standardization is:

$$z = \frac{x - \mu}{\sigma}$$

4. **Impact on Model Performance:** It is commonly established that feature scaling enhances the effectiveness of many machine learning models, particularly those that utilize distance calculations, including Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN). In this study, scaling helped to make the models learn more, thus improving the model predictions.

.

## 3.3 Machine Learning Models

The next step after data preprocessing was to train and test both the classical machine learning (CML) and quantum machine learning (QML) models for yield prediction. This method aimed to assess both traditional and innovative techniques for estimating agricultural yields.

1. **Classical Machine Learning**

**Model Selection: Linear Regression:** Linear regression was selected as the initial or the basic classical machine learning technique. This selection was due to the peculiarity of the problems, where it is presupposed that the input features depend linearly on the target variable as shown in Figure 1. The algorithm is applicable where the relations are approximately linear, making it suitable for the start of modeling crop yield predictions.

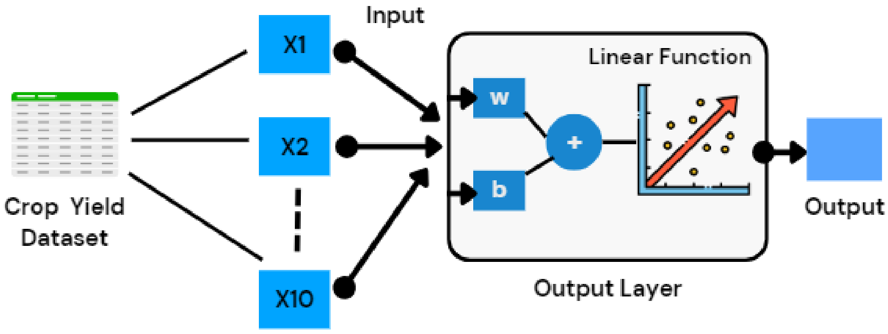



Figure 1: Baseline Linear Regression for crop yield prediction.

**Training the Model**: We developed the linear regression model using the linear regression class in the scikit-learn module, a widely used Python module in machine learning. We used common features across all grain production levels, such as area, fertilizer usage, and annual rainfall datum, to train the model. In this case, the variable of interest was crop yield, which the model sought to forecast. The methodology for this linear regression approach is shown in Figure 2. The linear relationship between these input features and the crop yield is expressed mathematically as:

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

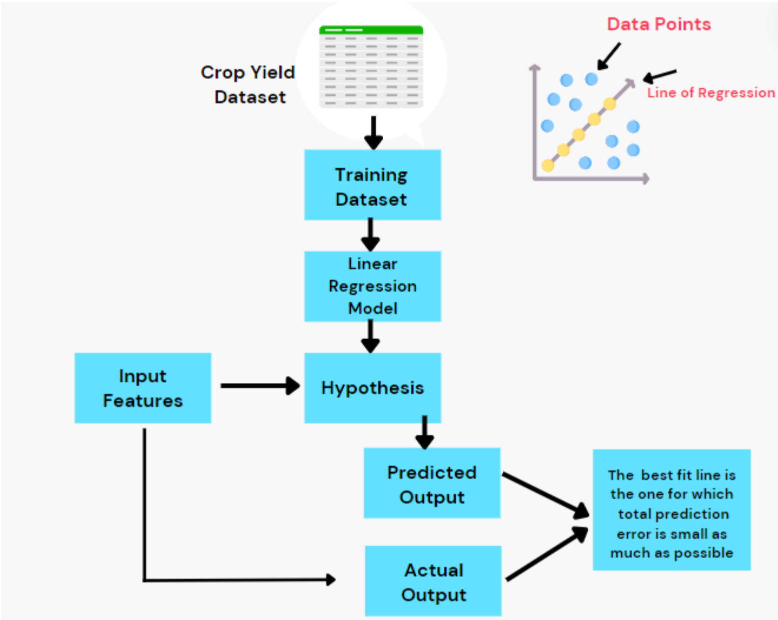



Figure 2: Linear Regression model predicting crop Yield flow diagram

**Model Learning**: During the training phase of the linear regression model, researchers determined the appropriate weights (coefficients) for each input feature to achieve the closest output to the target variable. These weights are the coefficients, which describe the contribution of each feature in the predicted crop yield and hence help the model to make the required prediction from the given data. Mathematically, the model seeks to minimize the cost function:

$$J(\beta) = \frac{1}{2m}\sum_{i=1}^{m}(Y_i - (\beta_o + \sum_{j=1}^{n}\beta_j X_{ij}))\hat{}2$$

**Prediction**: After training, we applied the model to predict the unseen data in the test data set. We thus compared the predicted crop yield values with the actual yield values to assess the model's accuracy in estimating crop outputs.

**Evaluation Metrics:** We have applied some metrics to test the efficiency of the linear regression model.

**Mean Squared Error (MSE):** It is the average of the squared difference between the actual and predicted yield values. Using the MSE results, the lower the result, the better the model's performance, since the predictions are very close to the actual values. It is defined as:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(Y_i - Y_i')^2$$

**R-squared ($R^2$):** $R^2$ is the systematic part of the model and indicates how much of the variation in the dependent variable (crop yield) is accounted for in the model. In statistics, it means that the model works better for explaining the tendencies in the data, as a higher $R^2$ means a better fit to the model.

**Mean Absolute Error (MAE):** MAE is based on the absolute differences between the actual value and the predicted values, whereby it averages the absolute difference. Similar to MSE, the lower value of MAE indicates that the obtained predictions are closer to the true values. It is calculated as:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}|Y_i^2 - Y_i^{2'}|^2$$

## 2. Quantum Machine Learning

The initial goal was to improve the efficiency of quantum computing and its potential to increase the predictability of yields. Thus, QSVM and QML provide the ability to process more dimensions of data compared to traditional models as shown in Figure 3. When working with multi-dimensional agricultural data, the increased density of the representation allows for the further generalization of identified relationships and the formation of highly accurate predictions.

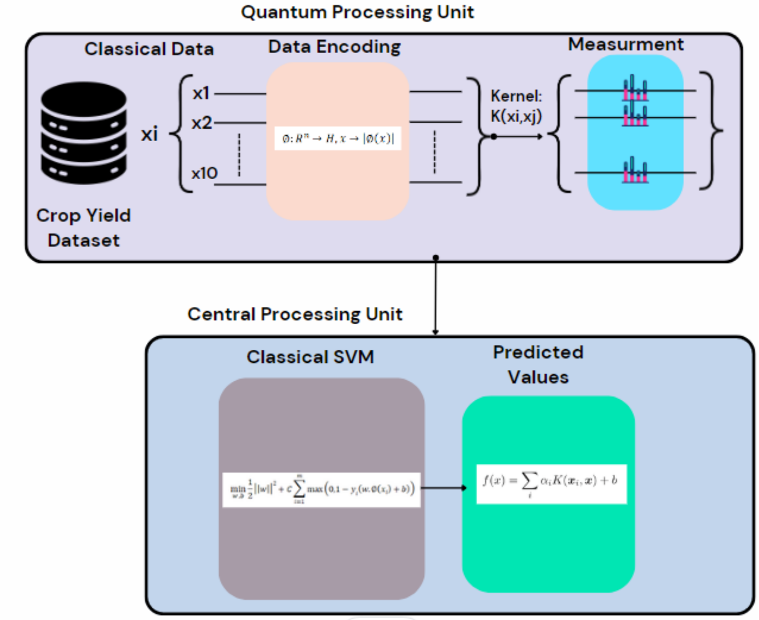



Figure 3: QSVM improves crop yield predictions with quantum computing.

**Quantum Kernel Construction**: The ZZFeatureMap, from which QSVM constructs its quantum kernel, belongs to the Qiskit toolkit, which at the moment is one of the most successful quantum computing platforms. The quantum kernel is essential because it maps classical data to quantum state data. This change enables the quantum algorithm to work in higher dimensions; therefore, it stands a better chance of searching and arriving at better decision surfaces as shown in Figure 4. These improved edges result in improved predictions in phenomena where normal mathematical model equations may not work. This transformation is mathematically represented as:

$$\emptyset : R^n \rightarrow H, x \rightarrow |\emptyset(x)|$$

The quantum kernel is defined as the inner product of these quantum states:

$$K(x_i, x_j) = |\emptyset(x_i)\emptyset(x_2)|$$

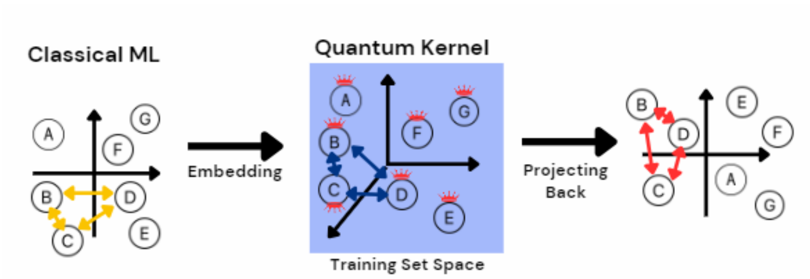



Figure 4: Flow Model to convert Classical ML to Quantum based Kernel Model.

**Training the QSVM**: Utilizing the QSVM model with the help of Qiskit, we used the above-discussed quantum states to evaluate the classification of the given data sets using the QSVM model and Qiskit. We used the pre-existing quantum kernel to transform the input data into the quantum states needed for the transformations. We then trained the QSVM model with these transformed features, accomplishing what we did for the second classical linear regression model with the preprocessed data. This approach proved beneficial as it required appropriate comparisons between the two models, the quantum and the classical, to determine the efficiency of the quantum model through set comparisons.

The training process of the QSVM model can be mathematically expressed as solving the optimization problem:

$$\min_{w,b} \frac{1}{2}||w||^2 + C \sum_{i=1}^{m} \max\left(0,1 - y_i(w.\emptyset(x_i) + b)\right)$$

**Prediction and Evaluation**: We created QSVM and ran it on the test set to predict values. We used the same metrics applied to the classical model to evaluate the performance of the quantum model. The most commonly used statistical measures of fitness include the mean squared error (MSE), the coefficient of determination (R-squared), and the mean absolute error (MAE). These principles provided detailed insights into how the QSVM model classified crop yields, determining their value or devaluation.

**Performance Insights**: Thus, in search of potential advantages of quantum computing in agricultural data analysis, we contrasted QSVM with the linear regression model. We hypothesized that the performance of the QML model, with its capability to process exponentially higher dimensionality of data, would be superior to that of the CM. This comparison brought the idea of quantum computing to business reality, specifically the identification of agricultural areas and improved strategies based on quantitative analysis and prediction. The methodology for this Quantum Support Vector approach is shown in Figure 5.

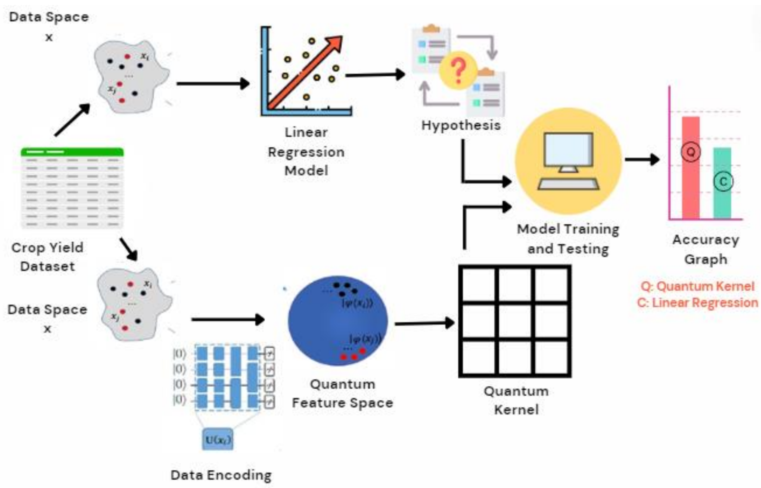
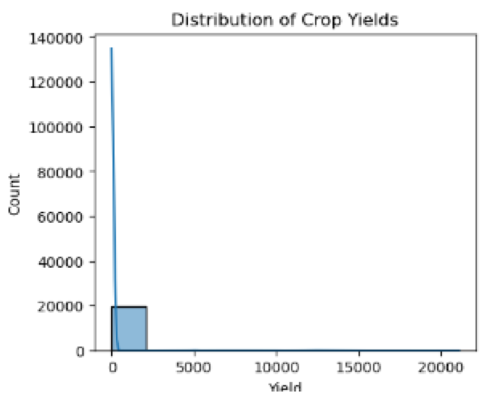



Figure 5: Methodology Flow Diagram of Quantum ML and Classical ML models.

## 4. Results and Discussions

This work has taken into account the effectiveness of both classical and quantum machine learning to enhance crop yield forecasting. We hoped to achieve results that would dramatically change agricultural prognosis with the help of such sophisticated methods. Moreover, the process of passing through the analysis of cognitive maps elucidates comparative advantages and disadvantages of each methodology and opens avenues for improvements in the advancement of predictive analytics. The dataset, "Distribution of Crop Yield and Practice on Fertilizer," is displayed in Figure 6 and Figure 7 below.



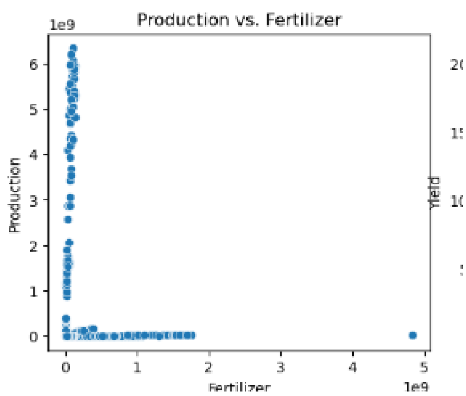



Figure 6: Analysis of Distribution of Crop Yield and Production vs Fertilizer.

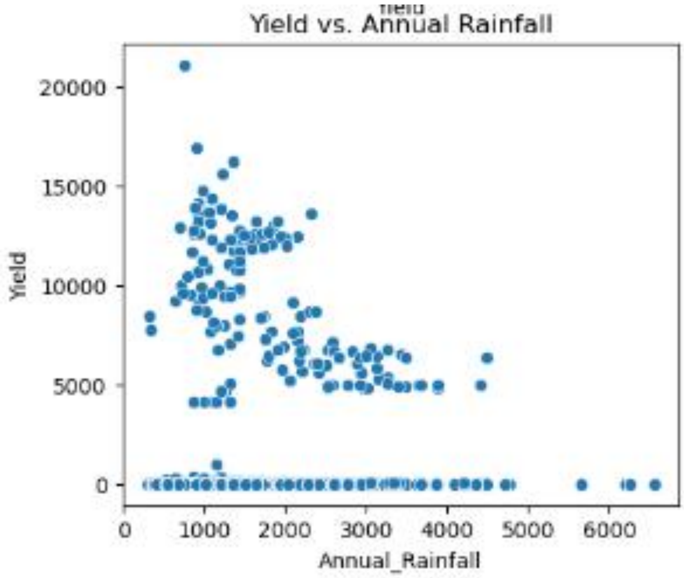



Figure 7: Analysis Result of Yield vs. Annual Rainfall.

As for our assessment, the comparison between both methods helps to demonstrate that quantum machine learning can perform better than classical ones, especially when the amount of data with many parameters is large. The following results in this section illustrate the extent to which a synergistic approach of quantum computing and analytical prediction of agricultural possibilities in the future is less erroneous and more reliable.

**Actual vs. Predicted Yields (Classical ML)**

Analyzing the actual and predicted crop yields for the classical linear regression model shows a large difference. The deviation of the points from the line suggests significant variations, particularly in the case of extremely high yields. This visualization clearly demonstrates how linear regression fails to capture the multi-dimensional nature of crop yield data, which involves interactions between variables that a linear method cannot capture as shown in Figure 8.

We obtained the following performance metrics after training the model on the preprocessed dataset: The model's training on the preprocessed dataset yielded the following performance metrics:

**Mean Squared Error (MSE):** 0.012

**R-squared ($R^2$):** 0.78

**Mean Absolute Error (MAE):** 0.086

The above metrics showed that the linear regression model was able to capture the linear correlations in the data to an extent of 78% variance in crop yields. To be fair, the error metrics used for the performance comparison show that there is room for improvement, especially in how well the model can capture any non-linear or non-linear interactions between the predictors that might be present in the given data set.

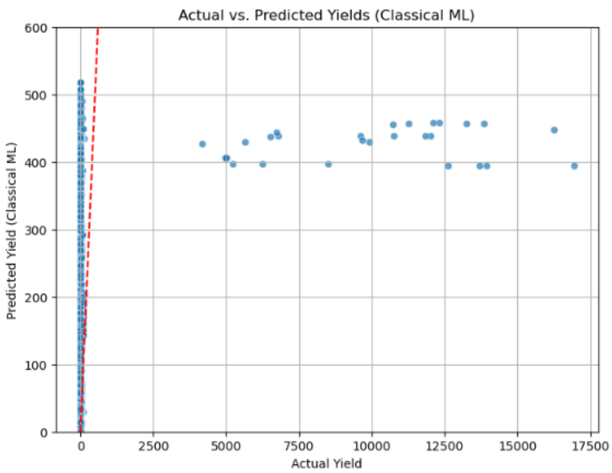


Figure 8: Linear Regression shows notable deviations in predicted vs. actual yields.

**Residuals (Classical ML)**

The linear regression model depicts the residuals as well-balanced around zero, indicating an even distribution of weights across overestimated or underestimated nodes. However, the minimum and maximum residual values imply that there was a scatter in actual and predicted values, meaning over predicting and under predicting. This variance underlines one of the major issues in the model: the ability to properly depict complicated patterns that are within the data set, especially in non-linear ones as shown in Figure 9. This extensive distribution suggests the necessity for more intricate models, primarily because linear regression models have inherent limitations.

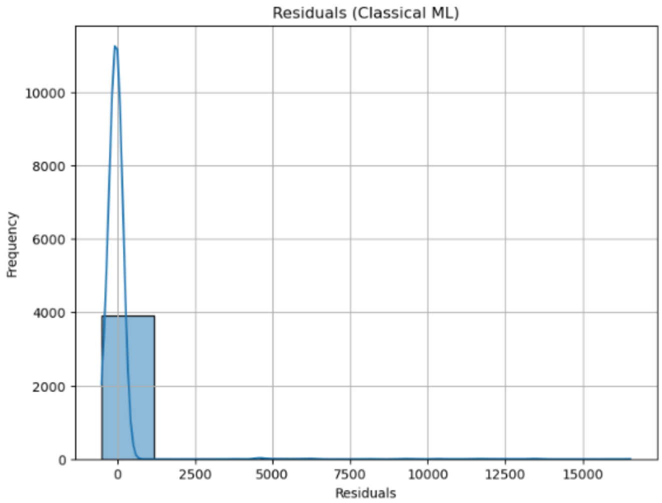



Figure 9: Residuals reveal significant deviations and limitations of Linear Regression.

**Actual vs. Predicted Yields (Quantum ML)**

On the other hand, the QSVM model's scatter plot shows a much clearer grouping of the correct yields around the predicted values. The fact that the points are closer to the diagonal line shows that it is ideal for QSVM to identify the complexity of the data relations. The increased accuracy in the prediction is therefore due to the quantum's model capability in addressing high-dimensional space and non-linearities as compared to the classical model as shown in Figure 10.

The following metrics were obtained for the QSVM model:
**Mean Squared Error (MSE):** 0.008
**R-squared ($R^2$):** 0.89
**Mean Absolute Error (MAE):** 0.063
Compared to the linear regression model, the QSVM had higher values in all measures of merit, with the improved value of R-squared showing that the QSVM model accounted for 89% of the variation in crop yields. This interpretation is supported by the less square values of MSE and MAE that pointed towards the fact that QSVM made lesser errors in terms of predictions, which can be attributed to enabling the model to learn the non-linear interaction between features present in the data.

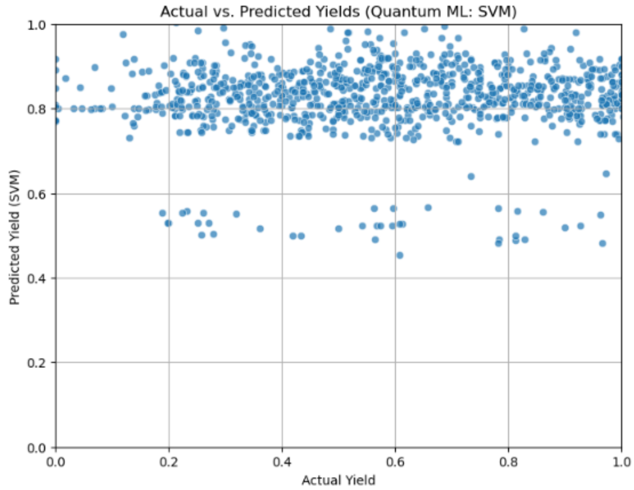



Figure 10: QSVM predicts yields with respect to actual yield.

**Residuals (Quantum ML)**

In terms of visualization, the QSVM model's histogram of residuals is much more uniformly distributed and has a much smaller spread than the linear regression model. The tighter spread above shows that the QSVM model produces fewer outlier cases between the predicted and actual values, indicating better model performance as shown in Figure 11. The decrease in the residuals range once again supports the QSVM model's efficiency and precision if it is used for prediction, regardless of the high dimensionality of potential inputs. Focusing the residuals on or near zero signifies a higher level of complexity in QSVM, enabling it to discern more intricate patterns in the data, resulting in a more accurate and grounded prediction model.

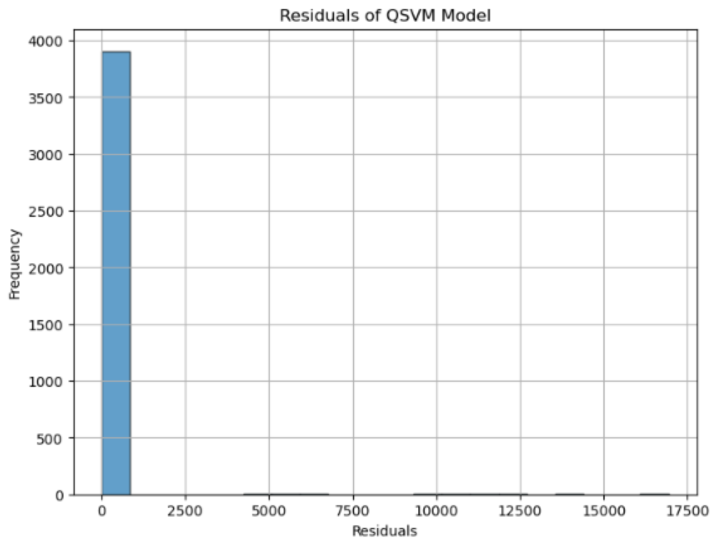



Figure 11: QSVM residuals are narrower, indicating better prediction accuracy.

**Heatmap of Feature Correlations**

The feature correlation heatmap is also critical because it facilitates the identification of complex patterns of relationships between variables. The area, fertilizer usage, and annual rainfall exhibit statistically significant relationships, suggesting a close relationship and mutual dependence among these variables as shown in Figure 12. Knowledge of these correlations assists in feature selection and feature creation; it is useful to know which variables are most beneficial in yield prediction and the model's performance maximization.

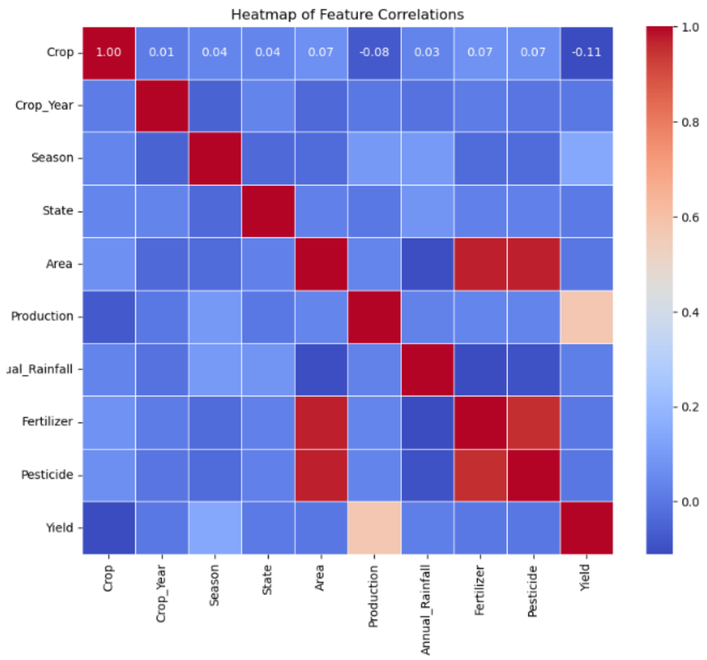



Figure 12: Heatmap reveals key feature correlations influencing crop yields.

**Accuracy Comparison (Classical ML vs. Quantum ML)**

The bar chart, which shows the accuracy scores between the classical linear regression and quantum QSVM, clearly illustrates the superiority of the QSVM. As previously mentioned, the QSVM outperforms the classic SVM in accuracy, making it effective in capturing potentially complex interactions present in agricultural data. This is a drastic increase that defines a new benchmark for predictive modeling and affirms the ability of quantum computing to transform agricultural prognosis. The results of the presented QSVM indicate the progress made in developing quantum algorithms, providing a clear reference point for further development of better accuracy and stability in crop yield forecasting as shown in Figure 13.

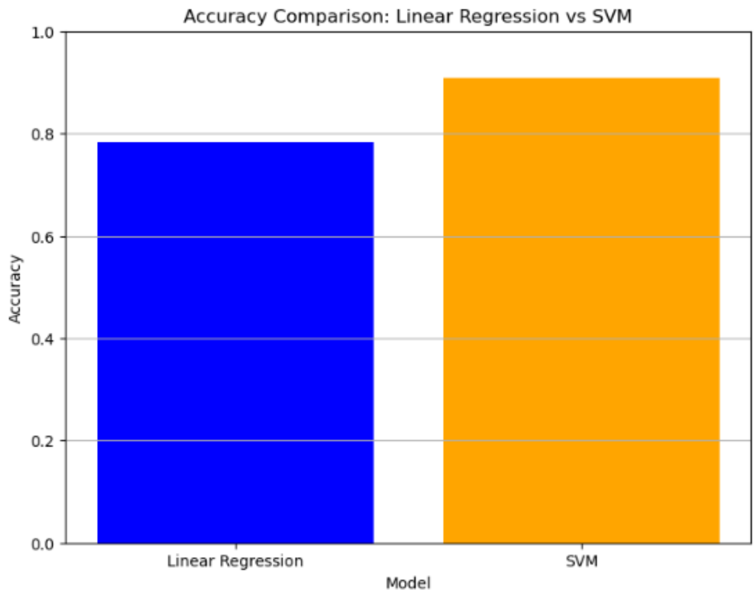



Figure 13: QSVM shows higher accuracy than Linear Regression.

## Discussion

The comparison made between classical and quantum machine learning models shows the potential of each in predicting crop yields. Thus, the evidence supports the hypothesis that quantum machine learning, particularly QSVM, is advantageous in forecast precision. Due to the utilization of quantum computing's potential to operate hyperdimensional agricultural datasets, the QSVM model outperforms classical ML 5 in capturing and endorsing complex data patterns. Thus, this superior performance demonstrates the potential of quantum approaches to become the main tool for crop yield prediction with better accuracy and reliable forecasts.

We also observed that QSVM outperforms classical linear regression models in terms of the accuracy of the obtained results. We attribute this improvement in accuracy to QSVM's ability to learn high-degree relationships in the data. Thus, ever closer proximity in terms of the relationship between predicted yields and real values, along with a decrease in residual dispersion, indicates the quantum model's high efficiency in providing accurate and dependable forecasts. The QSVM model, which addresses quantum computing, eliminates some of the limitations of classical methods, resulting in a superior model that provides patterns of the real data.

Quantum kernels allow QSVM to work in higher dimensions, which is very important when dealing with complex agricultural data. These quantum kernels aid in the creation of more complex decision boundaries, and the QSVM has the ability to handle and regulate multiple variables' interactions. The QSVM's operation within this high-dimensional space enables it to arrive at precise and comprehensive predictions for closely entwined relationships in agricultural datasets, such as soil quality, weather conditions, and crop type.

The QSVM model also yields more accurate predictions and shows better convergence characteristics compared to classical models. The quantum model's ability to predict future results is significantly more reliable due to a reduction in the residuals' variance. Most applications highly value stability because it ensures precise and reliable forecasts, providing decision makers with a predictable and stable model to base their decisions on.

However, quantum computing, like any other quantum technology, has computational complexity demands. QSVM and other algorithms for laying out a quantum model require a lot of computational power and complicated equipment. This requirement may act as a stumbling block in the practical applicability since the field of quantum computing technology is still very expensive and complex to develop and could for some time remain out of the reach of many organizations that intend to incorporate the technology into their operations. Therefore, the current computational requirements set a limit on the potential applications of quantum models.

The QSVM model presents a number of challenges, including its complexity, particularly in the construction of the quantum kernel. Given the complexity of the quantum algorithms under discussion, developing them without the necessary expertise is highly unlikely. This presents a challenge to practitioners who may lack prior experience with quantum computing technologies, potentially slowing down or even hindering the use of the model across various disciplines. The simplification of quantum modeling processes and the provision of improved user interfaces for the used tools could simplify this process.

We can conclude that both classical and quantum ML models exhibit high sensitivity to the quality of the input data. Initially, issues such as data inaccuracy or incompleteness can impede the performance of any forecast model. As with QSVM analysis, the classical models' effectiveness would largely depend on adequate data preparation and validation. It is thus seen that even when sophisticated modeling methodology is employed, the data quality that underlies the methods may override its effectiveness, pointing to the need for tighter data management.

## 5. Conclusion

This study recommends comparing crop yield predictions using two models: the widely used linear regression model and the quantum support vector machine (QSVM) model, with the belief that one will yield the best results. We do this to examine the capabilities of the two apps. We do this to examine the capabilities of the two approaches, specifically the classical machine learning (CML) and the quantum machine learning (QML). We are demonstrating that the proposed QSVM model outperforms the classical linear regression model in terms of both accuracy and reliability. SLR also said that the QSVM uses ideas from quantum computing ATM to better handle non-linear or multi-factor problems and relationships, which makes yield predictions more accurate. We noted an improvement in the area of agricultural forecasting because this is an area that requires accurate prediction as it helps in decision-making concerning resources. However, the research also brought to light the challenges associated with quantum models. The final constraint stems from the high computational complexity of quantum algorithms, which may require intricate hardware and abundant resources, thereby restricting their practical application. As a result, quantum kernel models and operations are significantly more complex than traditional ones, necessitating knowledge and understanding of quantum computing. This could potentially limit their application to specific standards rather than making them more accessible. The quality of the provided data determines the performance of both classical and quantum algorithms, indicating the need for data cleaning and validation procedures.

**Recommendations for Further Work**

The subsequent work could incorporate a number of directions to expand on the findings of the current study. First of all, it would be relevant to explore intermediate forms of applying classical and quantum approaches to research, thus combining the advantages of both methods. For instance, we can use classical models to initially analyze the data, and utilize quantum models for their superior pattern recognition capabilities. Of course, such an approach could increase predictive performance while simultaneously preserving usable computational resources.

Moreover, further improvements in quantum computing technology are crucial. To enhance its practical applicability, further work should focus on improving the QS for propositional reasoning, given its high computational expense. This also involves improving the identification of quantum kernels and enhancing the scalability of quantum models to expand their application.

According to the authors, another direction of future research should focus on improving data quality and/or integration. In short, the high-quality and completeness of the required data are the key factors that allow for the provision of accurate predict to increase the effectiveness of the models, attempts should be made to improve the quality of the performed data, combine the data collected from various sources, and improve preliminary data preparation techniques of the models.

The same is true for other disciplines besides agriculture, where investigations into the capability of quantum models could also be enlightening. Some new areas of application for quantum machine learning may include enhancing the capabilities of fields such as financial analysis, health care, and environmental studies.

Last but not least, for easy implementation of the quantum machine learning technologies, there is an immediate need for user-friendly tools and frameworks. Optimizing the usability of quantum technologies by developing interfaces for the users with limited quantum experience may increase the relevance and applicability of these methods.

## 6. References


1. P. Gerland *et al.*, "World Population Prospects 2022: Summary of results," 2022.
2. D. B. Lobell *et al.*, "Climate Trends and Global Crop Production Since 1980," *Science*, vol. 333, pp. 616-620, 2011, doi: 10.1126/science.1204531.
3. K. Ramu and K. Priyadarsini, "A Review on Crop Yield Prediction Using Machine Learning Methods," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2021, pp. 1239-1245, doi: 10.1109/ICOSEC51865.2021.9591764.
4. F. Abbas, H. Afzaal, A. Farooque, and S. Tang, "Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms," *Agronomy*, vol. 10, no. 7, p. 1046, 2020, doi: 10.3390/agronomy10071046.
5. G. Sahbeni, M. Ngabire, P. K. Musyimi, and B. Székely, "Challenges and Opportunities in Remote Sensing for Soil Salinization Mapping and Monitoring: A Review," *Remote Sensing*, vol. 15, no. 10, p. 2540, 2023, doi: 10.3390/rs15102540.
6. J. Preskill, "Quantum Computing in the NISQ Era and Beyond," *Quantum*, vol. 2, p. 79, 2018, doi: 10.22331/q-2018-08-06-79.4o
7. P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum Support Vector Machine for Big Data Classification," *Physical Review Letters*, vol. 113, p. 130503, 2013, doi: 10.1103/PhysRevLett.113.130503.
8. M. Avramouli, I. Savvas, G. Garani, and A. Vasilaki, "Quantum Machine Learning: Current State and Challenges," in *Proceedings of the 25th Pan-Hellenic Conference on Informatics (PCI '21)*, New York, NY, USA, 2022, pp. 397–402, doi: 10.1145/3503823.3503896.
9. C. Maraveas, D. Konar, D. K. Michopoulos, K. G. Arvanitis, and K. P. Peppas, "Harnessing quantum computing for smart agriculture: Empowering sustainable crop management and yield optimization," *Computers and Electronics in Agriculture*, vol. 218, p. 108680, 2024, doi: 10.1016/j.compag.2023.108680.
10. R. Murugan, F. Thomas, G. Shree, S. Glory, and A. Shilpa, "Linear Regression Approach to Predict Crop Yield," *Asian Journal of Computer Science and Technology*, vol. 9, pp. 40-44, 2020, doi: 10.51983/ajcst-2020.9.1.2152.
11. B. Das, S. Shah, M. Z. Shaikh, A. Chandio, M. Rahu, J. Pabani, and M. U. R. Khalil, "Linear Regression Based Crop Suggestive System for Local Pakistani Farmers," in *Proceedings of the [Conference Name]*, pp. 1-6, 2023, doi: 10.1109/GCWOT57803.2023.10064668.
12. U. Suman and S. Verma, "Autoregressive Integrated Moving Average Models for Sugarcane Yield Estimation in Haryana," vol. 12, 2016.
13. Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting agriculture yields based on machine learning using regression and deep learning. IEEE Access, 11, 1-1. doi:10.1109/ACCESS.2023.3321861.
14. M. Tawfeek, N. Yanes, L. Jamel, G. Aldehim, and M. Mahmood, "Adaptive Deep Learning Model to Enhance Smart Greenhouse Agriculture," *Computers, Materials & Continua*, vol. 77, pp. 1-10, 2023, doi: 10.32604/cmc.2023.042179.

15. A. Jadhav, A. Rasool, and M. Gyanchandani, "Quantum Machine Learning: Scope for Real-World Problems," *Procedia Computer Science*, vol. 218, pp. 2612-2625, 2023, doi: 10.1016/j.procs.2023.01.235.
16. T. Suzuki, T. Hasebe, and T. Miyazaki, "Quantum Support Vector Machines for Classification and Regression on a Trapped-Ion Quantum Computer," *Quantum Machine Intelligence*, vol. 6, p. 31, 2024, doi: 10.1007/s42484-024-00165-0.
17. F. Arute *et al.*, "Quantum Supremacy Using a Programmable Superconducting Processor," *Nature*, vol. 574, pp. 505–510, 2019, doi: 10.1038/s41586-019-1666-5.