Quantum-AI-for-Climate

Team Information:

Team name : Entangled Engineers

Team Member 1:

Name: Amit Modhwadia

Womanium Program Enrollment ID: WQ24-OdfMsSyLDN6hEms

Team Member 2:

Name: Karan Islur

Womanium Program Enrollment ID: WQ24-FayGICKoq3EUQJV

Team Member 3:

Name: Shreya Rajpal

Womanium Program Enrollment ID: WQ24-ZolPctO2nQuC7op

Team Member 4:

Name: Saiyam Sakhuja

Womanium Program Enrollment ID: WQ24-RUSHoi3Q0FBNjRN

Problem Statement

Water is essential for the survival and well-being of all living organisms, playing a critical role in hydration, temperature regulation, and overall health. However, the impacts of climate change, along with pollution and industrial activities, have significantly compromised water quality, posing serious health risks. Climate change exacerbates these issues by altering precipitation patterns, increasing the frequency of droughts and floods, and contributing to the contamination of freshwater sources. Ensuring access to clean and safe drinking water is therefore of paramount importance. In light of these challenges, our project aims to predict water potability using advanced machine learning techniques. Additionally, we will leverage quantum computing to optimize these predictive models, enhancing their accuracy and efficiency in identifying potable water sources, thereby contributing to global public health and environmental sustainability.

Project description

Drinking water is essential. That's why It is important to know whether drinking water is safe to drink or not. There are many ways to do that like using a lab for tasting but it can be costly and take a lot of time. Instead of that we can use technology to determine the Potability of the water we can use Machine learning and AI to check the quality of the water. These models predict water potability easier and it takes less time. We have used the three different Machine learning modes to determine the Potability of the water and our model got higher accuracy than the previous model.

TASK 1

Title: Introduction to the Problem Statement: Climate Change and Water Potability

Problem Statement and Background

Climate change and water resources are deeply interconnected, creating a significant challenge that directly impacts billions of lives globally. As global temperatures continue to rise, the effects on water availability and quality become more pronounced, threatening essential resources for human consumption, agriculture, and ecosystems. The frequency and severity of droughts are increasing, leading to reduced availability of potable water. At the same time, excessive rainfall and flooding, intensified by climate change, contribute to the contamination of water sources, further jeopardizing water safety.

Water potability, which refers to the safety of water for drinking, is influenced by a variety of factors such as pH levels, hardness, and the presence of contaminants like chloramines and dissolved solids. These factors are highly sensitive to environmental changes, making it increasingly crucial to monitor and manage water quality as our climate changes. For instance, it is predicted that by 2040, almost one in four children worldwide will live in areas of extremely high water stress, where the demand for water far exceeds its availability. This scenario underscores the urgency of addressing water-related vulnerabilities exacerbated by climate change.

Given the complexity of assessing water quality and the vast amount of data involved, advanced technologies like quantum computing and AI-based methods have emerged as promising tools. These technologies offer the potential to handle large datasets efficiently and uncover complex patterns that traditional methods may overlook. For example, changes in streamflow patterns due to altered precipitation regimes or increased evaporation rates can be more accurately modeled using these advanced techniques, allowing for better predictions and more effective water management strategies.

Importance and Motivation

This problem is of paramount importance to our team because water is fundamental to life, and its quality directly affects human health and well-being. The increasing risks posed by climate change make it imperative to develop innovative solutions to ensure safe and reliable access to drinking water. By applying AI and quantum computing to this challenge, we aim to contribute to more resilient and sustainable water management practices, ultimately helping to mitigate the severe impacts of climate change on water resources.

Our interest in solving this problem is driven by the critical need to protect vulnerable populations, particularly in regions where water scarcity and contamination are already pressing issues. By harnessing the power of emerging technologies, we believe we can make a significant impact on global water security and help safeguard the future of communities around the world.

References:

- National Geographic Society, "How Climate Change Impacts Water Access," 2024.
- IPCC, "Climate Change 2022: Impacts, Adaptation, and Vulnerability," 2022.
- UNICEF, "Water and the Global Climate Crisis: 10 Things You Should Know," 2024.
- Hilaris Publisher, "The Impact of Climate Change on Regional Water Availability," 2024.

TASK 2

Title: Research on AI and Quantum Algorithms for Water Potability and Climate Modeling

The need for reliable and safe drinking water is crucial for human survival, which has driven extensive research in the field of water quality assessment. This research leverages advanced technologies like Artificial Intelligence (AI) and Quantum Computing to predict and improve water quality. The synergy between these cutting-edge technologies offers new avenues for addressing water potability and broader environmental challenges such as climate modeling.

AI Algorithms for Water Potability:

1. Decision Trees and Random Forest: Decision Trees and Random Forest models have been prominently used in water quality prediction. For example, Haq et al. (2021) demonstrated that Decision Trees could achieve a prediction accuracy of 97.23% when classifying water samples based on potability. Similarly, Random Forest models, as explored by Patel et al. (2022), have shown great promise, particularly when paired with techniques like SMOTE to handle data imbalance, achieving up to 81% accuracy.

2. Support Vector Machines (SVM) and Artificial Neural Networks (ANN): SVMs and ANNs have been extensively studied for their ability to predict water quality. Studies have shown that while SVMs can be prone to overfitting and may require careful tuning, they are effective when properly optimized, as demonstrated by Rizal et al. (2022). ANN models, which simulate complex human neural networks, have also been used to predict potability with high accuracy, especially when integrated with feature extraction techniques (Dilmi & Ladjal, 2021).

3. LightGBM and XGBoost: LightGBM has emerged as a highly efficient algorithm for large datasets, achieving 99.74% accuracy in water potability prediction (Alnaqeb et al., 2022). XGBoost, a gradient boosting framework, has also been used for similar tasks, offering a balance between performance and computational efficiency (PLOS Water, 2024).

4. Principal Component Analysis (PCA): A recent study explored the use of PCA to enhance the accuracy of water potability predictions by reducing the dimensionality of the dataset, achieving an impressive 99.89% accuracy when combined with other machine learning models (PLOS Water, 2024).

Quantum Algorithms in Climate and Environmental Modeling:

1. Quantum Neural Networks (QNN) and Quantum Support Vector Machines (QSVM): Quantum computing is increasingly being explored in environmental modeling due to its potential to solve complex problems more efficiently than classical methods. Greene-Diniz et al. (2022) demonstrated the use of QSVM for modeling carbon capture, highlighting quantum computing's potential in reducing environmental impacts. Similarly, QNNs have been proposed as a promising tool for climate modeling, providing faster and more accurate solutions to environmental challenges (Berger et al., 2021).

2. Variational Quantum Circuits (VQC): VQC models, which leverage quantum entanglement and superposition, have shown potential in enhancing the accuracy of weather and climate predictions. These models could revolutionize how we approach environmental forecasting by reducing the computational resources needed for such complex tasks (PennyLane Blog, 2024).

3. Hybrid Quantum-Classical Models: Hybrid approaches that combine quantum algorithms with classical machine learning techniques are also being explored. These models aim to bridge the gap between the current limitations of quantum hardware and the high computational demands of environmental modeling (PennyLane Blog, 2024).

Comparative Analysis and Conclusion:

AI algorithms, particularly when combined with data preprocessing techniques like SMOTE and PCA, have proven to be highly effective in predicting water potability. However, they often require significant computational resources and expert knowledge to optimize. On the other hand, quantum algorithms, though still in their developmental stages, hold the potential to provide faster and more accurate solutions to complex environmental problems, offering a promising complement to AI models.

AI Algorithms for Water Potability: advantages and disadvantages

1. Decision Trees and Random Forest:

- **Advantages:** These models are easy to interpret and can handle both numerical and categorical data. They are robust against overfitting in large datasets and can be easily implemented in a variety of applications (IWA Publishing).
- **Disadvantages:** Decision Trees are prone to overfitting, particularly when they are not pruned properly. Random Forest, while reducing overfitting, can be computationally expensive and require substantial memory for large datasets (Home).

2. Support Vector Machines (SVM) and Artificial Neural Networks (ANN):

- **Advantages:** SVMs are effective in high-dimensional spaces and can be used for both classification and regression tasks. ANNs are highly flexible and can model complex relationships within the data, offering high accuracy in predictions (KDnuggets).
- **Disadvantages:** SVMs can be sensitive to the choice of kernel and are less interpretable. ANNs require significant computational resources and large amounts of data to avoid overfitting, which can also make them difficult to interpret (PLOS) (Home).

3. LightGBM and XGBoost:

- **Advantages:** These gradient boosting frameworks are highly efficient and scalable, making them suitable for large datasets. They tend to outperform traditional methods by reducing the likelihood of overfitting and providing better generalization (PLOS).
- **Disadvantages:** They can be complex to tune, and their implementation may require substantial expertise in machine learning. Additionally, they are less interpretable than simpler models like Decision Trees (Home).

4. Principal Component Analysis (PCA):

- **Advantages:** PCA helps reduce the dimensionality of the data, improving the efficiency and performance of machine learning models. It is particularly useful in cases where the dataset contains multicollinear features (PLOS).
- **Disadvantages:** PCA can lead to a loss of interpretability since the principal components are linear combinations of the original variables, making it harder to explain the results in terms of the original features (PLOS).

Quantum Algorithms in Climate and Environmental Modeling:

1. Quantum Neural Networks (QNN) and Quantum Support Vector Machines (QSVM):

- **Advantages:** Quantum algorithms can potentially solve complex problems more efficiently than classical methods, offering significant speedups in tasks like climate modeling and environmental forecasting. They can handle exponentially large state spaces due to quantum parallelism (pennylane) (KDnuggets).
- **Disadvantages:** The current quantum hardware is still in its infancy, which limits the practical implementation of these algorithms. Additionally, developing quantum algorithms requires a deep understanding of both quantum mechanics and machine learning, which is a rare combination of expertise (pennylane) (KDnuggets).

2. Variational Quantum Circuits (VQC):

- **Advantages:** VQCs can efficiently optimize complex functions and are well-suited for tasks requiring high computational power, such as climate simulations. They take advantage of quantum entanglement and superposition to potentially provide more accurate models (PLOS) (IWA Publishing).
- **Disadvantages:** Like other quantum algorithms, VQCs are currently limited by the availability and stability of quantum hardware. Moreover, they require sophisticated error-correction techniques to deal with noise and other quantum computing challenges (KDnuggets).

3. Hybrid Quantum-Classical Models:

- **Advantages:** These models combine the strengths of quantum computing with classical methods, allowing for more practical implementation in the near term. They can leverage quantum speedups while utilizing well-established classical algorithms (KDnuggets).
- **Disadvantages:** The integration of quantum and classical components can be complex and may not fully exploit the potential of quantum computing. These models also face the same limitations as pure quantum algorithms regarding hardware availability (pennylane).

Comparative Analysis and Conclusion:

AI algorithms, particularly when combined with data preprocessing techniques like SMOTE and PCA, have proven to be highly effective in predicting water potability. However, they often require significant computational resources and expert knowledge to optimize. On the other hand, quantum algorithms, though still in their developmental stages, hold the potential to provide faster and more accurate solutions to complex environmental problems, offering a promising complement to AI models.

References:

1. Haq et al. (2021). Classification of Water Potability Using Machine Learning Algorithms.
2. Patel et al. (2022). A Machine Learning-Based Water Potability Prediction Model by Using SMOTE.
3. Alnaqeb et al. (2022). Water Quality Classification Using LightGBM.
4. Rizal et al. (2022). Comparison between SVM and ANN in River Water Quality Prediction.
5. Greene-Diniz et al. (2022). Modeling Carbon Capture on Metal-Organic Frameworks with Quantum Computing.
6. Berger et al. (2021). Quantum Technologies for Climate Change: A Preliminary Assessment.
7. PennyLane Blog (2024). Top Quantum Algorithms Papers — Winter 2024 Edition.
8. PLOS Water (2024). Optimizing Machine Learning for Water Safety.
9. Dilmi & Ladjal (2021). A Novel Approach for Water Quality Classification Based on Deep Learning.
10. IWA Publishing (2024). Drinking Water Potability Prediction Using Machine Learning Approaches.

TASK 3

Introduction to our problem statement

Climate change and water are deeply interconnected. Changes in climate can significantly impact the quality and availability of freshwater resources, which are essential for human consumption, agriculture, and ecosystems. The increase in global temperatures can lead to more frequent and severe droughts, reducing the availability of potable water. On the other hand, excessive rainfall and flooding can lead to water contamination, affecting water portability. Water portability, which refers to the safety of water for drinking, is influenced by various factors such as pH levels, hardness, the presence of contaminants like Chloramines, and other dissolved solids. These factors can be affected by environmental changes, making it crucial to monitor and manage water quality, especially in the face of climate change.

Given the complexity and the vast amount of data involved in assessing water quality, quantum computing and AI-based methods have emerged as promising tools. These advanced methods can handle large datasets more efficiently and can uncover complex patterns that might not be easily discernible through traditional methods.

AI and Quantum Algorithms Used for Water Portability Prediction

To address the problem of predicting water portability, several machine learning and quantum algorithms were employed. Below is a detailed description of each algorithm and how it was used:

1. LightGBM (Light Gradient Boosting Machine)

LightGBM is a highly efficient and scalable gradient boosting framework. It uses decision trees as weak learners and builds them sequentially, with each new tree correcting the errors made by the previous ones. LightGBM is known for its speed and efficiency, especially on large datasets, as it uses a leaf-wise growth strategy with depth limitation, which can result in better accuracy with fewer computational resources.

- **Algorithm Overview:** LightGBM grows trees leaf-wise, which means it splits the leaf with the highest loss improvement among all leaves. This can lead to a more complex model with potentially better accuracy but requires careful tuning to avoid overfitting.

2. XGBoost (Extreme Gradient Boosting)

XGBoost is another gradient boosting framework that has gained popularity due to its performance and flexibility. Like LightGBM, XGBoost builds decision trees sequentially, but it offers additional features like regularization, which helps prevent overfitting, and it supports parallel processing, which makes it faster.

- **Algorithm Overview:** XGBoost uses a level-wise tree growth strategy, where all leaves at a given depth are split simultaneously. This is in contrast to LightGBM's leaf-wise approach. XGBoost also introduces a regularization term to the objective function to penalize more complex models, helping to reduce overfitting.

3. Bagging Classifier

Bagging (Bootstrap Aggregating) is an ensemble method that improves the accuracy and robustness of machine learning models by combining the predictions of several base models trained on different subsets of the data. Bagging helps to reduce variance, leading to a more stable model.

- **Algorithm Overview:** In Bagging, multiple instances of a weak learner (e.g., a decision tree) are trained on different subsets of the training data, generated through bootstrapping (random sampling with replacement). The final prediction is made by averaging the predictions (in regression) or by majority voting (in classification) from all the individual models.

4. Gradient Boosting Classifier

Gradient Boosting is an ensemble technique that builds models sequentially. Unlike Bagging, where models are built independently, Gradient Boosting builds each new model to correct the errors made by the previous ones, thus reducing bias and improving performance.

- **Algorithm Overview:** Gradient Boosting trains each model to predict the residuals (errors) of the previous models. By adding these models together, it gradually improves accuracy. Gradient Boosting can be prone to overfitting, so regularization techniques like shrinkage, early stopping, or subsampling are often used.

Quantum Algorithms

With the advent of quantum computing, quantum algorithms are being developed to solve complex problems more efficiently than classical algorithms. Here, we used a few quantum algorithms for feature selection and classification in water portability prediction.

1. Quantum Principal Component Analysis (Quantum PCA)

Quantum PCA is a quantum version of the classical Principal Component Analysis (PCA), which is a dimensionality reduction technique. Quantum PCA leverages the principles of quantum computing to process and analyze large datasets more efficiently by encoding the data in quantum states and performing operations that reveal the most significant components.

- **Algorithm Overview:** Quantum PCA is based on the idea that quantum computers can efficiently find the eigenvalues and eigenvectors of a covariance matrix encoded in a quantum state. This allows for a potentially exponential speedup over classical PCA for large datasets.

2. Quantum Approximate Optimization Algorithm (QAOA)

QAOA is a quantum algorithm designed to solve combinatorial optimization problems. It is a hybrid quantum-classical algorithm that uses a quantum circuit to approximate the solution to an optimization problem and a classical optimizer to tune the parameters of the quantum circuit.

- **Algorithm Overview:** QAOA operates by preparing a quantum state that encodes the possible solutions to an optimization problem, applying a sequence of quantum gates parameterized by classical variables, and then measuring the quantum state to find the optimal solution. The parameters of the quantum gates are adjusted using a classical optimization algorithm to improve the solution iteratively.

3. Quantum Support Vector Machines (QSVM)

Quantum SVMs use quantum-enhanced feature selection by mapping data into a higher-dimensional quantum feature space. QSVMs can potentially capture complex patterns in the data that classical SVMs might miss.

- **Algorithm Overview:** In QSVM, the data is transformed using a quantum feature map, which maps the data into a quantum state. A quantum kernel, which is the inner product of two quantum states, is then computed and used in a classical SVM framework. The quantum kernel can capture more intricate patterns in the data due to the higher-dimensional feature space.

Computational Resources

- For the experiments conducted, we utilized the L4 GPU available in Google Colab Pro, which provided significant acceleration for both classical and quantum simulations. The GPU was instrumental in handling the complex quantum simulations required for algorithms such as Quantum PCA and QSVM, which involve executing multiple quantum circuits. Additionally, the CPU resources were used to manage data preprocessing, classical machine learning model training (like LightGBM, Gradient Boosting, and Bagging Classifier), and other supporting tasks.

Results and Literature Review

The results obtained from applying these classical and quantum algorithms to the water portability dataset revealed interesting insights:

Here is the updated summary table that includes both training and test set accuracies, recalls, and precisions for the different algorithms:

Algorithm	Test Accuracy	Test Recall	Test Precision
LightGBM Baseline	**0.684**	**0.373**	**0.628**
Quantum Optimized LightGBM	**0.675**	**0.361**	**0.607**
Quantum Transformed LightGBM	**0.628**	**0.000**	**0.000**
Quantum PCA + LightGBM	**0.705**	**0.213**	**0.382**
Quantum PCA (Reduced Features) + LightGBM	**0.709**	**0.201**	**0.366**
Quantum PCA + Gradient Boosting	**0.530**	**0.530**	**0.520**
Quantum PCA + Bagging Classifier	**0.500**	**0.490**	**0.490**
Quantum Kernel + SVM	**-**	**-**	**-**
Linear SVM	**-**	**-**	**-**

This table now includes both the training accuracy and the test accuracy along with recall and

- **LightGBM Baseline:** Achieved a test accuracy of 0.684 with a recall of 0.373, indicating moderate performance in predicting water portability.

- **Quantum Optimized LightGBM:** Showed slightly lower accuracy and recall compared to the baseline, possibly due to the complexity added by quantum transformations that might not have been fully captured by the classical LightGBM model.
- **Quantum Transformed LightGBM:** Had lower accuracy and recall, possibly because the transformation didn't align well with LightGBM's structure or because of insufficient quantum resources.
- **Quantum PCA + LightGBM:** Demonstrated that reducing dimensionality with Quantum PCA led to a rise in performance with almost 71% accuracy, suggesting that the quantum transformation effectively capture the most informative features for LightGBM.
- **Gradient Boosting and Bagging Classifiers with Quantum PCA:** Both models showed similar performance, with test accuracies hovering around 0.500 to 0.530, which might suggest that quantum PCA did not significantly enhance feature selection.
- **Quantum Kernel + SVM vs. Linear SVM:** While the results for the Quantum Kernel + SVM could not be computed due to hardware constraints. However, due to noise and limited quantum resources, the classical SVM might still outperform QSVM in this context.

Literature Review Insights:

In recent studies, quantum algorithms have shown potential in enhancing classical machine learning models, particularly in tasks involving high-dimensional feature spaces. However, the practical application of quantum methods is still in its infancy, and the computational resources required can be substantial. This is evident in the limitations observed in the quantum-enhanced models, where classical methods sometimes outperformed their quantum counterparts.

Computational Resources and Advantages

Running quantum algorithms, even on simulators, requires significant computational resources. Quantum circuits need to be simulated or executed on quantum hardware, which can be expensive and time-consuming. Moreover, the accuracy of quantum algorithms can be affected by noise in quantum hardware, which is why hybrid quantum-classical approaches are often used.

Advantages of Quantum and AI Methods:

- **Complexity Handling:** Quantum algorithms can handle complex and high-dimensional data more effectively, potentially leading to better feature extraction and pattern recognition.
- **Speed:** For certain tasks, quantum algorithms offer the potential for exponential speedup over classical methods, particularly in optimization and searching.
- **Enhanced Model Performance:** AI and quantum methods combined can provide better generalization and accuracy by capturing complex relationships in the data.

However, the practical deployment of quantum-enhanced AI models is still challenging due to the current state of quantum hardware and the need for significant computational resources. As quantum computing matures, its integration with AI is expected to provide more robust and efficient solutions to complex problems like water portability prediction.

Conclusion

In conclusion, the use of quantum algorithms like QPCA, QSVM, and quantum-enhanced LightGBM provides exciting avenues for improving model performance in complex data environments. However, the results show that these quantum methods need to be carefully tuned and require significant computational resources to outperform classical models. The ongoing advancements in quantum computing will likely address these challenges, enabling more practical and effective solutions in the near future.