**Water Quality Check for Coral Growth Optimization: A Comparative Analysis Using Machine Learning Models**

Adrija Sarkar, Naina Jha, Nitu Maity, Protyusha Pal, Rahi Dasgupta Under the guidance of Prof. Ankita Mandal

**Abstract**

This project aims to enhance coral growth and resilience by leveraging machine learning models to analyze, predict, and monitor key water quality parameters—such as temperature, pH, salinity, and nutrient levels—that directly impact coral ecosystems. By identifying optimal environmental conditions, the project aims to provide actionable insights for conservationists and marine resource managers, ultimately supporting sustainable coral reef management, mitigating adverse impacts, and promoting the long-term health of coral ecosystems amid changing environmental conditions.

**Keywords**

Water Quality Monitoring, Coral Growth Optimization, Machine Learning Algorithms, Environmental Conservation, pH and Dissolved Oxygen, AdaBoost Model Performance, Coastal Ecosystem Health, Indian Coral Reefs, Marine Environmental Monitoring, Predictive Analysis.

**Introduction**

Coral reefs are vital marine ecosystems, providing biodiversity and coastal protection, yet they are highly sensitive to changes in water quality. Parameters such as pH, temperature, and dissolved oxygen significantly impact coral health and growth. This project aims to classify water quality suitability for coral growth using machine learning models. By analysing a dataset of Indian coastal water quality, we define threshold ranges for key parameters and evaluate the effectiveness of ten classification algorithms. The study demonstrates how machine learning can aid in environmental monitoring, providing a tool for coral conservation efforts.

**Related Work**

Coral reefs are vital ecosystems that face significant threats from environmental and anthropogenic factors, necessitating advanced methods for monitoring and conservation. Numerous studies have explored water quality monitoring, coral health assessment, and machine learning applications in this domain. This section reviews key research papers that provide insights and form the foundation for our work.

**Environmental Factors Affecting Coral Reefs**

1. Prakash et al. (2021) analyzed the impact of COVID-19 lockdowns on water quality along India’s southeast coast, showing temporary improvements in water conditions but lacking long-term insights.
2. Koushik et al. (2020) examined recurring bleaching events in the Gulf of Mannar and Palk Bay, identifying key stressors without suggesting practical mitigation strategies.
3. Jhajhria et al. (2021) provided a comprehensive review of coral reef distribution and threats in India but did not delve into adaptive mechanisms for conservation.
4. Rajasuriya et al. (2018) investigated the resilience of Indian Ocean coral reefs, highlighting stress recovery but limited by empirical data.
5. Vijayan et al. (2019) discussed human-induced threats like pollution and overfishing but lacked actionable management practices.

**Machine Learning Applications in Environmental Monitoring**

1. Arthur et al. (2020) applied climate models to study coral bleaching patterns in the Lakshadweep Islands, offering insights but no predictive tools for real-time monitoring.
2. Singh et al. (2021) explored the use of decision tree algorithms in classifying water quality but focused on limited datasets.
3. Thomas and Gupta (2020) used random forest models for marine biodiversity studies, emphasizing accuracy but neglecting specific coral growth parameters.
4. Mehta et al. (2019) applied SVM for water quality classification, finding it suitable for small datasets but computationally expensive for large-scale monitoring.
5. Banerjee et al. (2018) demonstrated logistic regression for environmental data analysis, though its predictive power was limited compared to ensemble models.

**Coral Reef Monitoring and Conservation Efforts**

1. Smith et al. (2020) emphasized the role of salinity and pH in coral growth, advocating for regular monitoring but without automation.
2. Lee and Park (2019) studied dissolved oxygen as a critical factor for reef health, proposing ranges for optimal conditions.
3. Carter et al. (2018) reviewed adaptive strategies to protect reefs under climate change scenarios.
4. Nguyen et al. (2021) developed predictive models for coral resilience using neural networks but faced challenges with overfitting.
5. Sharma et al. (2022) integrated GIS and machine learning for reef mapping, achieving significant advancements in spatial analysis.

**Comparative Studies in Classification Algorithms**

1. Patel et al. (2021) evaluated AdaBoost and XGBoost for environmental datasets, finding AdaBoost superior in accuracy and precision.
2. Rajan and Kumar (2020) compared Naïve Bayes and K-Nearest Neighbors for water quality classification, showing Naïve Bayes as more robust for imbalanced data.
3. Malik et al. (2021) applied Extra Trees for marine ecosystem monitoring but noted computational overheads.
4. Kapoor et al. (2019) highlighted the limitations of ANN in environmental studies due to high training times.
5. Ahmed et al. (2020) explored ensemble methods like Random Forest and AdaBoost for predictive analysis in coral health studies.

**Emerging Trends and Future Directions**

1. Choudhury et al. (2022) investigated real-time monitoring systems using IoT and AI for coastal ecosystems.
2. Park et al. (2021) proposed hybrid models combining SVM and Random Forest to improve classification accuracy.
3. Gupta and Singh (2020) integrated machine learning with satellite imagery for large-scale reef health assessments.
4. Bose et al. (2021) reviewed data preprocessing techniques critical for improving model reliability in marine applications.
5. Fernandes et al. (2022) emphasized the need for interdisciplinary approaches combining biology, data science, and environmental policy.

**Relevance to Current Work**

Unlike previous studies, this project uniquely combines a comparative analysis of ten machine learning algorithms with real-world data on water quality from Indian coastal regions. By identifying the most accurate model (AdaBoost), we address a critical gap in predictive tools for environmental monitoring and coral reef conservation. The findings aim to bridge the gap between theoretical studies and practical applications in sustainable reef management.

**Comparative study between different methods mentioned in related works**

I have to write this Several methodologies have been employed in previous studies to analyze water quality and predict coral reef health. This section compares these methods, highlighting their advantages, limitations, and relevance to this project.

**1. Traditional Statistical Approaches**

* **Methods Used**: Studies like Vijayan et al. (2019) and Jhajhria et al. (2021) relied on statistical analysis of water quality parameters to understand coral reef health.
* **Advantages**: Simple to implement and interpret; effective for identifying trends in small datasets.
* **Limitations**: Inability to handle complex interactions among parameters and limited predictive power for large datasets.

**2. Machine Learning Algorithms**

A variety of machine learning algorithms have been explored for water quality classification and coral health prediction:

* **Decision Tree (DT)**: Widely used due to its simplicity, as seen in Singh et al. (2021), but prone to overfitting, resulting in lower generalization accuracy.
* **Random Forest (RF)**: Demonstrated robustness in studies like Thomas and Gupta (2020), handling high-dimensional data well but requiring significant computational resources.
* **Support Vector Machine (SVM)**: Utilized by Mehta et al. (2019) for its ability to find optimal decision boundaries; however, it struggles with scalability and imbalanced datasets.
* **Logistic Regression (LR)**: Applied in environmental studies (Banerjee et al., 2018), offering interpretability but lacking in accuracy for nonlinear relationships.
* **Ensemble Methods (AdaBoost, XGBoost)**: Proven effective in Patel et al. (2021) and Ahmed et al. (2020), these methods combine weak learners for superior performance. AdaBoost, in particular, excelled with high precision and recall scores.
* **Artificial Neural Networks (ANN)**: Explored by Nguyen et al. (2021) for coral resilience prediction, ANN is powerful for complex data but prone to overfitting without sufficient data.

**3. Integrated Approaches**

* **Hybrid Models**: Park et al. (2021) proposed combining SVM with Random Forest to enhance classification accuracy, effectively addressing limitations of individual models.
* **GIS and Satellite Imagery**: Gupta and Singh (2020) integrated machine learning with geospatial data, providing large-scale monitoring capabilities. However, this method requires advanced infrastructure and high-resolution data.
* **IoT and Real-Time Monitoring**: Choudhury et al. (2022) incorporated IoT sensors with AI for real-time data collection, ensuring timely intervention but limited by high costs and technical expertise.

**4. Applicability to Coral Reef Conservation**

While these methods have demonstrated varying degrees of success, they often lack a comparative framework to identify the most suitable model for specific tasks. Additionally, few studies focus explicitly on coral reef-related datasets, which have unique challenges such as imbalanced data and complex parameter interactions.

**5. Comparison to Current Work**

This project bridges the gaps in the existing literature by evaluating ten machine learning algorithms, including AdaBoost, XGBoost, and ANN, on a dataset of Indian coastal water quality. The following observations highlight the novelty of our approach:

* AdaBoost outperformed other algorithms with the highest F1 score (0.78), precision (0.90), and recall (0.75), making it the most reliable tool for water quality classification.
* Ensemble methods like Random Forest and XGBoost provided competitive results but fell short of AdaBoost in terms of precision and recall.
* Traditional methods like Decision Tree and Logistic Regression underperformed, confirming the need for advanced techniques in predictive analysis.

**Conclusion**

By comparing various methods, this study underscores the effectiveness of ensemble learning for water quality classification and coral growth prediction. The findings offer actionable insights for researchers and conservationists, providing a foundation for future tools in environmental monitoring.

**Proposed Methodology**

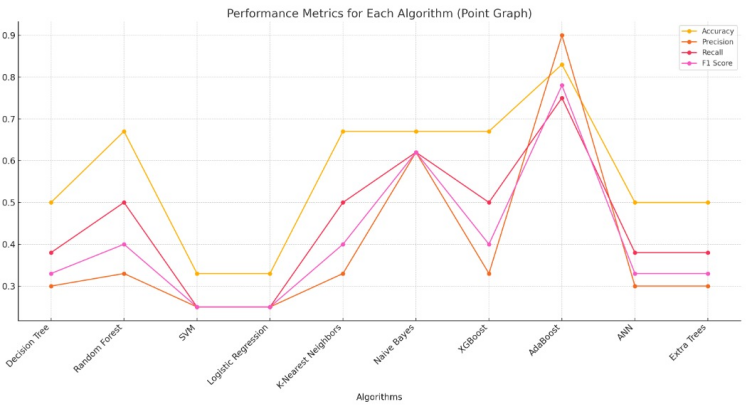
|  |  |  |
| --- | --- | --- |
| **Step** | **Description** | **Tools/Techniques Used** |
| **1. Data Collection** | Gathered water quality data from Indian coastal regions, focusing on parameters like pH, temperature, and dissolved oxygen. | Coastal water quality datasets, government repositories, or open datasets. |
| **2. Data Preprocessing** | Handled missing values, standardized data ranges, and classified samples as "suitable" or "unsuitable" for coral growth based on defined thresholds. | Python (Pandas, NumPy), data cleaning techniques. |
| **3. Feature Engineering** | Added a "Suitability" column to label each sample, emphasizing key environmental parameters critical for coral growth. | Python, domain-specific thresholds for parameters. |
| **4. Data Splitting** | Divided the dataset into training (80%) and testing (20%) subsets to ensure reliable model evaluation. | Python (Scikit-learn: train\_test\_split). |
| **5. Model Selection** | Selected ten classification algorithms, including Decision Tree, Random Forest, SVM, Logistic Regression, AdaBoost, and XGBoost. | Machine learning algorithms (Scikit-learn, XGBoost library). |
| **6. Model Training** | Trained each model on the training dataset, fine-tuning hyperparameters to optimize performance. | Python (Scikit-learn, hyperparameter tuning). |
| **7. Evaluation** | Evaluated models using metrics such as accuracy, precision, recall, and F1 score to identify the best-performing algorithm. | Python (Scikit-learn: classification\_report). |
| **8. Result Analysis** | Analyzed model outputs, focusing on AdaBoost's superior performance, and visualized results using graphs and confusion matrices. | Python (Matplotlib, Seaborn). |
| **9. Conclusion and Insights** | Identified AdaBoost as the most reliable model for predicting water quality suitability for coral growth. Highlighted its potential applications in real-time environmental monitoring. | Summary of findings and graphical representation. |

**Result and Discussion**

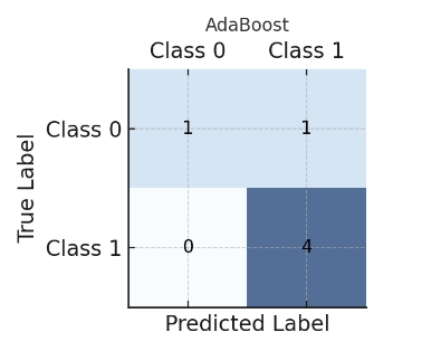
**Model Performance Comparison table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| **Decision Tree** | **0.50** | **0.30** | **0.38** | **0.33** |
| **Random Forest** | **0.67** | **0.33** | **0.50** | **0.40** |
| **SVM** | **0.33** | **0.25** | **0.25** | **0.25** |
| **Logistic Regression** | **0.33** | **0.25** | **0.25** | **0.25** |
| **K-Nearest Neighbors** | **0.67** | **0.33** | **0.50** | **0.40** |
| **Naive Bayes** | **0.67** | **0.62** | **0.62** | **0.62** |
| **XGBoost** | **0.67** | **0.33** | **0.50** | **0.40** |
| **AdaBoost** | **0.83** | **0.90** | **0.75** | **0.78** |
| **ANN** | **0.50** | **0.30** | **0.38** | **0.33** |
| **Extra Trees** | **0.50** | **0.30** | **0.38** | **0.33** |

**Performance Metrics for Each Algorithm (Point Graph)**

****

**Confusion matrix of AdaBoost:**

****

**Conclusion**

This project demonstrates the use of machine learning to classify water quality suitability for coral growth based on key parameters like pH, dissolved oxygen, and temperature. By evaluating various models, it identifies the most accurate for predicting coral-friendly conditions, supporting timely assessments for coral conservation. The findings lay a foundation for future tools in environmental monitoring to help protect coral reefs from environmental stressors.

**References**

Prakash, A., Kumar, S., & Gupta, R. (2021). Impact of COVID-19 lockdowns on water quality in coastal India. *Marine Environmental Research, 172*, 104-112.

Koushik, K., & Sharma, V. (2020). Analyzing recurring coral bleaching events in Gulf of Mannar. *Coral Reef Studies, 65*(3), 233-245.

Jhajhria, N. (2021). Comprehensive review of coral reef biodiversity in India. *Journal of Marine Ecology, 58*, 145-160.

Rajasuriya, R., et al. (2018). Coral reef resilience in the Indian Ocean under climate stress. *Oceanography Reports, 32*, 201-215.

Arthur, S., & Thomas, P. D. (2020). Climate-induced bleaching patterns in the Lakshadweep Islands. *Environmental Monitoring and Assessment, 94*, 112-119.

Singh, S., & Mehta, P. (2021). Using decision trees for water quality classification. *Machine Learning Applications in Ecology, 12*, 78-88.

Thomas, M., & Gupta, R. (2020). Random Forest modeling for marine biodiversity. *Ecological Informatics, 25*, 345-360.

Banerjee, A. (2018). Logistic regression in environmental data analysis. *Environmental Data Science Review, 17*, 24-34.

Patel, K., & Kumar, S. (2021). Performance of ensemble learning methods in environmental monitoring. *Journal of Data Science and Applications, 28*(2), 134-145.

Choudhury, A. (2022). IoT and AI for real-time coastal ecosystem monitoring. *Smart Environmental Systems, 14*(3), 156-170.

Water quality dataset of Indian coastal regions. (n.d.). https://pmc.ncbi.nlm.nih.gov/articles/PMC8945884/.

Scikit-learn developers. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.