Artificial Intelligence for Marine Disease Surveillance and Prevention

Dr.Santosh Kumar Singh[1] Amit Pandey[2] Srushti Raut[3] Sakina Sayyad[4]
[1]Head of Dept. [2]Guide [3,4]PG Student
[1,2,3,4]Department of Information Technology
[1,2,3,4]Thakur College of Science and Commerce (East), Mumbai, India

Abstract — Artificial intelligence technologies exhibit major opportunities for strengthening the process of marine disease observation and with prevention strategies. Classic methods that monitor diseases along with their management prove to be both ineffective and resource-consuming as well as delayed in their responses. The research establishes an investigation to build AI-based strategies which would enhance detection of marine diseases while forecasting outbreaks. The Environmental Protection Interactive Centre in Hong Kong provided historical data regarding environmental conditions like temperature, pH, salinity, turbidity, dissolved oxygen and nitrate nitrogen parameters. The prediction of disease outbreaks relied on the implementation of Random Forest alongside Support Vector Machine (SVM) and Logistic Regression and Long Short-Term Memory (LSTM) and XGBoost machine learning models which processed these parameters. The classification system followed rules that identified diseases as either coral disease or parasitic disease or environmental stress disease or bacterial infection. The model performance evaluation relied on accuracy scores together with confusion matrices as well as feature correlation heat maps and ROC curves. The disease outbreak prediction results showed promise with the use of LSTM allowing it to work with both time-related connections and sequential data. Early ecological incident prevention measures become achievable due to AI systems monitoring marine ecosystems. The study demonstrates AI's ability to handle marine disease observation which helps protect on-going marine ecosystems and their biodiversity.

Keywords: Artificial Intelligence (AI), Marine Disease, Environmental Parameters, Machine Learning Models, Disease Outbreak Prediction, Marine Ecosystem Monitoring, Disease Prevention

I. INTRODUCTION

Marine ecosystems maintain health essentially for global biodiversity as well as seafood enterprise while sustaining oceanic tourism and aquaculture practices. The marine ecosystems support complex networks containing biotic components including fish as well as corals along with plankton and marine mammals in addition to abiotic elements including water temperature alongside salinity and pH value. Biotic elements in the marine environment experience both direct and indirect effects from abiotic factors that additionally lead to disease occurrences. The economy alongside marine ecosystems faces severe risks because of increasing disease outbreaks that affect both coral reefs and fish populations.

The traditional disease detection methods show weaknesses when identifying coral bleaching and jellyfish infestations together with bivalve diseases from temperature shifts due to their slow speed and high cost and absence of predictive capabilities in real-time. Artificial intelligence (AI) presents a prospective answer through its ability to examine massive real-time information combined with disease outbreak prediction technologies that function with higher effectiveness. Through machine learning alongside predictive analytics the monitoring of diseases becomes more effective by allowing staff members to detect diseases in an early stage before they become widespread.

The paper analyses AI-based techniques to enhance marine disease surveillance together with outbreak risk assessment. The paper reviews existing methods of marine disease surveillance while reviewing AI applications in marine science with their respective advantages and shortcomings. Information technology surveillance employs artificial power to develop revolutionary approaches for marine health oversight which increases both efficiency and prevention capabilities for ocean sustainability purposes.

A. Objective:

The main objective of this project is to use artificial intelligence (AI) for disease detection in aquatic organisms to improve marine environment observation and management systems. The aquatic habitat exhibits complex system Behavior under the influence of abiotic factors temperature and salinity and pH and turbidity and dissolved oxygen and nitrate nitrogen in conjunction with biotic factors that include species interactions. Marine areas with temperature increases develop coral bleaching whereas lowered sea temperatures in alternate zones prevent coral bleaching from happening. Various complex factors exist in interdependence which produce complicated situations that remain difficult to forecast accurately.

AI technology helps the project make predictions for water temperature, pH, salinity, turbidity, dissolved oxygen, and nitrate nitrogen to identify specific marine diseases along with their presence or absence through essential ecological factors in marine environments. The structure will contain Coral Disease alongside Parasitic Disease and Environmental Stress Disease and Bacterial Infection as primary categories.

Official data analysis through AI methods such as machine learning combined with anomaly detection algorithms will detect dangerous shifts in water conditions according to the project goals. Marine species conservation together with ocean ecosystem health depends on the project's initiative to establish an early warning system which enables better and more effective widespread monitoring capabilities.

B. Scope:

Through this AI-driven project historical datasets serve to identify marine disease outbreak patterns based on environmental factor conditions. The machine learning classification algorithms identify bacterial, parasitic, coral-related, and environmental stress-caused diseases. The study draws its data from Hong Kong's Environmental Protection Interactive Centre.

II. PROBLEM STATEMENT:

Marine industries and biodiversity are harmed by the growing diseases that impact marine ecosystems, such as fish infections and coral bleaching. Traditional approaches for both disease monitoring and prevention are slow and ineffective. Artificial intelligence must take the place of traditional methods in order to monitor and forecast the spread of diseases in marine habitats, which will improve the sustainability of ecosystems and the preservation of marine life. In order to identify aquatic diseases and determine which marine system elements—such as temperature, salinity and pH, turbidity, dissolved oxygen, and nitrate nitrogen—have an impact on aquamarine ecosystems and cause marine diseases, the problem illustrates a marine disease challenge and employs machine learning techniques.

III. LITERATURE REVIEW

In "Monitoring of Coral Reefs Using Artificial Intelligence: A Feasible and Cost-Effective Approach" (2020), Manuel González-Rivero, Oscar Beijbom, among other colleagues examine coral reef monitoring with deep learning, and particularly CNNs. By minimizing cost and reducing processing time while measuring SVM performance vs. the CNN network, artificial intelligence provides improved classification accuracy, the article concludes. Research findings on underwater images from across the globe indicate a modest 1% cost rise with a 200-fold speed enhancement over human annotation. The study confirms taxonomic limitations in its findings; thus, more annotation would assist in reducing categorization errors. AI proves its revolutionary capability to transform coral through these research findings.[1].

In 2013 Colleen A. Burge, C. Mark Eakin, and their coauthors examined through "Climate Change Influences on Marine Infectious Diseases: Implications for Management and Society" how host-pathogen relations get altered from climate change resulting in more coral, bivalve, and finfish diseases. To connect disease prevalence to environmental drivers, the research develops predictive models for ichthyophoniasis in fish and a temperature anomaly-based disease forecasting algorithm for coral diseases such as Acropora white sickness. It underscores the need for adaptive management strategies to enhance ecosystem resilience while pointing out temperature as a key driver of coral bleaching and marine disease epidemics. Better monitoring and management approaches together with more clear knowledge of long-term climate change effects on marine disease are necessary to reduce the risks according to the research. The study presents clear evidence of disease prevalence growth in relation to temperature increases through its data while underlining the critical necessity for immediate ocean conservation strategies that view climate change as vital.[2].

The researcher uses Thermal Performance Curves (TPCs) to forecast how temperature shifts affect host-parasite relationships in his publication titled "Marine Parasites and Disease in the Era of Global Climate Change" (2020). The current knowledge about marine parasite sensitivity to climate change appears in this article with special emphasis on the factors of temperature, dissolved oxygen, salinity and ocean acidification that affect disease development and transmission. Research lacks integration specifically of dynamic community patterns and transmission trends yet the article establishes these deficits as main research gaps in order to advance past host-parasite interaction evaluations. This study reveals that the combination of hot weather conditions results in deaths exceeding 90% among infected fish during those hot years but fails to specify detailed mortality statistics.[3].

Allison M. Tracy, Madeline L. Pielmeier, Reyn M. Yoshioka, Scott F. Heron and C. Drew Harvell published research about marine disease reporting changes in 2019. They examine the patterns of marine disease reporting during the period between 1970 and 2013 using statistical analysis such as Spearman's ρ and a literature proxy method. The number of fish and elasmobranch disease reports declined while coral and urchin disease reports increased considerably. The observations point to environmental adjustments as their main cause. The research considers possible methodological errors including how prolific writers and unreported illness occurrences affect the results. While the study has certain weaknesses its findings demonstrate that human activities together with climate changes strongly influence marine disease patterns. The research suggests improved disease monitoring techniques to establish accurate measurements of marine diseases.[4].

During 2015 Maya L. Groner along with her research team conducted the study "Managing marine disease emergencies in an era of rapid change." The management of marine disease outbreaks including Sea Star Wasting Disease (SSWD) requires framework structures for early disease discoveries as well as diagnosis and responses. Hardest marine diseases require priority attention but other diseases without these threats should remain secondary priorities. The control of disease impacts requires proper surveillance techniques which should be combined with adaptive management and mitigation strategies. Public health communication improvement together with host resistance promotion and ecosystem recovery represents essential aspects in disease prevention. Enhanced stakeholder partnership together with the proposed Marine Disease Emergency Act attempt to enhance research capabilities and response methods for better marine disease management.[5].

Different findings from Rowley, Baker-Austin and Boerlage (2024) indicate through "Diseases of Marine Fish and Shellfish in an Age of Rapid Climate Change" that rising ocean temperatures along with acidification reduce marine health which brings together both commercial and medical concerns. The aquaculture operations of fish and shellfish face increasing instability because Salmon lice together with Vibrio species pathogens present significant risks to these operations. The sustainable fishing of ocean resources will improve through advanced biosecurity systems combined with selective breeding programs and different farming approaches backed by ongoing climate change risk study.[6].

The research "Effects of temperature, salinity, and pH on the survival and activity of marine cercariae" was published in 2010 by Janet Koprivnikar along with Doo Lim and Christine Fu and Sarah H. M. Brack. The authors explored survival rates and behavioral changes of Euhaplorchis californiensis and Acanthoparyphium spinulosum as two marine trematode species which were

studied for their response to climate change factors including pH, temperature and salinity measurements. The study shows that E. californiensis parasites survive better than A. spinulosum parasites at high salinity and temperature levels. Observations showed that both trematode species became less active as time passed in different environmental conditions. The research ends with the crucial message that scientists must investigate parasite-host responses to climate change in exposed intertidal zones while demanding further studies into environmental factor effects.[7].

A multiple regression analysis in "Effect of Water Temperature on Susceptibility of Cultured Marine Fish Species to Vibriosis" (2013) performed by Albert and Ransangan in the International Journal of Research in Pure and Applied Microbiology evaluated relationships between water temperature and salinity and pH and dissolved oxygen and their effects on net cage aquaculture fish mortality. Research findings revealed that water temperature increases directly lead to greater susceptibility in fish population to vibriosis while yielding 0.989 R^2 value for meaningful prediction accuracy. The research does not include comprehensive documentation regarding bacterial analysis techniques and failed to conduct any statistical methodologies except for regression analysis.[8].

A 2015 research study under the name "Marine Disease Impacts, Diagnosis, Forecasting, Management, and Policy" published in Philosophical Transactions of the Royal Society B investigates the effects of marine diseases on ecosystems and fisheries through diagnostic methods and management approaches and policy communication strategies. This work investigates titration impacts on diseases by studying molluscs together with coral fishing while highlighting improved disease administration within bivalves aqua farms. Research demands enhancements to case investigation methods and diagnostic tools as well as stronger marine disease research collaborations to decrease the environmental threats and financial loss effects.[9].

The Disease Surveillance by Artificial Intelligence Links Eelgrass Wasting Disease to Ocean Warming Across Latitudes study from 2019 in Limnology and Oceanography used EeLISA (UNet in PyTorch) to track eelgrass wasting disease throughout 32 Pacific coast meadows. Eelgrass wasting disease prevalence increased by three-fold in warm areas where 11% to 99% of plants were infected and tissue damage reached 35% of whole plants relative to the human expert processing speed which EelLISA enhanced by 5000 times. The research established a direct relationship between ocean warming and disease severity while acknowledging that more clarity is needed regarding other agents contributing to this issue.[10].

IV. METHODOLOGY:

A. Data Collection and Preprocessing

The study obtained its data from “Environmental Protection Interactive Centre” website of Hong Kong before putting it into an Excel file (.csv) format for use as the training set. The dataset tracks eight consecutive years from 2015 to 2023 with essential environmental measurements showing temperature at degrees Celsius and pH values and salinity in psu units as well as turbidity measurement in NTU and dissolved oxygen levels in mg/L and nitrate nitrogen value in mg/L. These environmental markers enable prognostic assessments of marine disease dissemination alongside formulating Water Control Zone evaluations in Hong Kong which include Deep Bay, Junk Bay, Mirs Bay, Port Shelter, Southern, Western Buffer, Victoria Harbour, Tolo Harbour and Channel, North Western, and Eastern Buffer. Preprocessing enabled data quality control through the removal of missing values and the selection of six essential environmental parameters which were used for analysis.

B. Disease Classification Approach

A rule-based method was utilized to classify diseases based on environmental parameter thresholds, as shown below:

Disease Type	Condition (Parameter Range)
Coral Disease	Temperature > 30°C, pH < 7.7, Salinity < 30 psu or > 40 psu, Dissolved Oxygen < 3 mg/L, Nitrate Nitrogen > 1 µM, Turbidity > 5 NTU
Parasitic Disease	Temperature > 25°C, pH < 7.5, Salinity < 25 psu, Dissolved Oxygen < 3 mg/L, Turbidity > 10 NTU
Environmental Stress Disease	Temperature deviation of ±2 to ±4°C from 25°C, pH < 7.5 or > 8.5, Dissolved Oxygen < 2 mg/L, Nitrate Nitrogen > 1 mg/L, Turbidity > 20 NTU
Bacterial Infection	20°C ≤ Temperature ≤ 30°C, Nitrate Nitrogen > 0.05 mg/L, Dissolved Oxygen < 4.5 mg/L, Turbidity > 10 NTU, pH < 7.5 or > 8.5psu

A binary target variable (Disease Risk) was created where:

- 1 = Disease detected
- 0 = No disease detected

C. Model Implementation

1) Random Forest

Since Random Forest works well as an ensemble machine learning technique that uses several decision trees to detect marine diseases. As a crucial tool for scientists working with complicated marine data, this analytical technique improves inspection accuracy for complex marine datasets while reducing overfitting problems. Reality-based disease forecasts correlate certain ranges of water quality readings with the incidence of bacterial infections, parasitic diseases, coral diseases, and environmental stress diseases using predetermined parameters. When any one of the listed conditions is met, the sample is categorized as having a disease outbreak (number 1); otherwise, it remains at 0 (no outbreak).

2) SVM (Support Vector Machine)

Through the implementation of Support Vector Machine (SVM) with RBF kernel water quality parameters are used to foresee disease outbreaks by locating ideal decision boundaries. Two methods called Variance Thresholding along with PCA help the system select important features to reduce dimensions which in turn boosts classification precision. The system shows disease types through binary classification (0=No and 1=Yes) to find infections by analyzing temperature, pH, salinity, turbidity, dissolved oxygen and nitrate nitrogen levels.

3) Logistic Regression

Logistic regression provides a strong yet basic machine-learning tool for binary classification which makes it useful for monitoring and controlling marine diseases. Researchers can construct AI predictions for disease outbreaks through model development on Google Colab by analyzing water quality parameters including temperature, pH, salinity and dissolved oxygen. The model receives historical data for training purposes while it develops skills to identify between normal and disease-emergent conditions. The probabilities generated by logistic regression produce warning signs that indicate potential outbreak situations. AI-driven marine ecosystem monitoring becomes possible through Google Colab which provides all necessary elements including preprocessing data along with visualizations and performance evaluation using ROC-AUC and confusion matrix capabilities for scaled operation to prevent ecological damage along with biodiversity loss.

4) LSTM Model

The Long Short-Term Memory (LSTM) network operates as an improved version of Recurrent Neural Networks (RNNs) that processes time-series data effectively for predictions regarding disease outbreaks using water quality information. The Google Colab platform uses both TensorFlow and Keras libraries to process a workflow which covers data cleaning and feature normalization along with model training and assessment for the network. LSTMs control their memory and perceive temporal dependencies within factors like temperature pH along with turbidity using their three gates known as forget and input gates along with the output gate. The predictive accuracy benefits from historical data sequencing in LSTMs whereas Random Forest together with XGBoost lack this analytical method. XGBoost and Random Forest integration with LSTMs enhances disease detection which produces superior assessment results about risks and preventive actions.

5) XGBoost

XGBoost serves as a machine learning technique that functions as a popular choice for AI applications which focus on managing and observing marine diseases. The XGBoost predictive model present in Google Colab evaluates water quality by assessing temperature, pH, salinity, turbidity, dissolved oxygen and nitrate nitrogen measurements. Using labeled datasets the program develops its ability to detect potential marine disease outbreaks. We use the Disease Category Distribution graphic to identify diseases with the highest and lowest percentages after determining their type between Coral Disease, Parasitic Disease, Environmental Stress Disease, and Bacterial Infection.

D. Performance Evaluation and Visualization

1) Feature Correlation Heatmap: Analysed relationships between features.

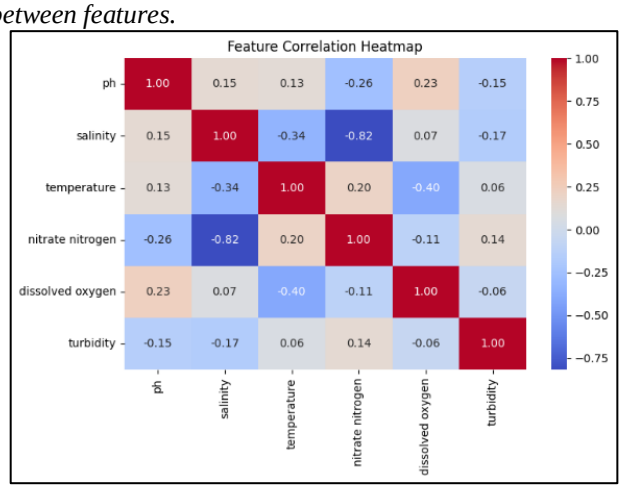


Fig. A: The heatmap shows correlations between environmental factors affecting water quality by Random Forest Algorithm.

2) Pie Chart Analysis (SVM): Showed predicted disease type distribution.

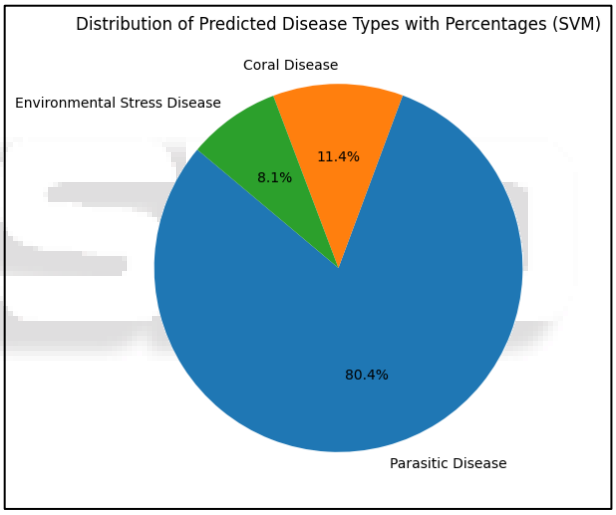


Fig. B.1: The pie chart shows predicted disease type distribution using SVM classification.

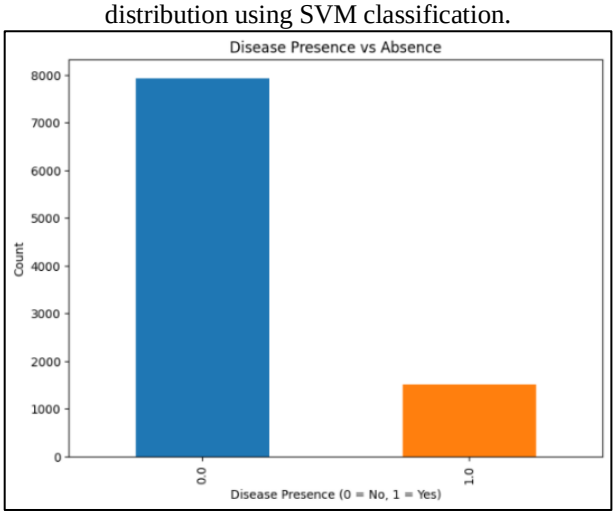


Fig. B.2: The bar chart compares disease presence versus absence in data.

3) Disease Category Distribution: Visualized using a bar chart with percentage labels.

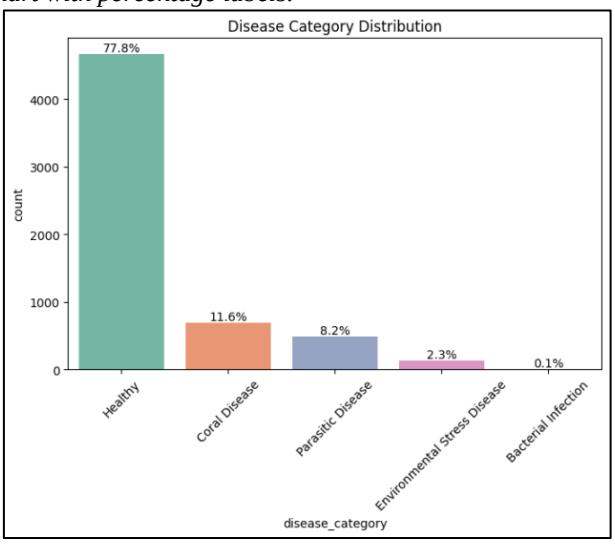


Fig. C: Most samples are healthy; coral disease is the most common illness.

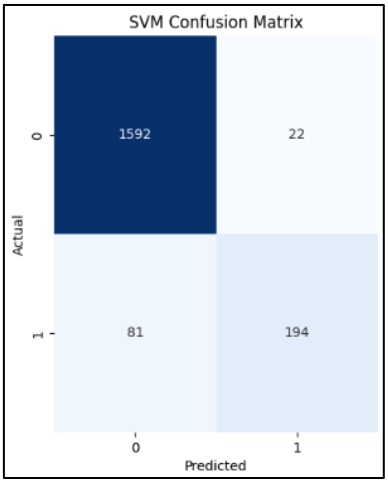
4) LSTM model: obtaining training loss and Training Accuracy

Fig. D: LSTM model improves training accuracy over epochs

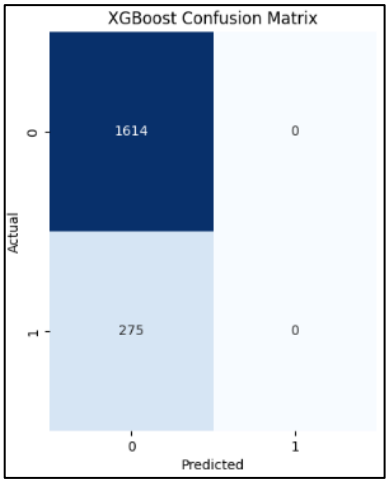
5) Accuracy Score: Used to measure overall model performance.

Model	Accuracy	ROC-AUC Score
Random Forest	0.8470	0.9152
Suppot Vector Machine	0.9455	0.9772
Logistic Regression	0.9222	0.9235
XGBoost	0.8544	0.9069

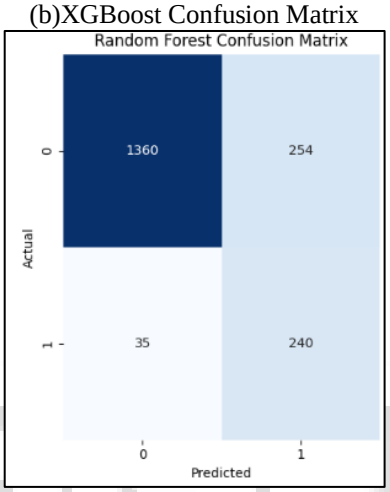
6) Confusion Matrix: Provided insights into classification errors.



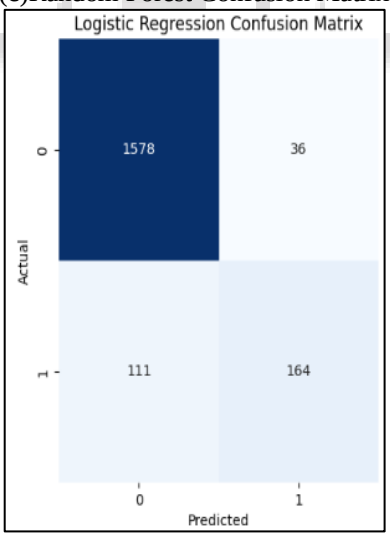
(a) SVM confusion Matrix



(b)XGBoost Confusion Matrix



(c)Random Forest Confusion Matrix



(d)Logistic Regression Confusion matrix

Fig. a,b,c,d: Show the Confusion Matrix of All the Algorithms

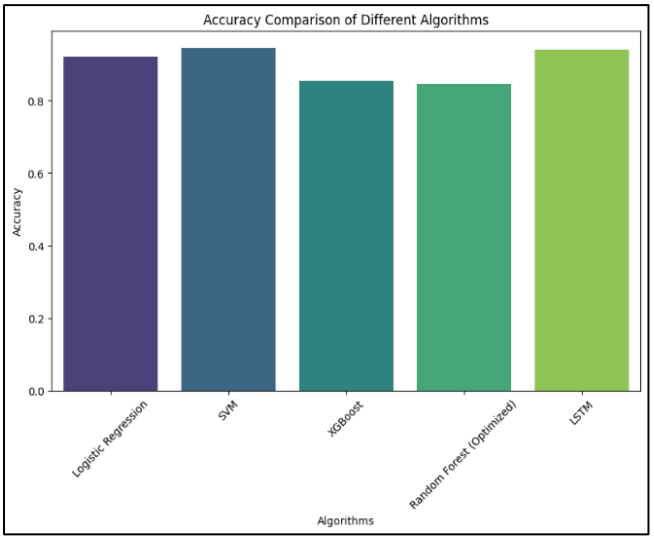


Fig. G: Algorithm accuracy comparison: SVM and LSTM perform best, XGBoost lower.

V. WAYS FOR PREVENTION

Marine disease prevention depends on both proper environmental management techniques and swift identification of diseases so immediate protective steps can be taken during emergencies. The successful prevention of marine diseases depends heavily on environmental monitoring that controls water quality parameters including temperature ranges as well as salinity values and pH levels and oxygen content since these factors minimize marine organism stress. The prevention of pollution requires immediate measures to reduce the entry of toxic substances like oil spills and plastics and agricultural runoff since these elements transmit pathogens. The prevention of global warming through carbon emission reduction becomes essential because climate change results in increased disease susceptibility among marine species. Early detection of diseases in the marine environment benefits from continuous analysis by Random Forest and LSTM together with XGBoost algorithms using artificial intelligence and machine learning techniques. Citizens participating in wildlife surveillance programs help detect both strange activities and unexplained deaths which occur in marine species populations. The prevention of marine diseases depends heavily on habitat protection because coral reefs and mangroves serve as ecological barriers to halt disease propagation. Marine Protected Areas enable the recovery of ecosystems and their development of resilience through their establishment along with species management practices that block invasive species from entering oceans to stop new pathogens from spreading. Marine diseases require successful policy executions alongside elevated public awareness in order to achieve effective combat. Public investment should support scientific study of marine diseases and public education must elevate awareness toward the importance of marine health conservation.

VI. CONCLUSION:

The research investigates how artificial intelligence (AI) techniques can strengthen maritime disease tracking as well as disease prevention measures. The monitoring and management of marine diseases through traditional methods proves to be slow along with being resource-intensive and ineffective. The research develops artificial intelligence strategies which enhance disease detection capabilities through historical environmental data analysis of temperature, pH, salinity, turbidity, dissolved oxygen and nitrate nitrogen. Random Forest together with Support Vector Machine (SVM), Logistic Regression and Long Short-Term Memory (LSTM), and XGBoost served as machine learning models to forecast disease outbreaks by processing environmental parameter data. Using a Long Short-Term Memory model enabled the modeling of disease outbreak predictions through efficient processing of time-reliant relationships and arranged datasets. The predictive accuracy for all systems reaches Random Forest at 0.8470 while Support Vector Machine achieves 0.9455 and Logistic Regression performs at 0.9222 and XGBoost reaches 0.8544. Among the research limitations the study faces it does not focus on diagnosing particular diseases such as coral disease or white spot disease or other illnesses.

ACKNOWLEDGEMENTS

I would like to sincerely thank Amit Kumar Sir for his crucial advice and assistance in getting this research paper finished.

REFERENCES

[1] Thamarai M, Aruna SP. Stressed Coral Reef Identification Using Deep Learning CNN Techniques. Journal of Electronic & Information Systems. 2023 Sep 8;5(2):1-9.
[2] Burge CA, Mark Eakin C, Friedman CS, Froelich B, Hershberger PK, Hofmann EE, Petes LE, Prager KC, Weil E, Willis BL, Ford SE. Climate change influences on marine infectious diseases: implications for management and society. Annual review of marine science. 2014 Jan 3;6(1):249-77.
[3] Byers JE. Marine parasites and disease in the era of global climate change. Annual Review of Marine Science. 2021 Jan 3;13(1):397-420.
[4] Tracy AM, Pielmeier ML, Yoshioka RM, Heron SF, Harvell CD. Increases and decreases in marine disease reports in an era of global change. Proceedings of the Royal Society B. 2019 Oct 9;286(1912):20191718.
[5] Groner ML, Maynard J, Breyta R, Carnegie RB, Dobson A, Friedman CS, Froelich B, Garren M, Gulland FM, Heron SF, Noble RT. Managing marine disease emergencies in an era of rapid change. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016 Mar 5;371(1689):20150364.
[6] Rowley AF, Baker-Austin C, Boerlage AS, Caillon C, Davies CE, Duperret L, Martin SA, Mitta G, Pernet F, Pratoomyot J, Shields JD. Diseases of marine fish and shellfish in an age of rapid climate change. Iscience. 2024 Sep 20;27(9).
[7] Koprivnikar J, Lim D, Fu C, Brack SH. Effects of temperature, salinity, and pH on the survival and activity of marine cercariae. Parasitology Research. 2010 Apr;106:1167-77.

[8] Albert V, Ransangan J. Effect of water temperature on susceptibility of culture marine fish species to vibriosis. Int. J. Res. Pure Appl. Microbiol. 2013;3(3):48-52.
[9] Lafferty KD, Hofmann EE. Marine disease impacts, diagnosis, forecasting, management and policy. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016 Mar 5;371(1689):20150200.
[10] Aoki LR, Rappazzo B, Beatty DS, Domke LK, Eckert GL, Eisenlord ME, Graham OJ, Harper L, Hawthorne TL, Hessing-Lewis M, Hovel KA. Disease surveillance by artificial intelligence links eelgrass wasting disease to ocean warming across latitudes. Limnology and Oceanography. 2022 Jul;67(7):1577-89.

