The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES

20(4): 1002-1008,2025 www.thebioscan.com

Machine Learning-Powered Diagnosis of Coral Reef Disease Factors Using PCA and FF-ANN Techniques

Dr. S. Kother Mohideen, Mr. V Praveen Kumar

1. Professor, Department of Computer Science, Acharya Bangalore B-School, Bengaluru.
2. Head, Department of Computer Science, Acharya Bangalore B-School, Bengaluru

DOI: 10.63001/tbs.2025.v20.i04.pp991-1001

KEYWORDS

Received on:

30-07-2025

Accepted on:

05-10-2025

Published on:

08-12-2025

Abstract

Coral reefs represent one of the most vital marine ecosystems, supporting a wide range of aquatic life. However, rising ocean temperatures and environmental fluctuations increasingly threaten their survival. These changes not only damage the corals themselves but also disrupt the numerous organisms that depend on them. To ensure effective conservation, it is crucial to identify and analyse the factors responsible for coral diseases. This study focuses on predicting the causal agents of coral reef diseases using a data-driven approach. The proposed model investigates two types of datasets: metagenomic sequence data associated with lesions of Coral Patch (CP) and Black Band Disease (BBD), and metatranscriptomic data capturing the biological activity within CP and BBD lesions. Both datasets contain varying occurrences of the two disease conditions, enabling a comprehensive examination of the underlying microbial contributors.Two computational techniques—Principal Component Analysis (PCA) and Feed-Forward Artificial Neural Networks (FF-ANN)—are employed to enhance prediction performance. PCA is used for dimensionality reduction, extracting the most informative features, while the FF-ANN utilizes a multilayer perceptron with backpropagation to classify the organisms responsible for coral infections and identify the most severe disease impacts. Experimental results demonstrate that the proposed PCA–FF-ANN framework outperforms conventional SVM and CNN models, offering a more effective solution for diagnosing coral reef disease causality.

INTRODUCTION

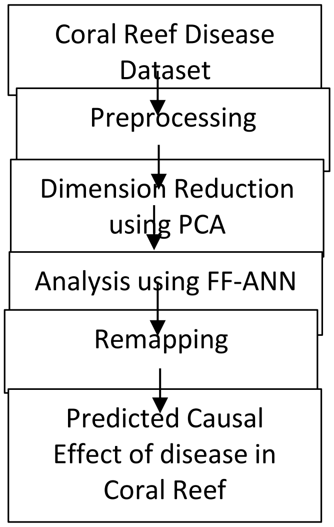
The Coral reefs are one of the most diverse ecosystem that are prevalent in the marine[1,2, 4]. They plays an important role in coastal production that are most valuable due to its different nature of sources[2, 5]. Due to the drastic change in the temperature of the sea the intensity and bleaching of coral reefs tends to increase tremendously. This cause the coral to lose its primary energy source such as endosymbiotic algae. The bleaching of coral will lead to mortality of reefs which cause the corals to lose its cover[4]. This may affect the fish communities of coral reefs. The different programs are been conducted for conservation and monitoring of coral reefs as it is most important biodiversity of marine body[7]. Monitoring is one of the important factor for absorbing the exact status of the organisms. Many monitoring phases are been conducted so as to determine the different organism's population that are prevalent in the reef area. This includes corals that are live or dead, algae etc. Paucity of data is the most valuable point for lack of conservation of natural habitats[2] . The data about the natural habitat should be considered as the major obstacles which supports us to understand the life style of the habitat. Usually the data will be collected using the remote sensing process. The warning signs should be detected before the structural loss happens. The corals are most prevalent in pacific oceans but there is a lack of data which may reduce the real time prediction, global analysis, adaptations and calibration models for predicting the future[3, 4]. Hence the preservative measures have to be taken for reducing the mortality of coral reefs[11]. In this proposed model the coral reef dataset is been analysed and the reason for the cause or disease is predicted. Here the dataset is been collected from mandeley repository and are analysed. The major theme of the research is to predict the most affecting organism that cause disease in the corals. Two different data types are been analysed such as the dataset based on the sequence in the metagenome the lesion that cause Cyanobacterial Patches (CP) and Black Band Disease (BBD) and the second dataset is based on the metatranscriptomes the lesion that cause cyanobacterial patches and Black Band Disease. Both the dataset includes different number of occurrences of the CP and BBD. Two different algorithms are used such as Principal Component Analysis (PCA) and Feed Forward Artificial Neural Network (FF-ANN). Here the PCA is used to reduce the dimension of the dataset and the FF-ANN is used to predict the type of organism that cause disease in the corals as well as the most affecting disease in the corals. The FF-ANN follows Multi Layer perceptron model with back propogation strategy. The organization of the paper is as follows, the section 2 discuss about the related works of the coral reef disease analysis model, the section 3 discuss about the proposed model, section 4 discuss about the experimentation and results of the proposed model and section 5 concludes the paper.

2. Related work

Cote et al [7] have developed a meta analysis model for integrating change estimations of macroalgal and coral covers. They used the in situ surveys for generating the patterns of ecological changes that happens in a large scale. They have examined the biases that are observed in the meta analysis model usingthe Caribbean reefs dataset. This model has variety of sampling methods and also depicted that the changes that are estimated over the coral cover is nearly same as the survey that was conducted during 1991 in carribbean. They have developed a cost effective method for getting the estimates of coral status. Jin Chao Wang et al [9] a digital borehole camera technology is been used for analyzing the coral reefs. The rock mass integrity index is been used for analyzing the coral integrity. The size of the cavities of corals are been estimated using the coefficients of the RMDI. They have also used a correlation model for getting the correlation between the coral integrity and the RMDI values. Thus the proposed model support the coral integrity, classification based integrity evaluation. Elizabeth et al [8] have developed a model to predict the changes in the coral covers and preventive measure for coral reef preservations. They have compiled more than 8000 coral cover data for comparing the annual changes in the covers of the coral reefs. They have compared the unprotected and protected coral reefs for observing the differences. They have found that the older MPAs are good for preventing the loss of coral reefs. John et al [10] have processed indo pacific reefs that occurs between 1968 to 2004. They have estimated that only few of the total sample have more than 60 % cover and also predicted that between the period of 1997 to 2003 the coral reef's mortality rate increases by 2%. Thus they estimate the rate of coral reef loss that happens in the indo pacific area. They have used regression analysis for estimating the annual loss of the coral reefs. Joseph Pallock et al [1] have established molecular specific coral reef pathogen detection model. They have also outlined the technologies that are emerging for diagnosing the diseases in the coal reefs and also addressed the challenges related to these techniques with the derived coral reef samples. Ainsworth et al [6]have used different studies such as microbial and structural based studies for differentiating the existing various types of black band diseases. They have differentiated two types of black band diseases as typical and atypical black band disease. In previous studies the external signs of atypical and typical black band diseases are not well defined. In this study they have differentiated the structure of tissue by detecting the white disease by extracting the white syndrome. These white syndromes are bacteria free that are extensive cells with lesions. Whereas the white diseases are with bacterial habitats. Thus this work concentrates on extracting the microscopic signs of various diseases in coral reefs for diagnosing the exact disease over it.

3. Proposed Model:

The proposed model includes several steps as shown in fig. 1. The dataset of coral reef is given as input. Here two type of inputs are given as dataset based on sequence of metagenome and dataset based on sequence of metatranscriptomes. The input data is preprocessed for removing the empty values and replacing the missing values. Then the correlated coral data are processed with principle component analysis for removing the unrelated variables and for getting the significant variable, then this PCA processed data will be given as input for feed forward Artificial Neural network for predicting the causal effect of disease in coral reefs. The remapping is the section where the analysis is carries with different functions so as to improve the accuracy of the model. The proposed framework includes two different classification paradigms such as Principal component analysis and two layer feed forward Artificial Neural Network. The principal Component Analysis is type of multivariate analysis which is based on the eigenvector. The main purpose of the PCA is to reduce the number of variables that exists in the considered dataset or data matrix. It is widely used in statistics so as to reduce the dimensionality of data. The dataset considered for analysis will be in multidimensional and it is too hard to understand the structure of the data. This PCA converts the correlated variables into uncorrelated smaller variables which are commonly known as principle components.


Coral Reef Disease Dataset
Preprocessing
Dimension Reduction using PCA
Analysis using FF-ANN
Remapping
Predicted Causal Effect of disease in Coral Reef

Here the disease that are associated with the coral reefs are to be analyzed. Two different datasets are used such as one is based on the taxonomically annotated sequences in the metagenomes and another is based on the taxonomically annotated sequences in the metatranscriptomes. This dataset includes different variables such as domain, species, family, class, order, cyanobacterial patches and black band disease. The PCA is used to reduce the correlated variables from datasets of metagenome and metatranscriptomes of coral reef. This thereby makes the data to low dimensional space from high dimensional space. It also supports the framework to find the difference between the different samples that are taken to process the dataset and figure out the most significant variable that is dominant for the difference. This filtered and preprocessed data can be visualized as graphs. The first principle component of the data is domain and it is succeeded by number of occurrences of cyanobacterial patches and black band disease and its relative frequency. The two layer feed forward Artificial Neural Network is used for analyzing the preprocessed data. It is used to capture the valuable information from highly complex system that are non linear in nature. Usually it finds the relationship between the dependent and independent variables of raw data. The Multi Layer Perceptron with back propagation is used for analyzing the data. It includes different layers with interconnected neurons that have a feed forward structure. The dependent and independent variables are denoted as neurons that exists in different layers. The operation of MLP includes two different steps as training and testing. During the training operation the input parameters such as Domain, CP, BBD, and its relative frequency are trained. These parameters are provided with a weight and are fed to hidden layer. Then the learning is done by associating these inputs. The final operation is testing where the output of causal effect of disease in coral reef will be generated. The output will be the prediction of most significant disease causing organism in coral reefs.

4. Experimentation and Result:

The experimentation is carried out in tool and python is used as a language for processing the dataset. The data set is collected from mandeley repository. The dataset is of two types as taxonomical annotated sequences in the metagenomes and taxonomical annotated sequences in the metatranscriptomes. These two types of datasets are microbial lesions that cause cyanobacterial patches and black band disease. The metagenome based dataset includes 384 different samples of dieases and the metatranscriptome based dataset includes 387 different samples of diseases in coral reefs. The dataset includes different parameters as domain, species, family, class, order, cyanobacterial patches and black band disease and the relative frequency of the cyanobacterial patches and black band diseases. The 182 samples are taken for analyzing the causal effect. The dataset is been processed using the principal component analysis for acquiring the most significant variables that has higher impact for predicting the disease causal effect. It is then given as input to the feed forward artificial neural network. The hidden neurons are determined at the first phase. Here nearly 45% of the data is been used for testing set. The data set includes three main domains as Archaea, Bacteria and Eukaryote. The analysis is carried out based on these three domains that which organism cause one or both of the disease such as CP and BBD. For these 182 samples the number of occurrences of CP and BBD is calculated and are averaged likewise the relative frequency of both the CP and BBD are calculated and are averaged. Separate comparison is done to predict the most disease causing domain. The analysis is carried out with both type of dataset i.e. with the metagenome dataset and metatranscriptome dataset. Both the data are compared to get the most affecting organism on coral reefs.

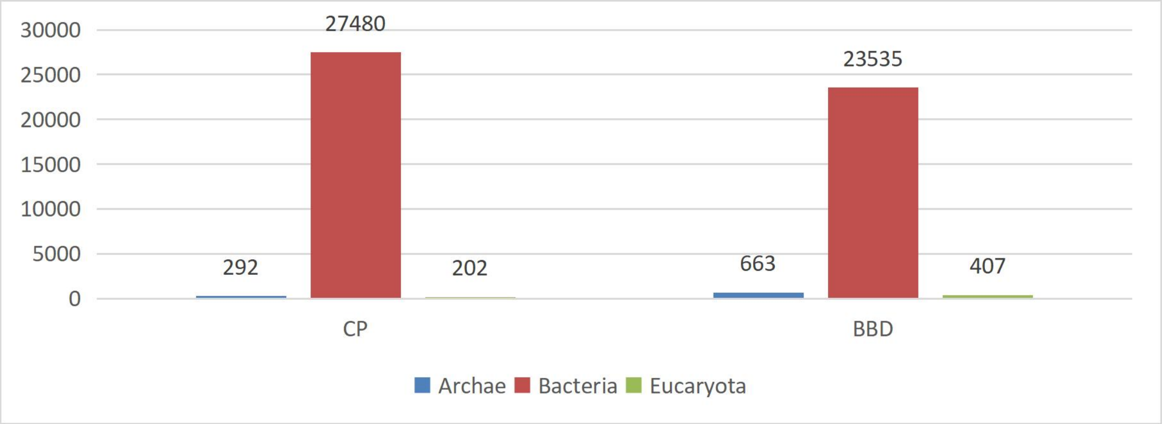


Fig 2. Number of Occurrences of CP and BBD in D1

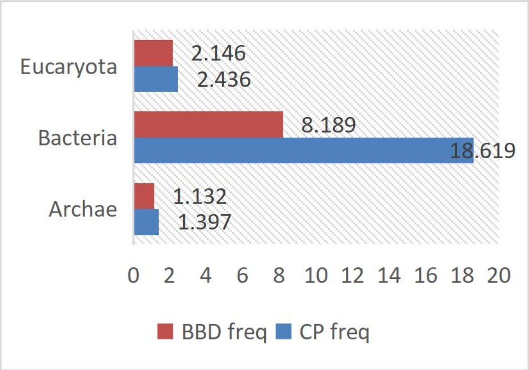


Fig 3. Frequency of CP and BBD in D1

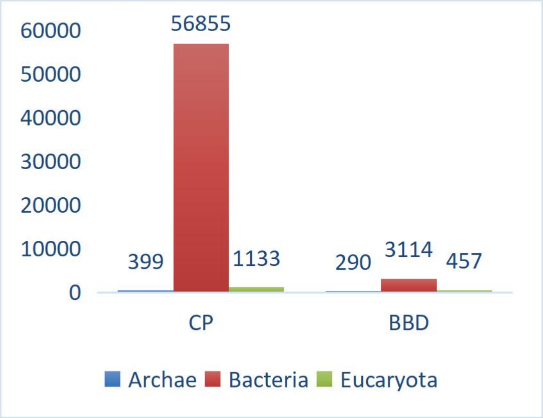


Fig 4. Number of Occurrences of CP and BBD in D2

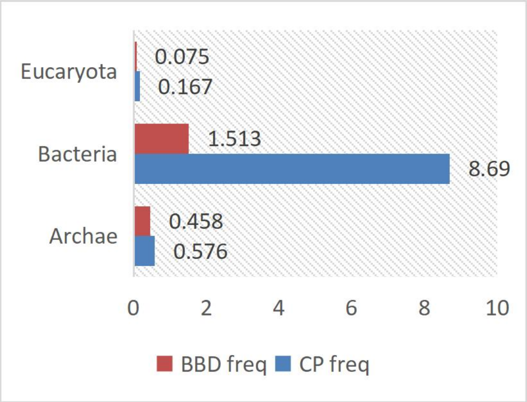


Fig 5. Frequency of CP and BBD in D2

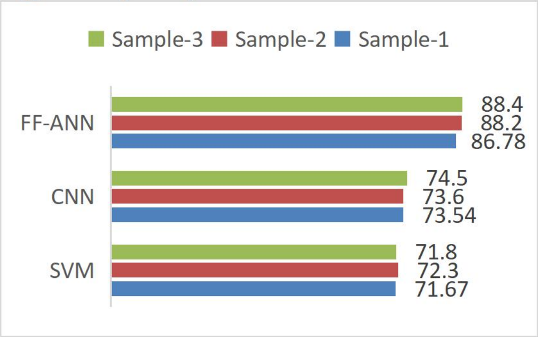


Fig 6. Comparison of Accuracy

The fig 2. Shows the number of occurrences of CP and BBD in D1 where D1 denotes the dataset that is based on the sequences in the metagenome. Out of 182 samples considered for analysis it is found that the CP seems to be the most defected disease in the coral reefs and BBD seems to be second largest causing disease of coral reefs. It is shown that both the disease are mostly caused due to the bacteria than other domains. Hence we can predict that bacteria is the most disease causing organism of coral reefs. The Fig 3. Shows the relative frequency of the CP and BBD in D1. This graph shows that the relative frequency of CP and BBD due to bacteria is 18.619 and 8.189 respectively which shows that the occurrence of CP is higher than the occurrence of the BBD and bacteria is the most causing effect of disease in corals. The fig 4. Shows the number of occurrences of CP and BBD in D2 where D2 denotes the dataset that is based on the sequences in the metatranscriptomes. We can observe that the occurrence of CP is higher than the BBD as in D1. It is shown that both the disease are mostly caused due to the bacteria than other domains. Hence we can predict that bacteria is the most disease causing organism of coral reefs. The Fig 5. Shows the relative frequency of the CP and BBD in D2. This graph shows that the relative frequency of CP and BBD due to bacteria is 8.69 and 1.513 respectively which also shows that the occurrence of CP is higher than the occurrence of the BBD and bacteria is the most causing effect of disease in corals. On anlysing both the datasets we can observe that the bacterias are the most disease causing organisms which cause CP in the coral reefs. We can also observe that CP is the most affecting disease in the corals. In fig 6. The accuracy of the proposed model with the existing algorithms is presented. It shows that the proposed model is too good when compared with the existing algorithm such as Support Vector Machine (SVM) and Convolution Neural Network.

5. Conclusion

The proposed model is analysed using the coral reef disease dataset and the reason for the cause or disease is predicted. This research work is carried out to predict the most affecting organism that cause disease in the corals. Two different data types are been analysed such as the dataset based on the sequence in the metagenome the lesion that cause CP and BBD and the second dataset is based on the metatranscriptomes the lesion that CP and BBD. Both the dataset includes different number of occurrences of the CP and BBD. Two different algorithms are used such as Principal Component Analysis (PCA) and Feed Forward Artificial Neural Network (FF-ANN). Here the PCA is used to reduce the dimension of the dataset and the FF-ANN follows Multi Layer perceptron model with back propogation strategy for predicting the type of organism that cause disease in the corals as well as the most affecting disease in the corals. The proposed model is more accurate than the other existing models. It is observed from the result that the cyanobacterial patches are the most prevalent disease found in corals and it is also observed that the bacterias are the organisms that mostly cause cyanobacterial patches in corals. Hence the bacteria that affects the corals have to be determined so as to reduce the mortality rate of corals.

References

1. F. Joseph Pollock, Pamela J. Morris, Bette L. Willis, David G. Bourne, “The Urgent Need for Robust Coral Disease Diagnostics”, Plos Pathogens, Vol. 7, No. 10, pp. 1-10, 2011.

2. Nurhalis B. Wahidina, Vincentius P. Siregarc, Bisman Nababanc, Indra Jayac, Sam Wouthuyzend, “Object-based image analysis for coral reef benthic habitat mapping with several classification algorithms”, Procedia Environmental Sciences, Vol. 24 (2015), pp. 222 – 227.

3. F. Javier González-Barrios, Lorenzo Álvarez-Filipa, “A framework for measuring coral species-specific contribution to reef functioning in the Caribbean”, Ecological Indicators, 95 (2018) 877–886.

4. Ivor D. Williams, Courtney S. Couch, Oscar Beijbom, Thomas A. Oliver, Bernardo Vargas-Angel, Brett D. Schumacher and Russell E. Brainard, “Leveraging Automated Image Analysis tools to transform our capacity to Access Status and Trends of Coral Reefs”,

Automated Analysis of Benthic Imagery, Vol. 6, pp. 1-14, 2019.

5. Daniel F. Martinez-Escobar, Jennie Mallela, Assessing the impacts of Phosphate mining on coral reef communities and reef development, Science of the Total Environment, 692, (2019), 1257-1266.

6. T. D. Ainsworth, E. Kramasky-Winter, Y. Loya, O. Hoegh-Guldberg and M. Fine, “Coral Disease Diagnostics: What’s between a Plague and a Band?”, Applied And Environmental Microbiology, Feb. 2007, p. 981–992, Vol. 73, No. 3.

7. I. M. Cote, Jennifer A. Gill, T. A. Gardner, Andrew Richard Watkinson, Measuring coral reef decline through meta-analyses, Philosophical Transactions of the Royal Society B, 360, 385-395, 2005.

8. Elizabeth R. Selig, John F. Bruno, “A Global Analysis of the Effectiveness of Marine Protected Areas in Preventing Coral Loss”, PloS ONE, Vol. 5, No. 2, 1-7, 2010.

9. Jin-chao Wang & Chuan-ying Wang (2017) Analysis and evaluation of coral reef integrity based on borehole camera technology, Marine Georesources & Geotechnology, 35:1, 26-33

10. Bruno JF, Selig ER (2007) Regional Decline of Coral Cover in the Indo-Pacific: Timing, Extent, and Subregional Comparisons. PLoS ONE 2(8), 1-8, 2007.

11. Shihavuddin,A.S.M.,Gracias,N.,Garcia,R., Gleason,A.,andGintert,B.(2013). Image-based coral reef classification and thematic mapping. Remote Sens. 5, 1809–1841.