

Original Article

Generative Data Modelling to Improve Marine Data Predictions

Agnes Nalini Vincent¹, K. Sakthidasan¹, Nassirah Laloo², Uhoze Bagurubumwe³

¹Department of Electronics and Communication Engineering, Hindustan Institute of Technology and Science, Chennai, India.

²Department of Creative Arts, Film and Media Technologies, School of Innovative Technologies and Engineering, University of Technology, Mauritius.

³Faculty of Information Technology, AMITY Institute of Higher Education, Pierrefonds, Mauritius.

¹Corresponding Author : vanalini@mauritius.amity.edu

Received: 18 November 2024

Revised: 25 September 2025

Accepted: 07 October 2025

Published: 31 October 2025

Abstract - The ability to project changes in benthic communities based on environmental parameters is vital for constructing resilience in marine ecosystems. This study employs a Deep Recurrent Neural Network (RNN) to predict hard corals and fish assemblages based on water quality parameters. Marine Data of Flic en Flac Lagoon, located in Mauritius, is used for this purpose. The use of Generative Adversarial Network (GAN) and its variants, including Wasserstein GAN, Conditional GAN, and Climatic GAN, to improve the prediction accuracy of Deep RNN is investigated. A state-of-the-art Marine Data GAN (MGAN) has been proposed and investigated. Empirical evidence proves that MGAN minimizes the Wasserstein distance Jensen-Shannon divergence that can exist between the generated and original data distribution, than any other GAN. In contrast, for the pH of water, the Kullback-Leibler (KL) divergence of MGAN is much higher than WGAN, highlighting WGAN's superior performance in capturing the pH distribution. Generated data from MGAN is then used as input to the Deep RNN to perform predictions. This hybrid MGAN Deep RNN model shows substantial improvements across evaluation metrics compared to the basic Deep RNN model, which uses the actual dataset. Specifically, MAE improved by 7.44, RMSE by 8.07, and R^2 from a negative to a positive value, demonstrating the enhanced predictive accuracy of the hybrid model. Thus, this research identifies MGAN-Deep RNN as the best model for the prediction of marine data under consideration. As an outcome, this research provides valuable insights into the administration of marine ecosystems in the Flic en Flac Region of Mauritius.

Keywords - Deep Recurrent Neural Network, Generative Adversarial Network, Marine data, Prediction, Mauritius.

1. Introduction

The Marine Ecosystem supports human health, provides livelihood, drives economic growth, serves as a habitat for a considerable number of marine species, promotes biodiversity, regulates the climate, and contributes to food security. The aquatic species biodiversity includes marine mammals, corals, fish assemblages, mollusks, crustaceans, algae, and microbes. However, the health of these benthic environments is associated with some unprecedented threats and essential challenges. Globally, coastal ecosystems have declined during the past two decades, resulting in a 35–85% reduction in the amount of these species [1]. The main threats and stressors to marine species biodiversity are multifaceted and include rising Sea Surface Temperatures (SST) [2], ocean acidification or practical salinity [3], coral bleaching [4], pH of ocean water [5], dissolved oxygen content termed as Chemical Oxygen Demand (COD) [6], phosphate and nitrogen concentration [7], pollution [8], agricultural runoff [9], depletion by overfishing [10], coastal development and tourism activities [11]. Further challenges include the physical

stresses imposed on benthic organisms by wave action, variations in water levels in littoral zones, and the complexity of accurately monitoring these enormous, intricate ecosystems and their interdependencies. At present, it is acknowledged that Artificial Intelligence (AI) has a constructive contribution towards incorporating machine learning, deep learning, and data analytics for understanding and addressing the severity of these threats and challenges, thereby leveraging the efficiency of benthic conservation mechanisms. There are various extensive studies that showcase the individual impacts of these stressors, and the importance of data analysis in identifying their consequences is also well-established. However, a significant research gap persists. Most existing models and research focus on examining individual threats or a limited number of parameters in isolation. Such models can aid broader syntheses. However, most often these stressors, rather than acting alone, occur coincidentally, creating complex, intricate, cumulative, synergistic effects that are poorly understood and difficult to quantify. Without an integrated, holistic analytical framework, it is challenging to



classify and assess the cumulative and synergistic impact of these multiple and co-occurring stressors on marine biodiversity. Moreover, this gap in comprehensive frameworks limits the ability to accurately predict benthic coastal ecosystem responses and to prioritize the most effective conservation interventions.

Data-driven insights that facilitate initiative-taking management of marine resources, early detection of environmental concerns, informed decision-making, and predictive analytics [12] play a critical role in advancing towards ocean sustainability. Restoration of marine ecosystems is essential for maintaining human health and well-being. Preserving marine biodiversity for the future helps

in advocating for ocean sustainability. Therefore, this study i) addresses the problem of predicting benthic biodiversity decline subjected to multiple, co-occurring stressors, and further ii) proposes the model development and validation of an integrated predictive framework that can synthesize key oceanic parameters including SST, salinity, pH, COD, phosphate and nitrogen concentration to forecast its effects on existence of marine coastal biodiversity, and iii) conducts a thorough ecological feature importance and analysis. This study directly supports the restoration and preservation of marine ecosystems, which is crucial for preserving human health and well-being, by facilitating a more nuanced understanding of these interrelated threats. Various key related works in marine data classification and prediction have been explored and are tabulated in Table 1.

Table 1. Related work: key areas of research in marine data prediction

Key Areas of Research in Marine Data Prediction	Methods/Algorithms Used	References
Framework for forecasting and assessing the carbon sink of marine fisheries	Introduction to a novel nonlinear grey Bernoulli model that incorporates time-varying parameters for forecasting to address nonlinear, complex, and uncertain prediction challenges.	[13]
Prediction of fish-Rastrelliger kanagurta habitat preferences	Generalized Additive Model (GAM).	[14]
Forecasting appropriate habitats for blue whales' migration	Gradient Boosted Model (GBM).	[15]
The estimation/forecasting of sea surface temperature trends	Convolutional Neural Network (CNN), Long Short-Term Memory networks (LSTM), Recurrent Neural Network (RNN), and a Context Fusion Spatiotemporal Deep Learning network (CFSDL).	[16, 17]
Classification and forecasting of coral reefs, reef bleaching	Naïve Bayes, decision trees, support vector regression, and random forests, K-means clustering, and a convolutional neural network.	[18, 19]

According to Table 1 and based on the literature research carried out on related work had facilitated this study to identify the following: i) Need for Generative Adversarial Networks (GANs): Ahmed et al. [20] have discussed that the effectiveness of ML is less compared to Deep Learning (DL) techniques because of ML technique's large data requirement to produce meaningful results. This challenge of data sparsity is being attended by Yue et al. [21] using Generative Adversarial Networks (GAN) algorithms which can eventually augment new, high-quality data, ii) Need for conduct of feature engineering and statistics: use of ensemble modelling comprising of feature engineering and statistical analysis along with machine learning approaches can achieve accuracy to existing state-of-art algorithms as validated by [22]. A study by [23] assessed the consequence of environmentally heterogeneous variables, each with varying spatiotemporal resolutions, on the abundance of coastal fish using the tree-enabled Naive Bayes model. According to Rubbens et al. [24], tree-based techniques are better than linear models at predicting metrics of coral reef health-including species richness, biomass, and diversity- from a set of measurable habitat features. Therefore, while doing a

predictive analysis on species abundance, an ecological assessment of species richness, feature importance, and diversity analysis must be conducted [25]. iii) Preference of Recurrent Neural Networks over other algorithms for marine data predictions [12, 26-28].

From all the algorithms discussed above, this study identified the implementation of RNN for its prediction task. A key advantage of RNN over standard Artificial Neural Networks (ANN) is their inherent capacity to explain dependencies and sequential patterns in time-varying information, in which it is possible to presume that each data instance is dependent on the previous occurrences. But RNNs do acknowledge a difficulty in learning long sustained dependencies [29]. To overcome this weakness of the traditional RNN, LSTM [30] and Deep RNN architecture [31], [32] have been used. Thus, from the study of related works, this research identified a Deep RNN for the prediction of its considered marine dataset. According to a study [33], the predictive accuracy of algorithms is affected by several factors. Primarily, prediction accuracy is poor for imbalanced datasets whose class variable distribution is asymmetric [34].

Another major challenge in deep learning prediction accuracy is overfitting [35], which occurs when a model performs badly on new data not seen during training due to too much focus on training data. To reduce overfitting, strategies like dropout, regularization, cross-validation, and data augmentation are employed [36]. Further constraints for prediction accuracy are a lack of a rich dataset, data unavailability, or sparse data due to limited access to climate data. When data is available, they are frequently incomplete and of low quality [37]. Various initiatives are being taken to address these challenges. One of these initiatives is data enhancement or augmentation using Generative Artificial Intelligence (GAI).

Data Augmentation (DA) serves as a critical strategy towards intensifying the performance of Deep Learning (DL) models. Predictive model performance is significantly enhanced through data augmentation, a technique that artificially expands datasets by creating modified versions of existing samples [38]. Data augmentation is efficiently achieved by the use of generative modeling contributed by Variational Autoencoders (VAE) [39] and by Generative Adversarial Networks (GAN) [40]. VAE works better with larger sets, while GAN is the recommended model for smaller ones. Considering every previously mentioned aspect in this introduction section, this research aims to use variants of GAN for producing generated data for a Deep Recurrent Neural Network. This research paper is structured to discuss the architectures of GANs and Deep RNN in its literature review section, which is then followed by research design section that identified the need for GAN, compared the variants of GAN such as ordinary GAN, Wasserstein GAN (WGAN), Conditional GAN (CGAN), and Climatic GAN (ClimGAN) and proposed Marine Data GAN (MGAN). The discussion and findings section provided empirical findings and model implementation. The Divergence metrics of all the variants of GAN were compared in this section.

Through this comparison, MGAN has proved to be superior. Thus, the data obtained from MGAN was used as an augmented climatic and benthic dataset of the Flic en Flac coastal region belonging to the Republic of Mauritius. Further, predictive analytics using Deep RNN was conducted over the augmented dataset. The findings and discussions section recorded the predicted variables and prediction accuracy of the proposed MGAN-Deep RNN hybrid model. It was determined that the MGAN-Deep RNN is a better predictive model than the basic Deep RNN model. The limitations section identified the limitations of this study and provided ways to overcome the limitations.

2. Review of Literature

Research on predicting multi-faceted time-series marine data, consisting of climatic and benthic variables, primarily presents significant challenges like non-linearity, long-term interdependence, and high volatility [41]. Liu et al. [42] discussed that prediction models using time series data suffer

from poor model generalization and low prediction accuracy. Moreover, unprecedented climatic conditions can create extreme events in marine data and, hence, are not uniformly distributed events leading to an imbalance in data that negatively impacts the model's prediction accuracy [43]. Furthermore, the data collection of climatic data involves in situ measurements in deep remote ocean regions, data collected through satellites, where failure of sensors, orbital irregularities, or cloud cover may challenge the quantity of data collected. Hence, climatic data may be sparse [44].

Moreover, data collected manually through buoys and ships may also lead to data sparsity due to data collection at unavoidable uneven intervals. Conventional statistical approaches, including regression models, are frequently inadequate for modeling these complex, non-linear relationships and often struggle with imbalanced data distributions in the dataset [45]. Consequently, deep learning architecture has emerged as a powerful tool for this task, where [46] has successfully applied the Synthetic Minority Oversampling Technique (SMOTE) such that it enhances prediction accuracy of a dataset that had class imbalance, and research by Jafarigol et al. [47] has handled data sparsity by augmenting data using SMOTE and GANs. Based on this stated background, this review of literature section discusses two important deep learning families that are essential to the current study: i) GAN and their variations for Synthetic Data Generation, and for handling imbalance in class variables, and ii) Deep Recurrent Neural Networks (Deep RNN) for sequence prediction.

2.1. Related Work Regarding GAN as a Generative AI

Generative AI is defined by [48] as a class of computational techniques that derive new, semantically valid data from initial training samples. Generative AI is a powerful engine for creating novel data and a transformative tool for intelligence augmentation, adept at processing and answering user questions. Potential use cases for this methodology encompass the information technology help desk [49]. A cited analysis [48] indicates that generative AI could displace up to 300 million jobs held by knowledge workers while simultaneously boosting global GDP by 7%. Generative AI models are defined by their capacity to learn the underlying patterns of a training dataset and then synthesize novel data that reflects those patterns. Unlike most discriminative models, which only classify or recognize data, generative models can create new instances of the data by learning its underlying distribution. Researchers in Generative AI are increasingly leveraging GANs to perform data augmentation. The core objective of the GAN architecture, illustrated in Figure 1, is to analyze training data to learn its underlying probability distribution. GAN then uses this computed probability distribution to produce more instances that are similar to the actual data supplied. GAN consists of i) a generator and ii) a discriminator, which are constructed using a deep CNN.

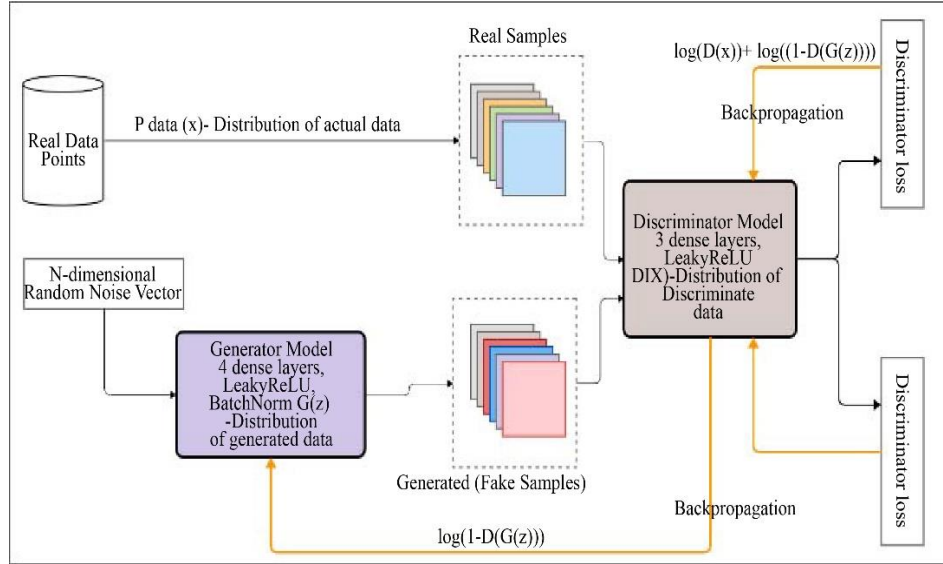


Fig. 1 GAN architecture diagram

Functionally, the generator intends to start with a randomly sampled latent vector, which, when applied to a sequential convolutional architecture, produces a synthesized output. These convolutional layers progressively up-sample (mapping to the higher dimension space) this noise input and create a synthetic sample whose dimension equals that of the data. Normalizing the inputs of each layer using batch normalization is carried out to ensure stabilization during training the network and to prevent vanishing gradients. Non-linearity is introduced to the system via a ReLU activation function applied after every layer, a critical step that enables the network to capture complex feature representations. A hyperbolic tangent activation function is applied in the output layer to constrain the model's final outputs to a predefined $[-1, 1]$ range.

The discriminator model functions as follows: It is built with standard convolution layers, which work as a classifier. It considers the data output of the generator and distinguishes the statistical measure of spread between synthetically produced samples and the authentic sample distribution. Suitable activation functions are used in each layer of the discriminator model to ensure consistent and effective learning throughout the training process. As a regularization technique, during the training process, a few subsets of neurons are randomly deactivated using a method called dropout. This is staged to mitigate overfitting and reduce the tendency of the model under consideration to over-memorize the training samples. Further, the output layer yields a value ranging from zero to one that is limited due to the action of the sigmoid activation function. Consequently, GAN undergoes an iterative training process characterized by an adversarial feedback loop: the generator, by creating results that closely resemble reality over time, complements the working of the discriminator, which concurrently equips itself to differentiate these fakes from authentic samples.

The training process of GAN is governed by the loss functions of each network; the functioning of the generator is focused on cutting down its loss by fooling the discriminator. On the other hand, the discriminator's work is aimed at lessening its own loss by correctly identifying real and synthetic data. Decreasing generator loss signifies that the generator is becoming effective in fooling the discriminator. Whereas the discriminator loss function quantifies the model's error in separating the generator's fake data from actual training data. A small discriminator loss suggests that the discriminator is performing its task efficiently and correctly identifying both real and generated samples. As stated in [50], training in GAN may have a limitation of mode collapse, where the generator may ignore the variation in the dataset and produce only the same result repeatedly. Moreover, vanishing gradients leading to poor learning may occur if the discriminator is too good, and training could become unstable.

Recent studies concerning use of GAN for augmenting environmental climate data to enhance the effectiveness of climatic prediction models involves the following steps: i) Wang et al. [51], used a variant of GAN called cycle consistent GAN to achieve higher accuracy in the model by modifying the SST through simulations thereby reducing biases, decreasing root mean square error, and increasing coefficients of correlation. However, this study did not discuss the generalizability of the approach, ii) whereas [52] established a conditional GAN derived from a geographic downscaling technique that efficiently turns very low-resolution inputs into high-resolution climate information. The method precisely addresses the downscaling process's intrinsic uncertainty, iii) Study proposed by Zhuang Li et al. [53], generated synthetic climatic data using a variant of GAN, aiming to increase the reliability of the model forecasts while enhancing spatial resolution.

2.2. Related Work on Evaluation Metrics of GAN

2.2.1. Divergence Measures of GAN

Within the framework of GAN [54], the utility of Jensen-Shannon (JS) and Kullback-Leibler (KL) Divergence is highlighted as measures for assessing the distance between real and generated distributions. Fundamentally, these two divergences calculate the difference between a probability distribution P and A second, reference distribution Q.

The differentiation between these metrics is that JS-divergence quantifies the divergence between P and Q using symmetric and finite values, whereas KL divergence is asymmetric. For a successful GAN, the JS Divergence should be the least. The mathematical representation of GAN is provided by Equation (1).

$$\max_{Di} \min_{Ge} V(Di, Ge) = E_{a \sim p(a)} [\log Di(a)] + E_{z \sim p(g)} [\log(1 - Di(Ge(c)))] \quad (1)$$

In the above equation, $p(a)$ denotes the probability variations of the real data, $Ge(c)$ the dispersion derived from

the generator, $p(g)$ the sample chosen from the generator, $Di(a)$ represents a discriminator network, and $Ge(c)$ represents the generator network.

2.2.2. Related Work on Variants of GAN

WGAN

Figure 2 provides a representation of the WGAN architecture. Fundamentally, there exists a real data distribution (P_r) which is fed to a GAN to achieve synthetic data generation. Thus, the GAN outputs generated data (P_g). There can be a difference between these two distributions, say $P_g - P_r$. The distance between the entire P_g and P_r is referred to as the Wasserstein distance [55]. As introduced in [56], WGAN aims to stabilize GAN training by lowering the Wasserstein distance. Rather than classifying values between a probability of 0 and 1 as in the GAN discriminator, the critic in the WGAN discriminator uses the Wasserstein distance (score) to identify the relative closeness of synthetic data from the actual one. As the Wasserstein distance shrinks, the generator creates data that aligns more accurately with the actual distribution.

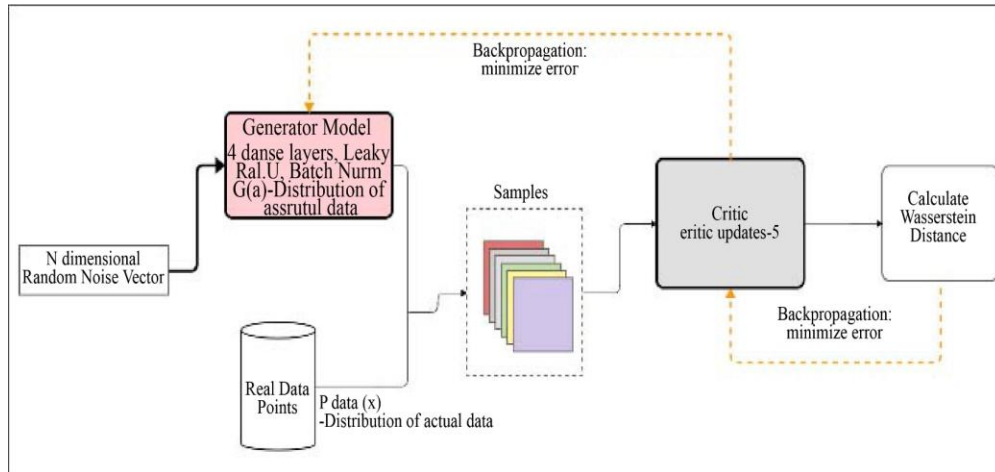


Fig. 2 WGAN architecture diagram

Training a conventional GAN to generate synthetic data may be quite unstable. When the sample distributions produced by the discriminator and generator are different and lack shared traits, then the JS divergence reaches saturation. This implies that it no longer gives the generator a useful signal to learn from, which causes the GAN to become unstable and experience mode collapse (a shortage of gradient that results in GAN failure) [57]. At these instances, the discriminator or critic could become too skilled, too quickly to evaluate the generated data that has been generated by the generator. A 'Lipschitz constraint' can be imposed to limit and sustain the discriminator's fast execution. This restriction is a condition that guarantees the discriminator can't change too quickly. A key constraint on the critic network in WGAN is that it must satisfy the 1-Lipschitz continuity condition. This keeps the critic from becoming too strong prematurely. By functioning as a regularizing condition, this constraint makes

training considerably more stable. Techniques like weight clipping, which merely clamp the neural network's weights to a restricted range, and the gradient penalty method, which imposes a direct constraint on the norm of the critic's gradients to be near 1, can be used to enforce this constraint. Weight clipping in WGAN is that which keeps the critic's weight within a predetermined limit (e.g., [-0.01, 0.01]) by applying a Lipschitz constraint [58]. Although this restriction stabilizes training, excessive clipping may also lead to issues such as WGAN not being able to extract rich features from the data, and weak gradients / weak learning signals may be sent to the generator, slowing down the convergence process. WGAN is being used to produce real climatic data [59]. GAN and WGAN thus differ in the way they manage their discriminator /critic, loss functions, and methods for determining how the generated data distribution deviates from the real data.

CGAN and CLIMGAN

Conventional GANs generate data from random noise, offering limited control over the output. To address this limitation, a Conditional GAN (CGAN) or a Climatic GAN (CLIMGAN) can be used. Conditioning transforms the generator's task from creating generic data (as in standard GANs) to producing output guided by specific parameters [60-62]. However, review by [63] states that this conditioning technique might make it more challenging for GANs to approximate the actual data distribution, which could affect the efficiency of the generative performance. As presented in Figure 3, CLIMGAN follows an architecture like GAN; however, it directs the data generation process by using

conditioning data input as additional guiding information. This conditioning data is specifically a class label from the training dataset, which is integrated with the noise vector fed into the generator [64]. Since CLIMGAN is conditioned on time-dependent variables (e.g., SST, pH, salinity), it allows it to produce physically consistent, multivariate climatic data trajectories that correspond to the chosen initial conditions [61]. The degree of realism in the generated outputs is often improved by conditioning the model with appropriate time-dependent variables, thus providing more diverse and realistic outputs. Study [65] states that deep CLIMGAN achieves high-quality climatic data generation.

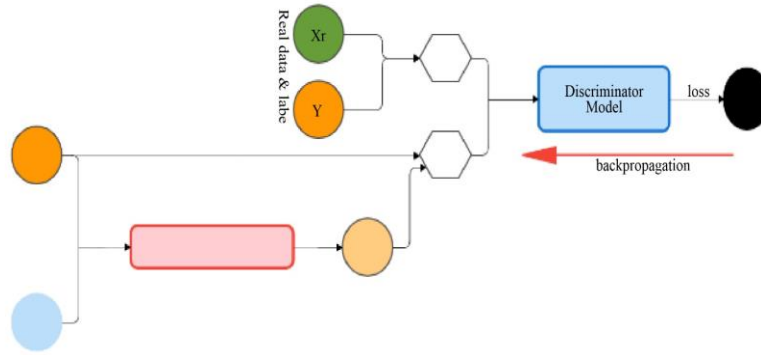


Fig. 3 CLIMGAN architecture diagram

2.3. Related Work Regarding Deep RNN

Zhang et al. [66] have proposed an RNN called an adaptive gated recurrent network (GRU), which efficiently learnt nonlinear time series data (which had both regular and severe event patterns), leading to improved predictions. The research presented in [67] focused on utilizing RNN to capture sophisticated temporal patterns in data points and to improve the accuracy of forecasting. A defining feature of RNNs is their capacity to incorporate information from prior steps into the processing of current ones. This is achieved by the loop

architecture consisting of cyclic connections present in RNN, which allows data to be transferred from one phase to the next [68]. The loop creates a kind of memory which permits the network to retain information about past inputs and use that information to predict future inputs, thus leading to improved prediction accuracy [69]. Though RNN has been used in a variety of applications where the input data is sequential or time dependent, training RNN is challenging due to its difficulty in learning long-term patterns and due to phenomena called vanishing gradients.

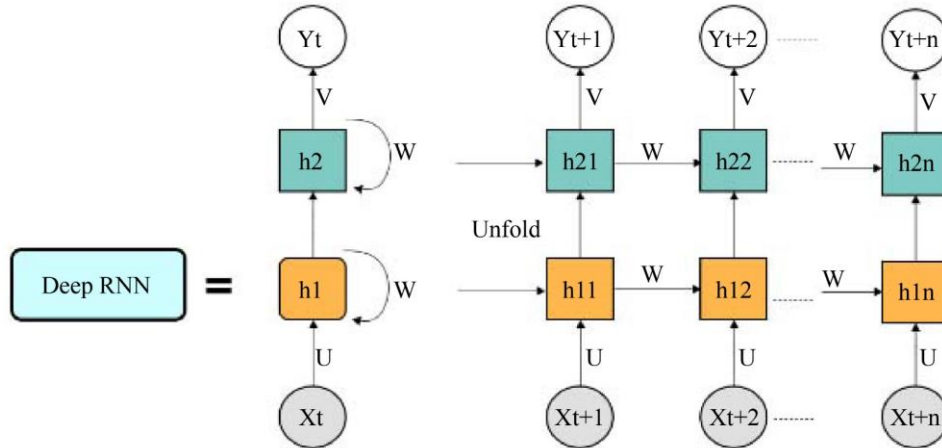


Fig. 4 Deep RNN architecture diagram

Layers of Gated Recurrent Unit (GRU) [70], as well as LSTM [71], solve the challenges of RNN. However, activation functions like sigmoid and hyperbolic tangent used in these networks lead to a persistent problem of gradient decay [72]. Therefore, constructing an effectively trainable deep network is difficult. An innovative type of RNN, termed Deep RNN Figure 4, is being used to resolve these limitations. Deep RNN is constructed by stacking hierarchically multiple RNN layers. Because of this depth, the Deep RNN can not only capture short- and long-term dependencies but can capture hierarchical features and more intricate trends within the data, because each constructed layer processes the entire sequence and the output generated by one layer becomes the input to the next [73]. This improved model is specifically designed to identify long-duration dependencies inherent in temporal and sequential data.

Computation in the Deep RNN architecture includes nonlinear/hidden layers (h_{11} to nn) between the input layer (x_1 to n) at time $k < t$ and the output layer (y_1 to n) at time t . Temporal dependencies are processed by hidden layers that make up each RNN layer. The architecture of Deep RNN Figure 4 considers a sequence of input, say $\{x_t\}$, where t varies from 1 to T , that is passed into the first layer of the Deep RNN. For instance, when considering layer l , the input of the current time step, the product of the weight matrix $U^{(l)}$ and x_t is calculated (that connects the input x_t to layer l), and this product is added to the product of the preceding hidden state value $hd_{t-1}^{(l)}$ and matrix representing weight $We^{(l)}$ that connects the preceding hidden state of the first layer. Further parameters called bias $b^{(l)}$ is added, and to this combination, activation function ϕa is being applied. This leads to the formation of subsequent $hd_t^{(l)}$ which is the current hidden state. Taking the time factor (t) into consideration, the hidden unit of the first layer can be denoted as below, Equation 2.

$$h_t^{(l)} = \phi a (We^{(l)} hd_{t-1}^{(l)} + U^{(l)} x_t + b^{(l)}) \quad (2)$$

The result y_t from the ultimate layer of Deep RNN for We , the matrix representation of weights of the final layer, σ , the function serving for activating the ultimate layer, and c being output bias, are denoted by the following Equation 3.

$$y_t = \sigma(Vhd_t^{(L)} + c) \quad (3)$$

The Deep RNN can model higher-level temporal patterns by adding depth, with top layers capturing more abstract, complicated representations and lower layers learning simpler patterns. Deep RNN has its application in household load forecasting, where the data has significant levels of uncertainty and volatility. However, adding layers to the neural network caused overfitting, which was addressed by increasing the volume and diversity of the data [74]. For

applications like video classification, language modeling, and prediction [75], Deep RNN outperforms shallow RNN. Hence, Deep RNN, when compared to conventional feed-forward networks, can manage sequenced data, including language, time series, and other temporal data types.

Deep RNN for climatic data prediction is evident in the following recent studies: i) compared to a deep neural network, Deep RNN (RNN with LSTM) had outperformed climatic predictions [76], ii) Study by [77] and [78] investigated ocean parameters using an augmented data DeepRNN approach that had achieved higher performance than RNN. However, studies [79, 80] recommended the superiority of Deep RNN for climatic data prediction, highlighting the need for dataset expansion and additional data from the data sources.

The present investigation intends to forecast benthic and climatic data using Deep RNN since climatic data is a time series.

3. Methodology

3.1. Overview of the Dataset

Water quality parameters used in this research are SST, salinity, pH, COD, which denotes dissolved oxygen units, Nitrate-Nitrogen, and Phosphate. These parameters are considered independent features or water quality parameters. This study sought to forecast the results of the above-mentioned independent variables on the target benthic variables, like Hard Corals (HC) and fish assemblages including Labridae (Wrasses), Chaetodontidae (Butterflyfishes), Scaridae (Parrotfishes), Acanthuridae (Surgeonfishes), Pomacentridae (Damsel fishes), and Algae.

3.2. Novelty of Research

Designing a hybrid computational model called MGAN-Deep RNN to achieve improved predictions of benthic variables, thereby leading to data-driven decision making towards habitat preservation, sustainable resource exploitation, and ecosystem-based management, is the prime aim of this paper.

The novel values proposed in this study are the following: i) introduction of a novel GAN called Marine Data GAN (MGAN) for domain-specific marine data generation to overcome the challenge of unbalanced sparse dataset and further using these augmented data for improving the predictive task, ii) proposing an integrated, holistic predictive framework using DeepRNN model to evaluate the combined effects of multiple, simultaneous stressors on marine biodiversity health, iii) conduct of benchmarking exercise on the variants of GAN and predictive models considered in this study to prove their superiority and iv) conduct of various effective statistical tests to quantify water quality parameters and their interventions on the status of health of benthic. To

the best of our knowledge, the benthic community dynamics in a Mauritius lagoon are forecasted for the first time in this study using the combined influence of multiple water quality metrics. Thus, the hybrid model MGAN-Deep RNN, being developed by this research for the prediction of benthic variables, has several innovative aspects, such as data augmentation, enhanced climatic data generation, and hybrid joint model training to achieve improved prediction accuracy of marine data under consideration.

3.3. Rationale of GAN and the Selection of Variants

This study compared its novel MGAN with the following number of GAN models: i) Firstly, the standard GAN was considered as the baseline GAN. This GAN, with its distinctive conceptual design, has an ability to generate extremely realistic samples while dealing with fundamental problems, including mode collapse and instability. Mode collapse signifies a failure of the generator to capture the true data diversity, resulting in the generation of repetitive or limited samples. Furthermore, GAN experience training instability caused by vanishing gradients, which occur if the discriminator becomes too proficient and provides almost minimal information to the generator, thereby halting the generator's learning ability. Climatic data comprises high-dimensional, multi-modal seasonal distributions.

Due to its propensity for mode collapse, the baseline GAN may provide samples from a subset of modes, which would seriously distort the actual diversity of environmental conditions and result in a biased augmented dataset. ii) Secondly, WGAN was selected as a comparison in this study due to its improvements in training stability. WGAN gives the generator a more meaningful gradient by using the Wasserstein distance as a loss function.

Training stability and meaningful gradient generation are especially crucial for continuous, multi-variate marine data parameters (such as SST, pH, and salinity). Thus, WGAN was used in this study with the objective that it may ensure better coverage of the actual data distribution though it may face computational overhead limitation due to gradient penalty, iii) Thirdly, The Conditional GAN was chosen as comparison in this study because it introduces conditional information (e.g., class labels of the dataset) into generator as well as into discriminator.

This conditioning is extremely pertinent to our application because it enables the creation of data samples conditioned on water quality indicators, so that the model can simulate the direct, conditional interactions between variables (e.g., generating benthic data given a specific SST and nitrate level). The effectiveness of CGAN is strongly influenced by how complete, high-quality, and representative its conditioning data is. However, the CGAN model may not learn a meaningful mapping if the conditioning variables are noisy, lacking, or don't have a strong, consistent relationship

with the target data. Lastly, the CLIMGAN model, though it has a limitation of focusing on spatial correlation rather than temporal dynamics, is proposed in this study as it was created especially for geoscientific and climatic data.

3.4. Pre-Processing

A comprehensive preprocessing workflow, coded in Python, was applied to the raw dataset to guarantee compatibility and data quality with the considered deep learning models. Initial data profiling identified and removed the following physiologically implausible outliers: i) in algae distribution, the general range existed was from 30.9 to 35.6. However, outliers of value 83.0 were present in the dataset and were removed. ii) In abiotic distribution, the general range was from 44.3 to 54.8. But outliers of value 12.0 were present in the dataset, which were removed. Key water quality metrics like SST, pH, and practical salinity values were originally stored as intervals. These were transformed into scalar values by using an `interval.split()` method, providing a single representative value for these variables. Addressing missing values was a multi-step process. `SimpleImputer`, which involved replacing missing values by the mean of the column, was used to impute isolated, single-point gaps. For longer gaps of less than five consecutive records, linear interpolation was employed to maintain the original temporal patterns within the marine dataset. Extended gaps exceeding five records were listwise deleted to prevent the introduction of spurious patterns. To ensure robust and effective neural network training, all features were subsequently normalized to a [0, 1] range using the `MinMaxScaler`, resulting in all parameters on the same scale. Thus, this pre-processed dataset served as the foundation for all subsequent modeling in this study.

3.5. Proposed Model Architecture: State-of-the-Art MGAN-Deep RNN Hybrid Model

This research is grounded in generative AI GANs for data generation to overcome the data sparsity of benthic data under consideration. The generated data is then applied to a predictive model with the motive of assessing whether this data, instead of actual data, improves the accuracy of prediction.

The model architecture proposed by this study Figure 5 has key components, including data preprocessing, and deduction of the most effective GAN out of an ordinary GAN, WGAN, CGAN, and CLIMGAN. This deductive learning is then used to inductively propose and assess a state-of-the-art GAN known as Marine GAN (MGAN). Based on model evaluation parameters such as KL-Divergence, JS-Divergence, and Wasserstein Distance, the superiority of MGAN is proved for the marine data under consideration. The architecture then led to predictive analysis using the Deep RNN algorithm for the MGAN-generated data. Through appropriate findings, this research proposed a hybrid model named 'MGAN-Deep RNN' as the model for the prediction of marine data under consideration.

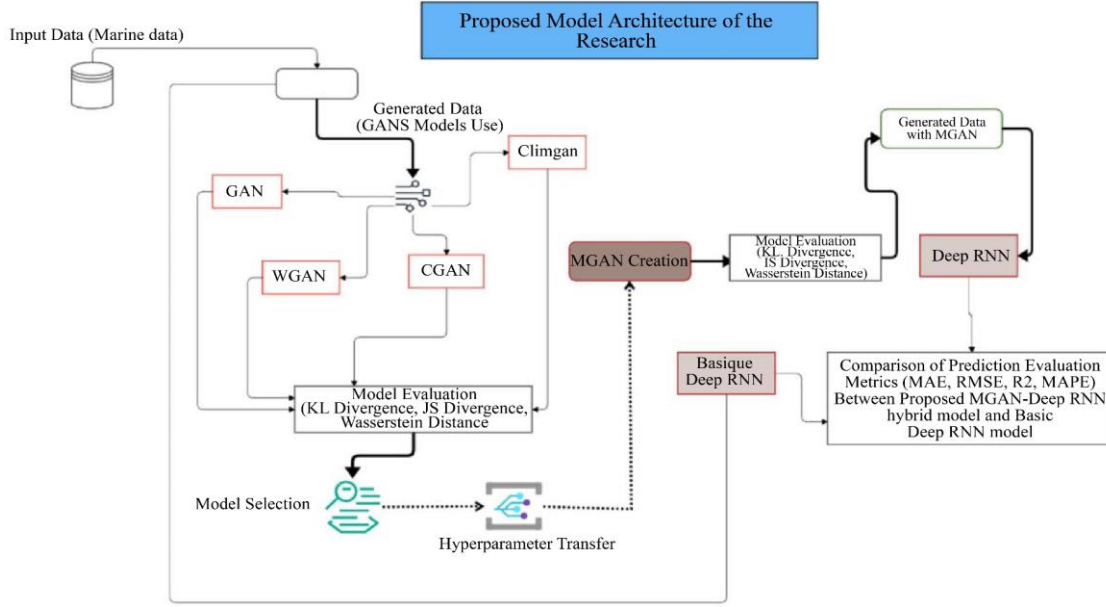


Fig. 5 Proposed model architecture

MGAN Workflow	
Step-1: Initialization - Randomly initialize the weights of the Generator (G_g) and Discriminator (D_d). - Define loss functions of the generator and discriminator. - Set up: - Discriminator activation functions, optimizer and weight clipping (to improve stability) - Generator optimizer. - Hyperparameters, including learning rate, and batch size. Step-2: Sampling - Take a sample of data ($D_d(x^{(n)})$) from the real data distribution - Generate fake data by sampling random noise vectors $z^{(n)}$ from a Gaussian or uniform distribution (P_z), then passing it through the Generator ($G_g(z^{(n)})$). Step-3: Conditioning - Add water quality parameters as conditional inputs (e.g., SST, pH, PSU, COD). - Concatenate these conditional inputs with noise vectors, then pass them through dense and convolutional layers to produce high-quality generated samples. Step-4: Training the Discriminator - Calculate the Discriminator Loss: $L_D = -\frac{1}{m} \sum_{n=1}^m [\log(D_d(x^{(n)})) + \log(1 - D(G_g(z^{(n)})))]$ (4) - Update discriminator parameters to minimize the L_D.	Step-5: Training the Generator - Generate a new batch of noise and pass it through the generator to create fake data. - Compute the Generator loss: $L_G = -\frac{1}{m} \sum_{n=1}^m \log(D(G_g(z^{(n)})))$ (5) - Update generator parameters to minimize L_G, aiming to reduce the Jensen-Shannon (JS) divergence. Step-6: Improvements - Train the discriminator and generator in a feedback loop to lower generating losses. - Update the generator parameters to minimize L_G, and Jensen-Shannon (JS) divergence. Step-7: Iterations - Continue training for a fixed number of iterations, updating the generator and discriminator at each step until the generator produces realistic data with minimized loss. - Use batch normalization layers to stabilize training. Step-8: Determination of MGAN - Measure performance using KL Divergence, JS Divergence, and Wasserstein Distance. - Compare these metrics for MGAN and WGAN. - Visualize the metrics to assess MGAN effectiveness on the marine dataset. Step 9: Augmented Data Application - Feed the realistic data generated by MGAN into a Deep RNN for prediction tasks.

Fig. 6 MGAN Workflow

Deep RNN Workflow	
Step-10: Initialization of Deep RNN: - Considering the MGAN generated data as input dataset. set the model parameters including learning rate, number of epochs, batch size, length of sequence - Define the RNN Architecture: - Hidden layer size, number of RNN layers, input features, output features Step-11: Define Deep RNN Class: - Initialize weights and bias for each layer - Implement a forward function to process inputs through RNN layers. - Produce predictions using the hidden state of the final RNN layer.	Step-12: Training the Deep RNN: - Loop through training data for the set number of epochs. - Compute loss and, using an optimizer, update weights by calculating gradients. Step-13: Evaluation: - Test the model and compute evaluation metrics (e.g., Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)) to measure accuracy. Perform comparison of evaluation metrics of MGAN-Deep RNN with basic Deep RNN to determine the superiority of MGAN-Deep RNN.

Fig. 7 Deep RNN Workflow

MGAN is the proposed novel GAN, which combines the approaches of WGAN for stability and CLIMGAN for climatic conditioning. The MGAN conditions the generator training with marine data variables while altering the discriminator training as in WGAN.

The improvements used in MGAN are the incorporation of residual layers to capture complex information, the addition of dropouts to prevent overfitting, the addition of 1D Convolutional layers with upscaling to capture the temporal relationships better, inclusion of gradient penalty to prevent exploding gradients and incorporating stabilization of the cost function, and Wasserstein loss being more effective means of quantifying the divergence between real observed and model-generated data distributions. Thus, to improve predictions, MGAN-generated synthetic data is inputted to the Deep RNN.

The workflow of MGAN and Deep RNN is shown in Figures 6 and 7, respectively. This sequence of workflow ensures a systematic approach to generating realistic data and training an accurate predictive model, enhancing both dataset quality and model performance.

4. Discussions and Findings

4.1. Multivariate Analysis of Benthic-Environmental Relationships

To identify the relationship of environmental variables (water quality parameters such as SST, pH, salinity, COD, N-N, and POS concentration) with benthic (species variables including hard corals, fish assemblages, and algae) of the year from 2017 to 2022, this study conducted a comprehensive multivariate analysis, including the following:

4.1.1. Average Benthic Abundance Representation

Figure 8 displays the mean benthic composition for the total survey period. HC showed the highest mean abundance (64.2), followed by Pomacentridae (39.7) and Chaetodontidae (33.7). All other groups each comprised less than 20% of the average recorded abundance. This representation was conducted to identify the benthic abundance over the total survey period from 2017 to 2022.

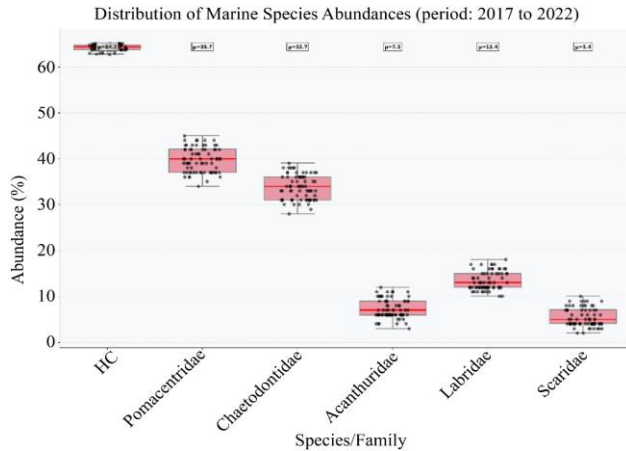


Fig. 8 Average benthic abundance representation

4.1.2. Benthic Compositional Analysis

For the study marine dataset under consideration, compositional analysis was conducted to observe the relative abundance of its biotic and abiotic factors. This analysis graphically depicts the health of marine ecosystems via an area chart Figure 9. Key observations from Figure 9 include i) The resilient core community is constituted by the stable presence of Hard corals from 2017 to 2022. Further, this stability with minor seasonal variations indicated that there existed no significant disturbance events like coral bleaching and that the health of the coral was not degraded to a noticeable extent. ii) The composition of fish communities showed a noticeable decline from November to April of a year and subsequent recovery during May to October of a year. This variation is thus considered seasonal, where the country Mauritius has two weather seasons: summer and winter [81].

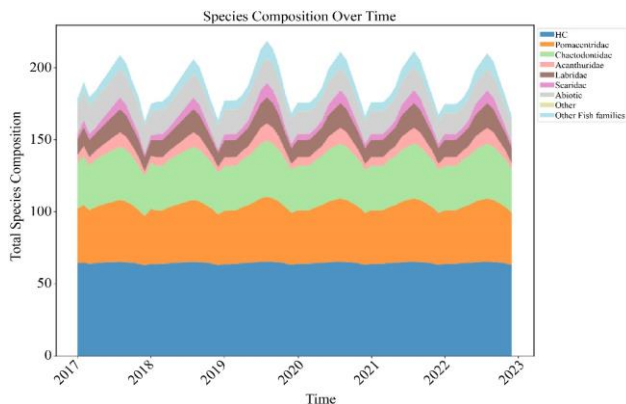


Fig. 9 Benthic compositional analysis

4.1.3. Correlation Between Benthic Abundance

Correlation analysis to identify the interrelationship between HC and five fish assemblages was conducted in this study, and the results are visualized in Figure 10. Given the highly positive association between HC and fish families depicted in Figure 10, the results carry meaningful ecological implications. The finding from the observed pattern may not necessarily indicate that these fish species directly depend on each other, but rather that they all may depend on a common factor: healthy hard coral cover.

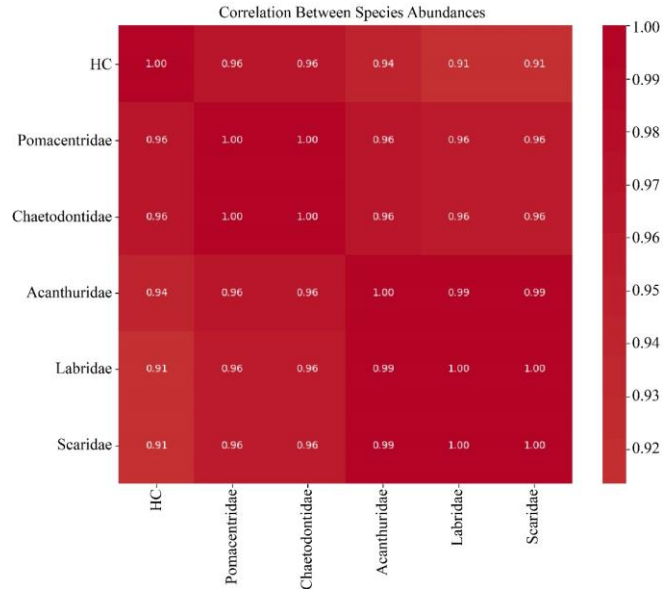


Fig. 10 Correlation Analysis between Benthic Abundance

4.1.4. Testing Environmental Community Relationships using PERMANOVA

To compare groups in multivariate data and determine if a variable accounts for data differences, this study conducted PERMANOVA [82] (Permutational Multivariate Analysis of Variance) test for its dataset: (i) 6 water quality parameters such as SST, pH, PSU, COD, N-N, POS as grouping variables and, (ii) HC + 5 fishes as benthic species. The findings of PERMANOVA were i) the F-statistic was 5.7829. This indicated that the grouping variables had a major influence on six benthic communities. The R^2 square value of 0.3480 indicated that water quality parameters significantly (as p-value was <0.001 for permutations = 999) explain 34.8% of variations in benthic composition.

4.1.5. R^2 Weighted Feature Significance Assessment

To recognize the most influential water quality variables, this work performed a feature importance analysis using a Random Forest Regressor. The resulting rankings, which quantify each variable's impact on performance, are shown in Figure 11. The findings revealed that nitrate levels (34.1%) are the most influential feature, followed by SST (30.0%), pH (18.1%), and salinity (9.0%) of water. The findings indicate that upwelling, terrestrial runoff, and various other factors that

lead to an increase in nitrogen concentration need to be monitored and managed to be at a nominal value, failing which may affect the health of the benthic. Rise in SST may

also deteriorate the health of benthic organisms, followed by the pH of the water. Therefore, these findings can inform coastal management towards resilience.

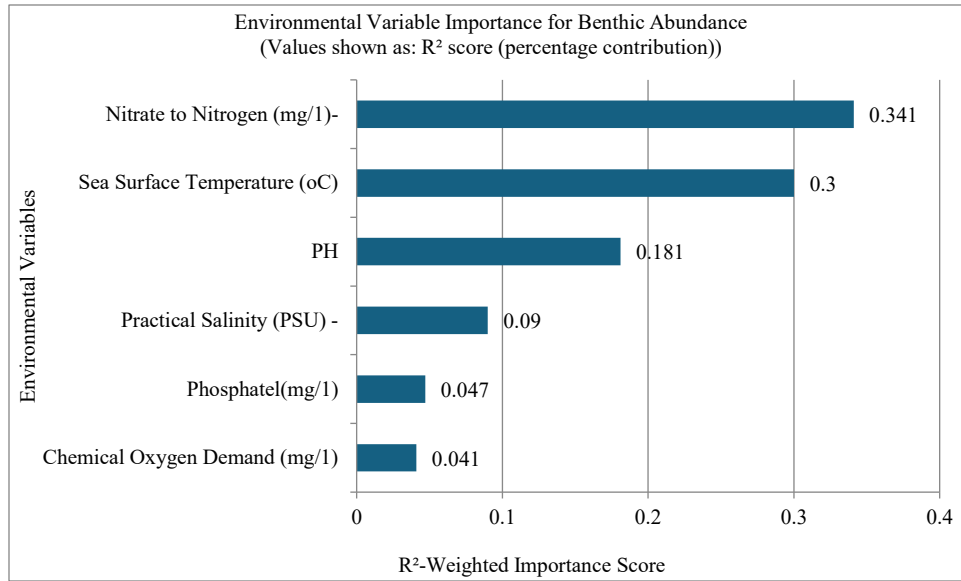


Fig. 11 R² Weighted feature significance plot

4.2. Identification of Unbalanced Datasets

Identification of an unbalanced dataset involves analyzing the distribution of classes/categories within the data.

The following are the methods used to detect unbalanced data from the study:

4.2.1. Class Distribution Analysis

Class imbalances are a fundamental feature of the unbalanced data set [83]. The number of samples in each class was determined. Histogram Figure 12 was used to visualize this distribution of classes. Discrepancies between class frequencies and their asymmetric behavior were found, which indicated an imbalance in the dataset.

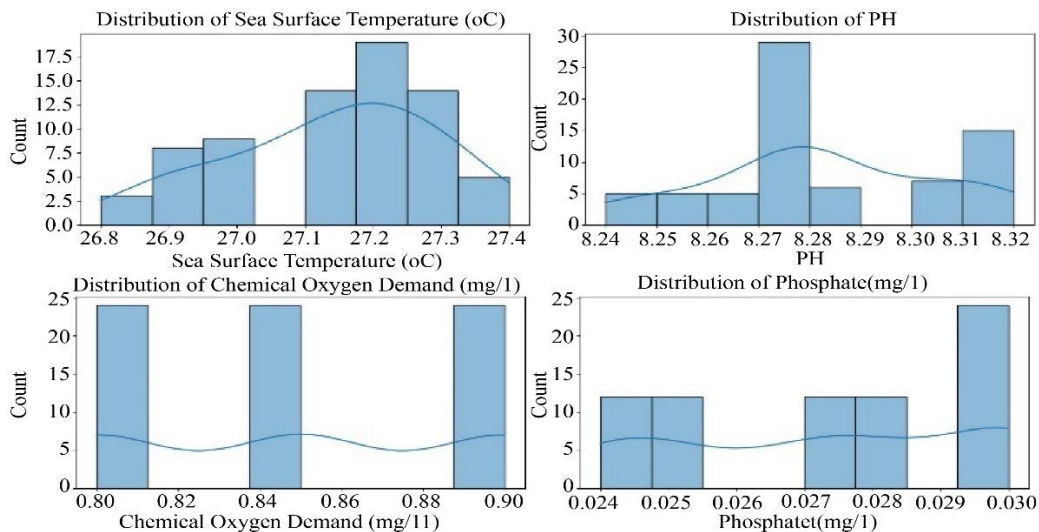


Fig. 12 Class distribution analysis of water quality parameters (SST, pH, COD, POS)

4.2.2. Mean-to-Median Ratio Analysis

The mean-to-median ratio is a statistic that compares the arithmetic mean of a dataset to its median. This ratio can provide insight into the shape of the distribution and the

skewness of the data. The mean-to-median ratio, if found to be significantly greater than 1, indicates disproportionate data and imbalance in the dataset [84]. Thus, alongside class distribution, central tendencies such as mean and median as

indicators of skewness were calculated for each class of independent variables (water quality variables) included in the dataset, and are stated in Table 2. As per the mean-to-median ratio, pH and COD indicated a symmetric distribution of data, as the value was found to be 1.0. But the value of the mean-to-median ratio of SST, PSU, N-N, and Phosphate variables is <1.0, which is an indicative measure of mean is less than median, and some values of these variables are pulling the mean to the left, suggesting the disproportionate data, and hence these variables exhibited a negatively skewed distribution. Thus, the mean-to-median analysis exhibited a negative skewness or an imbalance in the considered dataset.

4.2.3. Shannon Entropy

Shannon Entropy (H) values vary between 0 and 1. It is a measure to determine the imbalance in the dataset [85]. Shannon Entropy is indicated as Equation 4:

$$H = - \sum_{i=1}^k \frac{c_i}{n} \log \frac{c_i}{n} \quad (4)$$

where 'n' is the number of instances of the dataset and 'k' is the classes of size c_i .

The level of entropy serves as a gauge for uncertainty: high values signal disorder and unpredictability, while low values point to order and certainty.

From the Table. 2, it is evident that there exists an uncertainty in the dataset, as the Shannon entropy value is > 2.0 for almost all the variables, indicating that the dataset is imbalanced.

For example, the Shannon entropy for PSU is 4.915706, which indicates a high level of uncertainty or diversity in a dataset.

Table 2. Descriptive statistics of class data

	Mean-to-Median Ratio	Shannon Entropy	Shannon-Weaver Index	Coefficient of Variation
SST	0.997753	2.611447	2.611447	0.581971
pH	1.000285	3.009209	3.009209	0.275246
PSU	0.998816	4.915706	4.915706	1.589669
COD	1.000000	1.584963	1.584963	4.836626
N-N	0.995202	2.685256	2.685256	3.171682
POS	0.993939	2.251629	2.251629	8.419226

4.2.4. Shannon-Weaver Index [85]

The diversity function from the 'Vegan' package in R programming was used to determine the Shannon-Weaver Index as an indicator for diversity between various class variables (Equation 5).

$$H' = - \sum_{i=1}^n (p_i^* \ln p_i) \quad (5)$$

The term p_i Quantifies the fractional representation of a given species in the dataset. The value of this indicator varies from 1.5 to 3.5 (larger values signify higher diversity).

Determination of this value for the dataset under consideration (Table 2) indicates a value closer to the higher side, and hence an imbalance in the dataset is identified.

4.2.5. Coefficient of Variation (CV) [86]

This indicator can quantify the dispersion or spread of data. The equation for CV is given by Equation 6 below.

$$CV = \left(\frac{\text{Standard Deviation of Class Sizes}}{\text{Mean class size}} \right) \quad (6)$$

More variability in data is indicated by a greater CV, which suggests an imbalance in the dataset. The CV of water quality / physicochemical parameters involved in this work is indicated in Table 2. As tabulated, the values of CV were found to be marginally on the upper side, which indicates that the dataset is relatively uncertain and unbalanced. Hence, to

overcome this imbalance in the dataset, the use of various GANS for the dataset was tested, and the research proposed a novel MGAN for data generation.

4.3. Benchmark Testing GANs

Benchmarking computational algorithms can be facilitated by selecting the optimal or suitable algorithm for a particular scenario or dataset [87]. Under this section, this study has presented benchmark tests between GAN and its variants, such as WGAN, CGAN, and CLIMGAN, by running the same marine dataset in all the mentioned GANs.

4.3.1. Metric Analyses

Wasserstein Distance, JS, and KL divergence serve to quantify the similarity between the probability distributions of the generator and the discriminator. However, because of its symmetry, capacity to handle zero probability, and stability in producing finite values, JS divergence is typically a preferable option for uncertain climate data [88]. Hence, this research preferred JS divergence to KL divergence. Furthermore, KL divergence is frequently more informative for modeling and identifying the variation between real and approximate distributions, and hence was also used in the study. Wasserstein distance helps the model to converge better, making it a preferred metric for applications such as climatic data, where complete overlapping between distributions may not be possible. Therefore, in this study, KL Divergence, JS Divergence, and Wasserstein Distance were employed as metrics to ascertain the divergence.

The comparison of these three divergences for GAN and its variants is shown in Figures 13, 14a, and 14b. From Figure 13, WGAN is proven to be the best-performing one among GAN, CGAN, and CLIMGAN because of its lowest value of KL Divergence. For WGAN. The KL Divergence values of WGAN Figure 13 range from 0.220403 (for Sea Surface Temperature) to 0.944382 (for Practical Salinity). This suggests that WGAN usually produces distributions that are more identical to real marine data compared to other models, except in the case of practical salinity, where the value is higher. For GAN. The KL Divergence of GAN Figure 13 was higher, ranging from 0.077443 (for Nitrate to Nitrogen) to 1.454944 (for Practical Salinity). GAN performs better in Nitrate to Nitrogen but tends to struggle more with other variables, especially Practical Salinity. For CGAN, the KL Divergence values of CGAN Figure 13 were significantly higher, reaching 3.146695 (Sea Surface Temperature). This

suggests that CGAN struggles greatly to match the real data distributions. For CLIMGAN. High KL Divergence values were exhibited for CLIMGAN Figure 13, peaking at 4.471253 (Practical Salinity). It had the worst performance overall, especially for variables like Chemical Oxygen Demand and Practical Salinity. Thus, as per the comparison of KL Divergence Figure 13, WGAN outperforms GAN, CGAN, and CLIMGAN, with lower values across most variables. CGAN and CLIMGAN show significant divergence, with CLIMGAN performing the worst. Hence, through benchmark testing, WGAN is considered the winning GAN for the data under consideration. Further from Figure 14(a), WGAN is proven to be the best-performing one among GAN, CGAN, and CLIMGAN because of its lowest value of JS Divergence for most variables, such as SST, pH, COD, and POS. However, the Wasserstein distance graph in Figure 14(b) indicated that CLIMGAN was better performing than WGAN.

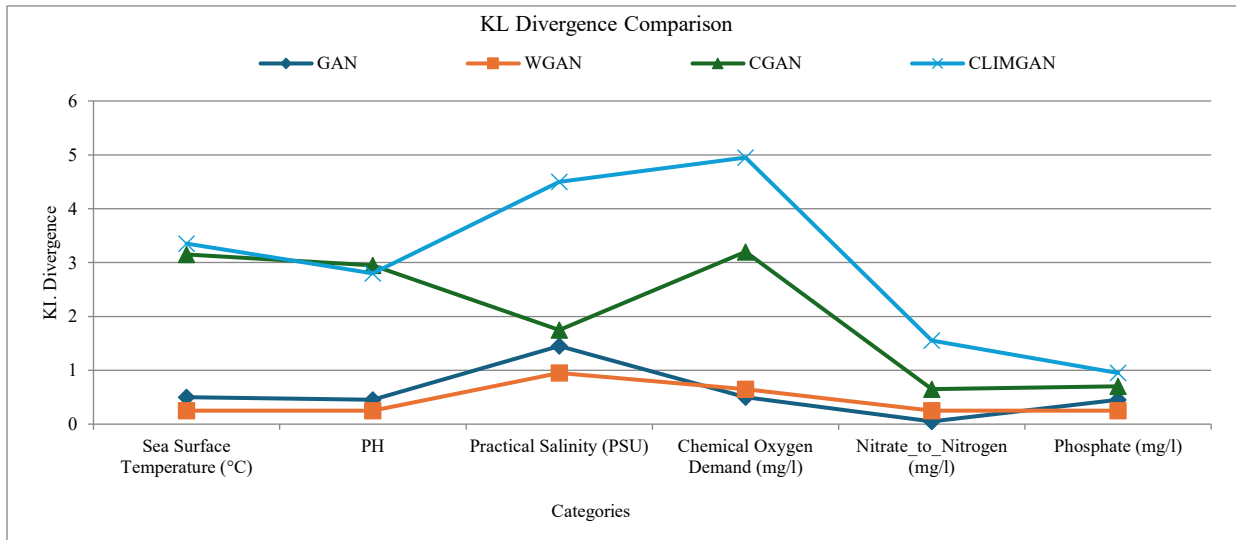


Fig. 13 Comparison: KL divergence metric analysis of GAN and its variants

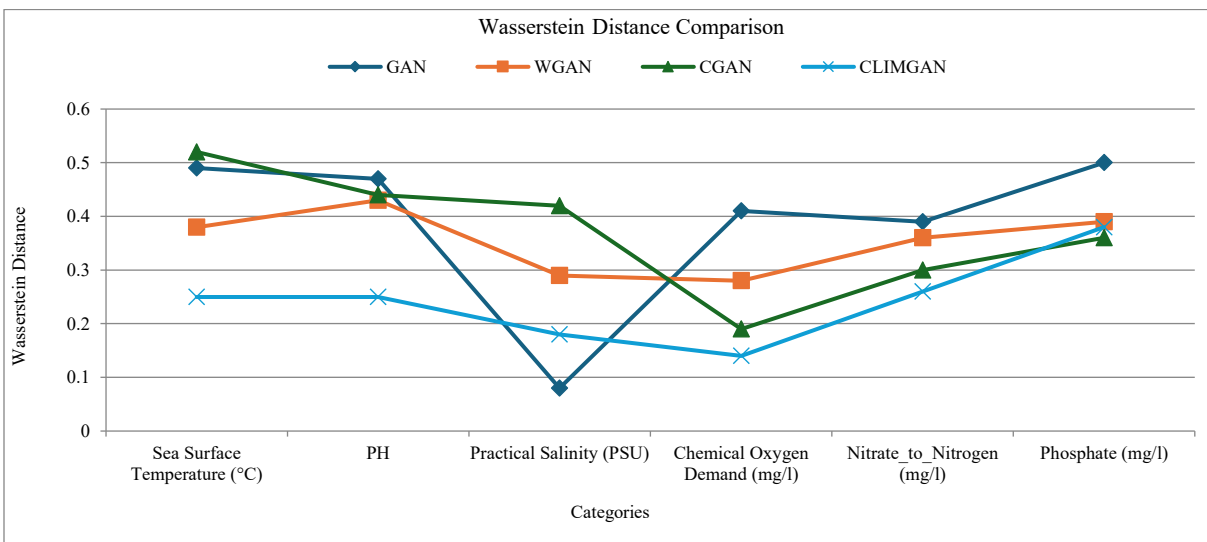


Fig. 14(a) JS divergence metrics analysis of gan and its variants

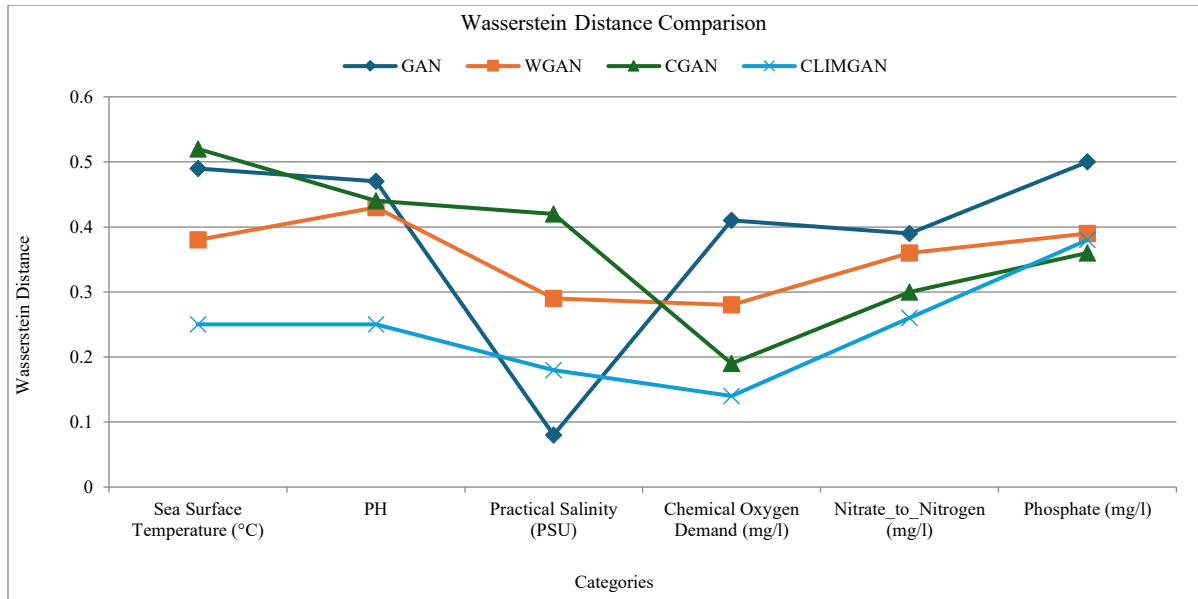


Fig. 14(b)Wasserstein distance metric analysis of GAN and its variants

But for most measures (KL Divergence and JS Divergence), it is proven that WGAN is the best-performing one among GAN, CGAN, and CLIMGAN.

4.3.2. Generator Loss and Discriminator Loss of WGAN

Plots of WGAN loss curves of its generator and discriminator components Figure 15) provide insight into the model's convergence behavior. From Figure 15, however, during the start of the training for say, until around 75 training epochs, the generator had been challenged to find a suitable

gradient, due to which the generator loss exhibited random performance.

Then, after one hundred epochs, as the generator improved, the discriminator's performance declined with less variation. It is evident that according to the WGAN training, the discriminator and generator worked against one another; thus, when the generator improved, the discriminator deteriorated.

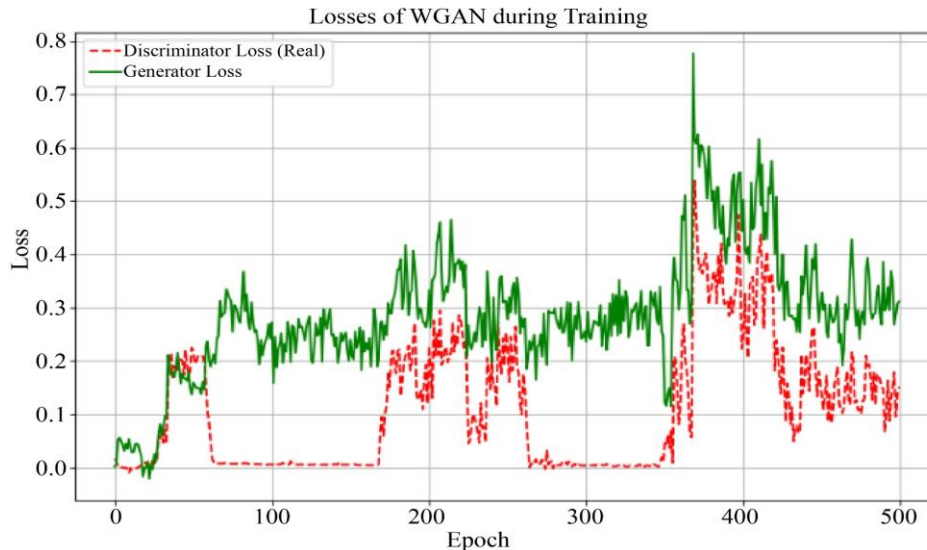


Fig. 15 Loss comparison of WGAN

4.3.3. WGAN Performance based on its Data Generation

As WGAN is an appropriate GAN for the dataset under consideration, differentiation between its actual marine data and the generated one has been plotted and exhibited in Figure 16. This Figure demonstrates the close alignment between the

distribution learned by the WGAN and the true data distribution. The generator's capacity of WGAN, to produce a broad range of realistic samples that encompasses the whole spectrum of potential data changes found in the actual dataset, exhibited generalization capability of WGAN.

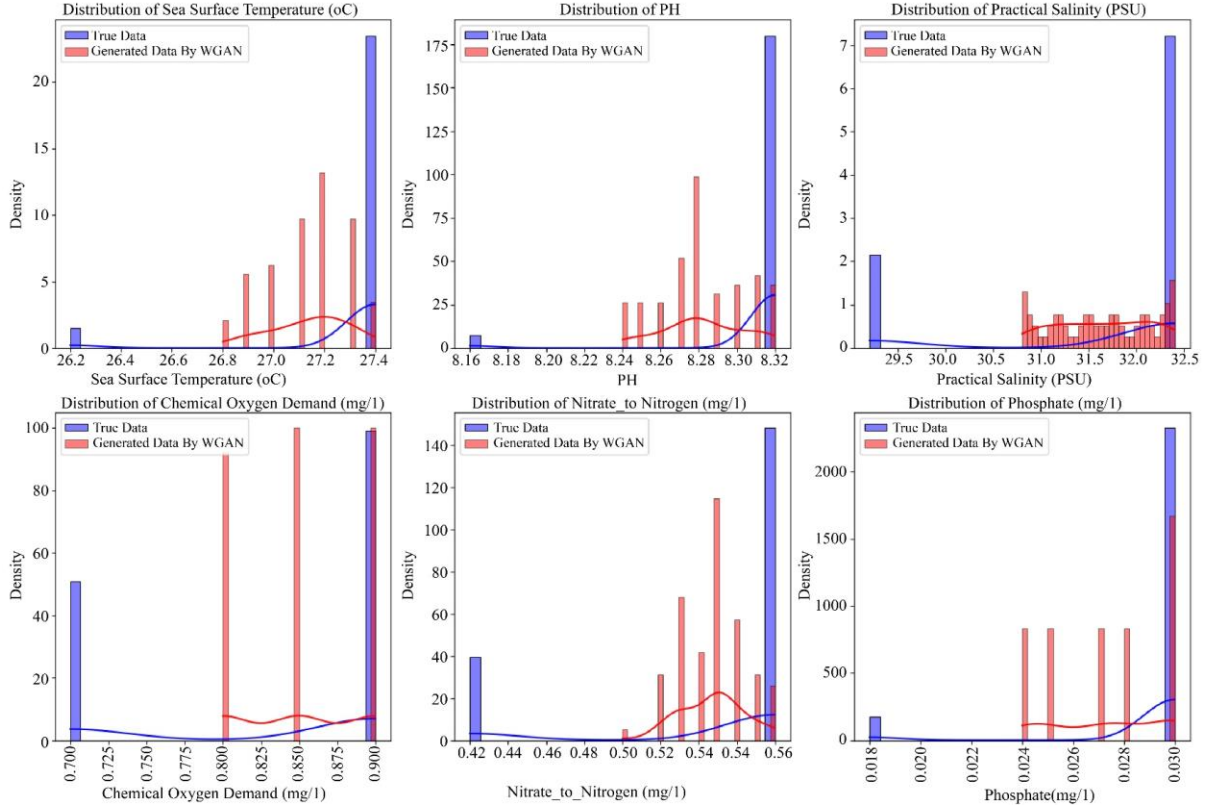


Fig. 16 Comparison of real and generated data for WGAN

4.4. Evaluation of State-of-the-Art Marine Data GAN (MGAN)

Through benchmark testing, this research identified WGAN to be superior in generating data when compared to other GAN variants. Instead of proceeding to predictions considering WGAN-generated data, this study focused on achieving high-quality climatic data generation by using CGAN features in combination with WGAN.

The use of CGAN was proposed because outputs can be conditioned by using climate-specific conditions, such as SST, PSU, and pH, as distinct labels. The use of WGAN was proposed due to its improved divergence metrics. These two research considerations led to the formation of a novel hybrid GAN known as Marine Data GAN (MGAN), which is a combined approach of WGAN and CGAN.

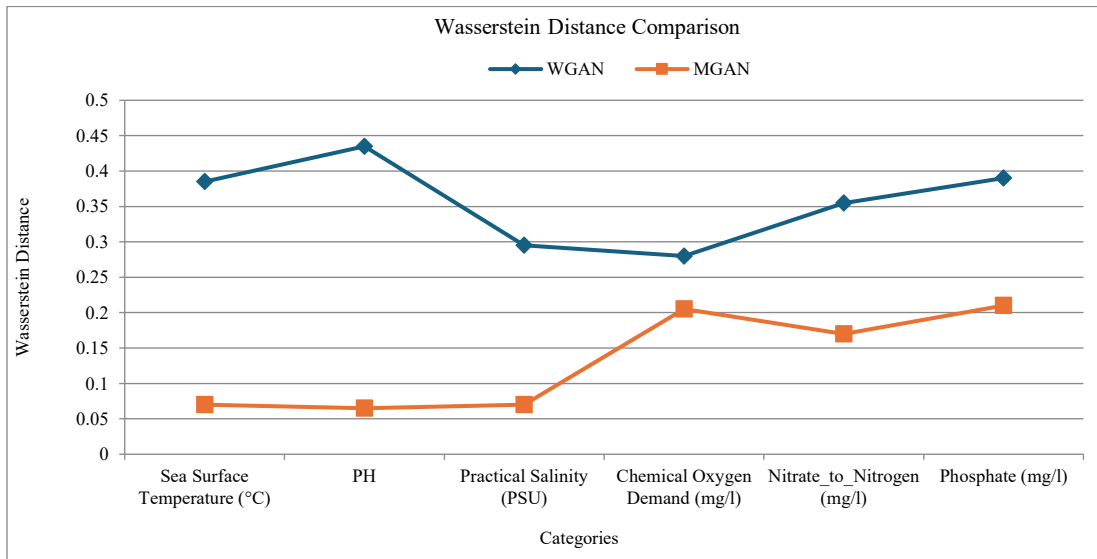


Fig. 17 Comparison: wasserstein distance metric analysis of MGAN and WGAN

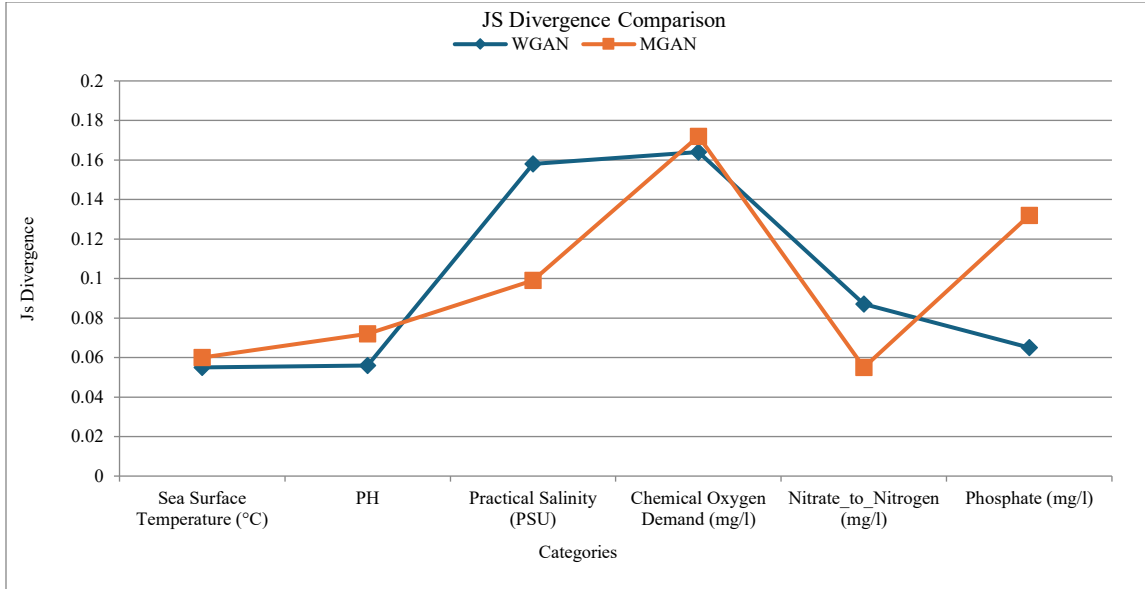


Fig. 18 (a) JS divergence metrics analysis of MGAN and WGAN

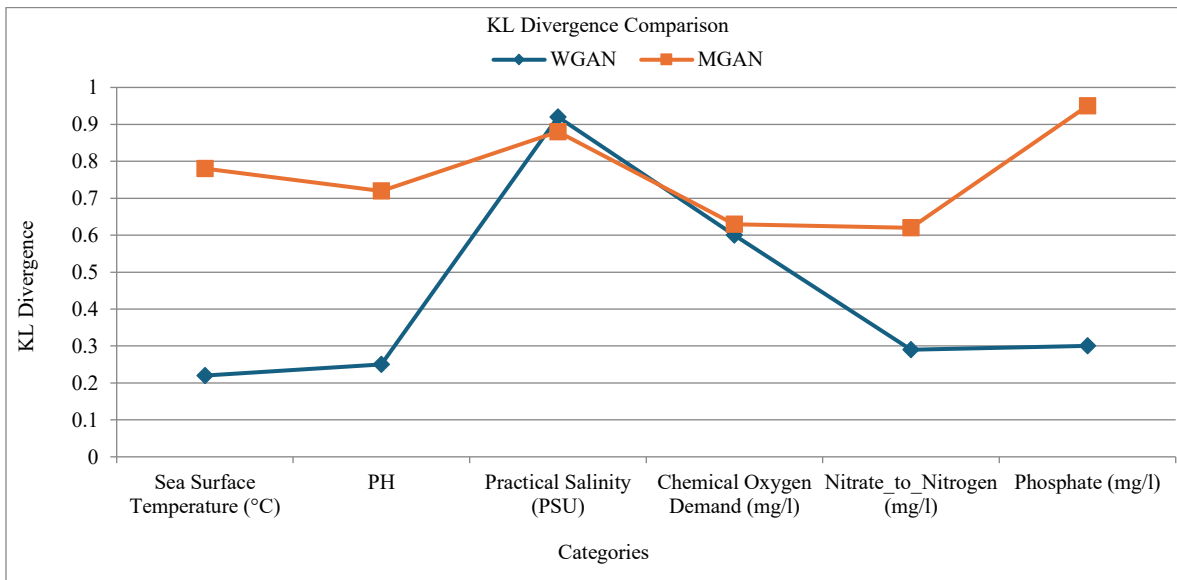


Fig. 18 (b) KL Divergence Metric Analysis of MGAN and WGAN

To determine the superior performance of MGAN, we implemented the recommended algorithm outlined in Figure 6, and the resultant divergence metrics were compared to the WGAN (which proved superior by section 4.3 – benchmark test). This comparison is produced as Figures 17, 18(a), and 18(b), respectively.

4.4.1. Analysis of MGAN Performance based on Wasserstein Distance

From Figure 17, it is evident that for Sea Surface Temperature (°C), MGAN performs much better with Wasserstein distance of 0.068704, compared to WGAN with Wasserstein distance of 0.383502, highlighting that MGAN generates a much closer approximation of the real distribution.

Similarly, for pH, MGAN's Wasserstein Distance of 0.062658 is far lower than WGAN's 0.433497, indicating that MGAN better captures the distribution of pH values.

With reference to Figure 17, it is determined that for Practical Salinity Unit (PSU), MGAN scores Wasserstein Distance of 0.070470, compared to WGAN's 0.293927, once again demonstrating MGAN's better fit for this parameter. For Nitrate_to_Nitrogen (mg/l), MGAN achieves a Wasserstein Distance of 0.171875, whereas for WGAN it is much higher at 0.353333, showing that MGAN better approximates this distribution. Thus, by considering the Wasserstein distance between MGAN and WGAN, it is illustrated that the proposed MGAN performance is superior to WGAN.

4.4.2. Analysis of MGAN Performance based on JS Divergence

From Figure 18(a), when considering the JS Divergence, for Sea Surface Temperature (oC), MGAN scores 0.059179, while WGAN has a slightly better score of 0.049734, indicating that WGAN generates a more accurate distribution for this parameter. For pH, MGAN has a JS Divergence of 0.067795, which is again higher than WGAN's 0.052342, further showing WGAN's ability to generate data closer to real-world values. On the other hand, MGAN outperforms WGAN in Practical Salinity (PSU) with JS divergence of 0.097572, as compared to WGAN's 0.155550.

4.4.3. Analysis of MGAN Performance based on KL Divergence

From Figure 18(b), when considering the KL Divergence, for Sea Surface Temperature (°C), MGAN has a KL Divergence of 0.804843, whereas WGAN performs significantly better with 0.220403, indicating a closer match to the real distribution for this parameter. In contrast, for pH, MGAN has a KL Divergence of 0.752175, which is much higher than WGAN's 0.242999, again highlighting WGAN's superior performance in capturing the pH distribution. However, for Practical Salinity Unit (PSU) and COD, MGAN performs slightly better than WGAN, say PSU of MGAN showed 0.885084, which is a better parameter than WGAN's 0.944382. Evaluation based on the Wasserstein distance revealed that MGAN typically surpassed WGAN, reflecting its stronger performance in generating samples that align with

the real data distribution. While MGAN also showed improvements in JS Divergence for a few categories, its performance was more variable compared to KL Divergence. These results suggest that MGAN is a promising generative model.

4.4.4. Performance Improvement of MGAN over GAN and its Variants based on Wasserstein Distance

The performance improvements of MGAN (Table 3) over the other considered GANs were calculated using the following Equation 7.

$$\% \text{ Improvement} = \left[\frac{\text{Metric of (GAN, or WGAN, or CGAN or CLIMGAN)} - \text{Metric of MGAN}}{\text{Metric of (GAN, or WGAN, or CGAN or CLIMGAN)}} \right] * 100 \quad (7)$$

From the Table 3, MGAN outperforms GAN and its variants by a large margin in almost all parameters. For example, for Sea Surface Temperature (°C), MGAN shows an improvement of +86.9% compared to CGAN, and +72.2% compared to CLIMGAN. For the COD variable, WGAN and CGAN showed competitive performance, and CGAN outperformed MGAN by illustrating a 9.9% improvement of 9.9%. Thus, in some cases, WGAN and CGAN are more competitive, particularly for Chemical Oxygen Demand (mg/l). However, MGAN excels in Wasserstein Distance for parameters like Sea Surface Temperature (°C), and PH, showing improvements of +86.9% to +86.0% compared to GAN and its variants.

Table 3. Comparison of performance improvements (in %) of MGAN with other variants of GAN for wasserstein distance

	% Improvement MGAN over GAN	% Improvement MGAN over WGAN	% Improvement MGAN over CGAN	% Improvement MGAN over CLIMGAN
SST	+86.0	+82.1	+86.9	+72.2
pH	+86.6	+85.6	+85.6	+86.2
PSU	+3.2	+76.0	+83.2	+61.3
COD	+51.0	+26.3	-9.9	+30.2
N-N	+55.7	+51.4	+42.4	+32.7
POS	+58.1	+46.1	+41.9	+45.4

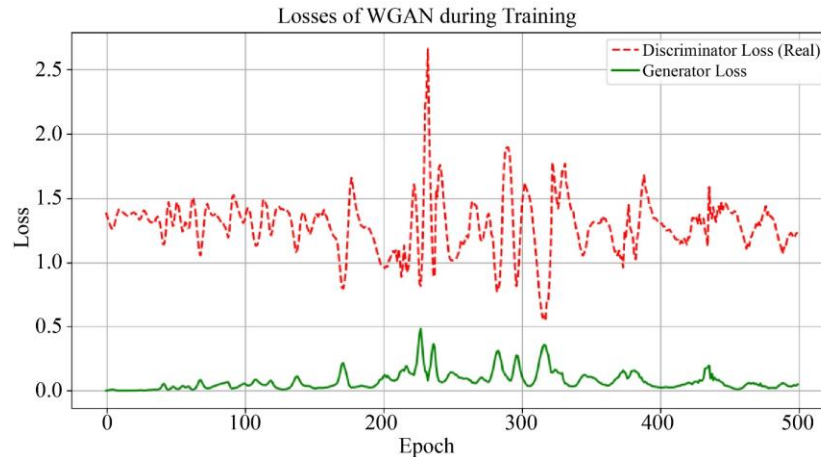


Fig. 19 Loss comparison of proposed MGAN

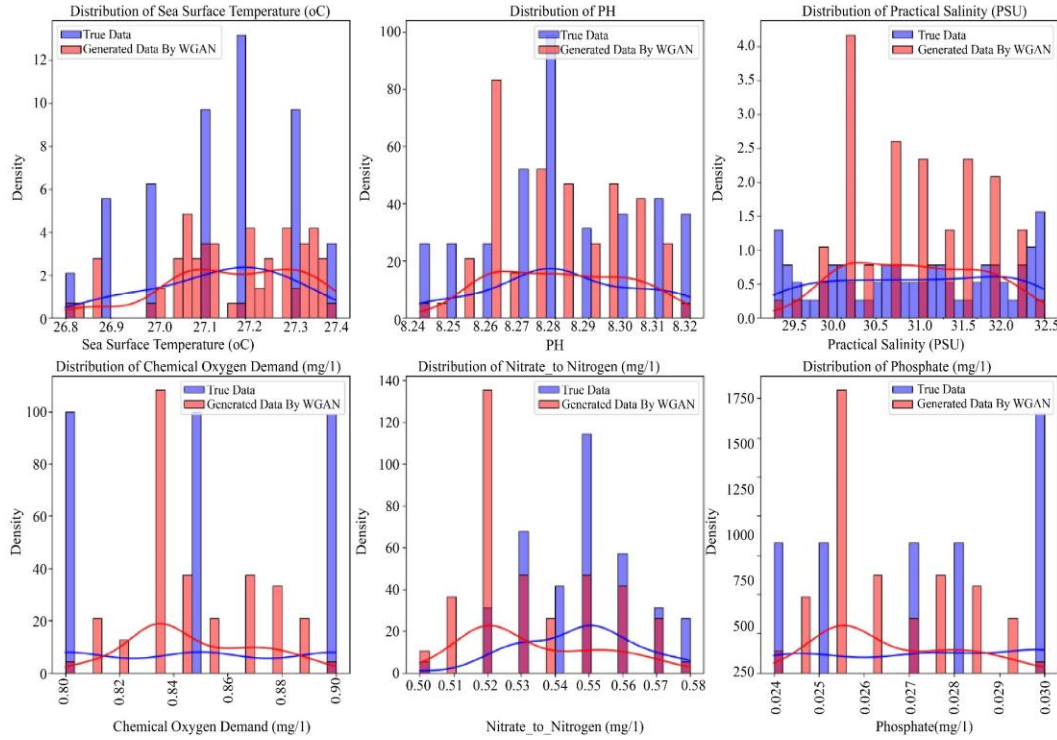


Fig. 20 Comparison of original data and generated data of the proposed MGAN

Parameter	Value
0	Hyperparameters {'BATCH_SIZE': 64, 'NUM_EPOCHS': 500, 'LEARNING_RATE': 0.0002, 'NOISE_DIM': 100, 'CONDITION_DIM': 6, 'OUTPUT_DIM': 7, 'train/test split': '50% train, 50% test', 'Data normalization': 'MinMaxScaler'}
1	Discriminator Architecture {'Input Layer': 13, 'Hidden Layers': '512, 256, 128 units', 'Activation Function': 'LeakyReLU (0.2 slope)', 'Output Layer': '1 unit (Sigmoid)'}
2	Generator Architecture {'Input Layer': 106, 'Hidden Layers': '128, 256, 512 units', 'Activation Function': 'ReLU', 'Output Layer': '7 units (Tanh)'}
3	Loss Functions {'Discriminator Loss': 'Binary Cross-Entropy (BCE)', 'Generator Loss': 'Feature Matching'}
4	Optimizers {'Optimizers': 'Adam', 'betas for Adam': '(0.5, 0.999)'}
5	Training Configuration {'Gradient Clipping': 'No', 'Random Noise for Generator': 'Gaussian noise (mean=0, std=1)', 'Batch Normalization': 'No'}

Fig. 21 Hyperparameters' settings of proposed MGAN

4.4.5. Generator and Discriminator Loss of MGAN

To monitor the training progression of the proposed MGAN, both the losses in Figure 19 of the generator and the critic were considered, and the quantized dispersion level of these losses was visualized as plots. From this Figure, the generator loss is relatively stable and low. Low fluctuations suggest that it is successfully minimizing its loss function, which is an indicator of stable training, and that MGAN creates realistic data. The discriminator loss is higher and fluctuates, which is common in GAN training.

These fluctuations indicate that the discriminator is adjusting to new data produced by the generator. Overall, the balance acknowledged in the losses suggests that both generator and discriminator are improving but still found to challenge each other, which is a sign of a healthy adversarial training process.

4.4.6. Generated Data of MGAN

Figure 20 is visualized to exhibit generated or synthetic data (represented in red color bars) by the proposed MGAN and its comparison with the actual physicochemical data (represented in blue color bars). The synthetic indicators successfully reflected both the distribution and variability of the real dataset. For instance, in some variables (like SST, pH, PSU, and Phosphate), the generated data aligned closely with the actual physiochemical data fed as input to MGAN, whereas in others, like N-N, there were found to be minor discrepancies. Overall, Figure 20 shows that the proposed MGAN efficiently learned the synthetic data that exhibited strong correspondence with the actual dataset for the variables SST, PH, and PSU. For these variables, the generated data replicated distribution patterns of the real instances and successfully captured values of density and variability of the feature. However, only for COD did the data generated over-

represent certain ranges of values. Thus, the well-matched distributions developed by the proposed MGAN indicated potential for the synthetic data to be used in model training for Deep RNN.

4.4.7. Parameters Settings of MGAN

The hyperparameters' settings for the novel MGAN are indicated in Figure 21.

Statistical Model Validation of Proposed Hybrid MGAN Model

To compare two empirical distributions without parametric assumptions, the Two-Sample Kolmogorov-Smirnov (KS) test was used, which analyzes the cumulative distribution function of each parameter [89]. Alignment between the two distributions improves as the KS statistics decrease. Moreover, the KS test's null hypothesis is that the distributions under comparison are identical, and this null hypothesis can be accepted if the p-value, the probability of observed data > 0.05 , is less than the threshold significance level. This study conducted KS tests for each water quality parameter (pH, SST, salinity, COD, nitrogen, and phosphate concentrations). This evaluation sought to confirm, through quantitative measures, the superiority of MGAN in generating samples that are statistically resembling the original environmental indicators. This study assumed the null

hypothesis (H_0) as that both the synthetic and actual data are taken from the same continuous distribution. The hypothesis was assessed using the KS test. The outcome of the KS statistics and associated p-value for each parameter (**p > 0.05) is stated in Table 4. The KS test is found to be statistically significant for SST, salinity, and nitrogen concentration, whereas for other factors, the p-value allows the study to fail to reject the null hypothesis. Thus, this test validates and proves the effectiveness and performance of the proposed MGAN.

Table 4. KS test results

Feature	KS Value Statistic	p-value
SST	0.2222	0.0569
pH	0.2639	0.0130
PSU	0.0972	0.8889
COD	0.3333	0.0006
N-N	0.1528	0.3722
POS	0.3333	0.0006

4.4.8. Maximum Mean Discrepancy (MMD)

MMD computes a statistic that quantifies the dissimilarity between two sets of data points drawn from different underlying distributions [90]. MMD is a kernel-based minimum distance estimator. A lower MMD score indicates the similarity between the two distributions.

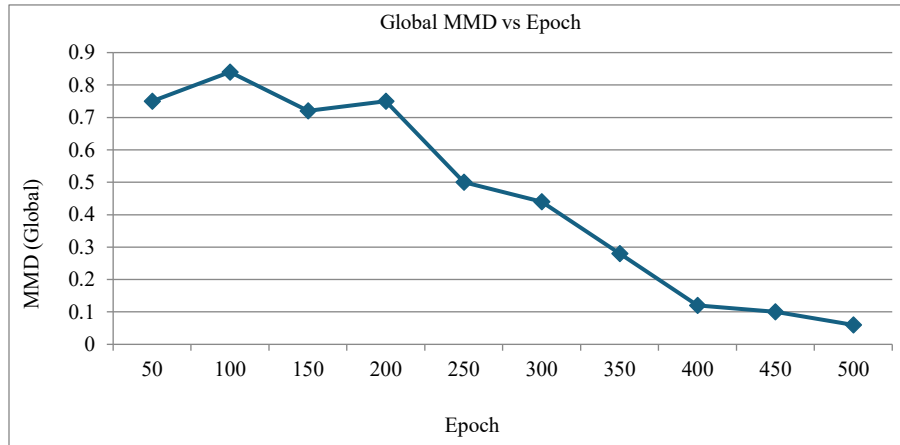


Fig. 22 MMD analysis

A global MMD analysis was performed in this work to assess the effectiveness of the proposed MGAN. MMD metric over 500 training epochs is illustrated in Figure 22. The MMD value exhibits a strong monotonic decrease, beginning at approximately 0.75 and converging to a stable value near 0.05 by epoch 500.

Out of which, for 300 to 400 epochs, the rate of decrease of the MMD value is drastic. Whereas for 400 to 500 epochs, the curve slowly approaches a minimum value, indicating that the model fine-tunes the derived distribution to better align with the finer details of the real data, thus proving the performance and robustness of the proposed MGAN.

4.9. Predictions by Deep RNN and Its Performance

This research used Deep RNN for the prediction of HC and fish assemblages upon the effects of SST, pH, PSU, and COD. With an objective of achieving improved predictions, modelling of Deep RNN was conducted in two phases as follows:

Phase 1: Model 1 - Predictions by basic Deep RNN for the real dataset of the Flic en Flac region of Mauritius

Phase 2: Model 2 - Predictions by hybrid MGAN-Deep RNN (Prediction by Deep RNN when using the MGAN-generated dataset)

4.9.1. Assessment of Predictive Models According to their Evaluation Metrics

A comparison of these two predictive models was carried out based on performance metrics and is shown as a heat map in Figure 23. These measures are multiple statistics of error calculation. For instance, if making large mistakes is undesirable, the Mean Square Error (MSE) is a superior statistic where a smaller value indicates a better result [91]. However, the R² metric is considered an exception; it quantifies the model's goodness-of-fit, representing the rate of variance with scores ranging from 0% (worst fit) to 100% (best fit). From the heat map (Figure 23) that illustrates the

values of evaluation metrics, it is identified that MGAN DeepRNN consistently shows lower Root Mean Square Error (RMSE) with a value of 0.21, which is << RMSE (of 5 in value) of basic Deep RNN. Also, values of MAE and MAPE suggest that the Deep RNN model enhanced with the use of MGAN for the generated dataset is better at making accurate predictions. Higher R² scores for MGAN DeepRNN also indicate that MGAN DeepRNN is more reliable in explaining data patterns. Thus, from observations on evaluation metrics, it is suggested that the hybrid MGAN–Deep RNN model performed better than the basic Deep RNN.

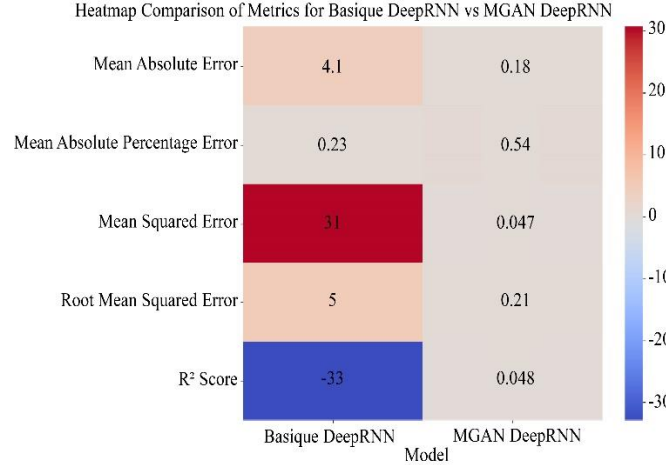


Fig. 23 Comparison of evaluation metrics of predictions of basic deep RNN and proposed MGAN Deep RNN

4.9.2. Comparison of Predictive Models based on Learning Curves

Learning curves plot the validation and training losses throughout epochs, and they determine the model's learning process. Figure 24 (a) and (b) depict the comparative results of learning curves of the baseline Deep RNN and the hybrid MGAN-Deep RNN.

As shown in Figure 24(a), the simple Deep RNN's loss decays gradually until the 50-epoch mark, indicating steady learning. Subsequently, the model enters a period of accelerated convergence, where the MSE plunges from 636.434 to 73.518 over the next 30 epochs. At the 100th epoch, the MSE was recorded as 35.5568, which indicates that the model is underfitting with high bias.

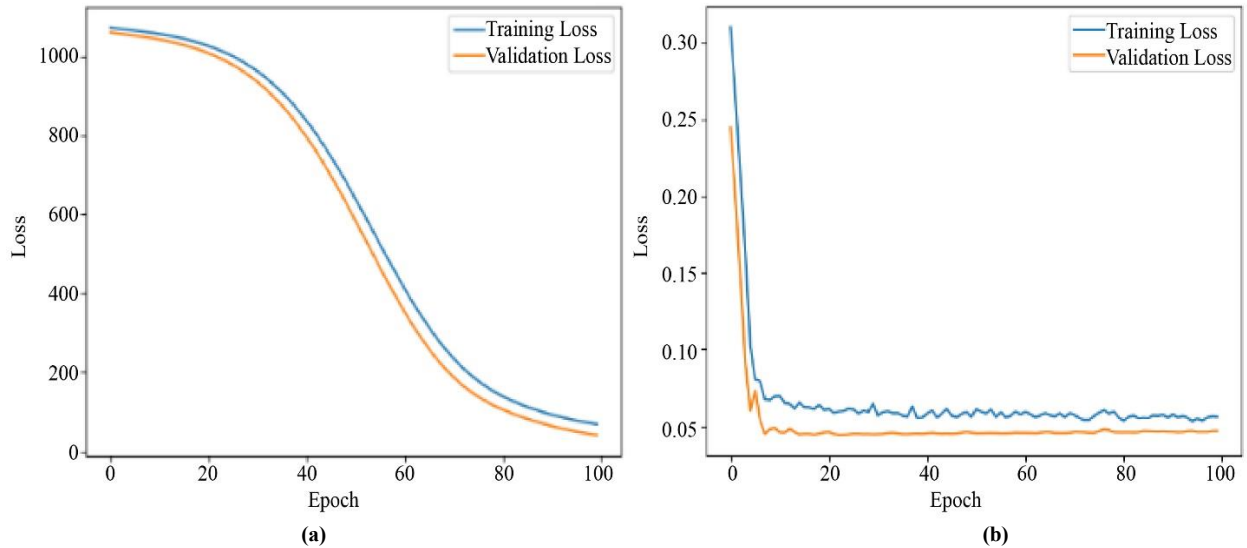


Fig. 24 (a) Learning curves of simple deep RNN, and (b) Learning curves of proposed MGAN Deep RNN.

It can be due to the linear RNN model with very few layers trying to fit a complex nonlinear function, or it can also be the model's ineffectiveness in forecasting outcomes due to poor input features. Figure 24(b) shows the loss curves for the proposed hybrid MGAN-Deep RNN model. The loss curves fall together, and the validation loss does not increase after a few epochs. This illustrates that there is no overfitting. Moreover, the validation loss closely tracked the training loss with little deviation, also following a declining trend. These losses being closer to each other (without much divergence) provides a clear sign that the model has generalized and avoided overfitting. Given that the model works well on unseen validation data, these curves demonstrate effective training with consistent improvement.

It is also observed that for the proposed hybrid MGAN-Deep RNN model, both training and validation curves stabilize around 50 epochs, indicating early convergence. But minimal values of MSE for the training: ~ 0.06 , and the validation: ~ 0.04 , are a good indication of generalization and better performance, demonstrating the superiority of the proposed hybrid model when compared to the benchmarked model.

Thus, MGAN-based augmentation has improved model efficacy and achieved enhanced generalization. This can be due to the i) Wasserstein distance used in MGAN to guarantee

meaningful gradients and improve realistic sample creation [92] and ii) leveraging conditional inputs (in MGAN, water quality parameters were used as condition inputs) to guide the data augmentation process [93].

4.9.3. Comparison of Predictive Models According to Test Loss

Performance assessment of a predictive model typically involves observing its test loss. The MGAN-Deep RNN achieved a very low test-loss of 0.046888 versus 35.5568 for the basic Deep RNN, indicating that the MGAN-Deep RNN model more accurately captures patterns in the unseen data.

The MGAN-Deep RNN yielded a significantly lower test loss, indicating a major improvement in performance compared to the basic Deep RNN. This significant reduction in test loss of the MGAN-Deep RNN model can be due to a refined MGAN for data augmentation prior to the prediction exercise, as indicated by Jouini et al. [94].

4.9.4. Comparison of Predictive Models based on Visualization of Real Vs Predicted Data

A comparison of real Hard Corals (HC) data and predicted values of hard corals by the two predictive models, namely basic Deep-RNN and MGAN-Deep RNN, can be made through the visualization, as shown in Figures 25 and 26.

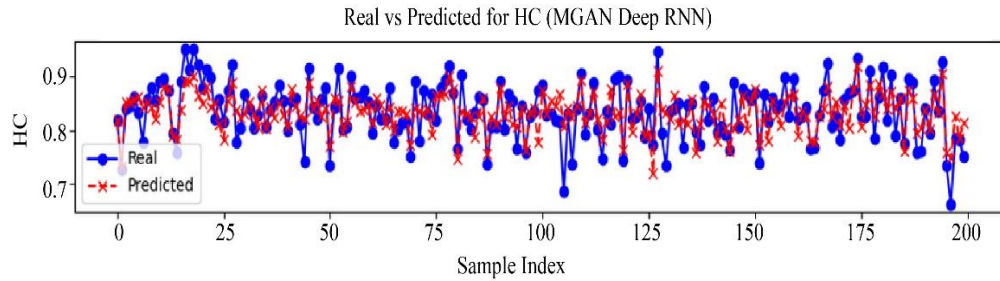


Fig. 25 Real Vs Predicted values of HC by MGAN-DeepRNN

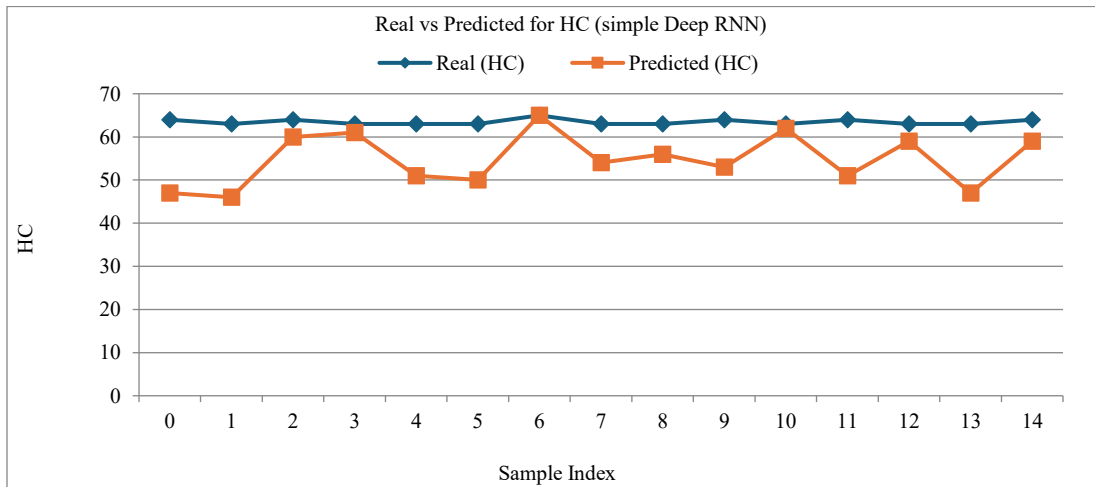


Fig. 26 Real Vs Predicted values of HC by basic DeepRNN

4.9.5. Sensitivity of the Model to Hyperparameter Tuning

The practical deployment of a predictive model often involves encountering data that deviates from the ideal training conditions.

Therefore, we evaluated its robustness and sensitivity by applying three distinct hyperparameter sets (set1, set2, set3) and executing the proposed MGAN-DeepRNN model for these parameters [95]. It then compared the performance of

these three MGAN-DeepRNNs with the default MGAN-DeepRNN. The metric RMSE was considered for comparison and is tabulated in the table below (Table 5). From Table 5, the RMSE variation of sets 1, 2, and 3 from the default model ranges from 0.0254 to 0.0319, which is minimal, thus proving that the proposed MGAN Deep RNN model is robust and stable. The following figure (Figure 27) visualizes this sensitivity test. The minor variation between the MSE values of different sets, as illustrated in Figure 27, proves the stability of the proposed MGAN Deep RNN model.

Table 5. Hyperparameter sensitivity test (for proposed model) through RMSE indications

	LSTM1 Units	LSTM2 Units	Dropout	Epochs	Batch Size	Learning Rate	RMSE
Set1	128	64	0.2	75	32	0.001	0.2354
Set 2	64	32	0.3	100	64	0.0005	0.2419
Set 3	256	128	0.1	75	16	0.001	0.2379
Default	128	64	0.2	50	32	0.001	0.2136

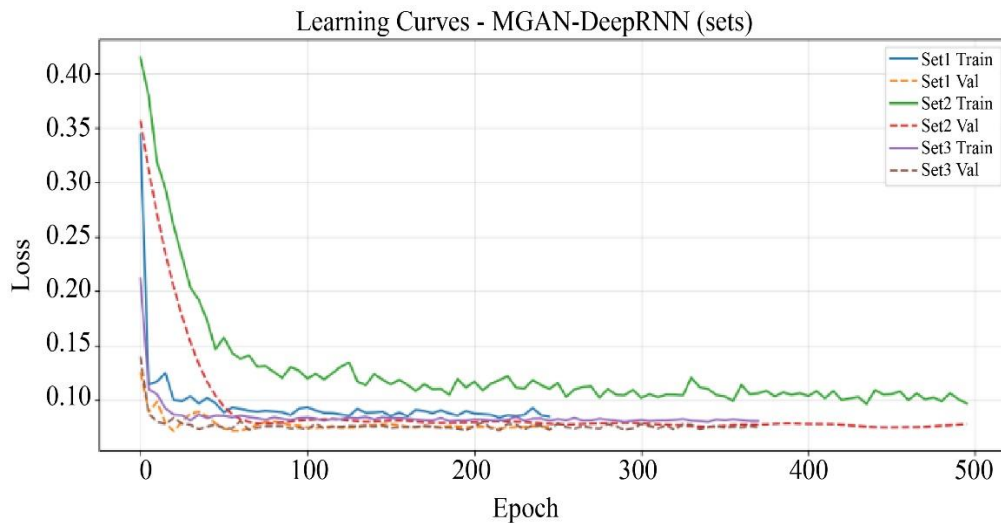


Fig. 27 Hyperparameter sensitivity test (for proposed model) using learning curves

4.9.6. Predictive Uncertainty Check of the Proposed Hybrid MGAN-DeepRNN Model using Monte Carlo Dropout

Monte Carlo (MC) dropout is a popular approach that can be used to represent model uncertainty in neural networks [96]. In this method, a dropout technique is used both in the training phase and during the inference phase. This dropout technique randomly masks the hidden units, which means that fractions of neurons and their connections are turned off.

After this masking, multiple forward passes are performed for the same input, every time using a distinct group of randomly dropped neurons. As an outcome of the training phase and inference phase, different sets of generated MC dropout outputs are obtained. These outputs are illustrated in Figure 28. The variability in these outputs is quantified using the mean as one of the descriptive statistics, while the other measure is the standard deviation. From these values, the model's prediction and its uncertainty can be estimated. The

Figure above Figure 28 exhibits true distribution (dotted lines), mean predictions (blue line), and their confidence intervals (shaded region of -1 standard deviation) obtained after implementation of MC dropout. It is found that the confidence intervals are narrow and 50% of the true values fall within $\pm 1\sigma$. This indicates marginally high confidence and certainty in prediction by the proposed model for benthic species such as HC, Pomacentridae fishes, Acanthuridae, Chaetodontidae, and Scaridae fishes. But the confidence intervals are broad for algae and Labridae fishes, which indicates that the proposed model shows underconfident/uncertain predictions. This can be improved by allowing the model to predict intervals by using both Mean Variance Estimation and MC Drop as proposed by [97]. From the findings of this predictive uncertainty test, this study emphasized the following practical implications: (i) Predictions with a high degree of confidence have been achieved for HC, and fishes such as Pomacentridae,

Chaetodontidae, and Acanthuridae. Regarding these benthic groups, the model is very reliable. Hence, in the future, if there is a significant population drop in the above groups, then an immediate investigation into water quality or coral health can be triggered. Moreover, any strategic and operational

decisions made for the above groups by marine coastal management can be implemented with great assurance. (ii) Meanwhile, predictions of fish Labridae and algae have low confidence, and this calls for a more intense field validation before a decision-making process.

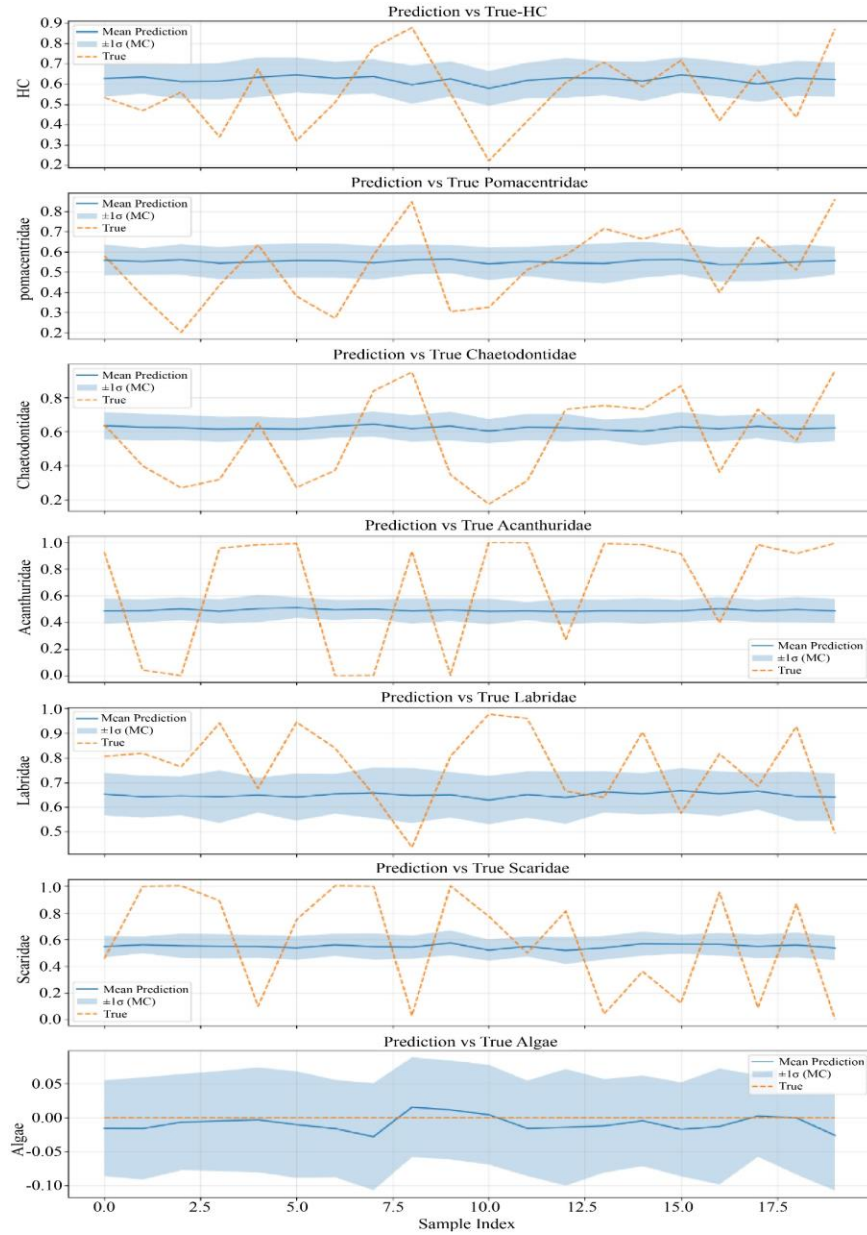


Fig. 28 Predictive Uncertainty Test (MC Dropout) of Proposed Model

4.10. Discussion on Prediction Outcomes of the Hybrid MGAN-DeepRNN

This study performed a robust analysis of feature significance of the predicted data employing (i) RF (Random Forest) Regressor (Figure 29), (ii) SHAP (SHapley Additive exPlanations) values method (Figure 30), (iii) Permutation Feature Importance with PLSRegression (Figure 31). This study proposed to use the first two methods based on the study

by [98]. This study preferred to use the PFI with PLSRegression with reference to [99], which reported that this method can be used for datasets that are high-dimensional and containing collinear predictors. The outcomes of the above three analyses are visualized as Figures 29, 30, and 31, respectively. From the figures, it is evident that nitrogen concentration, SST, salinity, and COD are prominent features that may influence the health of the benthic.

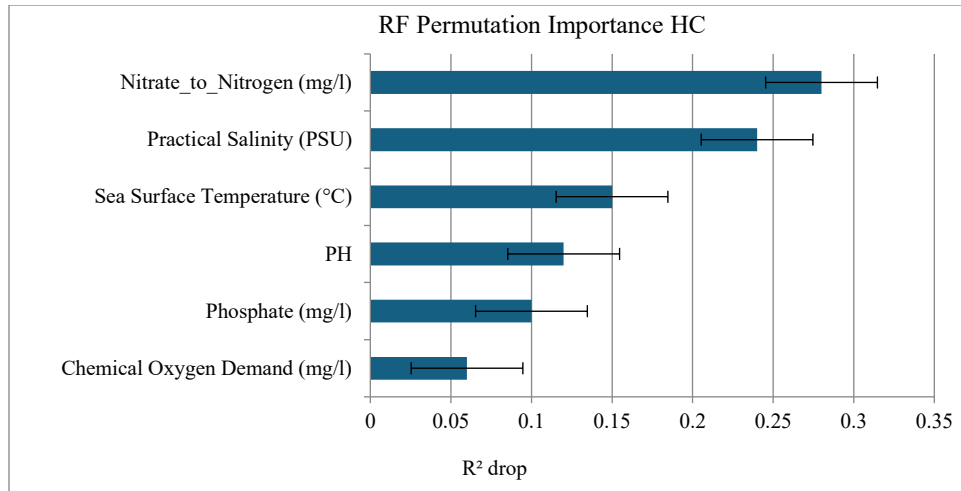


Fig. 29 Feature significance of predicted results employing RF regressor

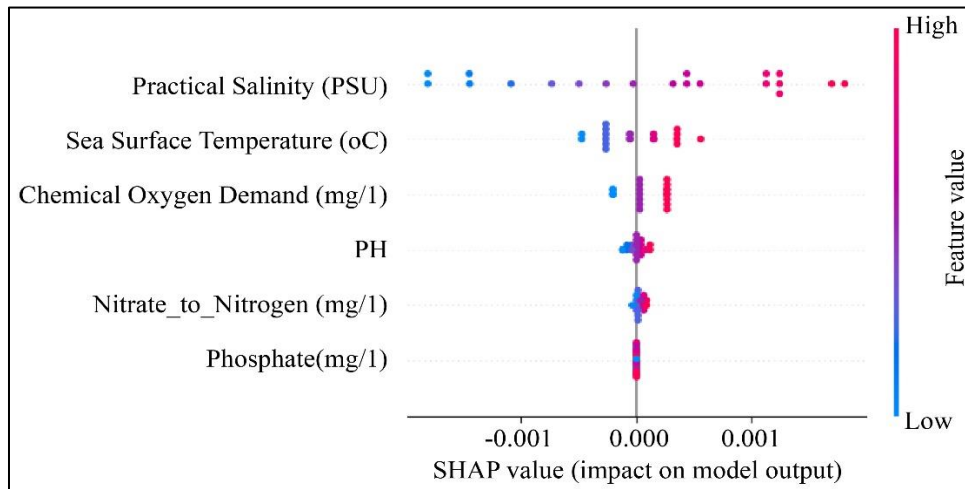


Fig. 30 Feature significance of predicted results by utilizing the SHAP values method

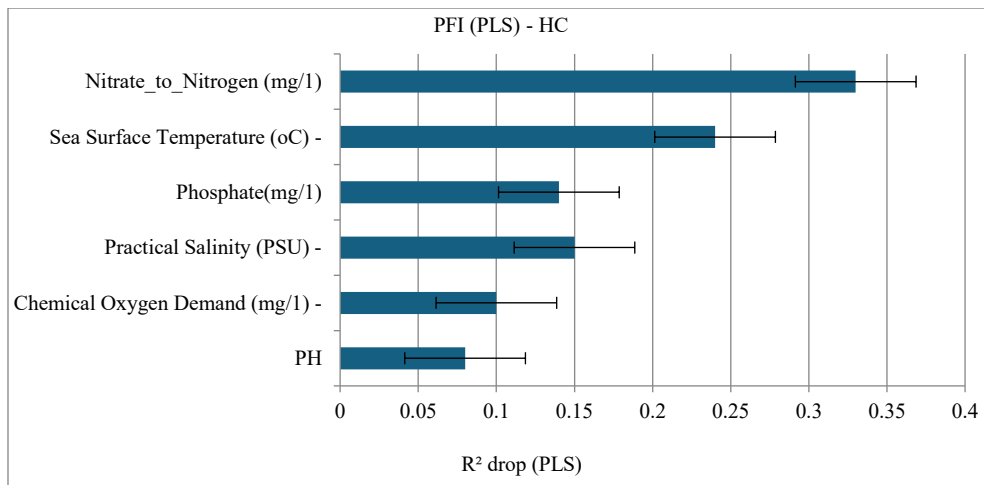


Fig. 31 Feature significance of predicted results by applying PFI with PLS regression

The feature importance analysis conducted on the predicted data indicated which environmental factors are most important in influencing the coral cover and fish forecasts

made by the hybrid model. From the findings of this analysis, the following practical implications are summarized: (i) The primary factor that impacts benthic health in the region under

consideration is water chemistry, particularly nitrogen levels/nutrient stress. Reducing nitrate pollution should be the top priority of marine coastal management plans to safeguard this region's coral reefs and fish, as existing literature suggests [100]. (ii) The second most important factor is SST.

Therefore, an increase or variation of SST over the nominal value (26°C to 30°C for tropical islands like Mauritius) may lead to the corals being more susceptible to bleaching because they may already be under nutrient stress. Reef restoration efforts can combat SST variations and promote benthic health [101]. (iii) Another significant property is salinity. Sharp variations of salinity can exert considerable influence on coral reefs. Heavy rainfall and freshwater flow from land, which can also carry sediments and contaminants, are common causes of low salinity. Therefore, one of the strategies as advised by the study [102] for coastal management might include the safeguarding and restoration of mangroves and coastal wetlands.

Informing effective marine management authorities requires accuracy in predictions as well as a transparent, comprehensible, and trustworthy understanding of the underlying drivers. By combining three different approaches, our study uses a strong, multifaceted feature importance analysis that aims to achieve this goal. It seeks to give strategic planners and decision-makers concise, useful information on the level to which water quality parameters (SST, pH, salinity, COD, nitrogen, and phosphate concentrations) influence changes in the marine ecosystem. The outcome of this analysis helps to identify priority areas for intervention, monitoring, and investment. These feature importance analyses are essential for building resilience and enhancing the sustainability of marine resources.

4.11. Summary on Novelty of the Research and its Relevance to Existing Research Findings

4.11.1. For Novel MGAN

Novel MGAN is a combined architecture of conditioning of GAN with four water quality parameters, namely SST, pH, salinity, and COD, followed by using WGAN architecture for achieving better training stability, thus generating high-quality data. The superiority of proposed MGAN over existing GAN, WGAN, CGAN and CLIMGAN (Table 3) is proven and it can be because the decision on conditioning parameters for novel MGAN was made based on the outcome of ecological analysis such as (i) R^2 Weighted Feature Importance analysis (Figure 11), (ii) quantification of environmental factors using descriptive statistics such as mean to median ratio, Shannon entropy, Shannon-weaver index and coefficient of variation (Table 2).

This study conducted Two-Sample Kolmogorov-Smirnov (KS) tests (Table 4) and MMD analysis Figure 22 to quantitatively confirm that the augmented data output by the proposed MGAN is statistically indistinguishable from the

input environmental data, thereby proving the statistical significance and robustness of the proposed MGAN over its variants.

4.11.2. For the Proposed Hybrid MGAN Deep RNN Model

The hybrid MGAN-Deep RNN model's ability to mitigate the fundamental challenge of data scarcity in the marine data set through its implementation of a novel MGAN is the main factor contributing to its performance improvement. The MGAN creates diverse and physically realistic synthetic samples by learning effectively the joint probability distribution of benthic communities and water quality indicators. Moreover, a Deep RNN is created by constructing an RNN with multiple layers. Because of this architectural advantage, the Deep RNN can capture hierarchical characteristics and more complex patterns in the data in addition to short- and long-term dependencies as explained in the literature review of this study. On the other hand, the basic Deep RNNs are compelled to learn from a small and possibly noisy dataset, which results in excessive variance (overfitting) and poor performance on unknown data, which moreover suffers from vanishing and exploding gradients due to its architecture. The sensitivity of the proposed MGAN Deep RNN model to hyperparameter tuning was determined by changing the hyperparameters three times and re-running the model. Comparison of the MSE values of the different runs of the MGAN-Deep RNN model illustrated minor variation between the MSE values of different sets Figure 26, thus emphasizing the stability of the proposed MGAN Deep RNN model.

Further, the study also carried out predictive uncertainty checks for the proposed MGAN-DeepRNN model using the Monte Carlo Dropout method Figure 27 to prove the proposed model's superiority.

The quality of the prediction data was further analyzed thoroughly by conducting feature importance analysis of predicted data using the following methods: (i) Random Forest Regressor Figure 28, (ii) SHAP values methods Figure 29 (iii) PFI-PLS Figure 30. These analyses have gained meaningful research insights whose findings can contribute as valuable inputs for strategic decision-making towards conservation of the benthic variables subjected to water quality parameters.

4.12. Limitations and Challenges

The results suggested that MGAN is a promising generative model, particularly for applications where capturing the overall distribution of the data is crucial. However, further investigation is needed to understand the specific reasons behind MGAN's strengths and weaknesses in different scenarios. This study recommends future research directions, such as refining the MGAN model for better accuracy or exploring other synthetic data generation methods for comparison.

5. Conclusion

The authors of this paper focused on a predictive analysis of marine data to contribute towards building the resilience of the marine ecosystem of the Flic en Flac region of the Republic of Mauritius. In this work, a detailed analysis was conducted on the prediction of the benthic, including hard corals and fish community, based on the effects of SST, pH, practical salinity of ocean water, dissolved oxygen, known as chemical oxygen demand, nitrates, and phosphate concentration in the ocean waters of the region under consideration. The novelty of this study includes the proven superiority of the proposed Marine Data GAN for the generation of data over other considered variants of GANs, the prediction of marine data under consideration using a basic Deep RNN, and comparing the predictions with MGAN-Deep RNN. The study compared the hybrid model with a basic Deep RNN based on its evaluation metrics, namely, MSE, MAE, RMSE, MAPE, and R^2 value, whose values demonstrated an improved prediction performance by the hybrid model. Importantly, despite the unbalanced dataset, this investigation successfully achieved improved marine data predictions using

the proposed novel hybrid model. The research as the way forward informs the marine data conservation of the Flic en Flac region about the predicted outputs, along with their evaluation metrics. Future research may determine the best Hyper-Performance Optimization (HPO) / tuning method for improving the predictions. Moreover, the research could further explore underlying mechanisms contributing to MGAN's strengths and weaknesses. Thus, by HPO of predictive models and by strengthening the MGAN, the researchers may arrive at valuable insights on marine data predictions, which would lead them to devise effective plans and strategies to mitigate the impact of stressors on marine benthic communities.

Acknowledgments

The researchers wish to appreciate the Ministry responsible for the country's marine resources, the Republic of Mauritius, for providing the independent variables, and MOI: The Mauritius Oceanography Institute, the Republic of Mauritius, for sharing the target variables required for this research.

References

- [1] Terry P. Hughes et al., "Spatial and Temporal Patterns of Mass Bleaching of Corals in the Anthropocene," *Science*, vol. 359, no. 6371, pp. 80-83, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] W.J. Skirving et al., "The Relentless March of Mass Coral Bleaching: A Global Perspective of Changing Heat Stress," *Coral Reefs*, vol. 38, no. 4, pp. 547-557, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] John M. Guinotte, and Victoria J. Fabry, "Ocean Acidification and its Potential Effects on Marine Ecosystems," *Annals of the New York Academy of Sciences*, vol. 1134, no. 1, pp. 320-342, 2008. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] M.S. Pratchett et al., "Effects of Coral Bleaching and Coral Loss on the Structure and Function of Reef Fish Assemblages," *Coral Bleaching*, pp. 265-293, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Scott C. Doney et al., "The Impacts of Ocean Acidification on Marine Ecosystems and Reliant Human Communities," *Annual Review of Environment and Resources*, vol. 45, pp. 83-112, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] John C. Davis, "Minimal Dissolved Oxygen Requirements of Aquatic Life with Emphasis on Canadian Species: A Review," *Journal of the Fisheries Research Board of Canada*, vol. 32, no. 12, pp. 2295-2332, 2011. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Laura Fernandes de Barros Marangoni et al., "Unravelling the Different Causes of Nitrate and Ammonium Effects on Coral Bleaching," *Scientific Reports*, vol. 10, no. 1, pp. 1-14, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Donald C. Behringer, Brian R. Silliman, and Kevin D. Lafferty, *Marine Disease Ecology*, Oxford University Press, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] J. Michael Beman, Kevin R. Arrigo, and Pamela A. Matson, "Agricultural Runoff Fuels Large Phyto plankt on Blooms in Vulnerable Areas of the Ocean," *Nature*, vol. 434, no. 7030, pp. 211-214, 2005. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Lord Wright, "The Impact of Over fishing on Marine Ecosystems," *Poultry, Fisheries & Wild life Sciences*, vol. 11, no. 1, 2023. [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Chaymae Boujnan "Sustainable Tourism Development and Coastal Management in the Marchica Lagoon (Eastern Moroccan Mediterranean Coast)," *2nd International Congress on Coastal Research (ICCR2023)*, vol. 502, pp. 1-4, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Ahmed Alietal., "Marine Data Prediction: An Evaluation of Machine Learning, Deep Learning, and Statistical Predictive Models," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, pp. 1-13, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Xuemei Lietal., "Prediction and Assessment of Marine Fisheries Carbon Sink in China Based on a Novel Nonlinear Grey Bernoulli Model with Multiple Optimizations," *Science of the Total Environment*, vol. 914, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Sahina Akter et al., "Deciphering the Distribution of Indian Mackerel, *Rastrelliger kanagurta* (Cuvier, 1817) along the North west Coasts of India," *Thalassas: An International Journal of Marine Sciences*, vol. 40, no. 3, pp. 1481-1493, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [15] Luciana C. Ferreira et al., "Predicting Suitable Habitats for Foraging and Migration in Eastern Indian Ocean Pygmy Blue Whales from Satellite Tracking Data," *Movement Ecology*, vol. 12, no. 1, pp. 1-20, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Yan Jiao et al., "Construction of Sea Surface Temperature Forecasting Model for Bohai Sea and Yellow Sea Coastal Stations Based on Long Short-Time Memory Neural Network," *Water*, vol. 16, no. 16, pp. 1-13, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Hailun He et al., "Forecasting Sea Surface Temperature During Typhoon Events in the Bohai Sea using Spatiotemporal Neural Networks," *Atmospheric Research*, vol. 309, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Simone Franceschini et al., "A Deep Learning Model for Measuring Coral Reef Halos Globally from Multispectral Satellite Imagery," *Remote Sensing of Environment*, vol. 292, pp. 1-12, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Nathaphon Boonnam et al., "Coral Reef Bleaching Under Climate Change: Prediction Modeling and Machine Learning," *Sustainability*, vol. 14, no. 10, pp. 1-13, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Shams Forruque Ahmed et al., "Deep Learning Modelling Techniques: Current Progress, Applications, Advantages, and Challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521-13617, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Yiwan Yue et al., "Three-Dimensional Spatio-Temporal Slim Weighted Generative Adversarial Imputation Network: Spatio-Temporal Slim Weighted Generative Adversarial Imputation Netto Repair Missing Ocean Current Data," *Journal of Marine Science and Engineering*, vol. 13, no. 5, pp. 1-28, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Stefan Wolff, Fearghal O' Donncha, and Bei Chen, "Statistical and Machine Learning Ensemble Modelling to Forecast Sea Surface Temperature," *Journal of Marine Systems*, vol. 208, pp. 1-36, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Annukka Lehtikoinen et al., "Evaluating Complex Relationships between Ecological Indicators and Environmental Factors in the Baltic Sea: A Machine Learning Approach," *Ecological Indicators*, vol. 101, pp. 117-125, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Peter Rubbens et al., "Machine Learning in Marine Ecology: An Overview of Techniques and Applications," *ICES Journal of Marine Science*, vol. 80, no. 7, pp. 1829-1853, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Richard E. Thomson, and William J. Emery, *Data Analysis Methods in Physical Oceanography*, 4th ed., 2025. [[Google Scholar](#)] [[Publisher Link](#)]
- [26] Ali Pourzangbar, Mahdi Jalali, and Maurizio Brocchini, "Machine Learning Application in Modelling Marine and Coastal Phenomena: A Critical Review," *Frontiers in Environmental Engineering*, vol. 2, pp. 1-27, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [27] S.G. Penny et al., "Integrating Recurrent Neural Networks with Data Assimilation for Scalable Data-Driven State Estimation," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 3, pp. 1-25, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [28] Nga Thanh Duong et al., "Prediction of Breaking Wave Height by Using Artificial Neural Network-Based Approach," *Ocean Modelling*, vol. 182, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [29] Hong Ye et al., "Light Recurrent Unit: Towards an Interpretable Recurrent Neural Network for Modeling Long-Range Dependency," *Electronics*, vol. 13, no. 16, pp. 1-18, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [30] Lalit Kumar, Mohammad Saud Afzal, and Ashad Ahmad, "Prediction of Water Turbidity in a Marine Environment using Machine Learning: A Case Study of Hong Kong," *Regional Studies in Marine Science*, vol. 52, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [31] Heng Shi, Minghao Xu, and Ran Li, "Deep Learning for House hold Load Forecasting-A Novel Pooling Deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271-5280, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [32] Debasmita Mishra et al., "Deep Recurrent Neural Network (Deep-RNN) for Classification of Nonlinear Data," *Advances in Intelligent Systems and Computing*, vol. 1120, pp. 207-215, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [33] The Importance of Accuracy in Machine Learning: A Comprehensive Guide, 2024. [Online]. Available: <https://www.artsyltech.com/blog/Accuracy-In-Machine-Learning>
- [34] Salim Rezvani, and Xizhao Wang, "A Broad Review on Class Imbalance Learning Techniques," *Applied Soft Computing*, vol. 143, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [35] Sanad Aburass, and Maha Abu Rumman, "Quantifying Overfitting: Introducing the Over Fitting Index," *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Sydney, Australia, pp. 1-7, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [36] Reza Moradi, Reza Berangi, and Behrouz Minaei, "A Survey of Regularization Strategies for Deep Models," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 3947-3986, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [37] Tufa Dinku, "Challenges with Availability and Quality of Climate Data in Africa," *Extreme Hydrology and Climate Variability: Monitoring, Modelling, Adaptation and Mitigation*, pp. 71-80, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [38] Francisco J. Moreno- Barea, José M. Jerez, and Leonardo Franco, "Improving Classification Accuracy using Data Augmentation on Small Data Sets," *Expert Systems with Applications*, vol. 161, pp. 1-14, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [39] Samuel Stocksieker, Denys Pommeret, and Arthur Charpentier, "Data Augmentation with Variational Autoencoder for Imbalanced Dataset," *International Conference on Neural Information Processing*, Auckland, New Zealand, pp. 354-370, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [40] Lamia Alhoraibi et al., “Generative Adversarial Network-Based Data Augmentation for Enhancing Wireless Physical Layer Authentication,” *Sensors*, vol. 24, no. 2, pp. 1-22, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [41] Liang duo Shen et al., “Long-Term Wave Height Forecasting using VMD-Informer,” *Anthropocene Coasts*, vol. 8, no. 1, pp. 1-12, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [42] Zhenyu Liu et al., “Forecast Methods for Time Series Data: A Survey,” *IEEE Access*, vol. 9, pp. 91896-91912, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [43] Abimbola Ogungbire, and Srinivas S. Pulugurtha, “Effectiveness of Data Imbalance Treatment in Weather-Related Crash Severity Analysis,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2678, no. 11, pp. 88-105, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [44] Carolina Gabarró et al., “Improving Satellite-Based Monitoring of the Polar Regions: Identification of Research and Capacity Gaps,” *Frontiers in Remote Sensing*, vol. 4, pp. 1-15, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [45] Murtadha D. Hssayeni, and Behnaz Ghoraani, “Deep Regression Modeling for Imbalanced and Incomplete Time-Series Data,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 6, pp. 3767-3778, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [46] Mariyam Irshada, and V. Kumar, “SMOTE and Extra Trees Regress or based Random Forest Technique for Predicting Australian Rainfall,” *International Journal of Information Technology*, vol. 15, no. 3, pp. 1679-1687, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [47] Elaheh Jafarigol, and Theodore B. Trafalis, “A Distributed Approach to Meteorological Predictions: Addressing Data Imbalance in Precipitation Prediction Models through Federated Learning and Gans,” *Computational Management Science*, vol. 21, no. 1, pp. 1-23, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [48] Artificial Intelligence: Generative AI could Raise Global GDP by 7%, Goldman Sachs, 2023. [Online]. Available: <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent.html>
- [49] Stefan Feuerriegel et al., “Generative AI,” *Business and Information Systems Engineering*, vol. 66, no. 1, pp. 111-126, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [50] Dorcas Oladayo Esan, Pius Adewale Owolawi, and Chunling Tu, *Generative Adversarial Networks: Applications, Challenges, and Open Issues*, Deep Learning-Recent Findings and Research, IntechOpen, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [51] Ya Wang et al., “Correcting Climate Model Sea Surface Temperature Simulations with Generative Adversarial Networks: Climatology, Interannual Variability, and Extremes,” *Advances in Atmospheric Sciences*, vol. 41, no. 7, pp. 1299-1312, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [52] Guiye Li, and Guofeng Cao, “Generative Adversarial Models for Extreme Geospatial Downscaling,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, PP. 1-12, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [53] Zhuang Li et al., “PSRGAN: Generative Adversarial Networks for Precipitation Downscaling,” *IEEE Geoscience and Remote Sensing Letters*, vol. 22, pp. 1-5, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [54] Ian Goodfellow et al., “Generative Adversarial Networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139-144, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [55] Yee-Fan Tan et al., “Graph-Regularized Manifold-Aware Conditional Wasserstein GAN for Brain Functional Connectivity Generation,” *Human Brain Mapping*, vol. 46, no. 12, pp. 1-16, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [56] Martin Arjovsky, Soumith Chintala, and Léon Bottou, “Wasserstein GAN,” *arXiv Preprint*, pp. 1-32, 2024. [[CrossRef](#)] [[Publisher Link](#)]
- [57] Tetsuya Saito, “TIOLI-GAN with Wasserstein-1 Distance and Structural Gradient Penalty: A Theory,” *Authorea Preprints*, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [58] Ishaan Gulrajani et al., “Improved Training of Wasserstein GANs,” *Advances in Neural Information Processing Systems*, vol. 30, 2024. [[Google Scholar](#)] [[Publisher Link](#)]
- [59] Camille Besombes et al., “Producing Realistic Climate Data with Generative Adversarial Networks,” *Nonlinear Processes in Geophysics*, vol. 28, no. 3, pp. 347-370, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [60] Jangho Lee, and Sun Young Park, “WGAN-GP-Based Conditional GAN (cGAN) with Extreme Critic for Precipitation Downscaling in a Key Agricultural Region of the Northeastern U.S.,” *IEEE Access*, vol. 13, pp. 46030-46041, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [61] Neelesh Rampal et al., “A Reliable Generative Adversarial Network Approach for Climate Downscaling and Weather Generation,” *Journal of Advances in Modeling Earth Systems (JAMES)*, vol. 17, no. 1, pp. 1-28, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [62] Jian Sha et al., “A Spatial Weather Generator Based on Conditional Deep Convolution Generative Adversarial Nets (cDCGAN),” *Climate Dynamics*, vol. 62, no. 2, pp. 1275-1290, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [63] Anis Bourou, Valérie Mezger, and Auguste Genovesio, “GANs Class-Conditioning Methods: A Survey,” *arXiv Preprint*, pp. 1-19, 2024. [[CrossRef](#)] [[Publisher Link](#)]

- [64] Yutian Pang, and Yongming Liu, "Conditional Generative Adversarial Networks (CGAN) for Aircraft Trajectory Prediction Considering Weather Effects," *AIAA Scitech 2020 Forum*, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [65] Alexandra Puchko et al., "DeepClimGAN: A High-Resolution Climate Data Generator," *arXiv Preprint*, pp. 1-5, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [66] Yifan Zhang et al., "A Novel Extreme Adaptive GRU for Multivariate Time Series Forecasting," *Scientific Reports*, vol. 14, no. 1, pp. 1-15, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [67] Milind Kolambe, and Sandhya Arora, "Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques," *Journal of Electrical Systems*, vol. 20, no. 2s, pp. 575-586, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [68] Muhammad Waqas, and Usa Wannasingha Humphries, "A Critical Review of RNN and LSTM Variants in Hydrological Time Series Predictions," *MethodsX*, vol. 13, pp. 1-19, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [69] Harsh Agarwal et al., "Predictive Data Analysis: Leveraging RNN and LSTM Techniques for Time Series Dataset," *Procedia Computer Science*, vol. 235, pp. 979-989, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [70] Akhil Sethia, and Purva Raut, "Application of LSTM, GRU and ICA for Stock Price Prediction," *Information and Communication Technology for Intelligent Systems*, pp. 479-487, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [71] Jiajing Wang et al., "Predicting Stock Market Trends Using LSTM Networks: Overcoming RNN Limitations for Improved Financial Forecasting," *Journal of Computer Science and Software Applications*, vol. 4, no. 3, pp. 1-7, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [72] Shuai Li et al., "Independently Recurrent Neural Network (IndRNN): Building a Longer and Deeper RNN," *arXiv Preprint*, pp. 1-11, 2018. [[CrossRef](#)] [[Publisher Link](#)]
- [73] Qingchun Guo, Zhenfang He, and Zhaosheng Wang, "Monthly Climate Prediction using Deep Convolutional Neural Network and Long Short-Term Memory," *Scientific Reports*, vol. 14, no. 1, pp. 1-17, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [74] Mohammad Mahdi Bejani, and Mehdi Ghatee, "A Systematic Review on Overfitting Control in Shallow and Deep Neural Networks," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 6391-6438, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [75] Razvan Pascanu et al., "How to Construct Deep Recurrent Neural Networks," *arXiv Preprint*, pp. 1-13, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [76] Mohcen El Makkaoui, Anouar Dalli, and Khalid Elbaamrani, "A Comparative Study of RNN and DNN for Climate Prediction," *2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST)*, Marrakesh, Morocco, pp. 1-6, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [77] Sundeep Raj, Sandesh Tripathi, and K.C. Tripathi, "ArDHO-Deep RNN: Autoregressive Deer Hunting Optimization based Deep Recurrent Neural Network in Investigating Atmospheric and Oceanic Parameters," *Multimedia Tools and Applications*, vol. 81, pp. 7561-7588, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [78] Sundeep Raj et al., "Hybrid Optimized Deep Recurrent Neural Network for Atmospheric and Oceanic Parameters Prediction by Feature Fusion and Data Augmentation Model," *Journal of Combinatorial Optimization*, vol. 47, no. 4, pp. 1-33, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [79] Owais Ali Wani et al., "Predicting Rainfall using Machine Learning, Deep Learning, and Time Series Models Across an Altitudinal Gradient in the North-Western Himalayas," *Scientific Reports*, vol. 14, no. 1, pp. 1-18, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [80] Denis Krivoguz et al., "Enhancing Long-Term Air Temperature Forecasting with Deep Learning Architectures," *Journal of Robotics and Control (JRC)*, vol. 5, no. 3, pp. 706-716, 2024. [[Google Scholar](#)] [[Publisher Link](#)]
- [81] Mauritius Weather, *Met Office*, 2025. [Online]. Available: <https://weather.metoffice.gov.uk/travel/holiday-weather/africa/mauritius-weather>
- [82] Tancredo Souza, *Permanova*, Advanced Statistical Analysis for Soil Scientists, Springer, Cham, pp. 1-13, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [83] Le Wang et al., "Review of Classification Methods on Unbalanced DataSets," *IEEE Access*, vol. 9, pp. 64606-64628, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [84] Wentong He et al., "Long-Tailed Metrics and Object Detection in Camera Trap Datasets," *Applied Sciences*, vol. 13, no. 10, pp. 1-14, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [85] Malek Benslama, and Hatem Mokhtari, "2-Shannon's Theorem in Quantum Data," *Compressed Sensing in Li-Fi and Wi-Fi Networks*, pp. 1-7, 2017. [[CrossRef](#)] [[Publisher Link](#)]
- [86] Michael A. Schillaci, and Mario E. Schillaci, "Estimating the Population Variance, Standard Deviation, and Coefficient of Variation: Sample Size and Accuracy," *Journal of Human Evolution*, vol. 171, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [87] Mihaela Oprea, "A General Framework and Guidelines for Benchmarking Computational Intelligence Algorithms Applied to Forecasting Problems Derived from a Man Application Domain-Oriented Survey," *Applied Soft Computing*, vol. 89, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [88] Yingxu Wang et al., "Spectral Clustering based on JS-Divergence for Uncertain Data," *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, AB, Canada, pp. 1972-1975, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [89] K. Ramakrishna Kini et al., "Enhanced Data-Driven Monitoring of Waste Water Treatment Plants using the Kolmogorov-Smirnov Test," *Environmental Science: Water Research and Technology*, vol. 10, no. 6, pp. 1464-1480, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [90] Kensuke Mitsuzawa, "Word Sense Detection Leveraging Maximum Mean Discrepancy," *arXiv Preprint*, pp. 1-12, 2025. [[CrossRef](#)] [[Publisher Link](#)]
- [91] Guy S. Handelman et al., "Peering into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods," *American Journal of Roentgenology*, vol. 212, no. 1, pp. 38-43, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [92] Tanujit Chakraborty et al., "Ten Years of Generative Adversarial Nets (Gans): A Survey of the State-of-the-Art," *Machine Learning: Science and Technology*, vol. 5, no. 1, pp. 1-36, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [93] Kai Lang, and Yonghua Li, "Task-Aware Conditional GAN with Multi-Objective Loss for Realistic and Efficient Industrial Time Series Generation," *Journal of Big Data*, vol. 12, no. 1, pp. 1-21, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [94] Oumayma Jouini et al., "Wheat Leaf Disease Detection: A Light Weight Approach with Shallow CNN based Feature Refinement," *AgriEngineering*, vol. 6, no. 3, pp. 2001-2022, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [95] Pedro Mendes, Paolo Romano, and David Garlan, "Hyper-Parameter Tuning for Adversarially Robust Models," *arXiv Preprint*, pp. 1-8, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [96] Tal Zeevi et al., "Monte-Carlo Frequency Dropout for Predictive Uncertainty Estimation in Deep Learning," *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, Athens, Greece, pp. 1-15, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [97] Cristián Serpell et al., "Probabilistic Forecasting Using Monte Carlo Dropout Neural Networks," *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Havana, Cuba, pp. 387-397, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [98] Meredith G.L. Brown et al., "Random Forest Regression Feature Importance for Climate Impact Pathway Detection," *Journal of Computational and Applied Mathematics*, vol. 464, pp. 1-38, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [99] Jean-Pascal Matteau, Pierre-Luc Chagnon, and Paul Célécourt, "Partial Least Squares Regression to Explore and Predict Environmental Data," *Intelligence Systems for Earth, Environmental and Planetary Sciences*, pp. 1-32, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [100] Jithya Wijesinghe, Shermila M. Botheju, and Nalika R. Dayananda, "Coastal and Marine Pollution from Agricultural Activities," *Coastal and Marine Pollution from Agricultural Activities*, pp. 89-110, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [101] Edwin A. Hernández-Delgado, and Yanina M. Rodríguez-González, "Runaway Climate Across the Wider Caribbean and Eastern Tropical Pacific in the Anthropocene: Threats to Coral Reef Conservation, Restoration, and Social-Ecological Resilience," *Atmosphere*, vol. 16, no. 5, pp. 1-45, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [102] T.K. Teena Jayakumar, and Uttam Kumar Sarkar, "Habitat Degradation in Coral Reef Ecosystems and Mangroves: Current Status and Management Measures, Sustainable Management of Fish Genetic Resources, Springer, Singapore, pp. 111-149, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]