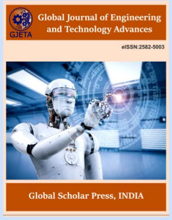
GJETA

Global Journal of Engineering and Technology Advances

eISSN: 2582-5003

Cross Ref DOI: 10.30574/gjeta

Journal homepage: https://gjeta.com/



(RESEARCH ARTICLE)

Check for updates

Classification of risk factors of climate change on infectious diseases using Bio-Inspired Algorithms

Kingsley Udonyah * and Imeh Umoren

Department of Computer Science, Akwa Ibom State University, Mkpat Enin, Nigeria.

Global Journal of Engineering and Technology Advances, 2023, 17(01), 029–044

Publication history: Received on 18 August 2023; revised on 30 September 2023; accepted on 03 October 2023

Article DOI: https://doi.org/10.30574/gjeta.2023.17.1.0201

Abstract

A new index of investigative approaches is being developed to help identify emerging infections and to detect the increased risks factors of infectious disease (IDs) occurrences that are expected to occur with climate change. The impacts of Climate Change with direct effects on human health through climate extremes and indirectly through infectious diseases are enormous. Hence, it requires study to determine correlations between Climate Change and Infectious Diseases using clinically validated data to improve behavioural health plans and early warning of public health risks. This work employs Meteorological and Infectious Disease data from the Figshare Data Repository, NiMET, NIH, WHO and literature, processes these data using SciKit Learn Preprocessing Package in Python and characterizes the risk factors of Climate Change Features: Humidity, Minimum Temperature, Maximum Temperature and Rainfall on Infectious Disease: Malaria, Pneumonia and Diarrhea using Nature-Inspired Algorithm using Artificial Neural Network (ANN) and Random Forest (RF) Algorithm with the R Statistical Programming Language. The work adopts Design Science Research (DSR) Methodology to analyze and classifies the risk factors of climate change on infectious diseases as well as evaluates the performances of selected Nature-Inspired Algorithms in classifying selected climatic factors and associated impacts on infectious diseases. Results obtained demonstrated that the RF performed better than ANN Algorithms with 96.9% and 95% accuracies respectively. Both models indicated that Rainfall and Temperature variations were common risks factors that indicated highest weights impacting the emergence and incidence of Infectious Diseases in Nigeria.

Keyword: Artificial Neural Networks; Climate Change; Infectious Diseases; Nature-Inspired Algorithms; Random Forests

1. Introduction

Within the last six years, the World Health Organization (WHO) have identified over 1000 epidemics of infectious diseases and issued warnings that Infectious Diseases (IDs) were spreading more swiftly than had previously been recorded and that novel infectious diseases were emerging at higher rates than at any other time in the history of man [1], [47]. As recently as July 23, 2022, the World Health Organization (WHO) made their highest level of alert, declaring monkeypox as a public health emergency of international concern (PHEIC) [2]. Essentially, IDs have become a major global driver of morbidity and mortality [3] with a huge burden on public health and the economic solidity of global communities [1], [4] and evidences are replete that infectious diseases have plague humanity throughout history [47] and in many cases have shaped human history [5],[6]. Several studies are unanimous in affirming Infectious Diseases as the leading causes of disability and death globally [1] and at the dawn of the 21st century, IDs have become responsible for one out of four deaths worldwide thus causing about 10 million deaths per year [7]. In 2019, Diarrheal Diseases and Lower Respiratory Infections were ranked among the top ten causative agents of deaths globally by the World Health Organization (WHO) [8]. This trend is a lot worse, especially in low-income generating countries of the world [1] with Fact Sheets published by WHO in 2020 showing huge incidences and impacts of Infectious Diseases in

* Corresponding author: Kingsley Udonyah

low-income countries such as Nigeria. These, coupled with past cases have resulted in researchers exploring the possibility of Climate Change having an impact on Infectious Diseases [9].

Climate change has been identified as responsible for climatic extremes and has also been pinpointed by WHO as one of the greatest threats to human health, particularly in this 21st century [10], [11], [12]. With the increase in climate change [11], coinciding with the rapid emergence of new infectious diseases [1], the research to find the possibilities of climate change having some sort of influence on the emergence or adaptation of infectious diseases are on the ascendency. The results from these researches indicates that climate change have some significant impacts on infectious diseases [11], [13]. Based on these evidences, [14] posits that "the crisis of emerging diseases will continue so long as global climate change occurs" and considering that an increasingly growing number of Infectious health outcomes have been associated with climate change [15], [16], this assertion is not misplaced and so are works to establish this fact. The impacts of Climate Change are likened to a double-edged sword having direct effects on human health through climate extremes [10], [11], [12], [16] and indirect effects through infectious diseases [16]. Several of the infectious diseases that have been documented are climate sensitive and the records show that climatic conditions have facilitated vector-borne disease outbreaks [17], [16].

Given that climate change is a naturally occurring phenomenon partly orchestrated by the activities of man, such as use and burning fossil fuels, pollution and deforestation [11], [15], [47], there is need to characterize the risk factors and analyze the impacts of climate change on Infectious Diseases using nature-inspired [18] optimization models that are collectively termed as Bio-Inspired Algorithms [19], [20], [21], [22], [19] as we seek to contribute our quota in finding ways to reduce the impacts of climate change and infectious diseases. Bio-Inspired Algorithms [23] have become very popular in solving optimization [22], [24], [19], [25] and classification [26], [18] problems and according to [18], they have become the "need of the hour for addressing complex real-world issues more efficiently and rapidly." Based on their applicability and efficiency, they have been applied to various complex problem domains with very commendable results recorded [27] so far. The success stories of Bio-Inspired Algorithms (BIAs) indicates that their applications to the problem space of analysing and classifying the impacts of Climate Change on Infectious Diseases is not out of place. Based on these evidences, we have sourced published data modelling correlation between climate change features: rainfall, humidity and temperature and infectious diseases prevalent in Nigeria: malaria, pneumonia and diarrhea and apply two robust nature inspired algorithms: Random Forest Algorithms (RF) and Artificial Neural Networks (ANN) to the classification and characterization of the risk factors based on their impacts on these infectious diseases.

2. Literature Review

2.1. Climate Change and Infectious Diseases

It has been established that several infectious disease pathogens, vectors and zoonotic agents, are strongly influenced by climatic factors including extreme whether phenomena, temperature variability and precipitation [28] which impact upon infectious diseases. Climate change causes variability in the climate system altering the atmosphere, the biogeochemical cycles (Carbon cycle, Nitrogen cycle and Hydrological cycle), the land surface, ice and both the biotic and abiotic components of the planet earth [29]. And given that Infectious Diseases, their causative or transmission agents and their habitats constitute part of the biotic and abiotic components, it is a valid hypothesis that climate change exudes some impacts on infectious diseases. Studies have shown that some common Infectious Diseases are climate sensitive [30]. For instance, two-thirds of Infectious Diseases pathogens that affect humans and domestic animals in Europe are climate sensitive [16], a snippet of the reality of climate change's impact on infectious diseases. Further evidences show that the environment is adversely affected by extreme weather conditions [31] thus affecting the Infectious Disease pathogens, vectors and host that live and interact in these environment [32]. All these evidences indicate that climate change directly or indirectly impacts upon infectious diseases. Climate Change causes changes in the environment resulting in alterations in the interactions among infectious disease hosts, reservoirs, vectors and pathogens [33]. There are evidences that climate change has some effects on malaria, arbovirus diseases such as dengue fever, parasitic and viral diseases such as Rift Valley Fever, Japanese Encephalitis, human African trypanosomiasis and leishmaniasis [34]. The United States Department of Health and the Human Services arm of the Center for Disease Control and Prevention (CDC) in September 2, 2020 published Fact Sheets indicating that changes in insect and arthropod ranges due to climate change have increased with a surge in venomous insect stings recorded. Also, changes in air quality, with increased levels of air pollution and pollen affecting respiratory health has been attributed to Climate Change and these affect the general behaviour of airborne infectious diseases [35]. Similarly, climate change has also been found to have significant consequences on human health, including increased risk of waterborne diseases such as cholera. Climate Change has been found to strongly affect the geographical distribution of insect vectors, and with the continued severity of extreme climate factors, these distributions are rapidly changing [34]. Infectious Disease Pathogens depend on certain environmental conditions to thrive and cause disease thus, alterations to the environment due to extreme weather events arising from climate change has the potential to support the survival of these pathogens and particularly increase their activities as well as the frequency at which they cause diseases and the severity of the diseases they cause [32]. Vector-borne and zoonotic infectious diseases have also been found to be directly influenced by climate change because, the very existence of vectors, such as mosquitoes, and pathogens, such as bacteria, viruses or parasites depend on certain climatic conditions [36] and with the above, the argument for infectious disease being impacted upon by climate change gains more momentum.

2.2. Risk Factors of Climate Change

Climate Change; unanimously recognized as one of the greatest challenges to humanity [37], [10], [30] has the potential to cause significant effects on humans, societies and natural systems, causing significant increases in temperature and changes in precipitation patterns, rising sea levels, and increased frequency and intensity of extreme weather events [31]. These changes are predicted to result in a wide range of risk factors that can have negative impacts on human health, ecosystems, world economies as well as the overall behaviours of infectious diseases. Climate change has been a great concern to scientists and researchers for more than half a century now [38], [39] due to its negative impacts across several key spheres of human endeavours to the extent that it has been predicted to impact upon humanity's psychological wellbeing [40]. Risk Factors of Climate change are documented to affect physical health, mental health, social relations [40] and particularly infectious diseases [10] arising from exposure to extreme whether events—including extreme rainfall, temperature variability, floods etc. With the unprecedented rate at which global warming is increasing over the past millennium [10], risk factors such as heat waves, droughts, floods, and storms have become more prevalent and they are among the most significant risk factors associated with climate change including extreme rainfall, temperature variabilities and strong winds. These events have a range of impacts on human health, including increased risk of heat-related illnesses, waterborne diseases, and injuries from flooding and landslides. Such is the impacts of climate Change that there continue to be projected increases in annual temperatures, and rainfall with these extreme weather events predicted to have both direct and indirect repercussions on the environment [32] as well as organisms in that environment. For instance, the incidence of malaria is high in South-South Nigeria because of the heavy rainfall [46] that results in floods that in turn sponsors the breeding of mosquitoes that cause malaria and with the overcrowding in Emergency Departments in many Nigerian healthcare facilities [45], mitigating the risks of infectious diseases presents the most optimal pathway.

2.3. Risk Modelling and Risk Factors Characterization

Several factors have been identified to play a part in the emergence, reemergence and incidence of infectious diseases, however, in recent times, climate change has become a major driver for infectious diseases and understandably drawing serious attention from governments and health professional and eliciting great research interests in the process and grouping these factors based on their characteristics and impacts is a direction worth considering. Characterization of the Risk Factors of Infectious Diseases is an approach that will prove effective in informing the general public and mitigating the effects of infectious diseases. Risk modelling is increasingly being used to inform public health actions to prevent, detect and mitigate climate change's impacts on infectious diseases [13]. The whole process of characterization begins with the processing of open source data beginning with extracting, analyzing and structuring information on what happened (diseases, its emergence and incidence characteristics), the location or region where it happened, the person or people it happened to and the time or season it happened [13]. Generally, risk modelling is a process for detecting and characterizing factors that increase an individual or a population's susceptibility to contracting a disease [13]. In the study of climate change's impact on infectious diseases, risk modelling detects and characterizes factors that increases possibilities for infectious diseases to be affected by climatic parameters. Using additional explanatory variables such as climate and meteorological data that outlines the presence, transmission and distribution of pathogens with open-source data for modelling has improved risk modelling in general. These further highlights why employing risk modelling and risk factors characterization in the management of infectious diseases is important. Adopting the risk modelling approach makes it possible to estimate if an infected population will become epidemic, and to characterize the prevalence of a disease over time [13].

3. Methodology

This work is based on the Design Science Research Methodology proposed by [41] and used by [42] and provides clarity on how research findings should be made, reported as well as the procedures adopted for data collection, literature analysis and review. The DSR methodology, popularized in 2019, is an incremental process that begins with the problem identification stage and progressively build artefacts that provide applicable solutions using methods and knowledge that have been defined by experts and universally accepted within that field [41], [42].

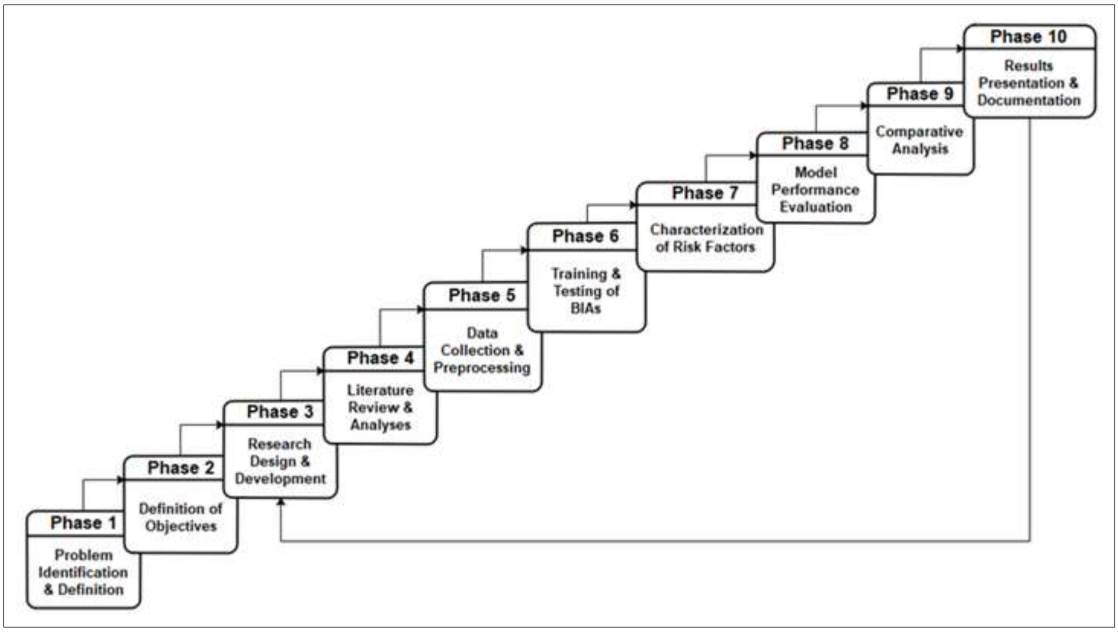


Figure 1 Design Science Research (DSR) Methodology

3.1. Model Considerations and Implementation

We adopted and implemented ***two*** Nature-Inspired Algorithms, both of which are machine learning techniques that are renowned for classification problems. With these, we classified and then characterized the risk factors of three climate change features: *mean temperature (structured as minimum and maximum temperature)*, *average rainfall* and *mean humidity* on the three infectious diseases: malaria, pneumonia and diarrhea using the SOMCH dataset. The Nature-Inspired Algorithms: *Artificial Neural Network (ANN)* and *Random Forest Algorithm* were chosen for this work based on their performance for classification tasks [43]. Using the preprocessed data, we implemented the models using the R Statistical Programming Language in RStudio. The design and implementation of the models could be summarized into the following steps.

- Preprocess (clean, normalize, scale) Dataset
- Pre-test six (6) Nature/Bio-Inspired Algorithms
- Pick two most applicable Algorithm based on Interpretability and Accuracy.
- Chose model coding tool: R Statistical Programing Language.
- Installed model libraries
- Implement the Proposed framework
- Structure the data to fit into the model
- Feed data into model for splitting on a ratio of 80:20
- Train models using the training data
- Test models using the testing data
- Fine tune models by altering model parameters
- Save trained models state
- Visualize model Results
- Evaluated and document results

3.2. Data Recording

Climate and Infectious Disease data, Bangladesh from the Figshare Data Repository published by Chowdhury in 2018 were obtained for this classification task. We focused on establishing a correlation between climate change risk factors and infectious diseases, particularly in Nigeria therefore, we obtained some climate data from the Nigerian Meteorological Agency (NiMET) which had windspeed, humidity, daily rainfall, minimum and maximum temperature as features for a 10-year period. However, this data lacked a target feature or output class (i.e. diseases) with which the features had some form of correlation. Considering this limitation, a new set of numerical climate dataset with recorded impact correlation to infectious diseases was necessary. The new Climate and Infectious Disease dataset was realized from an observational and exploratory study done at the Sylhet M.A.G. Osmani Medical College Hospital (SOMCH) and published in FigShare by [44]. The data examined the correlation existing between Temperature, Humidity and Rainfall on six infectious diseases (Malaria, Enteric Fever, Encephalitis, Diarrheal Disease, Pneumonia and Meningitis) based on clinical and laboratory diagnosis. The focus was to find a comprehensive perspective on the trend of these infectious diseases and their possible relation to climate change in Bangladesh [44]. The data indicated and supported a relationship between weather patterns and disease incidence thus providing a crucial reference point data for future studies. The infectious diseases classes were listed as climate sensitive based on the report of the IPCC and WHO in 2014 and 2016 [44]. In the Nigerian context, malaria, diarrhea and pneumonia were found to be the infectious diseases with increasing incidences in the rainy season, indicating a very high correlation to rainfall [44] and these were selected as the classes of diseases for this work.

```
     s_no           age              sex             season          disease           s_year          Jan_celc
 Min.   :   1   Min.   :12.00   Min.   :1.000   Min.   : 1.00   Min.   :1.000   Min.   :1.000   Min.   :16.80
 1st Qu.: 476   1st Qu.:22.00   1st Qu.:1.000   1st Qu.: 4.00   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:16.80
 Median : 951   Median :35.00   Median :1.000   Median : 7.00   Median :3.000   Median :4.000   Median :18.40
 Mean   : 951   Mean   :37.43   Mean   :1.422   Mean   : 7.12   Mean   :3.552   Mean   :3.455   Mean   :18.12
 3rd Qu.:1426   3rd Qu.:50.00   3rd Qu.:2.000   3rd Qu.:10.00   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:18.50
 Max.   :1901   Max.   :90.00   Max.   :2.000   Max.   :12.00   Max.   :6.000   Max.   :5.000   Max.   :19.30
 NA's   :3133
    Feb_celc        Mar_celc        Apr_celc        May_celc        June_celc       July_celc       Aug_celc
 Min.   :19.00   Min.   :23.80   Min.   :25.40   Min.   :26.80   Min.   :26.90   Min.   :27.60   Min.   :27.50
 1st Qu.:21.00   1st Qu.:24.50   1st Qu.:25.80   1st Qu.:26.80   1st Qu.:26.90   1st Qu.:28.10   1st Qu.:28.10
 Median :21.00   Median :24.50   Median :26.70   Median :26.90   Median :27.40   Median :28.10   Median :28.10
 Mean   :20.96   Mean   :24.88   Mean   :26.47   Mean   :27.19   Mean   :27.55   Mean   :28.24   Mean   :28.15
 3rd Qu.:21.30   3rd Qu.:25.90   3rd Qu.:27.00   3rd Qu.:27.50   3rd Qu.:28.10   3rd Qu.:28.50   3rd Qu.:28.40
 Max.   :21.90   Max.   :25.90   Max.   :27.40   Max.   :27.70   Max.   :28.10   Max.   :28.70   Max.   :28.50

    Sep_celc       Oct_celc        Nov_celc        Dec_celc        Jan_Hmd        Feb_Hmd        Mar_Hmd
 Min.   :27.7   Min.   :26.30   Min.   :22.60   Min.   :19.30   Min.   :72.0   Min.   :59.0   Min.   :59.00
 1st Qu.:27.7   1st Qu.:26.90   1st Qu.:22.60   1st Qu.:19.50   1st Qu.:73.0   1st Qu.:59.0   1st Qu.:59.00
 Median :28.2   Median :27.50   Median :23.60   Median :19.70   Median :74.0   Median :64.0   Median :61.00
 Mean   :28.2   Mean   :27.12   Mean   :23.38   Mean   :19.82   Mean   :74.7   Mean   :63.7   Mean   :62.71
 3rd Qu.:28.5   3rd Qu.:27.60   3rd Qu.:24.00   3rd Qu.:20.00   3rd Qu.:77.0   3rd Qu.:64.0   3rd Qu.:64.00
 Max.   :28.6   Max.   :27.60   Max.   :24.00   Max.   :20.70   Max.   :77.0   Max.   :69.0   Max.   :73.00

    Apr_Hmd         May_Hmd         Jun_Hmd         Jul_Hmd         Aug_Hmd         Sep_Hmd         Oct_Hmd
 Min.   :69.00   Min.   :77.00   Min.   :84.00   Min.   :83.00   Min.   :84.00   Min.   :82.00   Min.   :77.00
 1st Qu.:69.00   1st Qu.:78.00   1st Qu.:85.00   1st Qu.:85.00   1st Qu.:85.00   1st Qu.:82.00   1st Qu.:77.00
 Median :74.00   Median :80.00   Median :86.00   Median :86.00   Median :86.00   Median :82.00   Median :79.00
 Mean   :73.86   Mean   :79.59   Mean   :86.11   Mean   :85.29   Mean   :86.01   Mean   :83.47   Mean   :79.07
 3rd Qu.:79.00   3rd Qu.:82.00   3rd Qu.:89.00   3rd Qu.:86.00   3rd Qu.:87.00   3rd Qu.:87.00   3rd Qu.:80.00
 Max.   :79.00   Max.   :82.00   Max.   :89.00   Max.   :87.00   Max.   :88.00   Max.   :87.00   Max.   :81.00

    Nov_Hmd         Dec_Hmd        Jan_rainfall     Feb_rainfall    Mar_rainfall    Apr_rainfall    May_rainfall
 Min.   :71.00   Min.   :74.00   Min.   : 0.000   Min.   : 1.0   Min.   : 41.0   Min.   : 78.0   Min.   :403.0
 1st Qu.:72.00   1st Qu.:74.00   1st Qu.: 0.000   1st Qu.: 1.0   1st Qu.: 63.0   1st Qu.: 78.0   1st Qu.:403.0
 Median :75.00   Median :75.00   Median : 0.000   Median : 3.0   Median : 99.0   Median :427.0   Median :576.0
 Mean   :74.23   Mean   :75.69   Mean   : 2.502   Mean   :13.8   Mean   :106.5   Mean   :408.5   Mean   :571.8
 3rd Qu.:76.00   3rd Qu.:76.00   3rd Qu.: 0.000   3rd Qu.:20.0   3rd Qu.:147.0   3rd Qu.:804.0   3rd Qu.:728.0
 Max.   :78.00   Max.   :79.00   Max.   :19.000   Max.   :35.0   Max.   :165.0   Max.   :804.0   Max.   :728.0

  Jun_rainfall    Jul_rainfall    Aug_rainfall    Sep_rainfall  Oct_rainfall    Nov_rainfall     Dec_rainfall
 Min.   :469.0   Min.   :528.0   Min.   :578.0   Min.   :264   Min.   : 55.0   Min.   :  0.00   Min.   : 0.00
 1st Qu.:578.0   1st Qu.:528.0   1st Qu.:674.0   1st Qu.:323   1st Qu.: 55.0   1st Qu.:  0.00   1st Qu.: 0.00
 Median :648.0   Median :604.0   Median :722.0   Median :490   Median :142.0   Median :  0.00   Median : 0.00
 Mean   :694.4   Mean   :619.9   Mean   :713.2   Mean   :500   Mean   :148.5   Mean   : 18.17   Mean   :12.27
 3rd Qu.:946.0   3rd Qu.:673.0   3rd Qu.:767.0   3rd Qu.:732   3rd Qu.:231.0   3rd Qu.: 10.00   3rd Qu.:45.00
 Max.   :946.0   Max.   :786.0   Max.   :767.0   Max.   :732   Max.   :231.0   Max.   :144.00   Max.   :45.00
```

Figure 2 Summary of Features of chosen Dataset

3.3. Data Cleaning and Reduction

The obtained data were fed into the model as is but the accuracies averaged about 22.7% indicating that data needed further cleaning and preprocessing. The Random Forest Model determines number of tries (mtry) based on the square root of the number of features and when tuning the model using the *tuneRF* function, mtry = 2 gave the better results thus the temperature feature was split into minimum and maximum temperatures to bring the features to four that would yield a square root of 2 resulting in mtry of 2 and the target classes were reduced to three (3), i.e. malaria, pneumonia and diarrhea. The features had different ranges with humidity ranging between 20 and 100, temperature between 15 and 35 while rainfall ranged between 0 and 200 (Fig. 3) but this difference in range is known to affect the performances of the models therefore, we normalized the data to ensure each feature had the same range.

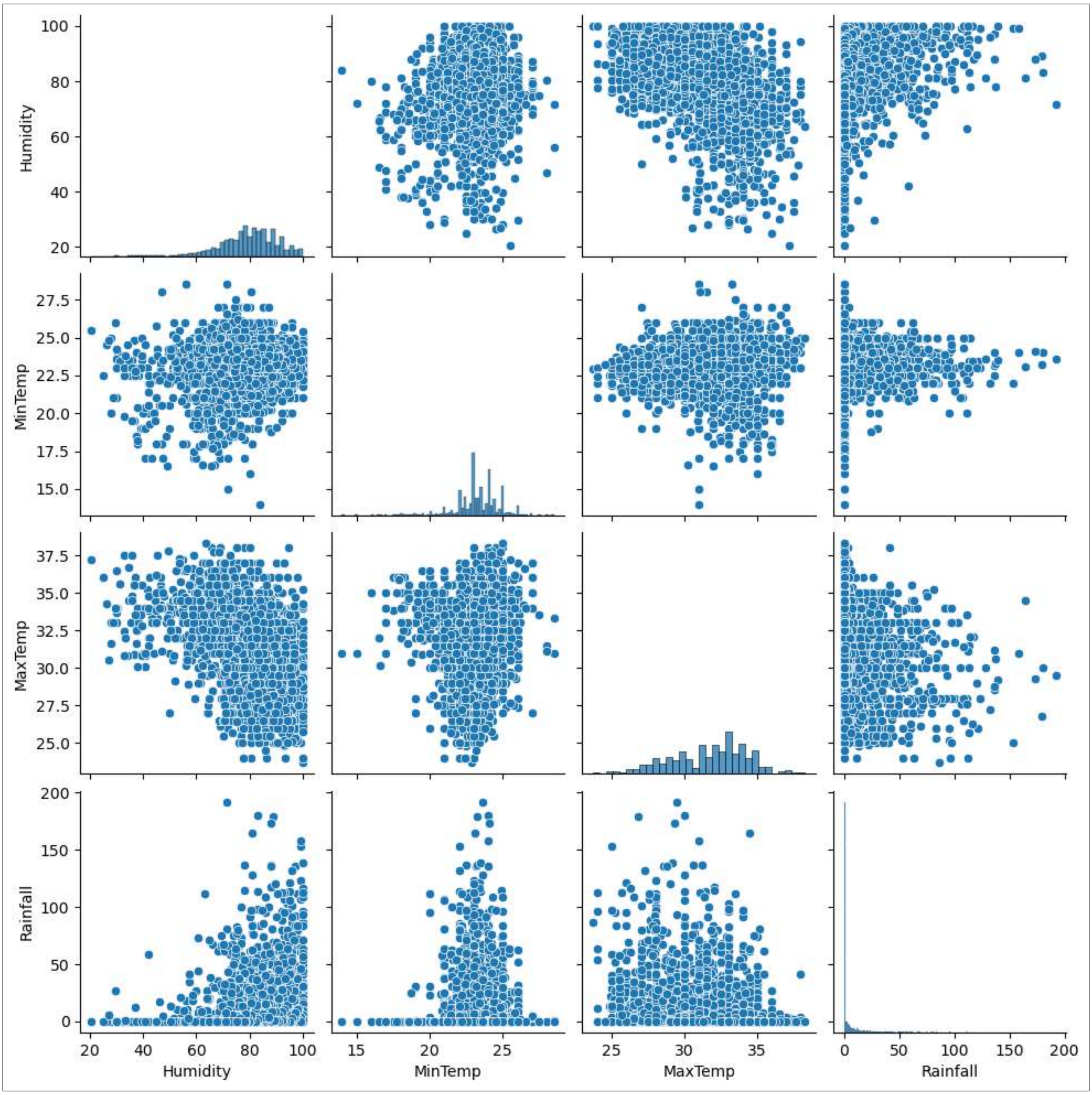


Figure 3 Pairwise Plot of Cleaned Dataset

3.4. Data Normalization

To avoid the possibility of one feature exerting more impact on the performance of the model, we normalized the dataset using the *MinMax Normalization method* (equation 1). MinMaxScaler of the SciKit Learn (SKLearn) Preprocessing Library in Jupyter Notebook to normalized the dataset.

$$\boldsymbol{x}' = \frac{x - \mathbf{min}(x)}{\mathbf{max}(x) - \mathbf{min}(x)} \quad \text{……….(1)}$$

Normalization scaled the values of all the features to the same range between 0 and 1 and the mean and standard deviation of all the features of the dataset lie between 0 and 1 after normalization.

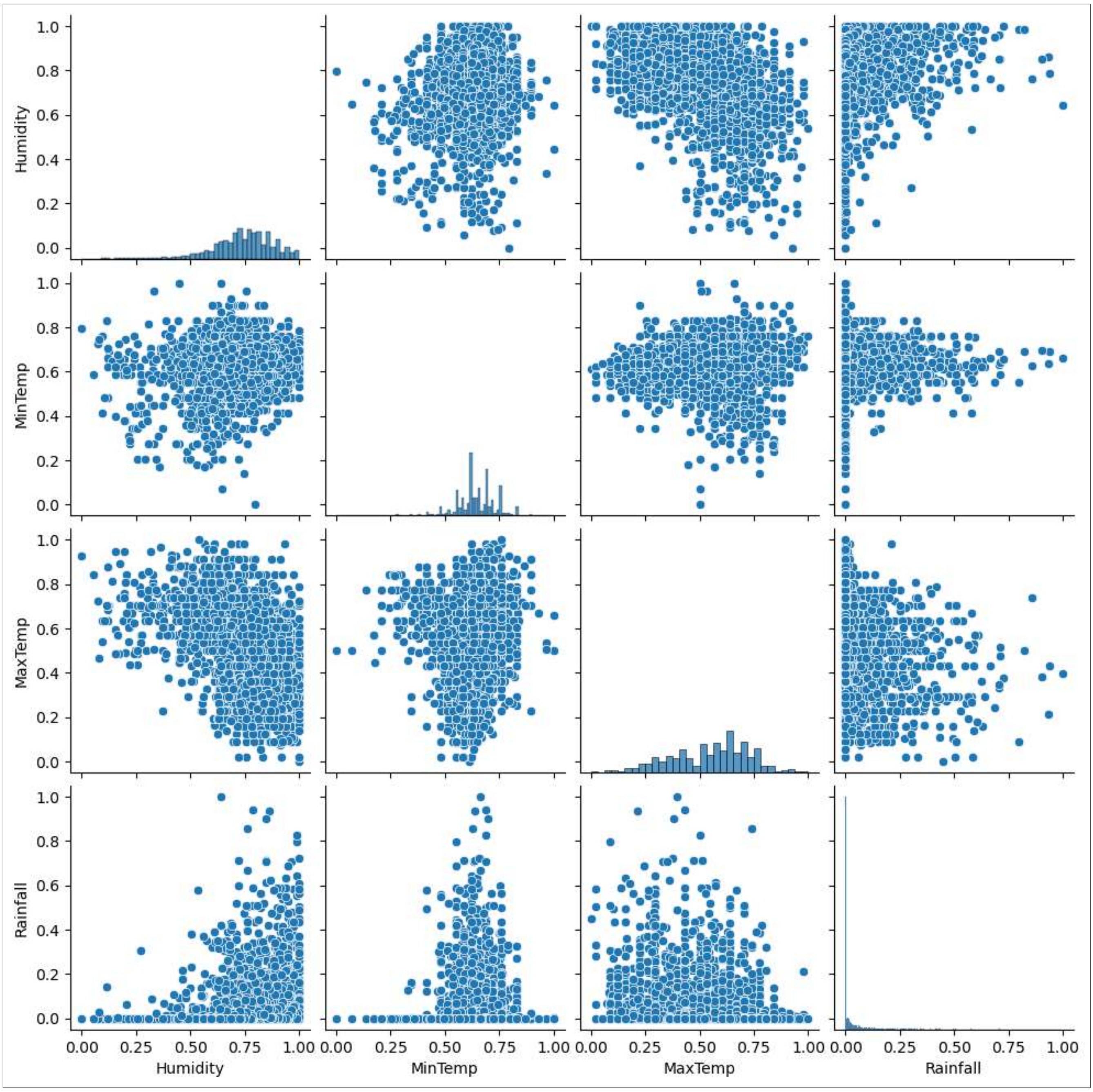


Figure 4 Pairwise Plot of Normalized Dataset

From findings, we reduced the infectious diseases from the six (6) recorded by Chowdhury *et al.*, 2018 to three (3) infectious diseases—Malaria, Pneumonia and Diarrhea, all of which better represented some of the most common Infectious Disease prevalent in Nigeria. From the initially cleaned dataset, we chose a sample size of ten percent (10%) having 110, 112 and 107 datapoints of Malaria, Pneumonia and Diarrhea respectively. A pairwise plot of the dataset shows that the data points are better distributed than compared to the previous datasets as shown by figure 5.

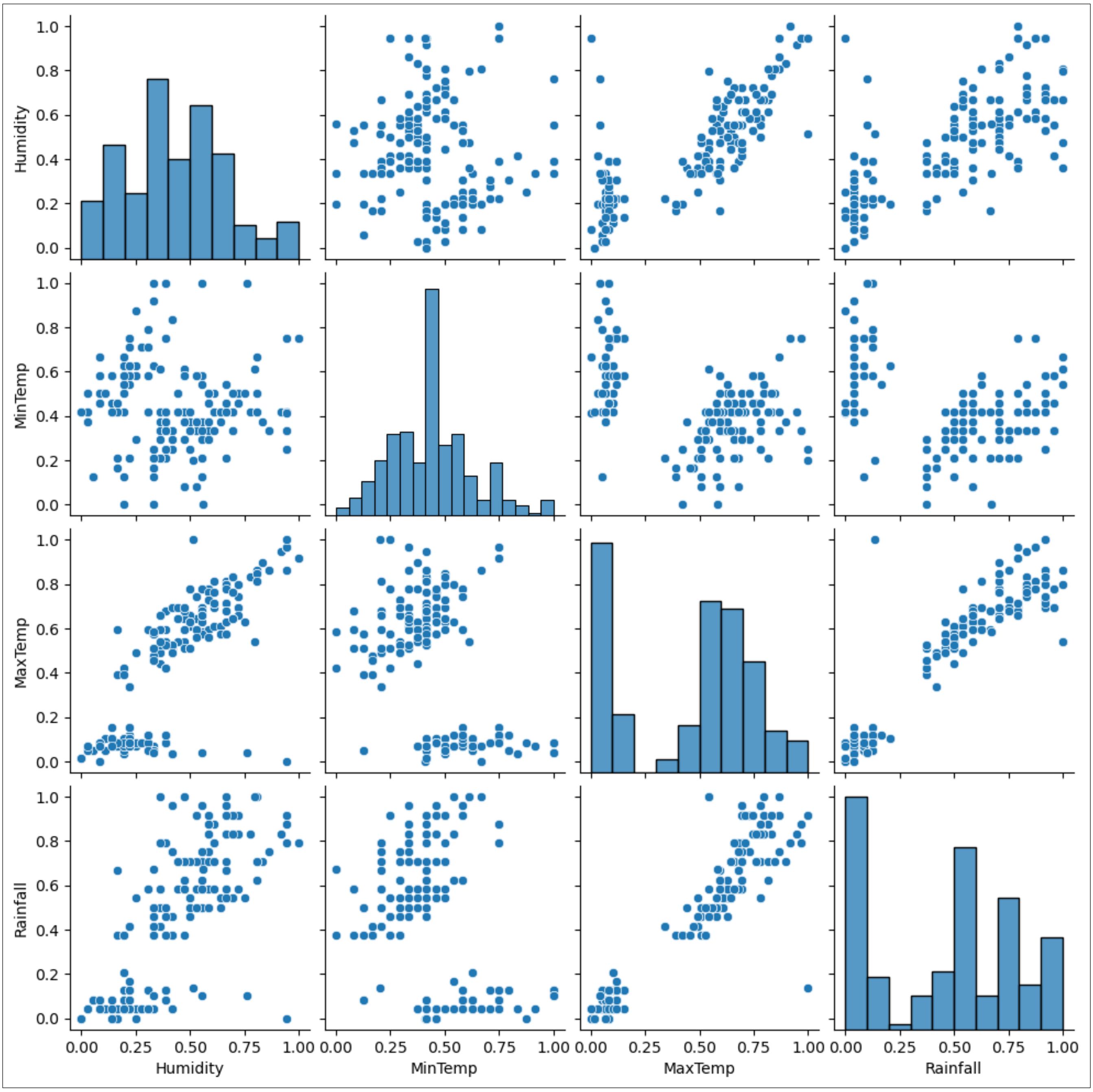


Figure 5 Pairwise plot of Normally Distributed Dataset

The now reduced dataset following the gaussian distribution resulted in better classification results in both models with Random Forest slightly edging Artificial Neural Network by 1.90%.

3.4 Training the Models

3.4.1. Random Forest (RF) Models

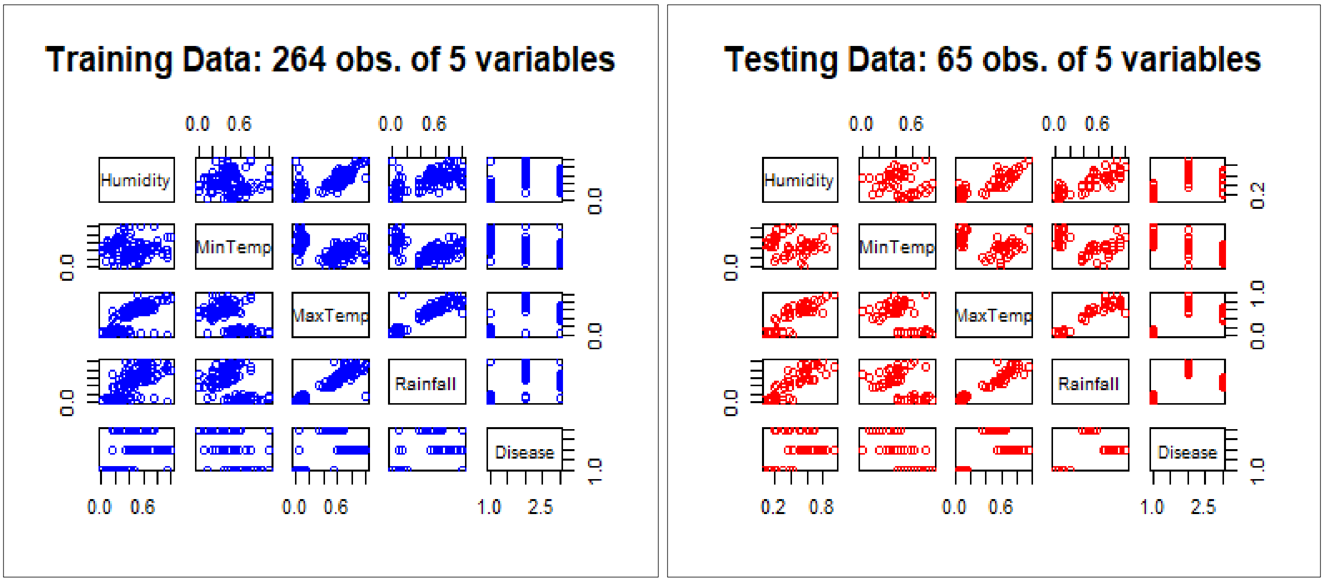


Figure 6 Random Forest Training and Testing Dataset

We fed the preprocessed data into the RF model as a .csv file containing five (5) variables: four (4) input (rainfall, minimum temperature, maximum temperature and humidity) and one target output variable (disease). The data set had 110, 112 and 107 observations for Malaria, Pneumonia and Diarrhea respectively. Required libraries were imported and the dataset was fed into the model and the splitting the dataset into the training and testing sets as shown in figure 6.

The model was tuned showing the number of tries at each iteration (mtry) at 2 (square root of 4 being the number of inputs variables) and number of trees (*ntree*) at 350 after several runs showed that more trees did not increase the performance of the model. The dataset was split into training and testing data on a ratio of 80:20.

3.4.2. Artificial Neural Network (ANN) Model

We used the multilayer perceptron (MLP) Artificial Neural Network (ANN) as the second NIA model for classifying the data. We fed the prepared dataset into the Artificial Neural Network model and encountered errors because the model could not converge due to the fact that the target feature (i.e. Disease) was textual. To solve this, each of the infectious diseases in the target output was assigned a number code (Table 1) and when this was fed back into the model, the model converged accordingly.

Table 1 Data Coding for Artificial Neural Network

SN	Infectious Disease	Assigned Code	Normalized Code
1	Malaria	1	0
2	Diarrhea	2	0.5
3	Pneumonia	3	1

4. Results

The Random Forest Algorithm performed well on the given dataset. It completely classified Diarrhea, Malaria and Pneumonia on the training data with an accuracy of 100% but missed out two classification for malaria on the testing data achieving an accuracy of 96.9% (Table 2).

Table 2 RF Confusion Matrix for Training and Testing Data

Confusion Matrix for Training Data				Confusion Matrix for Testing Data			
Classed/Actual	Diarrhea	Malaria	Pneumonia	Classed/Actual	Diarrhea	Malaria	Pneumonia
Diarrhea	84	0	0	Diarrhea	23	1	0
Malaria	0	86	0	Malaria	0	22	0
Pneumonia	0	0	94	Pneumonia	0	1	18
Classification Accuracy			100%	Classification Accuracy			96.9%
95% Confidence Interval			(0.9861,1)	95% Confidence Interval			(0.8932,0.9963)

The Random Forest Model recorded Out of Bag error = 0.15 using 350 tress (Figure 7) and an mtry = 2 (Figure 8) for all the disease classes. The out of bag error (black line) is lowest between 90 to 135 trees with no error change between 190 and 275 trees.

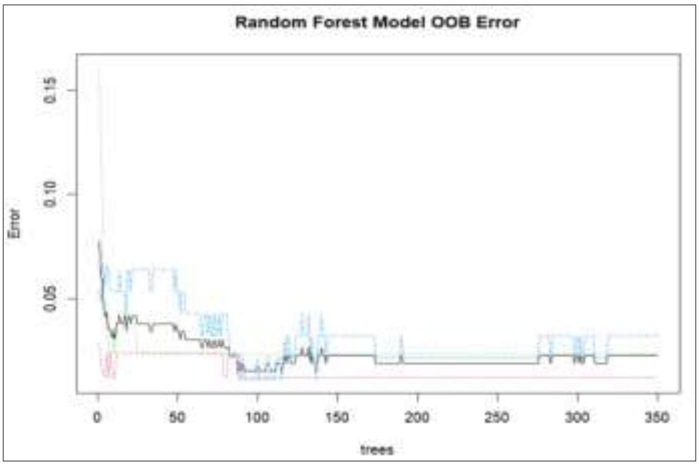


Figure 7 Random Forest Classification Error Plot

Tuning the model revealed that *mtry* of 1 or 2 resulted in an out of bag error of 0.027 (2.27%) with the error increasing as mtry increases thus mtry of 4 results in an OOB error of 0.030 (3.03%).

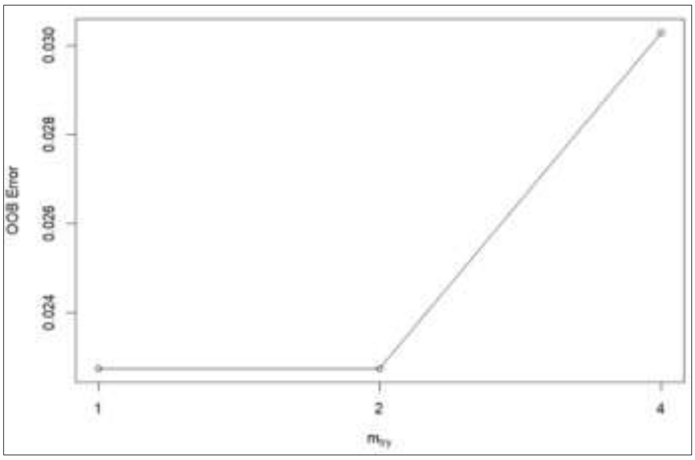


Figure 8 Out of Bag Errors relative to Number of Tries (Mtry)

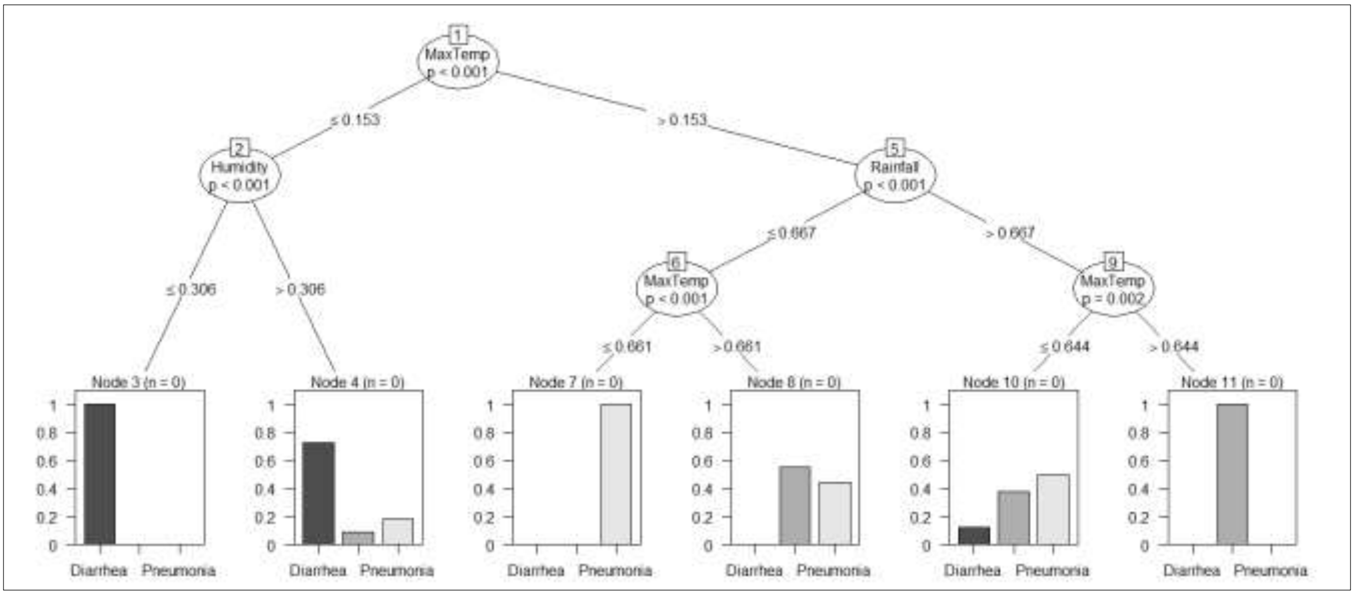


Figure 9 Random Forest Model Tree

Figure 9 presents a tree of the Random Forest model showing its classification based on the best split of the data. Searching through the data, it picked a best split at MaxTemp < 0.2288 and classifies Diarrhea if yes else performs another split at Humidity < 0.506 and continues the splits until all nodes are reached

Further samples of the Random Forest trees provide information that could enable us to decipher the characteristic behaviours of our features. The tree samples (Figure 9) presents findings that enables us to characterize the risks factors based on the weights and correlation with the particular infectious disease based on their p-values. *P-values* are probability values that lie between 0 and 1 and provide a measure of the strength of evidence against Null Hypothesis which in this work means climate change features impacts on infectious diseases. Trees 1 and 2 indicates that humidity impacts upon Diarrhea while Tree indicates it is Minimum Temperature that impacts upon Diarrhea with $p = 0.015$. The Random Forest model achieved an accuracy of 100% on the training dataset with a 95% Confidence Level (CL) and a Confidence Interval (CI) (0.9861, 1) and an accuracy of 96.9% on the testing dataset with a 95% Confidence Level (CL) with Confidence Interval (CI) (0.8932, 0.9963). Similarly, from figure 10 the RF model also presents the statistical information on all classes of infectious diseases contained in the data including sensitivity, specificity, detection rate as well as the balanced accuracy of each class for both the training and testing dataset.

```
Confusion Matrix and Statistics

           Reference
Prediction  Diarrhea Malaria Pneumonia
  Diarrhea        23       1         0
  Malaria          0      22         0
  Pneumonia        0       1        18

Overall Statistics

               Accuracy : 0.9692
                 95% CI : (0.8932, 0.9963)
    No Information Rate : 0.3692
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9536

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: Diarrhea Class: Malaria Class: Pneumonia
Sensitivity                   1.0000         0.9167           1.0000
Specificity                   0.9762         1.0000           0.9787
Pos Pred Value                0.9583         1.0000           0.9474
Neg Pred Value                1.0000         0.9535           1.0000
Prevalence                    0.3538         0.3692           0.2769
Detection Rate                0.3538         0.3385           0.2769
Detection Prevalence          0.3692         0.3385           0.2923
Balanced Accuracy             0.9881         0.9583           0.9894
```

Figure 10 Random Forest Model Tree

The findings from our Random Forest Model supports the findings in [43] which reports Random Forest as a highly interpretable Algorithm and also high on accuracy. The Artificial Neural Network (ANN) model recorded an error of 0.046 in 17150 steps (Figure 11) amounting 4.6% error which is a 102.64% increase from the error recorded by the Random Forest model. Notably, the ANN model is a black box like in implementation thus it offers less interpretability when compared to the Random Forest Model.

Table 3 ANN Model Test Confusion Matrix

Classed/Actual	Diarrhea	Malaria	Pneumonia
Diarrhea	25	0	0
Malaria	0	20	0
Pneumonia	0	0	31
Classification Accuracy			95%
95% Confidence Interval			(0.8769,0.9862)

Unlike the Random Forest Model which misclassified 2 points in the malaria class, the Artificial Neural Network model misclassified 2 points in the malaria class, 1 point in the Diarrhea class and 1 point in the Pneumonia class totaling four (4) misclassification and thus achieving an accuracy of 95% with a Confidence Level of 95% and a Confidence Interval between 0.8769 and 0.9862.

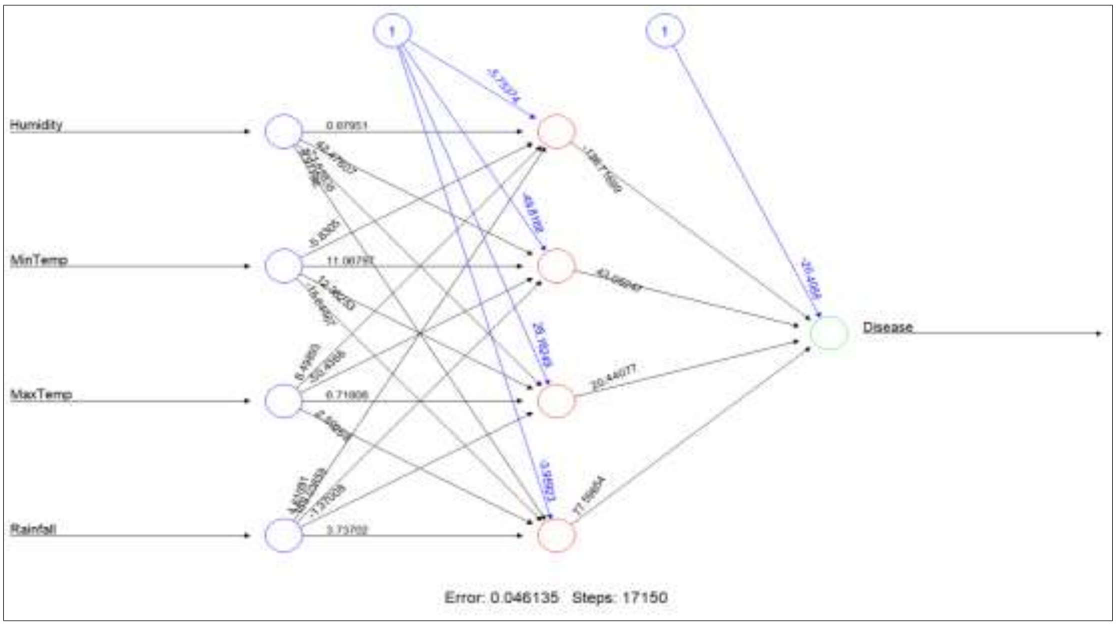


Figure 11 The ANN Model Architecture

Although both models performed well and showed great applicability on the reduced dataset, they both performed very poorly on the complete and original dataset using all the features as is, with accuracies ranging from 22% to 44% after numerous pass. There could be many reasons for this, one of which is that some advance processing is needed.

Temperature was found to be the most significant feature affecting the classification with a 78% importance and closely followed by rainfall with 70.65% importance and then followed by Humidity with 19.32% importance. The model indicates that minimum temperature with a meagre importance of 6.5% has very little impact of all the features determining the classes (figure 12). However, there could be some biases in this importance classification due to impurities that are removed from the Maximum Temperature not being removed in the other features thus affecting the importance classification. The feature having the most importance is likely to be responsible for impacting the diseases.

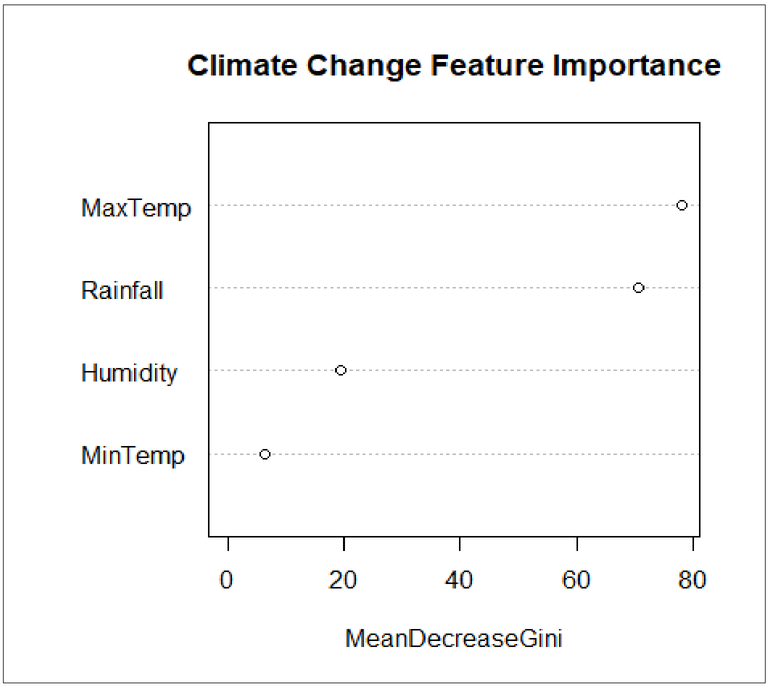


Figure 12 Climate Change Feature Importance

Findings show that both models indicated that rainfall and temperature significantly impact upon all three classes of infectious diseases of infectious diseases and this agrees with the current findings in research and the current reality in Nigeria where the increase in rainfall and temperature variations coincide with increase in the incidences of malaria and pneumonia and diarrhea. Figure 13 presents a summary of the performances of both models.

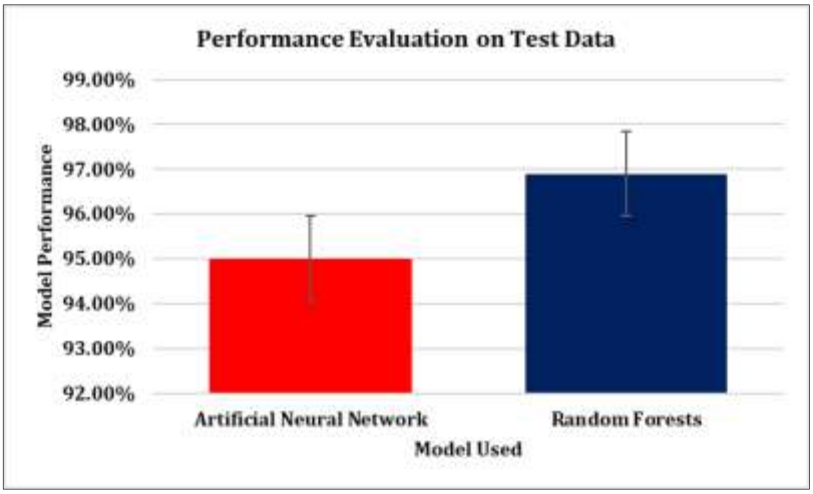


Figure 13 Performance of models on Test data

5. Conclusion

Both models indicated that Rainfall and Temperature variations were common risks factors that indicated highest weights impacting the emergence and incidence of Infectious Diseases in Nigeria. High Temperatures increased disease and pathogen transmission season and had high correlative impacts on Malaria and Diarrhea. Low Temperatures provided conducive atmosphere for flu pathogen that causes pneumonia and those that causes malaria. Heavy Rainfalls causes floods and creates ponds of stagnant water necessary for breading of vectors e.g. mosquitoes which causes malaria and leads to outbreaks of cold supporting pathogens that causes pneumonia thus rainfall was found to have a high correlative impact on malaria, pneumonia and diarrhea. These findings agree with literature [12], [16] hence affirming the argument that climate change features influence the emergence and incidence of infectious diseases. This study helps to highlight that proper drainage management and other related ethical practices could help reduce the incidence of malaria while curate health behavioural patterns could reduce the occurrence of diarrhea and pneumonia. Similarly. this study shows that Nature-Inspired Algorithms are applicable to classification climate change and infectious diseases data but also points out that dataset indicating proven clinical correlation between infectious diseases and climate change features in Nigeria are scanty thus collating clinical dataset on infectious diseases, particular in Nigeria where the current state of Healthcare Delivery is very low is vital to mitigating the challenges of infectious diseases in the current climate change reality.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Nii-Trebi, N. I. (2017). Emerging and neglected infectious diseases: insights, advances, and challenges. BioMed research international, 2017.

[2] Wei, F., Peng, Z., Jin, Z., Wang, J., Xu, X., Zhang, X., Xu, J., Ren, Z., Bai, Y., Wang, X., Lu, B., Wang, Z., Xu, J., & Huang, S. (2022). Study and prediction of the 2022 global monkeypox epidemic. Journal of Biosafety and Biosecurity, 4(2), 158–162. https://doi.org/10.1016/j.jobb.2022.12.001

[3] Kirtane, A. R., Verma, M., Karandikar, P., Furin, J., Langer, R., & Traverso, G. (2021). Nanotechnology approaches for global infectious diseases. Nature Nanotechnology, 16(4), 369-384.

[4] Hauck, K. (2018). The economics of infectious diseases. In Oxford Research Encyclopedia of Economics and Finance.

[5] Huremović, D. (2019). Brief history of pandemics (pandemics throughout history). Psychiatry of pandemics: a mental health response to infection outbreak, 7-35.

[6] Barbieri, R., Drancourt, M., & Raoult, D. (2021). The role of louse-transmitted diseases in historical plague pandemics. The Lancet Infectious Diseases, 21(2), e17-e25.

[7] Fenollar, F., & Mediannikov, O. (2018). Emerging infectious diseases in Africa in the 21st century. New Microbes and New Infections, 26, S10-S18.

[8] Baylor College of Medicine. (n.d.). Introduction to Infectious Diseases. https://www.bcm.edu/departments/molecular-virology-and-microbiology/emerging-infections-and-biodefense/introduction-to-infectious-diseases

[9] Casadevall, A. (2020). Climate change brings the specter of new infectious diseases. The Journal of Clinical Investigation, 130(2), 553-555.

[10] Caminade, C., McIntyre, K. M., & Jones, A. E. (2019). Impact of recent and future climate change on vector-borne diseases. Annals of the New York Academy of Sciences, 1436(1), 157-173.

[11] Howard, C., & Huston, P. (2019). Climate change and infectious diseases: The solutions: The health effects of climate change: Know the risks and become part of the solutions. Canada Communicable Disease Report, 45(5), 114.

[12] Singh, A. K., & Singh, B. J. (2020). Assessing the infectious diseases of students in different weather seasons in DIT University Dehradun, Uttarakhand (India). Asian Journal of Multidimensional Research, 9(3), 34-48.

[13] Rees, E., Ng, V., Gachon, P., Mawudeku, A., McKenney, D., Pedlar, J., ... & Knox, J. (2019). Early detection and prediction of infectious disease outbreaks. CCDR, 45(5).

[14] Brooks, D. R., & Boeger, W. A. (2019). Climate change and emerging infectious diseases: Evolutionary complexity in action. Current Opinion in Systems Biology, 13, 75-81.

[15] Cianconi, P., Betrò, S., & Janiri, L. (2020). The impact of climate change on mental health: a systematic descriptive review. Frontiers in psychiatry, 74.

[16] Semenza, J. C., & Paz, S. (2021). Climate change and infectious disease in Europe: Impact, projection and adaptation. The Lancet Regional Health-Europe, 9, 100230.

[17] Flahault, A., de Castaneda, R. R., & Bolon, I. (2016). Climate change and infectious diseases. Public health reviews, 37(1), 1-3.

[18] Yadav, A., & Vishwakarma, D. K. (2020). A comparative study on bio-inspired algorithms for sentiment analysis. Cluster Computing, 23(4), 2969-2989.

[19] Yang, X. S. (2020). Nature-inspired optimization algorithms: Challenges and open problems. Journal of Computational Science, 46, 101104.

[20] Abu Khurma, R., Aljarah, I., Sharieh, A., Abd Elaziz, M., Damaševičius, R., & Krilavičius, T. (2022). A review of the modification strategies of the nature inspired algorithms for feature selection problem. Mathematics, 10(3), 464.

[21] Pazhaniraja, N., Paul, P. V., Roja, G., Shanmugapriya, K., & Sonali, B. (2017, March). A study on recent bio-inspired optimization algorithms. In 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN) (pp. 1-6). IEEE.

[22] Li, G., Jin, Y., Akram, M. W., Chen, X., & Ji, J. (2018). Application of bio-inspired algorithms in maximum power point tracking for PV systems under partial shading conditions–A review. Renewable and Sustainable Energy Reviews, 81, 840-873.

[23] Sudhakar, A. V., & Karri, C. (2017, July). Bio inspired algorithms in power system operation: A review. In 2017 international conference on recent trends in electrical, electronics and computing technologies (ICRTEECT) (pp. 113-119). IEEE.

[24] Molina, D., LaTorre, A., & Herrera, F. (2018). An insight into bio-inspired and evolutionary algorithms for global optimization: review, analysis, and lessons learnt over a decade of competitions. Cognitive Computation, 10(4), 517-544.

[25] Valdez, F., Castillo, O., & Melin, P. (2021). Bio-inspired algorithms and its applications for optimization in fuzzy clustering. Algorithms, 14(4), 122.

[26] Nunoo-Mensah, H., Boateng, K. O., & Gadze, J. D. (2018). The adoption of socio-and bio-inspired algorithms for trust models in wireless sensor networks: A survey. International Journal of Communication Systems, 31(7), e3444.

[27] Fan, X., Sayers, W., Zhang, S., Han, Z., Ren, L., & Chizari, H. (2020). Review and classification of bio-inspired algorithms and their applications. Journal of Bionic Engineering, 17(3), 611-631.

[28] Beine, M., & Parsons, C. R. (2017). Climatic factors as determinants of international migration: Redux. CESifo Economic Studies, 63(4), 386-402.

[29] Talisuna, A. O., Okiro, E. A., Yahaya, A. A., Stephen, M., Bonkoungou, B., Musa, E. O., ... & Fall, I. S. (2020). Spatial and temporal distribution of infectious disease epidemics, disasters and other potential public health emergencies in the World Health Organisation Africa region, 2016–2018. Globalization and health, 16, 1-12.

[30] Semenza, J. C., Rocklöv, J., & Ebi, K. L. (2022). Climate change and cascading risks from infectious disease. Infectious Diseases and Therapy, 11(4), 1371-1390.

[31] Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., & Shukla, P. R. (2022). Global Warming of 1.5 C: IPCC Special Report on Impacts of Global Warming of 1.5 C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty. Cambridge University Press.

[32] Dennis, S., & Fisher, D. (2018). Climate change and infectious diseases: the next 50 years. Ann. Acad. Med. Singap, 47(10), 401-404.

[33] Bartlow, A. W., Manore, C., Xu, C., Kaufeld, K. A., Del Valle, S., Ziemann, A., ... & Fair, J. M. (2019). Forecasting zoonotic infectious disease response to climate change: mosquito vectors and a changing environment. Veterinary sciences, 6(2), 40.

[34] Fouque, F., & Reeder, J. C. (2019). Impact of past and on-going changes on climate and weather on vector-borne diseases transmission: a look at the evidence. Infectious diseases of poverty, 8(03), 1-9.

[35] Pöhlker, M. L., Krüger, O. O., Förster, J. D., Berkemeier, T., Elbert, W., Fröhlich-Nowoisky, J., ... & Mikhailov, E. (2021). Respiratory aerosols and droplets in the transmission of infectious diseases. arXiv preprint arXiv:2103.01188.

[36] Hinz, R., Frickmann, H., & Krüger, A. (2019). Climate change and infectious diseases. International climate protection, 269-276.

[37] Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... & Bengio, Y. (2022). Tackling climate change with machine learning. ACM Computing Surveys (CSUR), 55(2), 1-96.

[38] Doyle, J. (2016). Mediating climate change. Routledge.

[39] Hornsey, M. J., & Fielding, K. S. (2020). Understanding (and reducing) inaction on climate change. Social Issues and Policy Review, 14(1), 3-35.

[40] Clayton, S. (2020). Climate anxiety: Psychological responses to climate change. Journal of Anxiety Disorders, 74, 102263.

[41] Labazova, O. (2019). Towards a framework for evaluation of blockchain implementations.

[42] Six, N. (2021). Decision process for blockchain architectures based on requirements. arXiv preprint arXiv:2101.08998.

[43] Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 6308-6325.

[44] Chowdhury, F. R. (2018). Climate change and Infectious diseases in Sylhet, Bangladesh. [Data set]. figshare. https://doi.org/10.6084/M9.FIGSHARE.6354956

[45] Umoren, I., Udonyah, K., & Isong, E. (2019, August). A computational intelligence framework for length of stay prediction in emergency healthcare services department. In 2019 14th International Conference on Computer Science & Education (ICCSE) (pp. 539-549). IEEE.

[46] Umoren, I. J., Etuk, U. E., Ekong, A. P., & Udonyah, K. C. (2021). Healthcare logistics optimization framework for efficient supply chain management in Niger Delta Region of Nigeria. International Journal of Advanced Computer Science and Applications, 12(4), 594-604.

[47] Inyang, S., & Umoren, I. (2023). Semantic-Based Natural Language Processing for Classification of Infectious Diseases Based on Ecological Factors. International Journal of Innovative Research in Sciences and Engineering Studies (IJIRSES), *3*(7)

Authors' short Biography

	**Kingsley C. Udonyah** is holds a Bachelor of Science (BSc.) and Master of Science (MSc.) Degrees in Computer both from the Akwa Ibom State University, Nigeria. He has won several academic awards including Best Graduating Student, Department of Computer Science and Best Graduating Student, Faculty of Physical Sciences Akwa Ibom State University. He is a passionate lover of technology and its application and holds the IC3 Digital Literacy Certification. He is a lover of Artificial Intelligence with particular research interests in Machine Learning, Computational Intelligence, Cognitive Science and Brain Computer Interface. He has handled several local research projects in the Department of Computer Science, Akwa Ibom State University, Nigeria and has co-authored several articles in reputable journals.
	**Imeh J. Umoren** holds B.Sc., MSc. and PhD Degrees in Computer Science. Trained in the University of Calabar, Nigeria on Unix/Linux State Technologies and Data Processing from 1986-1988, he has won and held several scholarships including the Total E&P Graduate Scholarship from 2010- 2012. Currently, he is an academic staff in the Department of Computer Science, Akwa Ibom State University (AKSU), Mkpat Enin, Nigeria and has served in several administrative positions including: Acting Head, Department of Computer Science, Deputy Dean, Faculty of Physical Sciences and Departmental Post Graduate Coordinator etc. Dr. Umoren, has more than 12 years' experience in the Academia, Research and Global ICT industry. He is a strategic thinker, Analyst and a Multi-Tasking Development Specialist focusing on people, process and technology. He is passionate about the use of technology (Computational Intelligence) to transform lives, create wealth and accelerate development in Africa (Nigeria) and globally. Imeh Umoren is a member of several high-profile professional bodies including the IEEE (an elite group within the global engineering community), a recognized member of Polish Neural Networks Society, a member of Nigeria Computer Society (NCS), Internet Society etc. Dr. Imeh Umoren has over 60 published articles and conference papers. His research interests focus on Mobile Computing, Complexity Theory, Algorithms, Computational Intelligence, Machine Learning (ML) and Natural Language Processing (NLP).