AN IMPROVED RAINFALL PREDICTION MODEL FOR MINIMIZING THE NEGATIVE IMPACT OF INCREASED RAINFALL DAYS IN MINNA, NIGERIA USING ARTIFICIAL NEURAL NETWORK.

[1]Ibrahim, A. G., [1]Oyedum, O. D., [2]Abisoye, O. A., [1]Usman, M. N. and [1]Ughanze, J. I.
[1]Department of Physics, Federal University of Technology, Minna, Nigeria.
[2]Department of Computer Science, Federal University of Technology, Minna, Nigeria.
Correspondence: ibrahimaku@futminna.edu.ng +2348062382666.

ABSTRACT
This research aims at predicting the rainfall of Minna metropolis of Niger State, Nigeria using binary classification method of Artificial Neural Network (ANN). In this approach, four atmospheric parameters comprising those of rainfall, relative humidity, minimum temperature and maximum temperature spanning from January 2010 to December 2019 were acquired from the Geography Department of the Federal University of Technology, Minna. The default threshold classification method of ANN was investigated. The result revealed that: for the default threshold of 0.5, a prediction accuracy of 69%, sensitivity of 63%, specificity of 84% an error value of 1.3% and a total of 66 rainfall days were predicted as against 32 rainfall days in the data set. The implication of this result is that more rainfall days were anticipated in the metropolis which could lead to flooding in long run. It was recommended that for more accurate rainfall prediction, more robust data be used for network training.

Keywords: Rainfall prediction Model, Artificial Neural Network (ANN).

INTRODUCTION

Rainfall is a natural phenomenon whose prediction is challenging and demanding as the world continues to witness an ever-changing climate condition. Its forecast plays an important role in water resource management and therefore, it is of particular relevance to the agricultural sector, which contributes significantly to the economy of any nation Abdulkadir *et al*. (2012). The effect of rainfall on human civilisation is colossal. Rainfall means crops; and crop means life. Additionally, rainfall has a strong influence on the operation of dams and reservoirs, sewage systems, traffic and other human activities. Previous studies have shown that among the entire climate elements, rainfall is the most variable element both temporally and spatially which can have significant impact on economic activities (Pepin, 2017; Umar, 2012). Accurate and timely rainfall prediction can be very helpful for effective security measures for planning water resources management, issuance of early flood warning, construction activities, transportation activities, agricultural tasks, managing the flight operations and flood situation. Heavy rainfall can lead to numerous hazards, for example: flooding, including risk to human life, loss of crops and livestock, landslides which can threaten human life (Ifabiyi and Ashaolu, 2013).

In machine learning, classification can be referred to as task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain. Weather data consists of various atmospheric features such as wind, precipitation, humidity, pressure, and temperature among others. These parameters are related to one another in one form or the other and as such can serve as inputs for the prediction of a parameter of interest using data mining technique. Data mining techniques such as ANN can effectively predict rainfall by extracting the hidden patterns among available features of past weather data (Aftab and Ahmad, 2018). In classification prediction, the classifiers are used to find the class to which an unknown data belongs based on the information available from a set of data whose class is already known (Abisoye & Jimoh, 2018). Classification refers to a predictive modelling problem where a class label is predicted for a given sample of input data (Brownlee, 2020). From a modelling perspective, classification requires a training dataset with many samples of input and output from which to learn. A model will use the training dataset and calculate how to best map instances of input data to specific class labels. In this work, the input data are those of relative humidity, minimum temperature and maximum temperature while the class label to be predicted is Rainfall. The class label is often string values, for example “rain”, “no rain”, and must be mapped to numeric values for example, “rain” = 1, ‘no rain” = 0, before being provided to an algorithm for modelling. This is often referred to as label training, where a unique integer is assigned to each class label. Classification predictive modelling algorithms are evaluated based on their results. Researchers are developing and applying improved weather prediction models capable of accurately forecasting rainfall. Artificial Neural Network algorithm becomes an attractive inductive approach in rainfall prediction owing to the non-linearity, flexibility and data learning in building the models without any prior knowledge about catchment behaviours and flow processes.

It is important to be able to predict the magnitude of rainfall in order to minimise the negative impact of increased rainfall days capable of hazards such as flooding and erosion. This is the primary aim of this research. Processes involved include carrying out interpolation of missing data and min-max normalisation technique on the ten years meteorological data of daily Rainfall (mm), Relative Humidity (mmHg), Minimum Temperature (°C) and Maximum Temperature (°C) for Minna metropolis; and predicting the number of rainfall days using classification method of Artificial Neural Network as a predictive tool.

STUDY AREA

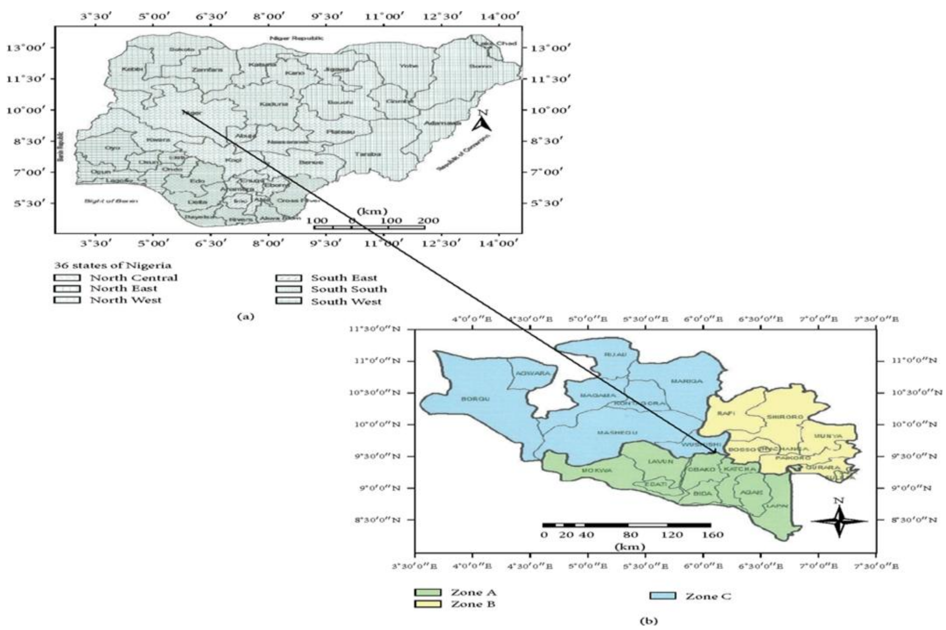

Figure 1: Map of (a) Nigeria and (b) Niger State showing the study area. (Geography Department, FUT Minna, 2021)

Minna is the headquarters of Chanchaga Local Government Area of Niger State, Nigeria. It is the capital city of the state. It lies between Latitudes 09°40′ 7.63″ N and 09° 39′ 59.72″ N and Longitudes 06° 30′ 0.32″ E and 06° 36′ 34.05″ E. Minna possesses the tropical continental wet and dry climate based on the Koppen Classification Scheme and is characterised with two distinct seasons namely the wet season which begins around March and runs through October and dry season which begins from October to March. The city has a mean annual rainfall of 1334 mm with September recording the highest rain of close to 330 mm on the average, while the least amount of rainfall occurs in December and January which can be as low as 1mm. Minna is about 150 km away from Abuja, the capital of Nigeria (Geography Department, FUT Minna, 2021).

Theoretical background: Escape velocity of water molecules from the surface of water bodies to form rain

the environment, the molecules on the surface of the water bodies gradually lose their molecular potential energy (P.E) given by (Halliday *et al.*,2006);

$$P.E = mgh \qquad (1)$$

Where m is the mass of the water molecules, g the acceleration due to gravity and h, the altitude at sea level.

As this happens, the molecules simultaneously gain molecular kinetic energy (K.E) given by (Jerzy, 2006);

$$K.E = \frac{3}{2}KT \qquad (2)$$

where k is the Boltzman constant and T, the temperature.

Both processes of equation (1) and (2) continues gradually till it reaches its maximum. Since energy can neither be destroyed nor created but can be transferred from one form to another. Molecular P.E equal molecular K.E.

$$mgh = \frac{3}{2}KT \qquad (3)$$

When the molecules gain maximum kinetic energy, they become excited and escape from the surface of the water bodies. At the point of evaporation, the higher the temperature of the molecules, the higher their molecular kinetic energy and simultaneously the velocity or rate of escape from the surface of the water bodies. The kinetic energy with which they escape is given by;

$$K.E = \frac{1}{2}mV_e \qquad (4)$$

where V_e is the escape velocity of water molecules and m, the mass of water molecules.

The K.E transports the excited molecules through the atmosphere. The molecules gradually lose their K.E and simultaneously acquire P.E due to decrease in temperature with height in the troposphere. Again, from the law of conservation of energy, that is, the total energy is constant in any process.

$$\frac{3}{2}KT = \frac{1}{2}mV_e^2 \qquad (5)$$

Making V_e the subject of the formula yields;

$$V_e = \sqrt{\frac{3KT}{m}} \qquad (6)$$

It is known that,

$$\rho = \frac{m}{v} \qquad (7)$$

where, ρ is the density of water and v, the volume of water molecules given by;

$$v = \frac{4}{3}\pi r^3 \qquad (8)$$

and m is the mass of water molecule. From (7) and (8), m is given by;

$$m = \frac{4}{3}\rho\pi r^3 \qquad (9)$$

Substitute *m* into (6) and solve for V_e;

$$V_e = \sqrt{\frac{9KT}{4\pi\rho r^3}} \qquad (10)$$

Equation (10) is the escape velocity with which a water molecule escape from the surface of water bodies into the atmosphere to form rain.

Review of Related Works

In Nigeria and on the worldwide scale, large numbers of attempts have been made by different researchers to predict rainfall accurately using various techniques. Some of the research efforts directed at prediction using methods in ANN are here presented:

Abdulkadir *et al.* (2012) developed a real time Artificial Neural Network based on rainfall forecasting model for Ilorin, Nigeria, using observed rainfall records. This ANN model is designed to run a real time task in which the input to the model is a consecutive data of the rainfall. The neural network was trained with sixty years (1952–2011) total monthly historical rainfall data. The trained network yielded 76 % and 87 % of good forecast for the training and testing data set respectively. The correlation coefficient of 0.88 was obtained which showed that the network is fit to be used for the subsequent quantitative prediction of rainfall in Ilorin.

Theethagiri *et al.* (2020) studied the different machine learning classification algorithms to predict the COVID-19 recovered and deceased cases. The k-fold cross-validation resampling technique was used to validate the prediction model. The prediction algorithm was evaluated with performance metrics such as prediction accuracy, precision, recall, mean square error, confusion matrix. The KNN algorithm predicts 92 % (true positive rate) of the deceased cases correctly, with 0.077 % of misclassification. Further, the KNN algorithm produces the lowest error rate as 0.19 on the prediction of accurate COVID-19 cases than the other algorithm. Also, it produces the receiver operator characteristic curve with an output value of 82 %.

Julie and Kannan, (2010) considered an approach to handle learning disability database to predict frequent signs and symptoms of the learning disability in school age children. Two classification techniques, decision tree and clustering were used for 125 real data sets with most of the attributes takes binary values, where each partition represents a cluster and it classifies the data into groups. Each group contains at least one object and each object must belong to exactly one group. The results obtained shows that the accuracy of the classifier was 77.6 % and incorrectly classified 22.4 %. Sensitivity of 84% and area under curve (AUC) of 71%.

Furthermore, Tyagi and Kumar (2017) predicted monthly rainfall by using Back Propagation, Radial Basis Function and Neural Network. For prediction, the dataset was collected from Coonoor region in Nilgiri district (Tamil Nadu). Performance was evaluated in terms of Mean Square Error. According to the results, higher accuracy was reported in Radial Basis Function Neural Network with smaller Mean Square Error. Moreover, the researchers also used these techniques for future rainfall prediction. However, devastating events such as flooding mostly caused by heavy rainfall are natural phenomenon that cannot be stopped from occurring but its effects can be minimised if effectively tackled and appropriate measures are taken to slow down its effects and frequency (Bukka *et al*. 2017).

METHODOLOGY

Data Acquisition

Ten years meteorological data of Rainfall (mm), Relative Humidity (mmHg), Minimum Temperature (°C) and Maximum Temperature (°C) for Minna metropolis, was acquired from the Geography Department, Federal University of Technology Minna for the period spanning from 2010 to 2019. The data were collected on daily and annual total basis.

Data preparation

To prevent error due to missing values, outliers, fields that are obsolete or redundant data from entering the network, data preparation is necessary. To reduce these errors, the missing values were replaced with the fields mean and for large gap, the fields were omitted. Min-max normalization was used. This normalization technique works by seeing how much greater the field value is than the minimum value min (x) and scaling this difference by the range. That is (Larose, 2005),

$$\dot{X} = \frac{x - min(x)}{range(x)} \quad (11)$$

where, x refer to the original field value, $\dot{X}$ is the normalized field value, min (x) is the minimum field value, min (x) is minimum field value and range (x), the difference of maximum field value and minimum field value.

Data Splitting

It is ethical to split the dataset to avoid over fitting or under fitting and usually decided according to the size and type of the data available (Kumar, 2020). In this work, splitting ratio of 70:15:15 was used to split the dataset, that is, the dataset was split into three parts: 70% for train data, 15% for testing data and 15% for validation data. The training set is the largest with 870 data set because a large data sample is required to fit the network so it can observe and learn from the data and optimise its parameters. 186 data sets were used for the testing and validation respectively.

Rainfall Prediction Analysis

Confusion matrix was used to analyse the prediction outcomes. In binary classification prediction, there are four major prediction (Confusion Matrix) outcomes that could occur as shown in Table 1.

Table 1: Confusion Matrix (CM) Outcomes in Binary Classification Prediction

Actual rain day	Predicted rain day	CM Outcome
0	1	False Positive
0	0	True Negative
1	0	False Negative
1	1	True Positive

The four Confusion Matrix (CM) outcomes are further explained as it applies to this work. True Positive (TP): this means that the actual target is positive (1) and the outcome of the prediction is positive (1). In relation to this work, True Positive implies that there was rainfall in a particular day of the data and that particular day was rightly predicted as a rainfall day by the network.

True Negative (TN): this means the actual target is negative (0) and the outcome of the prediction is negative (0). This implies that there was no rainfall in a particular day of the data and that particular day was rightly predicted as a dry (no-rainfall) day by the network.

False Positive (FP): this means the actual target is negative (0) and the outcome of the prediction is positive (1). Implying that there was no rainfall on a particular day of the data and that particular day was predicted as a rainy day by the network.

False Negative (FN): this means the actual target is positive (1) and the outcome of the prediction is negative (0). False

Negative implies that there was rainfall on a particular day of the data and that particular day was predicted as a dry (no-rainfall) day by the network.

Prediction Performance Evaluation

To evaluate the single threshold classification, the default threshold of 0.5 was maintained and the four outcomes of the prediction analysis were used to evaluate the classifier. The evaluation standards as applicable to this work are discussed below (Fokas et al., 2018):

Accuracy: Accuracy is defined as the percentage of correct predictions of the rain days and dry (no-rain) days. It can be calculated easily by dividing the number of correctly predicted rain days and no-rain days by the number of total predicted days.

$$Accuracy = \frac{TP+TN}{N} \qquad (12)$$

Sensitivity: Sensitivity is defined as the percentage of predicted rain days. This means when it actually rains, how often does the classifier correctly predicts rain days.

$$Sensitivity = \frac{TP}{TP+FN} \qquad (13)$$

Specificity: Specificity is defined as the percentage of predicted dry (no-rainfall) days. This implies that when it's actually a no-rain day, how often does the classifier correctly predicts dry (no-rainfall) day.

$$Specificity = \frac{TN}{TN+FP} \qquad (14)$$

Misclassification: Misclassification is defined as the percentage of wrong prediction of the rain days and dry (no-rain) days. It can be calculated by dividing the number of wrong predictions of rain days and no-rain days by the number of total predicted days. It is also known as error rate.

$$Error\ rate = \frac{FN+FP}{N} \qquad (15)$$

RESULTS AND DISCUSSION

Rainfall Prediction

Binary classification prediction method of ANN was used to predict the rainfall days and no rainfall days. The default threshold of 0.5 was used to classify which day will have rainfall and day with no rainfall. The network was trained using 70% (870) of the entire 1242 data set. After adequate training of the network, it was followed by testing using 15% (186) of the entire data. The summary of the result of the first iteration is presented in Table 2. The first column shows the individual day under consideration, the second column shows the true label for the day whether it rained or not. The third column shows the actual prediction of the network whether it rained or not while the fourth column is the confusion matrix (CM) outcome of the prediction.

The outcome of Table 2 is the first-round result obtained. As expected, several rounds of results were obtained by re-running the test data. The respective outcome was analysed and best results extracted. Same procedure as stated was carried out for nine more iterations. The individual prediction results (similar to Table 2). A summary of the confusion matrix obtained by summing the number of P, N, TP, FP, TN and FN for each round of the ten rainfall prediction results was compiled and presented in Table 3.

Table 2: Summary of Rainfall Prediction Result

DAYS	TRUE LABEL	PREDICTED LABEL	CM
1	0	1	FP
2	0	0	TN

3	0	0	TN
4	0	0	TN
5	1	0	FN
6	0	0	TN
7	0	0	TN
8	0	0	TN
9	1	0	FN
10	0	0	TN
:	:	:	:
183	1	0	FN

Table 3: Summary of Rainfall Prediction for Ten Iterations

S/N	P	N	TP	FP	TN	FN
1	59	127	18	41	113	14
2	29	157	10	19	135	22
3	25	161	13	12	142	19
4	66	120	20	46	108	12
5	32	154	17	15	139	15
6	35	151	14	21	133	18
7	42	144	13	29	125	19
8	53	133	16	37	117	16
9	38	148	12	26	128	20
10	73	113	16	57	97	16

Table 4 Rainfall Prediction Analysis

S/N	P	N	ACCURACY	ERROR	SENSITIVITY	SPECIFICITY
1	59	127	70%	1.3%	56%	86%
2	29	157	78%	2.2%	31%	93%
3	25	161	83%	2.6%	41%	92%
4	66	120	69%	1.3%	63%	84%
5	32	154	84%	2.0%	53%	89%
6	35	151	79%	1.9%	44%	90%
7	42	144	74%	1.7%	41%	91%
8	53	133	72%	1.4%	50%	88%
9	38	148	75%	1.8%1	38%	91%
10	73	113	61%	1.3%	50%	86%

Prediction Analysis

The objective of the prediction analysis is to help identify which of the ten rainfall predictions made is most accurate. The analysis involves those of Accuracy, Sensitivity, Specificity and Error, earlier presented in equations (12) to (15) respectively. P symbolise positive (number of rainfall days) and N symbolises negative (number of no-rain days). The results of rainfall prediction analysis for the ten iterations are presented in Table 4.

Sensitivity is the probability of a particular day to have rainfall, specificity is the probability of a particular day with no rainfall, accuracy tells how often the network classifies correctly while error rate is how often the network misclassifies. Sensitivity value is the best indicator since it gives the probability of having rainfall, a factor which this work is interested in predicting. The higher the sensitivity, the higher the probabilities of rainfall for those days. Inspecting the sensitivity column of Table 4, the 4th iterations result has the highest sensitivity of 63%, accuracy of 69% (not the best anyway) and the lowest error of 1.3%; proving to be the best choice for the rainfall prediction of Minna. 66 days were predicted with rainfall and 120 days with no-rainfall. However, the test target used for the prediction has 32 days of rainfall and 154 days with no rainfall. Implying that the prediction is 34 days in favour of rainfall.

CONCLUSION

This work predicted the occurrence of rain in Minna metropolis using the default threshold classification method of Artificial Neural Network with atmospheric parameters of relative humidity, minimum temperature and maximum temperature serving as input into the network. The result of the fourth iteration with a prediction accuracy of 69%, sensitivity of 63%, specificity of 84% and an error value of 1.3% was selected as the best rainfall prediction for Minna. It predicted a total of 66 rainfall days. This number is greater than the number of rainfall days captured in the data set. The implication of this result is that more rainfall days are expected in the metropolis which could lead to flooding. The magnitude of such flooding is anticipated to be severe when compared with those experienced in the year 2012, 2014 2017 and 2018. This result serves as an alarm to residents and a wakeup call to relevant stakeholders to embark on mitigation in order to minimise the negative impact of more rainfall days and to increase the societies resilience to hazards such as flooding and erosions. In view of the seriousness of this indicator, it is therefore recommended that, the atmospheric inputs be expanded to incorporate other relevant atmospheric parameters such as wind speed, wind direction, dew point, average temperature among others for a more encompassing rainfall prediction. Secondly, for higher accuracy of rainfall prediction in the metropolis, meteorological data for at least thirty years should be acquired. The performances of Artificial Neural Network improve when more data of relevant input data are employed for initial training.

The focus of future research is to attempt the prediction using the multiple threshold classification method of ANN. This will reveal if there exist a threshold between 0,1 and 1.0 which will perform better than the default threshold of 0.5 used in this work.

Acknowledgement

The authors want to acknowledge the Department of Geography, Federal University of Technology, Minna, Nigeria in respect of the data acquired for this research.

REFERENCES

Abdulkadir, T. S., Salami, A. W., & Kareem, A. G. (2012). Artificial Neural Network Modelling of Rainfall in Ilorin, Kwara State, Nigeria. *Journal of Research Information in Civil Engineering.*

Abisoye, O.A. & Jimoh, R.G. (2018). Comparative Study on the Prediction of Symptomatic and Climatic based Malaria Parasite Counts Using Machine Learning Models", *International Journal of*

Modern Education and Computer Science (IJMECS), Vol.10, No.4, pp. 18-25, 2018.DOI: 10.5815/ijmecs.2018.04.03
Aftab, S. and Ahmad, M. (2018). Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review (IJACSA) *International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5.*
Brownlee, J. (2020). Four Types of Classification tasks in Machine learning. Received from: www.machinelearningmastery.com
Bukka, U. A., Muhammad, B., & Yahaya, T. I. (2017). Effect of flooding on livelihood of Communities in Muwo District, Mokwa Local Government Area, Niger State, Nigeria. *International Journal of Scientific & Engineering Research Volume 8, 1615 ISSN 2229-5518*
Fokas, A., Grigoryan, A., Kubble, T., Zeggarhnski, B. (2018). Mathematical Physics. Imperial College Press. P 5-6.
Geo Dept. FUT Minna (2021). Department of Geography, Federal University of Technology, Minna.
Halliday, D., Sesnick, R. and Walker, J. (2006). Fundamentals of Physics. 5th Edition. Wiley, India. P 456.
Jerzy, L. (2006). Computational Chemistry. Volume 10, World Scientific Publishing. p 121
Julie M. D., and Kannan B., (2010). Significance of Classification Techniques in Prediction of Learning Disabilities. Cochin- 683 107, India.
Kumar, S. (2020). Data Splitting Technique to fit any Machine Learning Model. Received from: www.towardsdatascience.com
Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. Copyright C John Wiley & Sons, Inc. ISBN 0-471-66657-2
Pepin, N. C. (2017). A comparison of simultaneous temperature and humidity observations from the SW and NE slopes of Kilimanjaro: The role of slope aspect and differential land-cover in controlling mountain climate. Global and Planetary Change. Vol. 157, pp:244-258,
Umar, A. T. (2012). Analysis of Extreme Rainfall Events and Risk of Drought and Flood occurrences in Nigeria. *Nigeria Geographical Journal, Volume 8(2).*
Theerthagiri, P., Jeena, J., Usha, A, R., & Vamsidhar., (2020). Prediction of COVID-19 Possibilities using KNN Classification Algorithm. Received from: www. Researchsquare.com/articl/ rs-70985/v2 doi:10.21203/rs.3.rs-70985/v2
Tyagi, N. and Kumar, A. (2017). "Comparative Analysis of Backpropagation and RBF Neural Network on Monthly Rainfall Prediction, International. Conference Invention Computer Technology. ICICT vol. 1