Predicting rainfall runoff in Southern Nigeria using a fused hybrid deep learning ensemble

Arnold Adimabua Ojugo[1], Patrick Ogholuwarami Ejeh[2], Christopher Chukwufunaya Odiakaose[2], Andrew Okonji Eboka[3], Frances Uchechukwu Emordi[4]
[1]Department of Computer Science, College of Science, Federal University of Petroleum Resources Effurun, Warri, Nigeria
[2]Department of Computer Science, Faculty of Computing, Dennis Osadebay University, Asaba, Nigeria
[3]Department of Computer Science Education, Federal College of Education Technical, Asaba, Nigeria
[4]Department of Cyber Security, Faculty of Computing, Dennis Osadebay University, Asaba, Nigeria

Article Info

Article history:

Received Dec 26, 2022
Revised Apr 8, 2023
Accepted Apr 24, 2023

Keywords:

Deep learning
Nigeria
Optimization tasks
Profile hidden Markov
Rainfall runoff

ABSTRACT

Rainfall as an environmental feat can change fast and yield significant influence in downstream hydrology known as runoff with a variety of implications such as erosion, water quality, and infrastructures. These, in turn impact the quality of life, sewage systems, agriculture, and tourism of a nation to mention a few. It chaotic, complex, and dynamic nature has necessitated studies in the quest for future direction of such runoff via prediction models. With little successes in use of knowledge driven models, many studies have now turned to data-driven models. Dataset is retrieved from Metrological Center in Lagos, Nigeria for the period 1999-2019 for the Benin-Owena River Basin. Data is split: 70% for train and 30% for test. Our study adapts a spatial-temporal profile hidden Markov trained deep neural network. Result yields a sensitivity of 0.9, specificity 0.19, accuracy of 0.74, and improvement rate of classification of 0.12. Other ensembles underperformed when compared to proposed model. The study reveals annual rainfall is an effect of variation cycle. Models will help simulate future floods and provide lead time warnings in flood management.

Corresponding Author:

Arnold Adimabua Ojugo
Department of Computer Science, College of Science, Federal University of Petroleum Resources Effurun
PMB 1221 FUPRE road, Ugbomro, Effurun, 330102, Warri, Delta, Nigeria
Email: ojugo.arnold@fupre.edu.ng

1. INTRODUCTION

Rainfall runoff predictions have since become a critical issue, especially with the deluge around Benin-Owena River Basin of Nigeria that occurred from 2015 through 2018 and again in the second and third quarters of 2022. Many states witnessed the general displacement of their citizens across the nation [1], [2]. Thus, runoff prediction has become also a critical feat in planning and executing farming policies. Such predictions are possible via the use of mathematical models to yield knowledge and data-driven algorithms. Rainfall is often forecasted primarily via quantifying the runoff. The dynamic nature of environmental issues has become of great concern to its awareness. As such, these models must account for trending challenges and meet the new requirements to deal with its related tasks that includes (and not limited to) land degradation, pollution, erosion, flood resource management, land-use consequence, and climate changes [3], [4].

Rainfall has a significant influence on downstream hydrology and flooding resulting from runoff with a range of complications for water quality, land-use structures, agriculture, sewage system, tourism, and in general impacts on the quality of life [5]. With these, early warning of such is both critical and imperative in managing water resources [6]–[8]. The chaotic and complex nature of environmental processes makes runoff modeling and prediction a difficult task [6]–[8], despite the various advances in weather predictions and the accurate prediction of runoff is often challenging and germaine [9], [10] in operational hydrology.

Our study is motivated by [11]–[14] noting that: i) many models still have the issues of calibration and model validation resulting from the limited availability of datasets vis-à-vis the heterogeneity of the rainfall scheme that poises the model to relearn feats and parameters that are often difficult to understudy [15], [16] and ii) formulating such optimization tasks often requires carefully selected parameters–and yield an outcome that may amend previously considered variables. The careful selection of hyperparameters will yield an optimal solution, devoid a model of over-parameterization, and overfitting as well as vary with each problem domain [5], [17]. To overcome the stated pitfalls, we propose hybrid deep-learning runoff ensemble with the Benin-Owena River Basin development authority (BORDA) dataset retrieved from the National Metrological Centre in Lagos State, Nigeria.

2. LITERATURE REVIEW

2.1. Review of related literature

Globally, scientists in quest to actualize knowledge-driven and stochastic models, are often posied to ensure cum enhance an accurate prediction cum forecasts of rainfall [9]. Recent efforts are focused on using the auto-regressive moving average approach and in some cases yield such optimal solutions with the use of exogenous variables for multi-objective functions used to represent runoff hydrology datasets [18]. Ojugo *et al.* [19] used a gravitational search fused neural network model with observed data from the Chad River Basin for the period 1996-2007. Model had accuracy 0.97, sensitivity 0.68, and specificity 0.82 with 58%, 24%, 56%, and 42% respectively as computed coefficient of efficiency (COEs) for 4-stations being understudied. It was observed that rainfall results vary from long-term runoff with significant correlation between rainfall and runoff. The trained model thus, yields lead time warning for flood management and simulated future flood/runoff.

Durowoju *et al.* [20] used the autoregressive-3 (AR-3) model that yielded significant correlations of rainfall with cloud cover, humidity, and temperature difference. With sunshine sensitive to impulse response functions, they used a 4-TF model and forecasted rainfall with a 0.023 root mean square error (RMSE) as best suited for the model. This was found to outperform the univariate and multiple regression seasonal autoregressive integrated moving average (SARIMA) models. Ngene *et al.* [15] used generalized autoregressive conditional heteroskedasticity (GARCH) with the Chad Basin dataset from 1996 to 2007 on rainfall, temperature difference, relative humidity, sunshine, and cloud cover. The study established a significant association with rainfall for humidity, temperature difference, and cloud. Using impulse response functions GARCH (1, 0, 1) for predicting rainfall with RMSE of 2.3% as the most appropriate. When compared, we agree that the model performs better than both multiple regression and univariate SARIMA (1, 0, 1)*(1, 0, 1) models.

2.2. Data gathering

The selected area is the BORDA Nigeria. It has a land-mass of 22,045 km^2, a 1,023 mm annual mean rain, and 3.8 m/1.5 m^3/s perennial discharge for its dry/peak periods respectively. Figure 1 reflects the time-plot within the period 1999-2019, for which we see fragment starts during the period of constant low rainfall. Table 1 shows the detailed description of the BORDA dataset with the various features such as rainfall, temperatiure difference, and mean humidity.

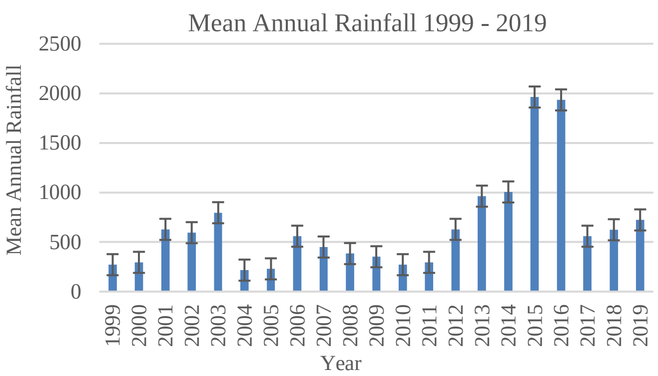

Figure 1. Clustered time plot of annual mean rain for the BORDA

Table 1. Detailed summary sheet of rainfall features for 1999–2019

Year	Rainfall in mm	Temperature difference TMax	TMin	Mean humidity	Mean sunshine in hours	Mean WindSpeed in mtrs/sec	Wind direction
1999	271.4	31.567	22.909	78.901	3.256	2.902	SW
2000	295.1	32.092	23.405	76.902	3.761	3.508	S
2001	628.9	31.533	23.508	83.000	3.021	2.892	W
2002	594.4	32.017	23.817	85.200	2.994	2.858	SW
2003	795.7	31.575	23.703	83.134	5.012	2.917	W
2004	216.4	31.733	24.442	79.013	4.561	3.375	SW
2005	229.4	31.567	23.468	85.301	4.092	2.935	SW
2006	558.8	32.092	24.501	79.34	4.432	3.451	SW
2007	449.6	31.917	23.908	81.211	3.895	3.209	S
2008	383.4	32.042	24.091	83.120	4.501	3.021	S
2009	351.7	31.575	23.508	83.753	4.458	3.508	NE
2010	271.4	31.733	23.717	83.917	5.067	2.892	W
2011	295.1	31.567	23.700	83.751	4.433	2.858	SW
2012	628.0	32.092	24.042	83.667	3.850	2.917	S
2013	963.0	31.533	23.458	83.667	4.042	3.375	SW
2014	1005.0	32.017	23.183	83.583	3.883	3.733	SW
2015	1963.1	31.458	23.617	81.501	2.933	3.3	S
2016	1934.1	32.142	23.842	84.751	4.358	3.058	SW
2017	558.8	31.917	23.317	85.167	4.001	2.825	S
2018	623.9	32.042	24.825	83.001	4.158	2.983	S
2019	723.1	31.558	23.483	81.333	4.575	3.15	W

2.3. Hybrid deep learning reinforcement ensemble

We use a hybrid deep learning ensemble as seen in the Figure 2 to be grouped as a component with 3-basic blocks as adapted from [14], [21]. The deep learning modular memetic network ensemble is divided into 3-basic models namely: i) the unsupervised deep learning Kohonen modular neural network, ii) the supervised cultural genetic algorithm (CGA), and iii) the knowledgebase, respectively.

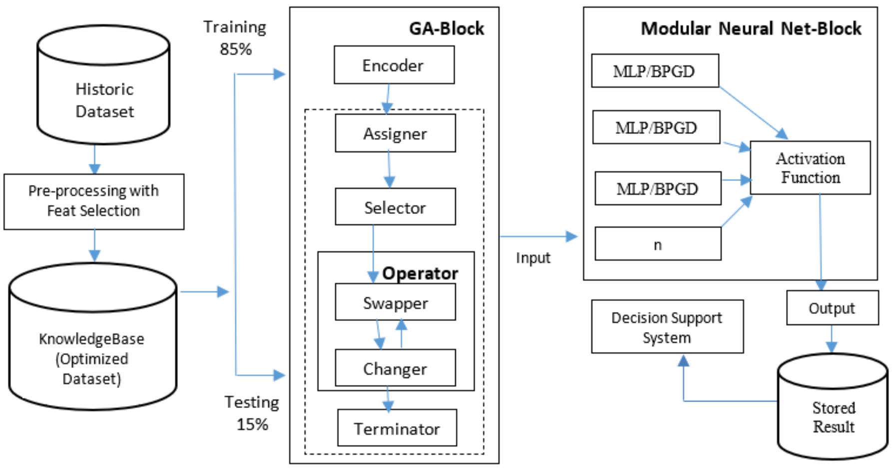

Figure 2. Deep learning modular memetic algorithm

2.3.1. Supervised cultural genetic algorithm

Genetic algorithm (GA) is inspired by the survival of the fittest (elitist) syndrome via a chosen population of potential solutions [22]–[25]. For a task, each candidate solution in the space is yielded using 4-basic operators [26], [27]. Candidates with genes (values) close to its optimal solution and/or the objective function) are said to be fit, as determined by its fitness function. The 4-basic operators can be found in [28], [29]. The variant cellular genetic algorithm (CGA) has 4-belief spaces namely: i) norm belief specifies particular values each rule must fall within, ii) domain belief shows information about the task, iii) temporal belief shows all available information about the task, and iv) spatial belief shows coverage data about the task as in [30], [31]. Also, CGA has an influence function ensures that candiates (rules) values must conform to thebelief-space(s). Afterwards, CGA generates new population with values that are bound to (i.e. do not violate) its belief space. These, in turn, helps reduce number of possible candidates generated till an optimum is reached [32]–[35].

2.3.2. Kohonen modular neural networ

The Kohonen modular neural networ (MNN) is a gridlike, feed-forward network whose first layer accepts input, and re-sends unbound to its second layer, which uses the transfer function to offer competitive computation. The competitive layer then maps similarity patterns into relations. Pattern relations noticed are used to determine the result after training [21], [36]. We modify the parameters and carefully create our deep-learning Kohonen MNN via a deep architecture [37]. Our deep learning is achieved by training the network via 2-stages: the pre-trained and fine-tuned processes [38] and is adapt from [39] and used as the experimental ensemble.

2.3.3. Experimental framework

The experimental framework is trained as [40]:

- Input data is received from the storage unit and sent to GA-unit (consisting of encoder, selector, swapper recombiner, swapper mutator, and belief terminator). Each phase yields a fundamental operation in the CGA to help train the dataset. In optimizing, dataset feats are held within a knowledgebase as operational data for the learning process [41]–[43].
- Our modular network receives the rules-dataset, which is then grouped as successive labelled instances (references). The classifier then passes the if-then rules values of selected parameters into data-point clusters. With rules modeled as a production system, the block has 4-components namely: i) a ruleset of rules, how each rule is patterned and the task therein, ii) knowledgebase of if-then rules selected as data features/parameters, iii) a control strategy to determine the order of execution of stored rules when it finds a match and how to resolve conflicts when/if several rules are matched simultaneously, and iv) a rule applier. The MNN as a component analyzer yields a self-learning block with rules optimized via crossover and mutation, enabling the trained ensemble to effectively, and predict the runoff values [44], [45].
- Lastly, the network acts as a decision support with predicted values (output) and the automatic update of rules-knowledgebase, as transactions are encountered with new data and thus classified.

Model is first, initialized with 30-selected if-then fit rules - which are then selected via the tourney approach as genes of the same parent. Ensemble uses the 2-point crossover to learn the dynamic, complex, and non-linear underlying feats of interest within the dataset. As we accept new off springs, a new pool emerges via mutation [46]. We then select 3-random rules and allocate new values (from 0-to-1) to confirm and not violate the ensemble's beliefs. With each time-stamped runoff data representing a value, selection is done for via the MNN to ensure the norm-domain-and-temporal beliefs is met. While, mutation (its number which determines how close model is to optimal solution) ensures spatial belief is met [1], [28]. All these are determined by the influence function (of rules and chromosome candidate) - to yield results as trained using the hybrid ensemble. These, impacts on how the ensemble is processed and stops when best rule has fitness score of 0.8 or higher.

3. FINDINGS AND RESULT DISCUSSION

3.1. Model evaluation

An ensemble's predictive capability is identified via 15-recorded/annotated labels for CGA-optimized runoff dataset. In prediction, we measure an ensemble's performance via confusion matrix as: i) sensitivity measures how good a model correctly classifies data with incorrectly classified labels present, ii) specificity measures how good a model will detect the absence of incorrect data-points when it is not present in the dataset, and iii) accuracy measures the proportion of true results seen as the degree of truth of a prediction. With TP=43, TN=3, FP=11, FN=5, and using (1)-(3), our computed values yields.

$$Sensitivity = \frac{TP}{TP+FN} = \frac{43}{43+5} = 0.90 \quad (1)$$

$$Specificity = \frac{TN}{FN+FP} = \frac{3}{11+5} = 0.185 \quad (2)$$

$$Accuracy\ F1 = \frac{TP+TN}{TP+TN+FP+FN} = \frac{43+3}{43+3+11+5} = 0.74 \quad (3)$$

Proposed ensemble resulted in a sensitivity of 0.9, with specificity 0.185, accuracy 0.74, and a 0.12 rate of improvement for data (not included from outset) used to train/test the ensemble as in Figure 3.

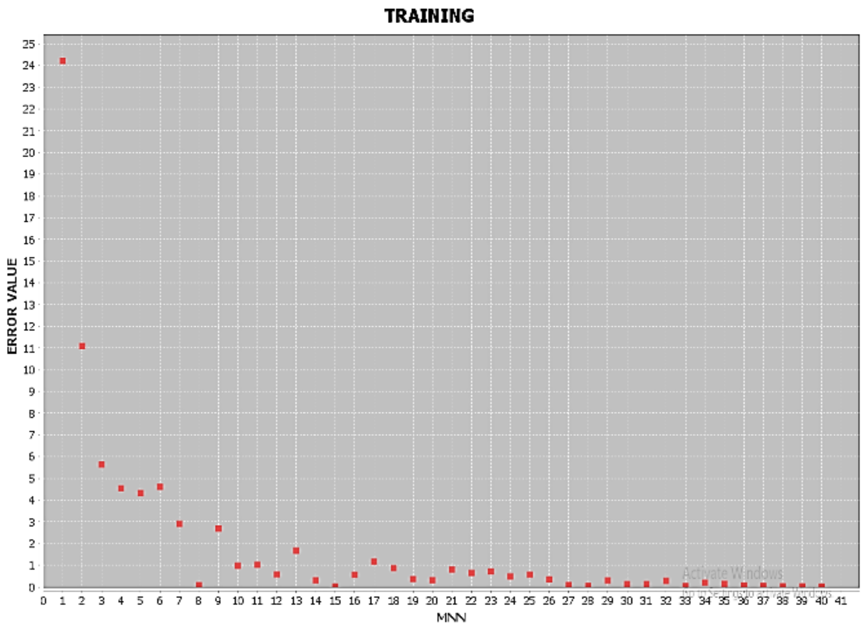

Figure 3. Hybrid ensemble training-phase result

3.2. Result findings

Training ensemble used the feedforward in time backpropagation learning algorithm for each phase until a finite epoch is reached. Training phase was noted to have reached its equilibrium at 40-epochs as in Figure 4 which represents training phase for the ensemble. Figure 4 show futures-rainfall prediction direction for the monthly forecast for 2023. For 2023, the ensemble shows a volatility varies between the ranges [0.412, 2.092] for the 12-months period (i.e. 52-weeks period). Thus, we witness an increase in rainfall rather than a drop in the runoff values in the near future. The results holds same for [3], [4], [16]. This may be have been possible through the change in condition due to model training via older dataset.

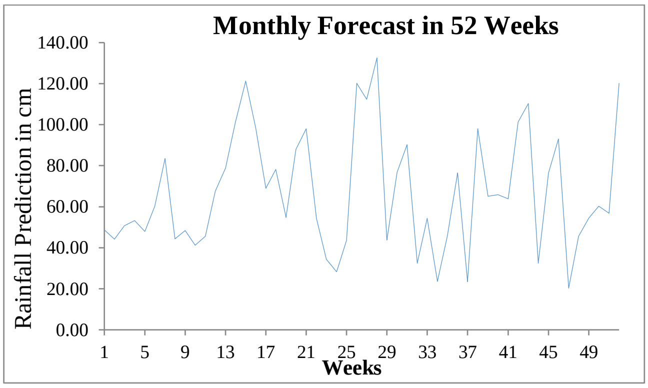

Figure 4. Futures rainfall runoff direction and volatility

3.3. Discussion of findings

Hybrid ensembles are quite challenging to implement due to a variety of issues such as: i) data encoding conflict from one algorithm to another within the ensemble, ii) there is also the issue of the underlying features of interest generated for each candidate solution, and iii) resolving of structural dependencies imposed on the ensemble by features in the dataset not contained from the outset. These, must be resolved for the ensemble to yield an optimal solution. Most modelers must select the requisite and appropriate parameter(s) to avoid ensemble over-fitting and over-training. Furthermore, the effects of such ensemble/hybrid is to prevent agents within a multi-goal tasks such as this, from creating and enforcing their own behavioral rules on the dataset during training.

4. CONCLUSION

Models are useful representations of a realistic system. Its primary goal is to posit an educational tool that provisions the right insight that helps a researcher to better understand a symmetric reflection of the reality the work portends. They also help advance existing knowledge to researchers yielding a new language that seeks to communicate hypotheses. Thus for this study, we only require a reasonably detailed and applicable model. To investigate hypotheses, parametric inputs are crucial and must be correctly estimated and finding the underlying probabilities. In addition, our interest must align with the ensemble's implementation as a feedback scheme via its prediction capabilities rather than its yielding numeric agreement for various observations.

REFERENCES

[1] Z. Abda, B. Zerouali, M. Chettih, C. A. Guimarães Santos, C. A. S. Farias, and A. Elbeltagi, "Assessing machine learning models for streamflow estimation: a case study in Oued Sebaou watershed (Northern Algeria)," *Hydrological Sciences Journal*, vol. 67, no. 9, pp. 1328–1341, Jul. 2022, doi: 10.1080/02626667.2022.2083511.
[2] A. T. Abdulrahim *et al.*, "Rainfall-runoff modeling for Challawa and Jakara catchment areas of Kano city, Nigeria," *Arid Zone Journal of Engineering, Technology & Environment*, vol. 17, no. June, pp. 211–220, 2021.
[3] A. Adegede, K. Ogbu, V. Ogwo, and C. Mbajiorgu, "Comparison of ACRU and HEC-HMS models in runoff prediction in a watershed, Southern Nigeria," *Journal of Agricultural Engineering and Technology*, vol. 22, no. 4, pp. 142–149, 2014.
[4] O. Adeola Fashae, A. Olusola, and V. Onyemaenu, "Rainfall–runoff in conterminous tropical river basins of Southwestern Nigeria," *African Geographical Review*, vol. 42, no. 1, pp. 14–28, 2023, doi: 10.1080/19376812.2021.1996250.
[5] A. Oludapo Olusola, O. Olumide, O. Adeola Fashae, and S. Adelabu, "River sensing: the inclusion of red band in predicting reach-scale types using machine learning algorithms," *Hydrological Sciences Journal*, vol. 67, no. 11, pp. 1740–1754, Aug. 2022, doi: 10.1080/02626667.2022.2098752.
[6] J. A. Aper, S. A. Iorkua, and N. J. Akegh, "Relationship between rainfall, surface runoff and soil loss on an experimental farm in the middle belt of Nigeria," *Nigerian Annals of Pure and Applied Sciences*, vol. 1, pp. 122–139, 2019, doi: 10.46912/napas.116.
[7] A. Awotwi, G. K. Anornu, J. A. Quaye-Ballard, and T. Annor, "Monitoring land use and land cover changes due to extensive gold mining, urban expansion, and agriculture in the Pra River Basin of Ghana, 1986–2025," *Land Degradation and Development*, vol. 29, no. 10, pp. 3331–3343, Oct. 2018, doi: 10.1002/ldr.3093.
[8] R. C. Carter and A. Parker, "Climate change, population trends and groundwater in Africa," *Hydrological Sciences Journal*, vol. 54, no. 4, pp. 676–689, Aug. 2009, doi: 10.1623/hysj.54.4.676.
[9] H. Chen, S. Guo, C. Xu, and V. P. Singh, "Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin," *Journal of Hydrology*, vol. 344, no. 3–4, pp. 171–184, Oct. 2007, doi: 10.1016/j.jhydrol.2007.06.034.
[10] G. C. Okafor, O. D. Jimoh, and K. I. Larbi, "Detecting changes in hydro-climatic variables during the last four decades (1975-2014) on downstream Kaduna River Catchment, Nigeria," *Atmospheric and Climate Sciences*, vol. 07, no. 02, pp. 161–175, 2017, doi: 10.4236/acs.2017.72012.
[11] A. Ifeka and A. Akinbobola, "Trend analysis of precipitation in some selected stations in Anambra State," *Atmospheric and Climate Sciences*, vol. 05, no. 01, pp. 1–12, 2015, doi: 10.4236/acs.2015.51001.
[12] C. M. Igwenagu, "Trend analysis of rainfall pattern in Enugu State, Nigeria," *European Journal of Statistics and Probability*, vol. 3, no. 3, pp. 12–18, 2015.
[13] A. I. Iwara, F. O. Ogundele, and M. S. Ajisegiri, "Rainfall-runoff-sediment relationship in vegetation fallows in the rainforest zones of Southern Nigeria," *Nigerian Journal of Geography and the Environment*, vol. 4, no. 1, pp. 88–100, 2019.
[14] A. A. Ojugo and O. Nwankwo, "Forging a spectral-clustering multi-agent hybrid deep learning model to predict rainfall runoff In Nigeria," *International Journal of Innovative Science, Engineering and Technology.*, vol. 8, no. 3, pp. 140–147, 2021.
[15] B. U. Ngene and N. Obianigwe, "Nigerian rain gauge station optimization and national development: the importance of head count," *IOP Conference Series: Materials Science and Engineering*, vol. 413, no. 1, pp. 1-7, 2018, doi: 10.1088/1757-899X/413/1/012022.
[16] O. A. Obiora-Okeke, J. R. Adewumi, and O. M. Ojo, "Impact of climate change on runoff prediction in ogbese river watershed," *FUOYE Journal of Engineering and Technology*, vol. 6, no. 4, pp. 392–396, 2021, doi: 10.46792/fuoyejet.v6i4.721.
[17] O. G. Chinwendu, S. O. E. Sadiku, A. O. Okhimamhe, and J. Eichie, "Households vulnerability and adaptation to climate variability induced water stress on downstream Kaduna River Basin," *American Journal of Climate Change*, vol. 06, no. 02, pp. 247–267, 2017, doi: 10.4236/ajcc.2017.62013.
[18] G. C. Okafor and K. N. Ogbu, "Assessment of the impact of climate change on the freshwater availability of Kaduna River basin, Nigeria," *Journal of Water and Land Development.*, vol. 38, no. 1, pp. 105–114, Sep. 2018, doi: 10.2478/jwld-2018-0047.
[19] A. A. Ojugo, R. E. Yoro, E. O. Okonta, and A. O. Eboka, "A hybrid artificial neural network gravitational search algorithm for rainfall runoffs modeling and simulation in hydrology," *Progress in Intelligent Computing and Applications*, vol. 2, no. 1, pp. 22–34, 2013, doi: 10.4156/pica.vol2.issue1.2.
[20] O. S. Durowoju, A. O. Olusola, and B. W. Anibaba, "Rainfall–runoff relationship and its implications on lagos metropolis," *Ife Research Publication in Geography*, vol. 16, no. 1, pp. 25–33, 2018.
[21] G. Behboud, "Reasoning using modular neural network," *Towards Data Science*, vol. 34, no. 2, pp. 12–34, 2020.
[22] N. Tomar and A. K. Manjhvar, "A survey on data mining optimization techniques," *IJSTE-International Journal of Science Technology and Engineering*, vol. 2, no. 06, pp. 130–133, 2015.
[23] Q. Li *et al.*, "An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–15, 2017, doi: 10.1155/2017/9512741.
[24] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *Journal of Big Data*, vol. 9, no. 1, p. 1-17, Dec. 2022, doi: 10.1186/s40537-022-00573-8.
[25] J. Jung, M. Maeda, A. Chang, M. Bhandari, A. Ashapure, and J. Landivar-Bowles, "The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems," *Current Opinion in Biotechnology*, vol. 70, pp. 15–22, 2021, doi: 10.1016/j.copbio.2020.09.003.

[26] J. R. Amalraj and R. Lourdusamy, "A novel distributed token-based access control algorithm using a secret sharing scheme for secure data access control," *International Journal of Computer Networks and Applications*, vol. 9, no. 4, p. 374-384, Aug. 2022, doi: 10.22247/ijcna/2022/214501.
[27] T. Ma, F. Wang, J. Cheng, Y. Yu, and X. Chen, "A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks," *Sensors*, vol. 16, no. 10, p. 1-23, Oct. 2016, doi: 10.3390/s16101701.
[28] A. A. Ojugo, A. O. Eboka, E. O. Okonta, R. E. Yoro, and F. O. Aghware, "Genetic algorithm rule-based intrusion detection system (GAIDS)," *Journal of Emerging Trends In Computing Information Systems*, vol. 3, no. 8, pp. 1182–1194, 2012.
[29] S. Chouhan, D. Singh, and A. Singh, "An improved feature selection and classification using decision tree for crop datasets," *International Journal of Computer Applications*., vol. 142, no. 13, pp. 5–8, May 2016, doi: 10.5120/ijca2016909966.
[30] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, "Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning," *IEEE Access*, vol. 7, pp. 115749–115759, 2019, doi: 10.1109/ACCESS.2019.2931637.
[31] J. Jung, M. Maeda, A. Chang, M. Bhandari, A. Ashapure, and J. Landivar-Bowles, "The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems," *Current Opinion in Biotechnology.*, vol. 70, pp. 15–22, Aug. 2021, doi: 10.1016/j.copbio.2020.09.003.
[32] A. F. Psaros, X. Meng, Z. Zou, L. Guo, and G. E. Karniadakis, "Uncertainty quantification in scientific machine learning: methods, metrics, and comparisons," *Journal of Computational Physics*, vol. 477, p. 111902, 2023, doi: 10.1016/j.jcp.2022.111902.
[33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp 1-48, 2019, doi: 10.1186/s40537-019-0197-0.
[34] Y. Kang, M. Ozdogan, X. Zhu, Z. Ye, C. Hain, and M. Anderson, "Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest," *Environmental Research Letter*, vol. 15, no. 6, p. 1-10, Jun. 2020, doi: 10.1088/1748-9326/ab7df9.
[35] S. Ju, H. Lim, and J. Heo, "Machine learning approaches for crop yield prediction with MODIS and weather data," *40th Asian Conference on Remote Sensing: Progress of Remote Sensing Technology for Smart Future, ACRS 2019*, no. Acrs, pp. 1–4, 2020.
[36] L. A. Belanche and F. F. González, "Review and Evaluation of Feature Selection Algorithms in Synthetic Problems," *Arxiv-Computer Science*, vol. 1, pp. 1-13, Jan. 2011.
[37] A. A. Ojugo and R. E. Yoro, "Forging a deep learning neural network intrusion detection framework to curb the distributed denial of service attack," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1498–1509, 2021, doi: 10.11591/ijece.v11i2.pp1498-1509.
[38] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, Sep. 2018, doi: 10.1016/j.jbi.2018.07.014.
[39] A. A. Ojugo and O. Nwankwo, "Modeling mobility pattern for the corona-virus epidemic spread propagation and death rate in Nigeria using the movement-interaction-return model," *International Journal of Emerging Trends in Engineering Research*, vol. 9, no. 6, pp. 821–826, Jun. 2021, doi: 10.30534/ijeter/2021/30962021.
[40] I. Ahmad *et al.*, "Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan," *Journal of the Indian Society of Remote Sensing*, vol. 46, no. 10, pp. 1701–1711, Oct. 2018, doi: 10.1007/s12524-018-0825-8.
[41] Y. Ampatzidis, V. Partel, and L. Costa, "Agroview: Cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence," *Computers and Electronics in Agriculture*, vol. 174, p. 1-12, Jul. 2020, doi: 10.1016/j.compag.2020.105457.
[42] A. Suruliandi, G. Mariammal, and S. P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 27, no. 1, pp. 117–140, 2021, doi: 10.1080/13873954.2021.1882505.
[43] S. Kumar and M. Singh, "Big data analytics for healthcare industry: impact, applications, and tools," *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.
[44] J. Segarra, M. L. Buchaillot, J. L. Araus, and S. C. Kefauver, "Remote Sensing for Precision agriculture: sentinel-2 improved features and applications," *Agronomy*, vol. 10, no. 5, p. 1-18, May 2020, doi: 10.3390/agronomy10050641.
[45] S. Nosratabadi, F. Imre, K. Szell, S. Ardabili, B. Beszedes, and A. Mosavi, "Hybrid Machine learning models for crop yield prediction," *Arxiv-Computer Science*, pp. 1-5, Mar 2020.
[46] D. H. Zala and M. B. Chaudhari, "Review on use of 'BAGGING' technique in agriculture crop yield prediction," *IJSRD - International Journal for Scientific Research & Development*, vol. 6, no. 8, pp. 675–676, 2018.

BIOGRAPHIES OF AUTHORS

Arnold Adimabua Ojugo was born February 28, 1980 to Williams and Queen Ojugo. He received his B.Sc. at computer science in 2000 from the Imo State University Owerri, M.Sc. at computer science in 2005 from Nnamdi Azikiwe University Awka, and Ph.D. at computer science in 2013 from the Ebonyi State University Abakiliki. He is a professor of computer science at the Department of Computer Science, Federal University of Petroleum Resources Effurun, Nigeria. His research interests include: intelligent systems computing, data science, cyber security in IoT, and graphs applications. He is a member of the Nigerian Computer Society, Computer Professionals of Nigeria, and International Association of Engineers. He is happily married to Dr. Prisca Ojugo with five children: Gregory Ojugo, Easterbell Ojugo, Eric Ojugo, Elena Ojugo, and Elizabeth Ojugo. He can be contacted at email: ojugo.arnold@fupre.edu.ng or arnoldojugo@gmail.com.

Patrick Ogholuwarami Ejeh received his higher national diploma (HND) in computer science from the Federal Polytechnic Auchi, Edo State in 2006, his M.Sc. in computer science from Northumbria University, Newcastle, United Kingdom in 2010, and his Ph.D. in computer science from Sunderland University, Sunderland, United Kingdom in 2017. He is currently a lecturer II with the Department of Computer Science in the Faculty of Computing at the Dennis Osadebey University, Asaba, Delta State. His research interests includes artificial intelligence, knowledge management, data science, and IoT. He is also a member Nigerian Computer Society and Higher Education Academic, United Kingdom. He is married to Dr. Chantal Ijeoma Ejeh with three children. He can be contacted at email: mejehpatrick@gmail.com.

Christopher Chukwufunaya Odiakaose received his B.Sc. from The Enugu State University of Science and Technology, Enugu and M.Sc. from the Federal University of Petroleum Resources Effurun in Delta State. He is currently undergoing his doctoral studies with the Department of Computer Science at the Federal University of Petroleum Resources Effurun in Delta State, Nigeria. He currently lectures at Department of Computer Science, Faculty of Computing at the Dennis Osadebay University, Asaba. He has several publications to his credit and his interest is in big-data, machine learning approaches, and user trust modeling. He can be contacted at email: osegalaxy@gmail.com.

Andrew Okonji Eboka received HND in computer science from Akanu Ibiam Federal Polytechnic in the year 1998, Ebonyi State, post graduate diploma from Ebonyi State University in 2013, B.Sc./Ed in computer science education from the Enugu State University of Science and Technology, Enugu in 2013. He received his M.Sc. in network computing from Coventry University, United Kingdom. He currently lectures at Department of Computer Science Education at Federal College of Education Technical, Asaba, Nigeria. His research interests include cyber security, ubiquitous computing, and forensics. He is a member of British Computer Society, Association of Computer Machinery, Computer Professionals of Nigeria, Nigerian Computer Society, and International Association of Engineers. He can be contacted at email: ebokaandrew@gmail.com.

Frances Uchechukwu Emordi received her B.Sc. computer science in 2015 from Michael Okpara University of Agriculture in Umudike in Abia State and her M.Sc. computer science in 2021 with special interest in cyber security from the Federal University of Petroleum Resources Effurun, Delta State in Nigeria, respectively. She currently lectures at Department of Cyber Security at the Faculty of Computing, Dennis Osadebay University, Asaba. Mrs. Emordi is currently a member of Nigerian Computer Society and Council for Registration of Computer Professionals of Nigeria. She can be contacted at email: emordi.frances@dou.edu.ng.