Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA
© Creative Research Publishers
Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

Hybrid Genetic Algorithm Trained Bayesian Ensemble for Short Messages Spam Detection

[1]Odiakaose Christopher, [2] Emordi Frances, [3]Ejeh Patrick, [4]Ashioba Nwanze, [5]Odeh Christopher, [5]Attoh Obiageli & [6]Azaka Maduabuchuku

[1,3,4,5,6] Dept of Computer Science, Dennis Osadebay University Anwai-Asaba, Nigeria.
[2,5]Department of Cybersecurity, Dennis Osadebay University Anwai-Asaba, Nigeria.
E-mails: osegalaxy@gmail.com, emordi.frances@dou.edu.ng, patrick.ejeh@dou.edu.ng, ashioba.nwanze@dou.edu.ng,odeh.christopher@dou.edu.ng, Attoh,obiageli@dou.edu.ng azaka.maduabuchuku@dou.edu.ng
,

ABSTRACT

Today's popularity of the short messages services (SMS) has created a propitious environment for spamming to thrive. Spams are unsolicited advertising, adult-themed or inappropriate content, premium fraud, smishing and malware. They are a constant reminder of the need for an effective spam filter. However, SMS limitations of 160-charcaters and 140-bytes size as well as its being rippled with slangs, emoticons and abbreviations further inhibits effective training of models to aid accurate classification. The study proposes Genetic Algorithm Trained Bayesian Network solution that seeks to normalize noisy feats, expand text via use of lexicographic and semantic dictionaries that uses word sense disambiguation technique to train the underlying learning heuristics. And in turn, effectively help to classify SMS in spam and legitimate classes. Hybrid model comprises of text preprocessing, feature selection as well as training and classification section. Study uses a hybrid Genetic Algorithm trained Bayesian model for which the GA is used for feature selection; while, the Bayesian algorithm is used as classifier.

Keywords: Hybrid Genetic Algorithm, Trained Bayesian Ensemble, Short Messages Spam Detection

Odiakaose, C. Emordi, F. Ejeh, P., Ashioba, N., Odeh, C., Attoh, O. & Azaka, M. (2024): Hybrid Genetic Algorithm Trained Bayesian Ensemble for Short Messages Spam Detection. Journal of Advances in Mathematical & Computational Science. Vol. 12, No. 1. Pp 37-52. Available online at www.isteams.net/mathematics-computationaljournal. dx.doi.org/10.22624/AIMS/MATHS/V12N1P4

1. INTRODUCTION

The advent of smartphones with enhance features has contributed to huge adoption of short messaging by users due to its portability, mobility, ubiquity of services and its low cost continues to promote SMSM to become the most used means of communication globally. , Short Message Service (SMS) is text service in mobile communications with protocols that allow exchange of short text messages between fixed line or mobile phone devices (Ojugo & Ekurume, 2021a, 2021b; Yao et al., 2022)Ojugo & Oyemade, 2021; Omede & Okpeki, 2023; Sahmoud & Mikki, 2022). An estimated 23-billion SMS was sent daily in 2014 globally; While, a total 8.3-trillion SMS was sent worldwide in the same year with net market revenue of over $128Billion in 2011. In 2022, the revenue was over $453Billion and has shown over 6.39billion SMS sent and received in Nigeria (Amalraj & Lourdusamy, 2022; Sobers, 2022). Its popularity and the consequent proliferation of SMS platforms, has also witnessed a corresponding rise in unsolicited SMS called spams (Sasikala et al., 2022; Udeze et al., 2022).

The ITU campaign witnessed a rise in the unsolicited commercial adverts as sent to mobile phones via SMS (Ileberi et al., 2022; Thorat et al., 2021). Recent drift from email to SMS spams is attributed to the availability of effective email filters, user awareness and industry collaboration (Ojugo & Eboka, 2018a, 2018b, 2018c, 2020a, 2020b). Spams are unwanted, unsolicited adverts or messages that includes emails, SMS etc – from a sender, sent indiscriminately with no prior relation to a user mostly for commercial reasons (Parker & Flowerday, 2020; Yildiz Durak, 2019). Spams can be adult-themed, inappropriate contents, unsolicited adverts, etc. They have become a great issue – causing great loss to mobile network operators, and users (Kumaraguru et al., 2010).

The overall growth of spams by 300% from 2011 to 2012 worldwide is attributed to the rise in adoption of SMS (Malasowe et al., 2023; Yeboah-Boateng & Amanor, 2014). Also, in Nigeria, it is estimated at 334million spam (Aghware et al., 2023b, 2023a) – which implies that many smartphone users are handicapped by the plethora of spams they receive (Oyemade et al., 2016). So that, aside being distractive and annoying, users need rest from such spam invasion (Ojugo & Nwankwo, 2021a, 2021b). Mobile network operators are geared towards reducing the number of spams over their network as such flooding makes the SMS channel more invasive and less secure (Linh, 2018; Rtayli & Enneya, 2020).

The tremendous rise in SMS is attributed to (Benchaji et al., 2021; Broadhurst et al., 2018; Jayatilaka et al., 2021): (a) trust in the SMS channel, (b) high open rate and average time for response, (c) its low cost of transaction, and (d) its ease and convenience. SMS has great benefit for both subscribers and operators in diverse ways cantered on convenience, flexibility, seamless integration of messaging services and data access. Others may include: (a) delivery of notifications, (b) guaranteed delivery, (c) reliable, low-cost for concise data, (d) ability to screen messages, (e) increases productivity, (f) more function with more user benefits, (g) delivery to many users at same time, (h) ability to receive many data, (i) e-mail generation, (l) creation of user groups, (m) integration with other Internet-apps, and (n) increase in revenue for MNO .

2. MATERIALS AND MACHINE LEARNING FRAMEWORKS

2.1 Spam Filters

SMS spam filters shares similar features and challenges with email spam filters. They are both saddled with the task of real-time filtering efficiency and the option to decide between client-side and or server-side filtering.

The mobile space is also faced with the challenge of overcoming misclassification cost and eliminate false-positives (genuine SMS incorrectly classified as spam by filter), and issue of concept drift in order to evade filters. Thus, most existing approaches of combating SMS spam are imported from successful email-solutions (Wang *et al.,* 2010). But, not all solutions to email spam are applicable to SMS due to the fact that established email spam filters are unable to tackle SMS Spam because performance of email spam filters is seriously degraded when used to filter SMS spam.

This is attributed to its limited 160-character of 140-bytes sized messages. Also, these messages are rife with slangs, symbols, emoticons and abbreviations that inhibit proper classification (Tiago et al, 2016)). To overcome the shortfall of email filters in handling SMS spam successfully, a combination of filtering techniques to reduce noise in SMS and expands the message size – is the focus of this research. Spam filters can be divided into a number of broad categories based on the method used to filter Spam.

They include: list based, challenge/response system, content based, collaborative and Heuristics Based filters.

1. Blacklist: This earliest spam-filtering method seeks to block unwanted messages from an already created list of senders. Blacklists are records of email addresses, Internet Protocol (IP) addresses and phone numbers that have been previously used to send spam. When incoming message arrives, spam filter checks if IP, email address or phone number is on a blacklist. If so, the message is considered spam and rejected. Blacklists ensure known spammers cannot reach users' inboxes. Their only demerit is that they can also misidentify legitimate senders as spammers (Hasib, Motwani and Saxena, 2012).

2. Whitelist: To block spams, whitelist does almost exactly the opposite of blacklist. Rather than scheme to specify which senders to block messages from, it specifies which senders to allow messages from. These addresses/phone numbers are placed on a trusted-users list. Most spam filters uses a whitelist in addition to other spam-fighting feats as a way to cut down on the number of genuine SMS that accidentally get flagged as spam. A very strict filter that uses just whitelist implies that anyone not approved is automatically blocked. Some anti-spams use a variation, called automatic whitelist – in which an unknown sender's email address is checked against a database; if they have no history of spamming, their message is sent to the recipient's inbox and they are added to the whitelist.
    a. Greylist filter works with the assumption that spammers sends batch of messages once. When message from unknown address is received, it blocks and revert a failure delivery to the sending server. If the message is resent, which most legitimate servers do, filter receives it and adds the address/phone number to the list. Although overhead of the filter is low, its demerit is the unjust delay delivery experienced by genuine messages to its recipient (Hasib, Motwani and Saxena, 2012).
    b. Challenge-Response: The filter forces a message sender to prove they are human via some test. This filter blocks undesirable messages by forcing the sender to perform a task before their message is delivered. With task success, the message (and future messages) will be delivered to the recipient; While, failure to complete the challenge after a certain time period, leads to message rejection (Akshay, Shraddha, Sakshi and Ajinkya, 2013).

The most common challenge consists of distorted images and text. To pass the challenge, a human must type the text or arrange the images correctly. With challenge/response false positives can be reduced to barest minimum. Another merit of this approach is in its low system resource requirements, since no CPU-intensive pattern matching is required. However, this approach causes more problems than it solves. For inexperienced or visual handicapped users, the challenges are completely unsolvable. Regular users are provoked by the challenges and choose not to do so since they view it as an unacceptable irritation. Also, automated email that a user would want to receive (travel confirmations, online purchase receipts, etc) are trapped by this approach and never delivered (Process Software report, 2006).

2.2. Motivation / Statement of Problem

Study is motivated (Ifeka & Akinbobola, 2015; Igwenagu, 2015) as thus:

1. Spams have continued to soar with the advent of SMS. The alarming growth rate of spams with SMS popularity have now created a propitious environ for spammers to exploit subscribers; Thus, causing both financial loss and emotional instability as consequences to users, corporate organs and mobile network operator(s).
2. Academic researches and companies are today, faced with the challenge of dealing with SMS spam. A major issue has been that existing approach(es) to resolving SMS spam are imported from successful email anti-spam solutions (Wang *et al.,* 2010). Thus, are quite unable to effectively and efficiently tackle SMS spam successfully – as their performance is seriously hampered and degraded by the parametric feats used to filter spams.
3. The formulation and design of an effective SMS filter has continued to suffered setback(s) due to the inherent reason that SMS filters by design are not as simple as email filters due to its limited size of 160-characters of 140bytes sized data. These amongst other constraints, continue to create rippled impediment in size of feature to be selected for training and consequently contributing to poor learning and classification of learning algorithm.
4. Also, SMS is rippled with abbreviations, slangs and emoticons that inhibit proper classification of words and/or text corpus (Tiago et al, 2016).

To overcome these, we deploy a hybrid SMS filter ensemble to reduce noise in form of slangs, emoticons, abbreviations as well as expand message size to enhance effective classification using text normalization and semantic expansion in SMS spam filtering.

2.3. Data Sampling

Dataset is retrieved from the Knowledge in discovery dataset (KDD-CUP 1999). Also, the dataset is split into: ***training*** (75%), and ***test*** (25%). (Ojugo & Eboka, 2021; Ojugo & Ekurume, 2021b, 2021a; Ojugo & Nwankwo, 2021b).

2.4. The Proposed Memetic Ensemble

GA is inspired by evolution, and consist of population based on potential solutions to a specific task. An individual with gene close to optimal, is fit. Fitness function determines how close an individual is to optimal solution. Each solution is an individual whose optimal is found via 4-operators (Mohd Ibrahim et al., 2022; Tomar & Manjhvar, 2015; Voke et al., 2023).

The basic operators are defined as thus (Ojugo, Abere, et al., 2013; Ojugo, Aghware, et al., 2015a, 2015b):

a. Initialize – Individual data are encoded into forms suitable for selection. Each encodings type used has its merit. Binary encodings are computationally more expensive. Decimal encoding has greater diversity in chromosome and greater variance of pools generated; float-point encoding or its combination is more efficient than binary. Thus, it encode as fixed length vectors for one or more pools of different types. The *fitness* function evaluates how close a solution is to its optimal – after which they are chosen for reproduction. If solution is found, function is *good*; else, is *bad* and not selected for crossover. The fitness function is the only part with knowledge of task. If more solutions are found, the higher its fitness value.

b. Selection – best fit individuals close to optimal are chosen to mate. The larger the number of selected, the better the chances of yielding fitter individuals. This continues until one is chosen, from the last two/three remaining solutions, to become selected parents to new offspring. Selection ensures the fittest individuals are chosen for mating but also allows for less fit individuals from the pool and the fittest to be selected. A selection that only mates the fittest is *elitist* and often leads to converging at local optima.

c. Crossover ensures best fit individual genes are exchanged to yield a new, fitter pool. There are two crossover types (depends on encoding type used): (a) ***simple*** crossover for binary encoded pool. It allows single- or multi-point cross with all genes from a parent, and (b) ***arithmetic*** crossover allows new pool to be created by adding an individual's percentage to another.

d. Mutation alters chromosomes by changing its genes or its sequence, to ensure new pool converges to global minima (instead of local optima). Algorithm stops if optimal is found, or after number of runs if new pools are created (though computationally expensive), or when no better solution is found. Genes may change based on probability of mutation rate. Mutation improves the much-needed diversity in reproduction.

Cultural GA is a variant with a belief space as thus: (a) Normative (has specific value ranges to which an individual is bound), (b) Domain (has data about task domain), (c) Temporal (has data about events' space is available), and (d) Spatial (has topographical data). In addition, an influence function mediates between belief space and the pool – to ensure and alter individuals in the pool to conform to belief space . CGA is chosen to yield a pool that does not violate its belief space and helps reduce number of possible individuals GA generates till an optimum is found (Supriya & Akki, 2021; Ying, 2019; Zhang et al., 2007).

2.5. The Experimental Framework

Figure 1 shows the schematics of the proposed experimental model (Ojugo & Eboka, 2021; Ojugo & Ekurume, 2021b, 2021a; Ojugo & Nwankwo, 2021b), which is explained in sections as:

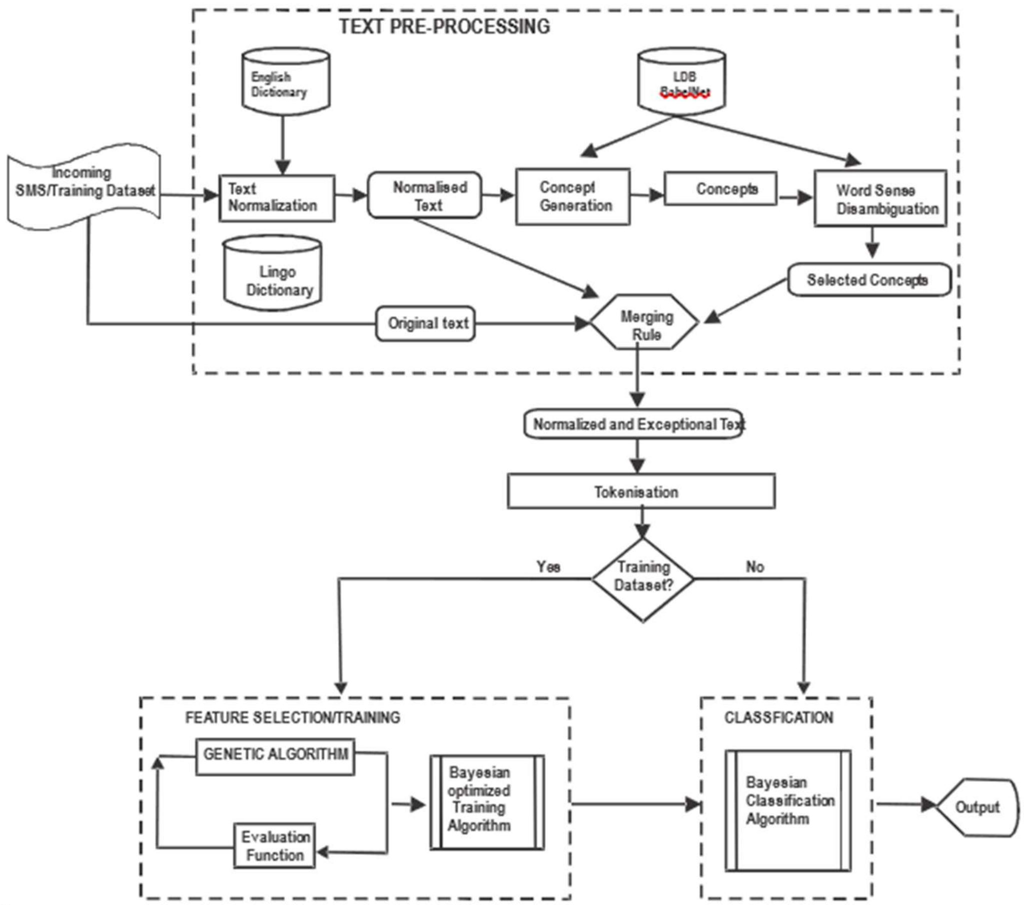

Figure 1. Proposed Memetic Model (Ojugo, Yoro, et al., 2013; Ojugo & Yoro, 2013, 2020, 2021)

Pre-processing Section: is accomplished as:

1. Raw text refers to the original text from the sender for normalization and expansion.
2. Text normalization uses two dictionaries. The first is an English dictionary to check if the text are English so as to then normalize text to its root form. The second is a slang dictionary used to translate slangs into English text. The basic operation of this stage is to replace slangs and abbreviation with standard English words from this dictionary. The Freeling English dictionary and No slang dictionary are proposed.
3. Concepts generation are semantically analyzed already normalized text to deduce their concept. The concepts are provided by Language Data Base BabelNet repository.

4. Word sense disambiguation (WSD): Here, from a variety of concept generated, this stage is used to find the concept that is more relevant according to the context of the original message, among all generated concepts related to a certain words. It equally relies on concepts are provided by Language Data Base (LDB) BabelNet repository
5. Tokenisation breaks down a text corpus into individual elements that serve as input for various natural language processing algorithms. Normalised texts are broken into individual words and stop words and punctuation characters are equally removed in this unit.
6. Merging: Features that define combines of result for pre-processing (original text, normalizing and disambiguation stage). Merging rule answers the question from each stage: (a) should it keep the original token(s)?, (b) should text normalization be performed?, (c) should it perform concepts generation?, and (d) should it perform the word sense disambiguation?
7. Normalized/Expanded text are combined text to obtained from various output from a variety of the pre-processing stages.

Feature Select/Train: The need to minimize the number of features as input parameters used often add to the computational complexity of a model. Thus, our memetic algorithm for feature selection obtained from the text pre-processing section. The input is the dataset (tokens obtained via tokenization of normalized and expanded text from text pre-processing section). It is done as follows:
1. GA Unit yield a rule-based, genetic scheme of the normalized and expanded test defined. The algorithm initializes with a random population, subjected to repetitive use of recombination, mutation, inversion and select operators to improve the generated (new) population.
2. Evaluation contains a fitness function that measures quality of solution. It computes optimal solution by comparing values of the chromosomes against each other using some predefined function.
3. Training Unit: Trains the filter based on Bayes Probability Theorem. It uses known SMS corpus of spam and genuine messages/texts. A collection of tokens appearing in each corpus and their total occurrences (scores) are maintained in the database – so that based on their occurrences, each set of spam and genuine data is assigned a criterion or probability score for its capacity of determining a text or message to either be a spam or genuine text (Yoro, Aghware, Akazue, et al., 2023; Yoro, Aghware, Malasowe, et al., 2023; Yoro & Ojugo, 2019a, 2019b).

Classification section – uses the frequency probability of occurrence of each word (tokens) as spam or legitimate, each incoming unseen normalized message data is processed and classified as either legitimate or spam by the Bayesian classifier. In the event of misclassification, users can rectify this classification by reading the message and re-adding the message to inbox. This will automatically correct and update the database for future classification. Thus, making Bayesian filters quite adaptive (Okonta et al., 2013, 2014; Wemembu et al., 2014).

Output section – yields classification result of the filter. which can either be spam or ham.

Ensemble Operations – The Genetic algorithm trained neural network (GANN) is initialized with (n-r!) combined if-then rules. An individual's fitness is selected from 30-individuals via the *tournament* method to determine new pool for mating. Crossover and mutation is applied to help the network learn the dynamic, non-linear underlying features of interest via the multipoint crossover to yield new parents to the dataset.

The new parents contribute to yield new individuals. Mutation is reapplied and solutions are allotted new random values that still conform to belief space. Mutation is applied depends on CGA progression on the net and how fit is the fittest individual (i.e. fitness of the fittest rule divided by 2). New individuals are made to replace old with low fitness so as to create a new pool. Process continues until individual with a fitness value of 0 (i.e. until an optimal solution) is found or reached (Ojugo et al, 2013; 2015).

Table 1: Rule-Based Encoded Score

Code	Rule Input Parameters	Genuine	Spam
P01	Message Size	0.50	0.50
P02	Message Character	0.50	0.50
P03	Message From	0.50	0.50
P04	Message To	0.50	0.50
P05	Subject	0.30	0.70
P06	Body of Message	0.25	0.75

Fitness function (*f*) is resolved with initial pool (Parents) using the genuine class as thus:
R1:50 R2:50 R3:50 R4:50 R5:30 R6:50

Table 2: 1st and 2nd Generation of Population

Code	Feature Selection	Chromosomes (Binary 0 or 1)			Fitness
		Parent 1st Gen	Retrain	Parent 2nd Gen	
P01	50	110010	1 and 6	110001	49
P02	50	110010	2 and 5	110010	50
P03	50	110010	3 and 6	110001	49
P04	50	110010	4 and 5	110010	50
P05	30	011110	5 and 6	011101	29
P06	25	011001	6 and 5	011010	26

Initialization/selection via ANN ensures first 3-beliefs are met; and mutation ensure fourth belief is met. Its influence function influences how many mutations occurs, and how close its solution is, and impacts how algorithm is processed. Model stop if best individual has fitness of 0.3 (Ojugo et al., 2014; Ojugo & Otakore, 2020a, 2020b, 2021). Model stops if stop criterion is met. GANN utilizes number of epochs to determine stop criterion (Dawson and Wilby 2001).

3. FINDINGS AND RESULT DISCUSSION

3.1. Proposed Ensemble Evaluation

In this study, accuracy, recall, error rate (ER) and specificity are used to evaluate the performance of the detection models. The formulas of the above criteria are calculated as follows (Ibor et al., 2023; Ojugo, Aghware, et al., 2015b; Ojugo, Oyemade, et al., 2015). To measure effectiveness and accuracy, we measure their rate of misclassification and corresponding improvement percentages in both training and test data sets as summarized in Tables 3. Equations for misclassification rate and its improvement percentage of unsupervised (B) model against supervised (A) model respectively, is calculated as follows:

$$Classification\ Rate\ \frac{No.of\ Incorrect\ class}{No.of\ Sample\ set} \quad (1)$$

$$Improvement\ \frac{MR(A) - MR(B)}{MR(A)}\ x\ 100 \quad (2)$$

Table 3. Summary Sheet For Benchmark Models

Model	Classification Error Rate		Improvement Percentage	
	Training	*Testing*	*Training*	*Testing*
GA	0.317	0.165	0.023	0.845
Naïve Bayes	0.525	0.172	0.063	0.646
Proposed Memetic	0.196	0.045	0.147	0.983

Tables 3 shows the classification error rate with Naïve Bayes, GA and the proposed Genetic Algorithm trained Bayesian Network (GABN) yielding 0.317, 0.525 and 0.196 (i.e. error rate in detecting and classifying false-positive and true-negative) respectively; It also in turn, consequently – yields and promises the various improvement rate as of 2.3%, 6.3% and 14.7% respectively. This implies that the proposed model outperforms the other benchmark models (i.e. Genetic Algorithm and Naïve Bayes) used respectively.

3.2. Findings and Discussion

A true positive (TP) is a case (rule) that correctly distinguishes spam from genuine text. True negative (TN) shows normal text message data classified correctly as normal. The false negative (FN) denotes a case in which a text is classified as normal data, and a false positive (FP) means that a normal text is classified as a spam (Oyemade & Ojugo, 2020, 2021). The accuracy rate shows the overall correct detection accuracy of the dataset, ER refers to the robustness of the classifier, recall indicates the degree of correctly detected attack types of all cases classified as attacks, and specificity shows the percentage of correctly classified normal data. In the above, higher accuracy and recall and lower ER indicate good performance (Ojugo & Eboka, 2018c; Ojugo & Otakore, 2018a; Okobah & Ojugo, 2018).

Using Naïve Bayes and GA (benchmark models) to ascertain how well our hybrid GABN algorithm performed, we obtain the results in fig 2 and fig 3 respectively. It shows that hybrid GABN outperforms Naïve Bayes and GA models. However, for the mean processing time required to converge – it is found that GABN performed least. This can be attributed to the fact that: (a) the hybrid model needs to first use GA as pre-processor to train Bayesian net, (b) for such hybrids, there are always structural dependencies with the underlying heuristics employed/merged and conflicts in data encoding that is required. These must be resolved for the model to perform appropriately.

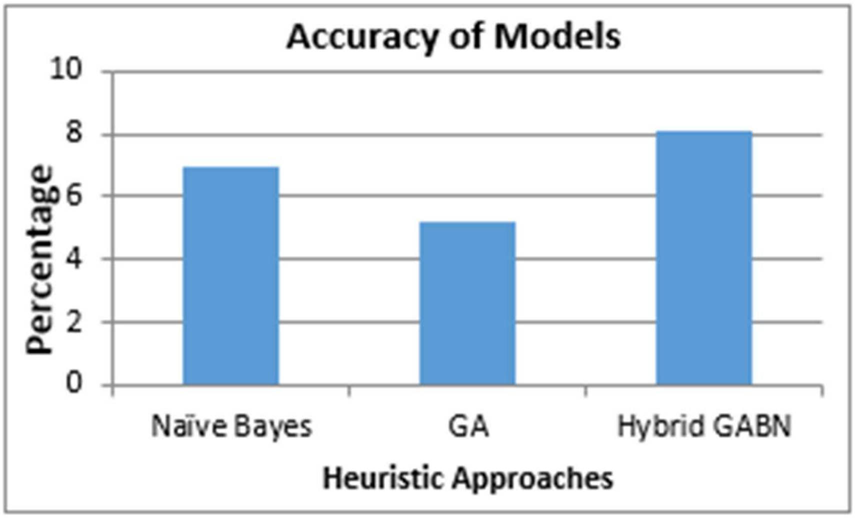

Figure 2. Comparative Accuracy Percentage

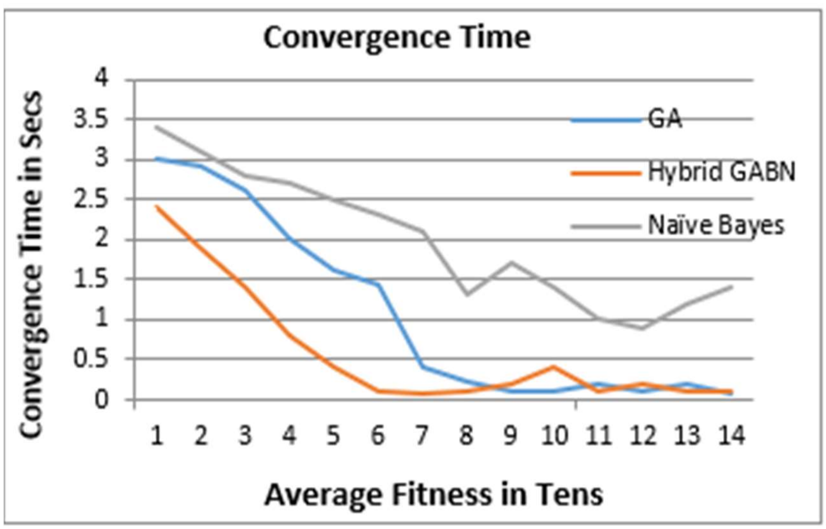

Figure 3. Convergence time in seconds

To harness the inherent benefits of these models, the conflicts must be resolved – so we can exploit historic data as well as explore the domain space to yield an optimal solution. We must select appropriate feats to devoid the model of poor generalization, overtraining and overfit (Akazue et al., 2023; Muslikh et al., 2023; Oladele et al., 2024).

4. CONCLUSION

From the consequences of SMS spam to users, corporate organizations, MNO and the world, several concerted efforts to detect the intrusion of spam in various communication media has paid off especially in combating email spam (Ojugo, Odiakaose, et al., 2023). Spam filter receives the message and classifies it as either ham, or spam. Its performance is measured by the number of incorrectly marked spam, and unidentified spams generated. An ideal filter will correctly classify all SMS with almost zero error rates of false positive/negative – via trade-offs between the number of false positives and false negatives (Ojugo & Okobah, 2017b, 2017a; Ojugo & Otakore, 2018b).

Models are useful to represent reality as their primary value is to serve as educational tools for insight to help us better understand and reflect upon reality (Aghware et al., 2023b, 2023a). They compile knowledge and are vehicles to communicate hypotheses. Its sensitivity analysis help modelers to reflect on a variety of system theories (Malasowe et al., 2023; Ojugo, Akazue, Ejeh, Ashioba, et al., 2023; Ojugo, Akazue, Ejeh, Odiakaose, et al., 2023; Ojugo, Eboka, et al., 2015).

REFERENCES (10 PT)

Aghware, F. O., Yoro, R. E., Ejeh, P. O., Odiakaose, C. C., Emordi, F. U., & Ojugo, A. A. (2023a). DeLClustE: Protecting Users from Credit-Card Fraud Transaction via the Deep-Learning Cluster Ensemble. *International Journal of Advanced Computer Science and Applications*, *14*(6), 94–100. doi: 10.14569/IJACSA.2023.0140610

Aghware, F. O., Yoro, R. E., Ejeh, P. O., Odiakaose, C. C., Emordi, F. U., & Ojugo, A. A. (2023b). Sentiment analysis in detecting sophistication and degradation cues in malicious web contents. *Kongzhi Yu Juece/Control and Decision*, *38*(01), 653.

Akazue, M. I., Debekeme, I. A., Edje, A. E., Asuai, C., & Osame, U. J. (2023). UNMASKING FRAUDSTERS : Ensemble Features Selection to Enhance Random Forest Fraud Detection. *Journal of Computing Theories and Applications*, *1*(2), 201–212. https://doi.org/10.33633/jcta.v1i2.9462

Amalraj, J. R., & Lourdusamy, R. (2022). A Novel distributed token-based algorithm using secret sharing scheme for secure data access control. *International Journal of Computer Networks and Applications*, *9*(4), 374. https://doi.org/10.22247/ijcna/2022/214501

Benchaji, I., Douzi, S., El Ouahidi, B., & Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. *Journal of Big Data*, *8*(1), 151. https://doi.org/10.1186/s40537-021-00541-8

Broadhurst, R., Skinner, K., Sifniotis, N., & Matamoros-Macias, B. (2018). Cybercrime Risks in a University Student Community. *SSRN Electronic Journal*, *May*. https://doi.org/10.2139/ssrn.3176319

Ibor, A. E., Edim, E. B., & Ojugo, A. A. (2023). Secure Health Information System with Blockchain Technology. *Journal of the Nigerian Society of Physical Sciences*, *5*(992), 1–8. https://doi.org/10.46481/jnsps.2022.992

Ifeka, A., & Akinbobola, A. (2015). Trend Analysis of Precipitation in Some Selected Stations in Anambra State. *Atmospheric and Climate Sciences*, *05*(01), 1–12. https://doi.org/10.4236/acs.2015.51001

Igwenagu, C. M. (2015). Trend Analysis of Rainfall Pattern in Enugu State, Nigeria. *European Journal of Statistics and Probability*, *3*(3), 12–18. www.eajournals.org

Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using GA algorithm for feature selection. *Journal of Big Data*, *9*(1), 24. https://doi.org/10.1186/s40537-022-00573-8

Jayatilaka, A., Arachchilage, N. A. G., & Babar, M. A. (2021). Falling for Phishing: An Empirical Investigation into People's Email Response Behaviors. *ArXiv Preprint ArXiv …*, *Fbi 2020*, 1–17.

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, *10*(2), 1–31. https://doi.org/10.1145/1754393.1754396

Linh, D. (2018). Insider threat detection: Where and how data science applies. *Cyber Security: A Peer-Reviewed Journal*, *2*, 1–8. https://www.ingentaconnect.com/content/hsp/jcs/2018/00000002/00000003/art00003

Malasowe, B. O., Akazue, M. I., Okpako, E. A., Aghware, F. O., Ojie, D. V., & Ojugo, A. A. (2023). Adaptive Learner-CBT with Secured Fault-Tolerant and Resumption Capability for Nigerian Universities. *International Journal of Advanced Computer Science and Applications*, *14*(8), 135–142. https://doi.org/10.14569/IJACSA.2023.0140816

Mohd Ibrahim, A., Venkat, I., De Wilde, P., Mohd Romlay, M. R., & Bahamid, A. (2022). The role of crowd behavior and cooperation strategies during evacuation. *Simulation*, *98*(9), 737–751. https://doi.org/10.1177/00375497221075611

Muslikh, A. R., Setiadi, I. D. R. M., & Ojugo, A. A. (2023). Rice disease recognition using transfer xception convolution neural network. *Jurnal Teknik Informatika (JUTIF)*, *4*(6), 1541–1547. https://doi.org/10.52436/1.jutif.2023.4.6.1529

Ojugo, A. A., Abere, R. A., Eboka, A. O., Yerokun, M. O., Yoro, R. E., Onochie, C. C., & Oyemade, D. A. (2013). Hybrid neural network models for rainfall runoffs: Comparative study. *Advancement in Scientific and Engineering Research*, *1*(2), 22–34.

Ojugo, A. A., Aghware, F. O., Yoro, R. E., Yerokun, M. O., Eboka, A. O., Anujeonye, C. N., & Efozia, F. N. (2015a). Dependable Community-Cloud Framework for Smartphones. *American Journal of Networks and Communications*, *4*(4), 95. https://doi.org/10.11648/j.ajnc.20150404.13

Ojugo, A. A., Aghware, F. O., Yoro, R. E., Yerokun, M. O., Eboka, A. O., Anujeonye, C. N., & Efozia, F. N. (2015b). Evolutionary Model for Virus Propagation on Networks. *Automation, Control and Intelligent Systems*, *3*(4), 56. https://doi.org/10.11648/j.acis.20150304.12

Ojugo, A. A., Akazue, M. I., Ejeh, P. O., Ashioba, N. C., Odiakaose, C. C., Ako, R. E., & Emordi, F. U. (2023). Forging a User-Trust Memetic Modular Neural Network Card Fraud Detection Ensemble: A Pilot Study. *Journal of Computing Theories and Applications*, *1*(2), 1–11. https://doi.org/10.33633/jcta.v1i2.9259

Ojugo, A. A., Akazue, M. I., Ejeh, P. O., Odiakaose, C., & Emordi, F. U. (2023). DeGATraMoNN: Deep Learning Memetic Ensemble to Detect Spam Threats via a Content-Based Processing. *Kongzhi Yu Juece/Control and Decision*, *38*(01), 667–678.

Ojugo, A. A., Ben-Iwhiwhu, E., Kekeje, O. D., Yerokun, M. O., & Iyawa, I. J. (2014). Malware Propagation on Social Time Varying Networks: A Comparative Study of Machine Learning Frameworks. *International Journal of Modern Education and Computer Science*, *6*(8), 25–33. https://doi.org/10.5815/ijmecs.2014.08.04

Ojugo, A. A., & Eboka, A. O. (2018a). Assessing Users Satisfaction and Experience on Academic Websites: A Case of Selected Nigerian Universities Websites. *International Journal of Information Technology and Computer Science*, *10*(10), 53–61. https://doi.org/10.5815/ijitcs.2018.10.07

Ojugo, A. A., & Eboka, A. O. (2018b). Comparative Evaluation for High Intelligent Performance Adaptive Model for Spam Phishing Detection. *Digital Technologies*, *3*(1), 9–15. https://doi.org/10.12691/dt-3-1-2

Ojugo, A. A., & Eboka, A. O. (2018c). Modeling the Computational Solution of Market Basket Associative Rule Mining Approaches Using Deep Neural Network. *Digital Technologies*, *3*(1), 1–8. https://doi.org/10.12691/dt-3-1-1

Ojugo, A. A., & Eboka, A. O. (2020a). An Empirical Evaluation On Comparative Machine Learning Techniques For Detection of The Distributed Denial of Service (DDoS) Attacks. *Journal of Applied Science, Engineering, Technology, and Education*, *2*(1), 18–27. https://doi.org/10.35877/454ri.asci2192

Ojugo, A. A., & Eboka, A. O. (2020b). Memetic algorithm for short messaging service spam filter using text normalization and semantic approach. *International Journal of Informatics and Communication Technology (IJ-ICT)*, *9*(1), 9. https://doi.org/10.11591/ijict.v9i1.pp9-18

Ojugo, A. A., & Eboka, A. O. (2021). Empirical Bayesian network to improve service delivery and performance dependability on a campus network. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *10*(3), 623. https://doi.org/10.11591/ijai.v10.i3.pp623-635

Ojugo, A. A., Eboka, A. O., Yoro, R. E., Yerokun, M. O., & Efozia, F. N. (2015). Hybrid Model for Early Diabetes Diagnosis. *2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, *50*, 55–65. https://doi.org/10.1109/MCSI.2015.35

Ojugo, A. A., & Ekurume, E. O. (2021a). Deep Learning Network Anomaly-Based Intrusion Detection Ensemble For Predictive Intelligence To Curb Malicious Connections: An Empirical Evidence. *International Journal of Advanced Trends in Computer Science and Engineering*, *10*(3), 2090–2102. https://doi.org/10.30534/ijatcse/2021/851032021

Ojugo, A. A., & Ekurume, E. O. (2021b). Predictive Intelligent Decision Support Model in Forecasting of the Diabetes Pandemic Using a Reinforcement Deep Learning Approach. *International Journal of Education and Management Engineering*, *11*(2), 40–48. https://doi.org/10.5815/ijeme.2021.02.05

Ojugo, A. A., & Nwankwo, O. (2021a). Forging a Spectral-Clustering Multi-Agent Hybrid Deep Learning Model To Predict Rainfall Runoff In Nigeria. *International Journal of Innovative Science, Engineering and Technology*, *8*(3), 140–147.

Ojugo, A. A., & Nwankwo, O. (2021b). Spectral-Cluster Solution For Credit-Card Fraud Detection Using A Genetic Algorithm Trained Modular Deep Learning Neural Network. *JINAV: Journal of Information and Visualization*, *2*(1), 15–24. https://doi.org/10.35877/454RI.jinav274

Ojugo, A. A., Odiakaose, C. C., Emordi, F. U., Ejeh, P. O., Adigwe, W., Anazia, K. E., & Nwozor, B. (2023). Forging a learner-centric blended-learning framework via an adaptive content-based architecture. *Science in Information Technology Letters*, *4*(1), 40–53. https://doi.org/10.31763/sitech.v4i1.1186

Ojugo, A. A., & Okobah, I. P. (2017a). Computational Solution for Modeling Rainfall Runoff Using Intelligent Stochastic Model: A Case of Warri in Delta State Nigeria. *Journal of Digital Innivations and Contemporary Res. in Science Engineering and Technology*, *5*(4), 45–58. https://doi.org/10.22624

Ojugo, A. A., & Okobah, I. P. (2017b). Hybrid Fuzzy-Genetic Algorithm Trained Neural Network Stochastic Model for Diabetes Diagnosis and Classification. *Journal of Digital Innovations & Contemp Res. In Sc., Eng & Tech*, *5*(4), 69–90. https://doi.org/10.22624

Ojugo, A. A., & Otakore, D. O. (2018a). Redesigning Academic Website for Better Visibility and Footprint: A Case of the Federal University of Petroleum Resources Effurun Website. *Network and Communication Technologies*, *3*(1), 33. https://doi.org/10.5539/nct.v3n1p33

Ojugo, A. A., & Otakore, O. D. (2018b). Improved Early Detection of Gestational Diabetes via Intelligent Classification Models: A Case of the Niger Delta Region in Nigeria. *Journal of Computer Sciences and Applications*, *6*(2), 82–90. https://doi.org/10.12691/jcsa-6-2-5

Ojugo, A. A., & Otakore, O. D. (2020a). Computational solution of networks versus cluster grouping for social network contact recommender system. *International Journal of Informatics and Communication Technology (IJ-ICT)*, *9*(3), 185. https://doi.org/10.11591/ijict.v9i3.pp185-194

Ojugo, A. A., & Otakore, O. D. (2020b). Intelligent cluster connectionist recommender system using implicit graph friendship algorithm for social networks. *IAES International Journal of Artificial Intelligence*, *9*(3), 497~506. https://doi.org/10.11591/ijai.v9.i3.pp497-506

Ojugo, A. A., & Otakore, O. D. (2021). Forging An Optimized Bayesian Network Model With Selected Parameters For Detection of The Coronavirus In Delta State of Nigeria. *Journal of Applied Science, Engineering, Technology, and Education*, *3*(1), 37–45. https://doi.org/10.35877/454RI.asci2163

Ojugo, A. A., & Oyemade, D. A. (2021). Boyer moore string-match framework for a hybrid short message service spam filtering technique. *IAES International Journal of Artificial Intelligence*, *10*(3), 519–527. https://doi.org/10.11591/ijai.v10.i3.pp519-527

Ojugo, A. A., Oyemade, D. A., Allenotor, D., Longe, O. B., & Anujeonye, C. N. (2015). Comparative Stochastic Study for Credit-Card Fraud Detection Models. *African Journal of Computing & ICT*, *8*(1), 15–24. www.ajocict.net

Ojugo, A. A., & Yoro, R. E. (2013). Computational Intelligence in Stochastic Solution for Toroidal N-Queen. *Progress in Intelligent Computing and Applications*, *1*(2), 46–56. https://doi.org/10.4156/pica.vol2.issue1.4

Ojugo, A. A., & Yoro, R. E. (2020). Predicting Futures Price And Contract Portfolios Using The ARIMA Model: A Case of Nigeria's Bonny Light and Forcados. *Quantitative Economics and Management Studies*, *1*(4), 237–248. https://doi.org/10.35877/454ri.qems139

Ojugo, A. A., & Yoro, R. E. (2021). Forging a deep learning neural network intrusion detection framework to curb the distributed denial of service attack. *International Journal of Electrical and Computer Engineering*, *11*(2), 1498–1509. https://doi.org/10.11591/ijece.v11i2.pp1498-1509

Ojugo, A. A., Yoro, R. E., Okonta, E. O., & Eboka, A. O. (2013). A Hybrid Artificial Neural Network Gravitational Search Algorithm for Rainfall Runoffs Modeling and Simulation in Hydrology. *Progress in Intelligent Computing and Applications*, *2*(1), 22–34. https://doi.org/10.4156/pica.vol2.issue1.2

Okobah, I. P., & Ojugo, A. A. (2018). Evolutionary Memetic Models for Malware Intrusion Detection: A Comparative Quest for Computational Solution and Convergence. *International Journal of Computer Applications*, *179*(39), 34–43. https://doi.org/10.5120/ijca2018916586

Okonta, E. O., Ojugo, A. A., Wemembu, U. R., & Ajani, D. (2013). Embedding Quality Function Deployment In Software Development: A Novel Approach. *West African Journal of Industrial & Academic Research*, *6*(1), 50–64.

Okonta, E. O., Wemembu, U. R., Ojugo, A. A., & Ajani, D. (2014). Deploying Java Platform to Design A Framework of Protective Shield for Anti– Reversing Engineering. *West African Journal of Industrial & Academic Research*, *10*(1), 50–64.

Oladele, J. K., Ojugo, A. A., Odiakaose, C. C., Emordi, F. U., Abere, R. A., Nwozor, B., Ejeh, P. O., & Geteloma, V. O. (2024). BEHeDaS: A Blockchain Electronic Health Data System for Secure Medical Records Exchange. *Journal of Computing Theories and Applications*, *2*(1), 1–12. https://doi.org/10.33633/jcta.v2i19509

Omede, E. U., & Okpeki, U. K. (2023). Use of Mathematical and Unified Modelling Language Tools to Provide Learning Motivation System for Students in Tertiary Institution. *Journal of Applied Sciences and Environmental Management*, *27*(6), 1135–1139. https://doi.org/10.4314/jasem.v27i6.11

Oyemade, D. A., Akpojaro, J., Ojugo, A. A., Ureigho, R., Imouokhome, F. ., Omoregbee, E., Akpojaro, J., & Ojugo, A. A. (2016). A Three Tier Learning Model for Universities in Nigeria. *Journal of Technologies in Society*, *12*(2), 9–20. https://doi.org/10.18848/2381-9251/CGP/v12i02/9-20

Oyemade, D. A., & Ojugo, A. A. (2020). A property oriented pandemic surviving trading model. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(5), 7397–7404. doi: 10.30534/ijatcse/2020/71952020

Oyemade, D. A., & Ojugo, A. A. (2021). An Optimized Input Genetic Algorithm Model for the Financial Market. *International Journal of Innovative Science, Engineering and Technology, 8*(2), 408–419. https://ijiset.com/vol8/v8s2/IJISET_V8_I02_41.pdf

Parker, H. J., & Flowerday, S. V. (2020). Contributing factors to increased susceptibility to social media phishing attacks. *SA Journal of Information Management*, *22*(1). https://doi.org/10.4102/sajim.v22i1.1176

Rtayli, N., & Enneya, N. (2020). Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. *Journal of Information Security and Applications*, *55*, 102596. https://doi.org/10.1016/j.jisa.2020.102596

Sahmoud, T., & Mikki, D. M. (2022). Spam Detection Using BERT. *The Frontiers of Society, Science and Technology*, *14*(2), 23–35. https://doi.org/10.48550/arXiv.2206.02443

Sasikala, G., Laavanya, M., Sathyasri, B., Supraja, C., Mahalakshmi, V., Mole, S. S. S., Mulerikkal, J., Chidambaranathan, S., Arvind, C., Srihari, K., & Dejene, M. (2022). An Innovative Sensing Machine Learning Technique to Detect Credit Card Frauds in Wireless Communications. *Wireless Communications and Mobile Computing*, *2022*, 1–12. https://doi.org/10.1155/2022/2439205

Sobers, R. (2022). 166 Cybersecurity statistics and trends. *Data Security*, *12*, 23–29.

Supriya, B. N., & Akki, C. B. (2021). Sentiment prediction using enhanced xgboost and tailored random forest. *International Journal of Computing and Digital Systems*, *10*(1), 191–199. https://doi.org/10.12785/ijcds/100119

Thorat, O., Parekh, N., & Mangrulkar, R. (2021). TaxoDaCML: Taxonomy based Divide and Conquer using machine learning approach for DDoS attack classification. *International Journal of Information Management Data Insights*, *1*(2), 100048. https://doi.org/10.1016/j.jjimei.2021.100048

Tomar, N., & Manjhvar, A. K. (2015). A Survey on Data Mining Optimization Techniques. *IJSTE-International Journal of Science Technology & Engineering* |, *2*(06), 130–133. www.ijste.org

Udeze, C. L., Eteng, I. E., & Ibor, A. E. (2022). Application of Machine Learning and Resampling Techniques to Credit Card Fraud Detection. *Journal of the Nigerian Society of Physical Sciences*, *12*, 769. https://doi.org/10.46481/jnsps.2022.769

Voke, O. D., Akazue, M. I., Omede, E. U., Oboh, E. ., & Imianvan, A. . (2023). Survival Prediction of Cervical Cancer Patients using Genetic Algorithm-Based Data Value Metric and Recurrent Neural Network. *International Journal of Soft Computing and Engineering*, *13*(2), 29–41. https://doi.org/10.35940/ijsce.B3608.0513223

Wemembu, U. R., Okonta, E. O., Ojugo, A. A., & Okonta, I. L. (2014). A Framework for Effective Software Monitoring in Project Management. *West African Journal of Industrial and Academic Research*, *10*(1), 102–115. http://www.ajol.info/index.php/wajiar/article/view/105798

Yao, J., Wang, C., Hu, C., & Huang, X. (2022). Chinese Spam Detection Using a Hybrid BiGRU-CNN Network with Joint Textual and Phonetic Embedding. *Electronics*, *11*(15), 2418. https://doi.org/10.3390/electronics11152418

Yeboah-Boateng, E. O., & Amanor, P. M. (2014). Phishing , SMiShing & Vishing : An Assessment of Threats against Mobile Devices. *Journal of Emerging Trends in Computing and Information Sciences*, *5*(4), 297–307.

Yildiz Durak, H. (2019). Human Factors and Cybersecurity in Online Game Addiction: An Analysis of the Relationship Between High School Students' Online Game Addiction and the State of Providing Personal Cybersecurity and Representing Cyber Human Values in Online Games. *Social Science Quarterly*, *100*(6), 1984–1998. https://doi.org/10.1111/ssqu.12693

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2). https://doi.org/10.1088/1742-6596/1168/2/022022

Yoro, R. E., Aghware, F. O., Akazue, M. I., Ibor, A. E., & Ojugo, A. A. (2023). Evidence of personality traits on phishing attack menace among selected university undergraduates in Nigerian. *International Journal of Electrical and Computer Engineering*, *13*(2), 1943. https://doi.org/10.11591/ijece.v13i2.pp1943-1953

Yoro, R. E., Aghware, F. O., Malasowe, B. O., Nwankwo, O., & Ojugo, A. A. (2023). Assessing contributor features to phishing susceptibility amongst students of petroleum resources varsity in Nigeria. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(2), 1922. https://doi.org/10.11591/ijece.v13i2.pp1922-1931

Yoro, R. E., & Ojugo, A. A. (2019a). An Intelligent Model Using Relationship in Weather Conditions to Predict Livestock-Fish Farming Yield and Production in Nigeria. *American Journal of Modeling and Optimization*, *7*(2), 35–41. https://doi.org/10.12691/ajmo-7-2-1

Yoro, R. E., & Ojugo, A. A. (2019b). Quest for Prevalence Rate of Hepatitis-B Virus Infection in the Nigeria: Comparative Study of Supervised Versus Unsupervised Models. *American Journal of Modeling and Optimization*, *7*(2), 42–48. https://doi.org/10.12691/ajmo-7-2-2

Zhang, Y., Egelman, S., Cranor, L. F., & Hong, J. (2007). Phinding Phish: Evaluating Anti-Phishing Tools. *In Proceedings of the Network & Distributed System Security Symposium (NDSS 2007)*, *March*, 1–16.