

*Corresponding author: Arnold Adimabua Ojugo, Department of Computer Science, Federal University of Petroleum Resources Effurun, Delta State, Nigeria

E-mail: ojugo.arnold@fupre.edu.ng

RESEARCH ARTICLE

Empirical Evaluation Using Intelligent Modeling in Prediction of Potential Cancer Problematics Cases in Nigeria

Arnold Adimabua Ojugo^{1*}, Chris Obaro Obruché², & Andrew Okonji Eboka³

¹Department of Computer Science, College of Science, Federal University of Petroleum Resources Effurun, Delta State, Nigeria

²Research Assistant, Department of Computer Science, College of Science, Federal University of Petroleum Resources Effurun, Delta State, Nigeria

³Department of Network Computing, Coventry University, Priory Street Coventry CV1 5FB, United Kingdom

Abstract: The rapid rate as well as the volume in amount of data churned out on daily basis has necessitated the need for data mining process. Advanced by the field of data science with machine learning approaches as new paradigm and platform, it has become imperative to provide beneficial support in constructing models that can effectively assist domain experts/practitioners – to make comprehensive decisions regarding potential cases. The study uses deep learning prognosis to effectively respond to problematic cases of cancer in Nigeria. We use the fuzzy rule-based memetic model to predict potential problematic cases of cancer – predicting results from data samples collected from the Epidemiology laboratory at Federal Medical Center Asaba, Nigeria. Dataset is split into training (85%) and testing (15%) to aid validation. Results indicate that age, obesity, environmental conditions and family relations (to the first and second degree) are critical factors to be watched for benign and malignant cancer types. Constructed model result shows high predictive capability strength compared to other models presented on similar studies.

Keywords: Cancer, memetic algorithm, cluster, epidemiology, reinforcement learning, Nigeria

1. Introduction

Cancer is viewed as a group of diseases characterized by unregulated division and spread of cells. The cancerous cells may occur in liquids, as in leukemia. Most, however, occur in solid tumors that originally appear in various tissues in various parts of the body. By their original locations they are classified into various types of cancer, such as lung, colon, breast, prostate cancer, etc (Dawane and Pandit, 2012). Localized tumors can be removed by surgery or irradiation with high survival rates. As cancer progresses, however, it metastasizes – invading the surrounding tissues, entering the blood stream, spreading and establishing colonies in distant parts of the body (Ferlay et al, 2008). Only a third of patients with metastasized cancer survive more than five years. Invasive distensions spreading crab-like from a tumor in the breast were described by Hippocrates. From the crab, *karkinos* in Greek and *cancer* in Latin, came the name of the disease and the name of its inducing agents, carcinogens (Ukah and Nwafor, 2017).

The causes of serious ill-health in the world are changing. Infection as a major cause is giving way to non-communicable diseases such as cardiovascular disease and cancer. There were over 10-million new cancer cases worldwide and over 6-million deaths attributed to



cancer just in 1996 alone. In 2020 there are predicted to be 20 million new cases and 12 million deaths. Part of the reason for this is that life expectancy is steadily rising and most cancers are more common in an ageing population. More significantly, the globalization of unhealthy lifestyles with reference to cigarette smoking and the adoption of many other features of the modern Western diet (high fat, low fibre content) will increase cancer incidence (Malcolm, 2001; Ferlay et al, 2017; GLOBOCAN, 2012).

1.1. Cancer Around The World

Cancer is potentially, a fatal disease caused by environmental factors that mutate genes-encoding critical cell-regulatory proteins. Thus, the resultant aberrant cell behaviour leads to an expansive masses of abnormal cells that destroy surrounding normal tissue and can spread to vital organs resulting in disseminated disease, commonly a harbinger of imminent death (Ikwo, 2013). Cancer is the second leading cause of death in the United States. About one-half of all men and one-third of all women in the US will develop cancer during their lifetimes. Today, millions of people are living with cancer or have had cancer (Sylla and Wild, 2011).

Over the next 15 years, there will be a dramatic increase in the number of people developing cancer and globally, there are over 10-million new cancer cases diagnosed yearly, which will become in the neighborhood of 30-million by the year 2025. Cancer is now the public most feared disease with billions of dollars are spent annually on research for drugs, cancer charities and governments. Even so, a cure for cancer appears elusive (Jedy-Agba et al, 2012). Cancer is a complex genetic disease that is caused primarily by environmental factors. Its causative agents (carcinogens) are present in food, water, chemicals, air, and even in sunlight that people are exposed to. Since epithelial cells cover the skin, line the respiratory and alimentary tracts, and metabolize ingested carcinogens, it is not surprising that over 90% of cancers occur in epithelia (Fritz et al, 2000). Cancer can also be caused due to certain polyunsaturated fatty acid that generate damaging free radicals; And the intake of antioxidants that can scavenge these harmful radicals is also a confounding factor. Reducing infection, particularly in the poorer countries, will lead to reductions in cancer incidence. Other infectious agents associated with increased cancer risk include hepatitis B (liver), certain subtypes of human papilloma virus (cervix), bacterium *Helicobacter pylori* (stomach) and human immunodeficiency virus (many sites) (Ikwo, 2013; Doll et al, 1968).

The management of patients with cancer is often very costly. But, huge steps in improving the prognosis of patients with cancer are almost immediately achievable with present-day technology and sufficient financial resource, and all essentially relate to early detection. Cancer though, does not develop overnight. It does however, evolve over time with detectable premalignant lesions presaging the development of full-blown malignancy. Malignant tumors not only invade surrounding tissue; But, they are able to colonize other vital, organs – a process known as metastasis (Bray and Parkin, 2009). Widespread metastatic disease is a harbinger of imminent death. Thus, immediate referral to the oncologist after detection of any suspicious lump or symptom is paramount; in many parts of the world with poor health education patients present with very advanced disease. In the same vein, cancer screening programmes are designed to detect not only early asymptomatic malignant tumors (Curado et al, 2009); But also premalignant lesions. Even in the richer countries, such programmes are a significant financial burden, and the more cost-effective programmes target the higher-risk groups denoted by age (e.g. cervical screening, mammography, and colonoscopy) or occupation (e.g. blood in the urine of dye workers for bladder cancer) (Mills et al, 2011).

1.2. Classifying Cancerous Symptoms and Tumors

Cancer often presents tumors in both benign and/or malignant cases. Benign tumors are slow-growing expansive masses that compress rather than invade surrounding tissue. As such they generally pose little threat, except when growing in a confined space like the skull, and can usually be readily excised. Many so-called benign tumors have malignant potential



especially those occurring in the large intestine. They should be removed before malignancy develops. Conversely, malignant tumors usually grows rapidly, invading surrounding tissue and colonizing distant organs. The ability of tumor cells to detach from the original mass (the primary tumor) and set up a metastasis (secondary tumor) discontinuous with the primary is unequivocal proof of malignancy. Tumors are also classified according to their tissue of origin; recognition of the parent tissue in a lymph node metastasis could establish the location of a hitherto undiagnosed primary tumor (Malcolm, 2001; Ikwo, 2013).

Cancers may be classified by their primary site of origin or by their histological or tissue types namely: (a) by site of origin yields breast, lungs, prostate, liver, carcinoma (kidney cancer), oral, brain etc, (b) by tissue types into sarcoma, carcinoma, myeloma, leukemia, lymphoma etc, (c) by grade abnormality of the cells with respect to surrounding normal tissues, and (d) by stage depending on the degree of the tumor size and its spread. Cancer is caused by the damage in DNA. These can be inherited via parents, or may be the spontaneous problem that occurs during the lifetime of a person or patient. This process is usually referred to as mutation (Ananya, 2012).

1.3. Motivation of The Study

Study is motivated (Stolfo et al, 2015; Ojugo and Eboka, 2021; Ojugo and Oyemade, 2021):

1. Unavailability of datasets and censored results – makes prediction of cancerous studies difficult to assess. Also, its has been found also that in some instances, the dataset have been found to consist of ambiguities, impartial truth and noise that must be resolved via robust search in the bid to classify observations and expected values effectively.
2. The non-reliability in performance with selecting network parameters, mismatched feats and anomalies has been attributed cum trigger by non-optimized data and lack of dataset. These have resulted in various incorrect prediction of results and allowed for some cases to evade detection via diagnosis. Eliminating noisy feats via an accurately optimized classifier will thus, foster a more efficient model cum algorithm to aid cancer studies prediction.
3. Cancer persist even with the adoption of several classifiers available. Thus, we need to explore parameter selection. The significance of a unified model capable of addressing optimization problems and machine learning has not been explored, thus the need to explore such model unification.

To overcome these pitfalls, we use a memetic algorithm to detect cancer problematic cases using KDD dataset as generated from Epidemiology Department of the Federal Medical Centre, Asaba in Delta State of Nigeria.

2. Materials and Methods

2.1. Cancer Around The World

Machine learning model requires the necessary instances as dataset to effectively train a model. These instances must be appropriately labeled so as to minimize error rates. The errors determine how effective a model has progressed in its training, and is able to mine the data features of interest. Errors in classification can result from group the data points as used to predict new objects in a class (Ojugo and Eboka, 2020). Our dataset is retrieved from the Federal Medical Centre Epidemiology Laboratory in Asaba, Delta State. It contains 4687-cancer cases with 34-attributes to include patient bio-data, symptoms patient suffers from, HIV and other tests, history of the disease, diagnostic tools used, treatment that includes regimens for the type of the disease and doses given, with its drug reaction, the follow-up results for the whole treatment period, also costs and hospitalization paid. However, attributes that are likely to affect the patient behavior towards treatment includes: chemotherapy completed, treatment failed, treatment discontinued, death, and transferred out. We use the fuzzy (rule-based) universe of discourse, which was used to generate the rules for the proposed fuzzy discourse equation as Eq.1 (Ojugo and Otakore, 2020):

$$Fuzzy\ Discourse = \sum(A, B, C, D, E) * X \quad (1)$$

where A,B,C,D,E = picked questions option; X(0.02) = Assigned questions option fuzzy range value, while X(0.00) = unpicked option

Table 1.Fuzzy Encoded Universe Discourse For Cancer Dataset

Code	Fuzzy Set (Parameters)	Membership Function Degree of Cancer	
		Malignant (Type-I)	Benign (Type-II)
P01	Tumor/Lump	0.50	0.00
P02	Bleeding / hormone imbalances	0.50	0.00
P03	Mouth Ulcer and skin sores	0.50	0.00
P04	Pain and difficulty swallowing	0.50	0.50
P05	Weight loss and poor appetite	0.50	0.00
P06	Fatigue and depression	0.20	0.40
P07	Sweats, burps and tastelessness	0.45	0.35
P08	Nausea and vomiting	0.50	0.15
P09	Diarrhea	0.50	0.20
P10	Constipation	0.50	0.20
P11	Abdominal bloating	0.50	0.35
P12	Neurological symptoms	0.50	0.20
P13	Cardio-vascular symptoms	0.20	0.20
P14	Respiratory Symptoms	0.20	0.20

2.2. Hybrid Memetic Algorithm (Genetic Algorithm Trained Neural Network)

Our Modular Neural Network (MNN) as detailed in (Ghazale, 2018) is an improved deep learning neural network with learning that features an independent series of intermediary components – forming a module operating under certain architecture. These intermediary acts as bridge to receive individual network module output as input that helps compute the final output, which is resolved via a tangent activation function. MNN seeks to reduce large network into potentially, smaller, more managerial network (Aleskerov et al, 1997). It enhances efficiency via connected units that exponentially increases, as independent networks are added. While, this complicates the network structure, it improves computational efficiency with reduced computational time on individual task assigned to segmented modules, and tasks are executed in parallel with module re-organization to improve flexibility and network adaptability (Bolton and Hand, 2001).

The network enhances intelligence and increases time efficiency by reducing the network’s learning time – achieved via an independent training algorithms applied at each module with training dataset implemented independently and more quickly. This makes the model more flexible, adaptable and robust as rules can be re-used independently at various networks. Re-usability of rules has been a tedious experienced with such large and complex networks. With appropriate data encoding and carefully selected feats – the network experiences improved performance, compartmentalization via removal of partitions of interfaces, greater flexibility and eliminates redundancy (Brause et al, 1999). Thus, our MNN architecture is one comprised of smaller network(s) – whose modularization allows for easy learning and understanding of data feats, grants model greater flexibility via task execution parallelism via compartmentalization, eases code reuse, flexibility and adaptability (Burge and Shawe-Taylor, 2001). MNN passes data via task decomposition and training modules via a multi-objective, multi-agent and multi-region support module that aids effective classification. MNN can be implemented using the multi-layered perceptron, adaptive resonance theory and self-organizing maps. The network is trained via either the supervised, unsupervised or reinforcement learning (Chiu and Tsai, 2004).

Our hybrid is divided into 3-units (a) supervised genetic algorithm, (b) an unsupervised Kohonen neural network, and (c) knowledgebase – as seen in figure 1 and described below.

1. The modular design as in figure 1 shows input is received and passed via GA-block (consisting of encoder, selector, swapper recombiner, swapper mutator, and lastly,

- belief terminator for CGA. Each phase performs an integral GA fundamental operator process to have the dataset trained. Upon completion of the optimization, the dataset feats are held within the knowledgebase as a special holding place for operational data during machine learning process (Ojugo and Yoro, 2021; Ojugo and Eboka, 2020).
- The MNN block receives optimized rule dataset, grouped as successive labeled/unlabeled transactions instances as in fig 1 (Ojugo and Otakore, 2021; Aleksey and Alexander, 2016; Phua et al, 2004; 2007; Stolfo et al, 1997; 2000). With this, our classifier propagates IF-THEN transaction rule values of selected, predefined variables into the varying classes for detection. Rules are modeled as a production system with 4-components: (i) **rule set** containing in each rule, the pattern of how rule(s) and operation(s) are applied, (b) **knowledgebase** of transaction rule-set of (genuine and cancer classes) IF-THEN rules as selected data feats, (c) **control** strategy to specify the order in which the rules are compared to those in the knowledgebase to find a match and it seeks also a way to resolve conflicts that arise when several rules are matched at the same time, and (d) a rule **applier**. The MNN provides a self-learning ability and acts as the principal component analyzer with rules optimized by CGA's crossover and mutation so that the trained model or network can effectively, autonomously classify transaction into both class types.
 - Last stage of the network acts as a decision support and recognition system, with predicted values (output) and the automatic update of rules-knowledgebase, as transactions are encountered with new data, and thus, classified.

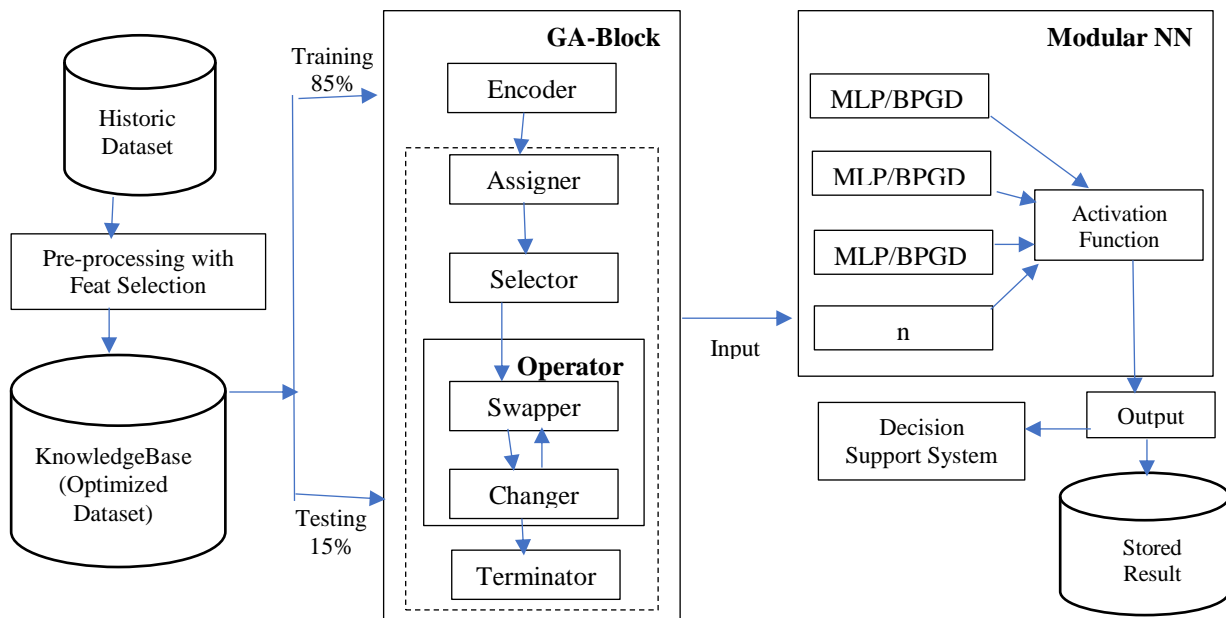


Figure 1. Hybrid Learning Ensemble

The model is initialized with IF-THEN rules, whose fitness is computed. 30-rules are selected via **tournament** method. Model uses a 2-point crossover that helps it to learn the dynamic and non-linear feats in dataset. 1-to-30 rules randomly generated via Gaussian distribution and correspond to these crossover points are selected (all genes of a single parent). As new parents contribute to yield a new pool of rules with genes of various parents (applied via mutation) – the model selects 3-random genes. These are then allocated new random values (between 0 and 1) – which still conforms to model belief space. The random values yield a score, generated for both cancer (benign and malignant) and normal rules from the universe discourse equation. Selection via MNN ensures that first 3-beliefs are met; mutation ensures the fourth belief is met. Its influence function determines how the number of mutations, how close a solution is and its impact on how algorithm is

processed. Model stops when best rule has fitness that equals the suspicion score or is higher than computed fitness function of transactions by each cardholder (Stolfo and Prodromidis, 1999; Syeda et al, 2002; Vasta et al, 2005; Wheeler and Aitken, 2019; Minahan, 2013; Xu et al, 2007).

3. Result Findings and Discussion

3.1. Result Findings

Dataset is divided into a ratio of 40% for training and 60% for testing. The predictive capability of the model is identified via fifteen-sign abnormalities labelled among GA-optimized (benign/malignant and normal) dataset. The training phase uses a feedforward training algorithm and approach with an epoch training cycle for each phase until a finite epoch is obtained or an equilibrium is reached. We obtained an equilibrium at 40-epochs as in the training phase interface with the dots. Figure 2 shows training phase interface; while figure 3 shows the test phase.

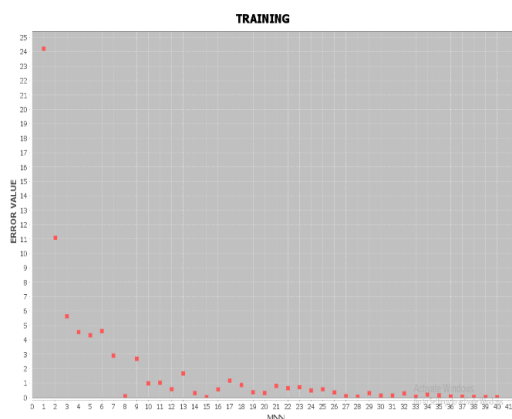


Figure 2. Training Phase Result

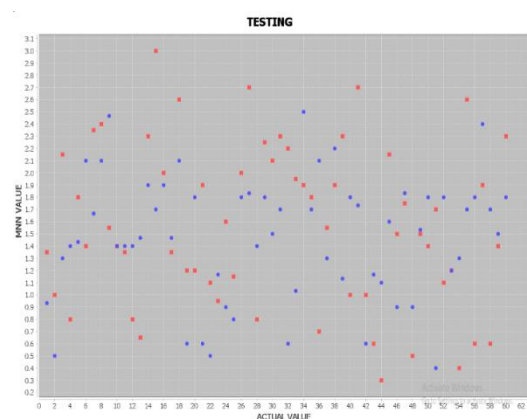


Figure 3. Testing Phase Result

From our confusion matrix, we compute: (a) sensitivity is the measure of how likely the model will predict the presence of all instances are cancerous (i.e. benign and malignant cases) when it is present, (b) specificity measures how likely model will detect the absence of cancerous (i.e. benign and malignant cases) when it is not present and/or not exhibited in the dataset, and (c) accuracy measures the proportion of true results seen as the degree of truth of a prediction. And given by Equations (2) – (4) respectively.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) - \text{where TP} = 43, \text{ and FN} = 5 \quad (2)$$

$$\text{Thus, we have } [43 / (43 + 5) * 100] \rightarrow [0.895 * 100] = 90\%.$$

$$\text{Specificity} = [\text{TN} / (\text{TN} + \text{FP}) * 100] \quad (3)$$

$$\text{We have that } [3 / (11 + 5) * 100] = 19\%.$$

$$\text{Accuracy} = [(\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) * 100] \quad (4)$$

$$\text{We also have } [(43 + 3) / (43 + 3 + 11 + 5)] * 100 = 74\%$$

The model is found to have a sensitivity of 90%, specificity value of 19% and prediction accuracy of 74% (0.74) with a rate of improvement of 12-percent for data inclusion that were not originally used to train the model as in figure 4.

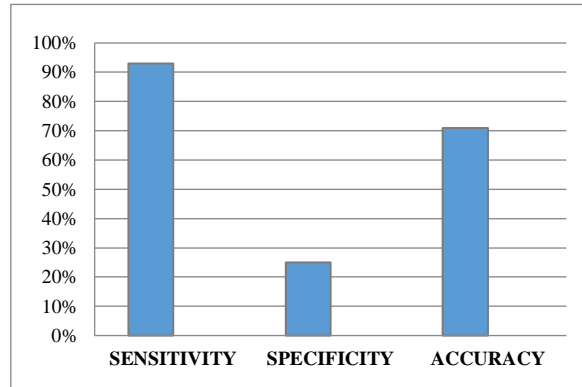


Figure 4. Statistical Analyses for the Model Performance

3.2. Discussion of Findings

If a patient exhibits at most three symptoms or less of a class of cancer alongside the tumor or lump THEN (C_1) it is benign; Else, if a patient exhibits exactly four symptoms of a class and above, it implies (C_2) that though it may be benign for now – it is likely to become malignant; whereas – if the patients is exhibits more than six of the symptoms of a class of cancer THEN (C_3) it is malignant. The IF-THEN Rules generated from fuzzy partitions of classification of either the benign and malignant cases are as thus:

R1: IF patient exhibits tumor or lump and its serious THEN class C_1 .

R2: IF patient exhibits tumor or lump and shows bleeding signs and/or hormonal imbalance and both symptoms are serious THEN class C_1 .

R3: IF patient exhibits tumor/lump, with bleeding and/or hormonal imbalance and mouth ulcer and both symptoms are serious THEN class C_1 .

R4: IF patient exhibits tumor/lump, with bleeding and/or hormonal imbalance, ulcer and difficulty swallowing and both symptoms are serious THEN C_1 .

R5-R6: IF patient exhibits tumor/lump, bleeding/hormonal imbalance, mouth ulcer, difficulty swallowing with weight loss and loss of appetite, as well as fatigue/depression and both symptoms are serious THEN class C_2 .

R7-R14: IF patient exhibits tumor or lump, hormonal imbalance and bleed, ulcer, difficulty swallowing, weight loss and loss of appetite, fatigue/depression, sweats/burps, nausea, diarrhea, constipation, abdominal bloating, cardio-vascular symptoms, neurological symptoms, respiratory symptoms and all symptoms are serious THEN class C_3 .

4. Summary and Conclusion

Our hybrid memetic algorithm employed the fuzzy universe discourse linguistics and fuzzy system as a preprocessor. In the design, building and implementing if such hybrid – we took cognizance that genetic algorithm will help speed up the ANN to avoid it being trapped at local maxima as well as in region of multi-modal local maxima. This will enable the model yield robust optima in the shortest amount of time. The fuzzy system will help better represent variables and data values in the model.

Hybrids are quite difficult to implement and explore – even though they always yield optimal and better solutions. However, care must be employed during parameter selection to avoid over-fitting, over-parameterization and over-training. Also, the correctly formatted (explored) historic dataset must be encoded through the underlying algorithm's structured learning for robustness and code reuse as well as allow for model's adaptability and flexibility. This will in turn help to address the inherent issues of statistical dependencies imposed on the model by the various models fused for hybridization. However, proper encoding schemes must be selected to help resolve the conflicts in the data feats of interest – as most systems may not adequately highlight the implications of such in a multi-agent

and multi-modal populated model. This is because the agents as they traverse the network or system – often can create their own behavioural rules on the dataset used – so that in most cases, they display results of complex chaos, non-linearity and dynamism (as expected) of the underlying probabilities of data feats of interest. To curb this, we employed Cultural-GA, which ensures via its belief functions that all conditions to yield better generation is met with the processes of crossover and mutation applied.

References

- Aleksey, A., Alexander, A., (2016). *Kohonen map to representative sample formation in training*, [web]: researchgate.net/publication/303635615_Kohonen_selforganizing_map_application_to_representative_sample_formation_in_training_of_multilayer_perceptron
- Aleskerov, E., Freisleben, B., Rao, B., (1997). *Cardwatch: A neural network based database Mining System for Credit Card Fraud Detection*, Proc. IEEE Computational Intelligence for Financial Eng., pp. 220-226
- Ananya, M., (2012) *Cancer Management*. News-Medical, Available on [web]: www.news-medical.net/health/Cancer-Management.aspx, p1-2
- Bolton, R., Hand, D., (2001). *Unsupervised Profiling Methods for Fraud Detection*. Credit Scoring and Credit Control VII, 22, pp149-178
- Bolton, R.J., Hand, D.J., (2002). *Statistical fraud detection: a review*, Statistical Science, 17(3), pp235-255
- Brause, R., Langsdorf, T., Hepp, M., (1999). *Neural Data mining for credit card fraud detection*, Proc. IEEE Int. Conf. Tools with Artificial Intelligence, pp. 103-106, 1999.
- Bray F, Parkin, D.M., (2009). *Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness*. European Journal of Cancer, Vol. 45, pp747–755.
- Burge, P., Shawe-Taylor, J., (2001). *Unsupervised Neural, Network Approach to Profiling the Behaviour of Mobile Phone, Users for Use in Fraud Detection*. J. of Parallel and Distributed Computing 61: 915–925
- Chiu, C and Tsai, C., (2004). *A Web Services-Based Collaborative Scheme for credit card fraud detection*, Proc. IEEE Int. Conf. e-Technology, e-Commerce and e-Service, pp. 177-181
- Curado, M.P., Voti, L, Sortino-Rachou, A.M., (2009). *Cancer registration data and quality indicators in low and middle income countries: their interpretation and potential use for the improvement of cancer care*, Cancer Causes Control, Vol. 20, pp751–756.
- Dawane J.S., Pandit V. A. (2012). *Understanding Redox Homeostasis and Its Role in Cancer*. *Clinical Diagnosis Research*, 6(10), pp1796-1802
- Delamaire, L and Abdou, H., (2009). *Credit card fraud and detection techniques: a review*, Banks and Bank Systems, 4(2), pp57-67
- Doll, R., Payne, P., Waterhouse, J.H., (1968). *Cancer incidence in five countries: A Technical Report*. 1st edition. Berlin: Springer-Verlag; 1968. p. 5-244.
- Dheepa, V., Dhanapal, R., (2009). *Analysis of Credit Card Fraud Detection Methods*, International Journal of Recent Trends in Engineering, 2(3), pp126-135.
- Duman, E.M., Ozcelik, H., (2011). *Detecting credit card fraud by genetic algorithm and scatter search*. Expert Systems with Applications, 38: pp13057–13063
- Fawcett, T., (1997). *AI Approaches to Fraud Detection and Risk Management*, AAAI Workshop. Technical Report WS-97-07, AAAI Press.
- Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin D., (2010). *Estimates of worldwide burden of cancer in 2008*, International Journal of Cancer, 127, pp2893–2917.
- Ferlay, J, Soerjomataram, I, Dikshit, R., (2017). *Cancer Incidence and mortality worldwide: Sources, methods and major patterns in Cancer Incidence in South-East Nigeria*, Orient Journal of Medicine, 29, pp1-12
- Fritz, A., Percy, C., Jack, A., (2000). *International Classification of Diseases for Oncology*. 3rd edition. Geneva: World Health Organization, p. 2-240.



- Ghazale, B., (2018). *Reasoning Using Modular Neural Network: an Innovative Solution to address question answering AI tasks*, Available from [web] towardsdatascience.com/reasoning-using-modular-neural-networks-f003cb6109a2?gi=7dbcd12eb7c, July 18, 2020
- Ghosh, S., Reilly, D.L., (1994) *Credit Card Fraud Detection with a Neural-Network*, Proc. Int. Conf. System Sciences: Decision Support and Knowledge-Based Systems, 3, pp.621-630
- GLOBOCAN., (2012).. *Int J Cancer* 2015; 136: E359-E386.
- Hand, D.J., Blunt, G., Kelly, G., Adams, N.N., (2000). *Data mining for fun and profit*, Statistical Science, 15(2), pp. 111-131,
- Ikawo, O.E., (2013). *A literature review of cancer in Nigeria*, Thesis submitted to Department of Biochemistry, College of Medicine, University of Lagos.
- Jedy-Agba, E., Curado, M., Ogunbiyi O., (2012). *Cancer Incidence in Nigeria: A report from population-based Cancer Registries*, *Cancer Epidemiology*, 36(5): e271–e278. doi:10.1016/j.canep.2012.04.007.
- Khin, E.M., (2019). *Employing artificial intelligence to minimize internet fraud*. *Int. Journal Cyber Society & Education*, 2(1), pp.61-72,
- Kim, M.J., Kim, T.S., (2002). *A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection*, Proc. Int. Conf. Intelligent Data Eng. and Automated Learning, p378-383
- Maes, S., Tuyls, K., Vanschoenwinkel, B., Manderick, B., (2017). *Credit Card Fraud Detection*, Vrije Universiteit Brussel – Department of Computer Sci., Pleinlaan 2, B-1050, Belgium. [web]:personeel.unimaas.nl/k-tuyls/publications/papers/maenf02.pdf
- Malcolm, R., (2001). *Cancer in Encyclopedia of Life Sciences*. Nature Publishing Group, Imperial College School of Medicine, London, UK. Pg 420-428
- Malek, W.M., Mayes, K., Markantonakis, K., (2008). *Fraud Detection and Prevention in Smart Card Based Environments Using Artificial Intelligence*. Int. Conf. CARDIS 2008, London, UK
- Marane, A., (2011). *Utilizing Visual Analysis for Fraud Detection, Understanding Link Analysis*, [web]: linkanalysisnow.com/2011/09/leveraging-visual-analytics-for.html
- Mills, E.J., Bakanda, C., Birungi, J., (2011). *Male gender predicts mortality in a large cohort of patients receiving antiretroviral therapy in Uganda*, *Journal of Int. AIDS Society*, 14, pp52-65.
- Minahan, T., (2013). *Fraud detection and prevention*. Available online and retrieved 2020 from [web]: nebhe.org/info/pdf/tdbank_breakfast/Fraud_Prevention_and_Detection.pdf
- Murad, U., Pinkas, G., (1999). *Unsupervised Profiling for Identifying Superimposed Fraud*. Proc. of PKDD99.
- Nigrini, M., (2011). *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigation*. Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-470-89046-2
- Ojugo, A.A., Eboka, A.O., (2018a). *Comparative evaluation for high intelligent performance adaptive model for spam phishing detection*, *Digital Technology*, Vol. 3, No.1: pp. 9-15, doi: 10.1269/dt-3-1-1
- Ojugo, A.A., Eboka, A.O., (2018b). *Modeling solution of market basket associative rule mining approaches using deep neural net*, *Digital Technology*, 3(1), pp.1–8, doi: 10.12691/dt-3-1-1
- Ojugo, A.A., Eboka, A.O., (2019). *Signature-based malware detection using approximate Boyer Moore string matching algorithm*, *Int. Journal of Mathematical Sciences and Computing*, 3(5): pp49-62, doi: 10.5815/ijmsc.2019.03.05
- Ojugo, A.A., Eboka, A.O., (2020). *Memetic algorithm for short messaging service spam filter text normalization and semantic approach*, *International Journal of Information & Communication Technology*, 9(1), pp. 13 – 27, doi: 10.11591/ijict.v9i1.pp9-18
- Ojugo, A.A., Eboka, A.O., (2020). *Empirical evaluation on comparative study of machine learning techniques in detection of DDoS*, *Journal of Applied Science, Engineering, Technology & Education*, 2(1), pp18–27, doi: 10.35877/454RI.asci2192
- Ojugo, A.A., Eboka, A.O., (2020). *Modelling behavioral evolution as social predictor for the coronavirus contagion and immunization in Nigeria*, *Journal of Applied Science, Engineering, Technology & Education*, 3(2): pp37–45, doi: 10.35877/454RI.asci130

- Ojugo, A.A., Eboka, A.O., (2021). *Empirical Bayesian network to improve service delivery and performance dependability on campus network*, International Journal of Artificial Intelligence, 10(3), pp623-635
- Ojugo, A.A., Otakore, O.D., (2018). *Improved early detection of gestational diabetes via intelligent classification models*, J. of Computer Science & Application, 6(2), pp. 82-90, doi: 10.12691/jcsa-6-2-5
- Ojugo, A.A., Otakore, O.D., (2020). *Forging an optimized Bayesian network model with selected parameter for detection of the Coronavirus in Delta State of Nigeria*, J. of Applied Sci., Engr., Tech. & Edu., 3(1): pp37–45, 2020, doi: 10.35877/454RI.asci2163
- Ojugo, A.A and Otakore, O.D., (2021). *Forging optimized Bayesian network model with selected parameter for detection of Coronavirus in Delta State Nigeria*, Journal of Applied Science, Engineering, Technology & Education, 3(1): pp37–45, doi: 10.35877/454RI.asci2163
- Ojugo, A.A., Oyemade, D.A., (2021) *Boyer Moore string-match framework for a hybrid short messaging service spam filtering technique*, IAES International Journal Artificial Intelligence, 10(3): pp519-527
- Ojugo, A.A., Yoro, R.E., (2020). *Empirical solution for an optimized machine learning framework for anomaly-based network intrusion detection*, Technology Report of Kansai University, TRKU-13-08-2020-10996, 62(10): pp6353-6364
- Ojugo, A.A., Yoro, R.E., (2021). *Forging a deep learning neural network intrusion detection framework to curb distributed denial of service attack*, International Journal of Electronics and Computer Engineering, Vol. 11, No. 2, pp 128-138
- Ojugo, A.A., Allenotor, D. Oyemade, D., Longe, O., Anujeonye, C.N., (2015a). *Comparative stochastic study for credit-card fraud detection models*, African Journal of Computing and ICT, 8(1-2): pp15 – 24, 2015.
- Ojugo, A.A., Eboka, A., Yoro, R.E., Yerokun, M., Efozia, F., (2015b). *Framework design for statistical fraud detection*, Mathematics and Computers in Sciences and Engineering, 50: 176-182, ISBN: 976-1-61804-327-6.
- Ojugo, A.A., Ben-Iwhiwhu, E., Kekeje, O., Yerokun, M., Iyawah, I.J., (2014). *Malware propagation on time varying networks: comparative study*, International Journal of Modern Education and Computer Science, 6(8), pp. 25-33, doi: 10.5815/ijmecs.2014.08.04
- Okobah, I.P., Ojugo, A.A., (2018). *Evolutionary memetic models for malware intrusion detection: a comparative quest for computational solution and convergence*, IJCAOnline International Journal of Computing Application. 179(39), pp34–43
- Phua, C., Alahakoon, D., Lee, V., (2004). *Minority Report in fraud detection: classification of skewed data*, ACM SIGKDD Explorations Newsletter, 6(1), pp. 50-59, 2004
- Phua, C., V. Lee, K. Smith, R. Gayler, (2007). *A comprehensive survey of data mining-based fraud detection research*, [web]: www.bsys.monash.edu.au/people/cphua/.
- Stolfo, S. J., Fan, D. W., Lee, W., Prodromidis, A and Chan, P. K. (2000). *Cost-Based Modeling for Fraud and intrusion detection: results from the JAM Project*, In Proc. DARPA Information Survivability Conf. and Exposition, vol. 2, pp. 130-144.
- Stolfo, S.J and Prodromidis, A.L, (1999). *Agent-based distributed learning applied to fraud detection*, Technical Report CUCS-014-99, Columbia University, 1999
- Stolfo, S.J., Fan, D.W., W. Lee, L.K. Prodromidis, P.K. Chan., (1997). *Credit card fraud detection using meta-learning: issues and initial results*, Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90
- Syeda, M., Zhang, Y. Q. and Pan, Y. (2002). *Parallel Granular Networks for Fast Credit Card Fraud Detection*, Proc. IEEE Int'l Conf. Fuzzy Systems, pp. 572-577.
- Sylla, B.S., Wild, C.P., (2011). *A million Africans a year dying from cancer by 2030: What can cancer research and control offer to the continent?*, International Journal of Cancer, 130(2), pp245–250.
- Tobiyama, S., Yamaguchi, Y., (2016). *Malware detection with deep neural network using process behaviour*, IEEE 40th Annual Computer Software and Applications Conf., Vol. 2, pp. 577-582, 2016
- Ukah, C., Nwafor, A.M., (2017). *Cancer incidence in south-east Nigeria: a report from Nnewi cancer registry*, Orient Journal of Medicine, 29(1-2), pp48 – 56

- Vatsa, V., Sural, S., Majumdar, A.K., (2005). *A game-theoretic approach to credit card fraud detection*, In. Proc. of Int. Conf. Information Systems Security, pp. 263-276.
- Voosoghi, R.B., Ghaffari, M and Razin, R., (2019). *Evaluation of the Efficiency of Adaptive Neuro Fuzzy Inference System in modeling of the Ionosphere Total Electron Content Time Series Case Study: Tebran Permanent GPS Station*, Journal of Geomatics Science and Tech., Vol. 8, no.4, Pp. 109-119
- Wheeler, R., Aitken, S., (2019). *Multiple Algorithms for Fraud Detection Artificial intelligence Applications*, The University of Edinburg, Scotland, pp. 1-12, [web]: <http://home.cc.gatech.edu/ccl/uploads/45/multiple-algorithms-for-fraud.pdf>
- Xu, J., Sung, A.H., Liu, Q., (2007). *Behaviour Mining for Fraud Detection*, Journal of Research and Practice in Information Technology. 39(1), pp. 3–18