

Diagnosis of Head and Neck Cancer in Developing Countries using a Stacked Ensemble Model

F. Akinbohun, Ambrose Akinbohun, Adekunle Daniel and Oghenerukevwe E. Oyinloye

Abstract—Head and neck cancers (HNC) are indicated when cells grow abnormally. The incidence of HNC is on the increase owing to several factors. There is often late presentation that can result in loss of lives (mortality) especially in Africa due to paucity of specialists. These challenges prompted the development of a stacked ensemble model for diagnosis of HNC to facilitate prompt referral. The data were collected which consists of 1473 instances with 18 features. Information Gain was used for selecting important features and three supervised learning algorithms were deployed for the base learners: Decision Tree (C4.5), K-Nearest Neighbors and Naïve Bayes. The predictions of the base learners were combined and passed to meta learners: Logistic Model Tree (LMT). The result showed that Information Gain method with stacked LMT was 95.11%. It was deduced that both Information Gain with stacked LMT produced higher accuracy than that of the base learners' results. Hence, this stacked model can be used for diagnosis of HNC in healthcare systems.

Index Terms—Decision Tree, Larynx, Naïve Bayes, Sinonasal.

I. INTRODUCTION

Head and neck cancer (HNC) are the types of cancer which occur as a result of uncontrollable growth of abnormal cells in head and neck regions of the body. Cancers can affect any part of human body but if they affect head and neck regions, they are known as head and neck cancers (HNC). The part of human body that HNC affect could be of various parts in the head and neck regions. These are: nasopharyngeal, thyroid, larynx, sinonasal etc. There are factors such as chemical or toxic compound exposure, ionizing radiation, pathogens, human genetics and so on that can predispose to HNC.

Patients with HNC present some symptoms which include weight loss, bleeding, swelling, hemoptysis, dyspnoea etc. Fatigue is a significant symptom of head and neck cancer patients [1]. These (symptoms) and signs are what we called features of HNC. The features were considered to predict the types of cancer around head and neck regions.

There are a large number of people who suffer from cancer [2]. Cancer mostly called malignant tumour has killed so many people in the world [3]. It is a deadly disease that needs utmost attention by both individual and government. National Cancer Institute [4] reported that patients with HNC worldwide is in excess of 550,000 cases and mortality rate was about 300,000 each year. The primary health workers are the first contact that the patients present to especially in developing countries; and the ability for the health workers to diagnose head and neck cancers is abysmally poor. This leads to delay in referral to the specialists in Tertiary Hospitals.

Healthcare organizations are going deep into the data analytics and clinical decision support environments to support population health management and value-based care. The healthcare industries have a large amount of patient records which are available from where information are to be extracted to assist the development of medical researches. These data are capable of reducing the problem being faced by primary health personnel. The use of predictive model for the analysis of medical data helps in diagnosis of HNC.

The application of machine learning in diagnosing head and neck cancer is a means of using available data which are meant to extract pattern and knowledge. Machine learning is of great significance to diagnose diseases including cancer. In machine learning, classification is a task or function that assigns items in a collection to target or classes. The goal of classification is to accurately predict the target class for each case in the data. A classification task begins with a data set in which the class assignments are known. In the model (training) process, different classification algorithms use different techniques for finding relationships between the values of the features and the values of the target or class. The task has the ability to predict the types of cancer around head and neck regions.

If there is no diagnosis of disease (HNC), there will be no prompt referral to the specialists in tertiary institutions. Hence, it is of great significance to employ machine learning algorithm to predict the types of cancer around head and neck regions especially at the primary healthcare level.

Cancers are needed to be diagnosed early so that treatment protocols can commence immediately. When HNCs are diagnosed early enough, then there would be good prognosis.

II. RELATED WORK

Some works by researchers which included the use of machine learning algorithms to diagnose diseases have been

Published on September 15, 2020.

F. Akinbohun (Corresponding Author), Rufus Giwa Polytechnic, Owo, Ondo State, Nigeria.

(e-mail: folakeakinbohun@yahoo.com)

A. Akinbohun (Principal Consultant), Department of Otorhinolaryngology, University of Medical Sciences Teaching Hospital, Akure, Ondo State, Nigeria.

(e-mail: akinbohunambrose@yahoo.com).

A. Daniel (Lecturer, Honorary Consultant), Department of Otorhinolaryngology, University College Hospital, Ibadan, Nigeria.

(e-mail: dkunle2013@yahoo.co.UK).

O. E. Oyinloye (Lecturer), Ekiti State University, Ado Ekiti, Ekiti State, Nigeria.

(e-mail: rukkivie@yahoo.com).

reviewed. They are, among others:

Two prediction models (Radiotherapy (RT) and End of Treatment (EOT)) at different time points were developed to predict weight loss ≥ 5 kg (yes/no) at 3 months post-Radiotherapy (RT) using Classification and Regression Tree (CART) algorithm [5]. The study indicated that CART could be employed for weight loss prediction in patients with head and neck cancer.

[6] evaluated the use of Classification and Regression Tree (CART) model in predicting weight loss following head and neck cancer radiation therapy. Two prediction modes were developed to predict weight loss ≥ 5 kg at 3 months' post-radiation therapy. The modes were (1) during radiation therapy planning using patient demographic, delineated dose data, planning target volume–organs at risk shape relationships data and (2) at the end of treatment (EOT) using additional on-treatment toxicities and quality of life data. This showed that CART was deployed to predict weight loss in head and neck cancers patients.

[7] evaluated a pattern of follow up visits among patients with head and neck cancers in Jos, North-Central, Nigeria. Data were collected from Jos University Teaching Hospital, Jos, North-central Nigeria and analyzed using the statistical software.

[8] considered a classifier model using Machine Learning Algorithms on differential diagnosis of suspicious thyroid nodules via Sonography (Ultrasound). The study included 970 histopathologically proven thyroid nodules in 970 patients. Two radiologists retrospectively reviewed ultrasound images, and nodules were graded according to a five-tier sonographic scoring system. Naïve Bayes, Radial Basis Function Neural Network, Support Vector Machine were employed for diagnosis of thyroid nodules.

[9] described competing causes of death in the head and neck cancer population. They identified patients with first mucosal head and neck cancer (HNC) from the Surveillance, Epidemiology and End Result database. The study showed that patients with head and neck cancer died from competing causes.

[10] developed a feature selection method for oral cancer using Apriori algorithm where the original algorithm of Boolean association rules of mining frequent item sets was employed. The data mining methods were explored to identify the suitable techniques for efficient classification of data.

[11] deployed Back Propagation Algorithm Neural Network to diagnose thyroid disease. The study indicated that developed neural network could be used as a diagnostic tool for earlier prediction of a thyroid disease.

[12] worked on the prediction of acute myeloid leukemia cancer using data mining. Data mining techniques such as Bayes Network, Jrip, J48, Multilayer Perceptron, IBK, Decision Tree were used on the dataset. The study showed that data mining algorithms could be used for myeloid leukemia cancer prediction.

[13] carried out a study where the objective of the study was to predict breast cancer survivability where different data mining techniques: Naïve Bayes, Back-Propagated Neural Network, and the C4.5 decision tree algorithms were adopted. It was established that C4.5 algorithm had a better performance than the other two techniques. The study was

limited to only one type of cancer (breast).

Prediction of survival in patients with esophageal carcinoma (cancer) using Artificial Neural Networks was conducted by [14]. It was observed that accurate estimation of outcome in patients with malignant disease was an important component of the clinical decision-making process

III. MATERIALS AND METHODS

The architecture of a stacked ensemble model to diagnose head and neck cancers is presented in Figure 1.

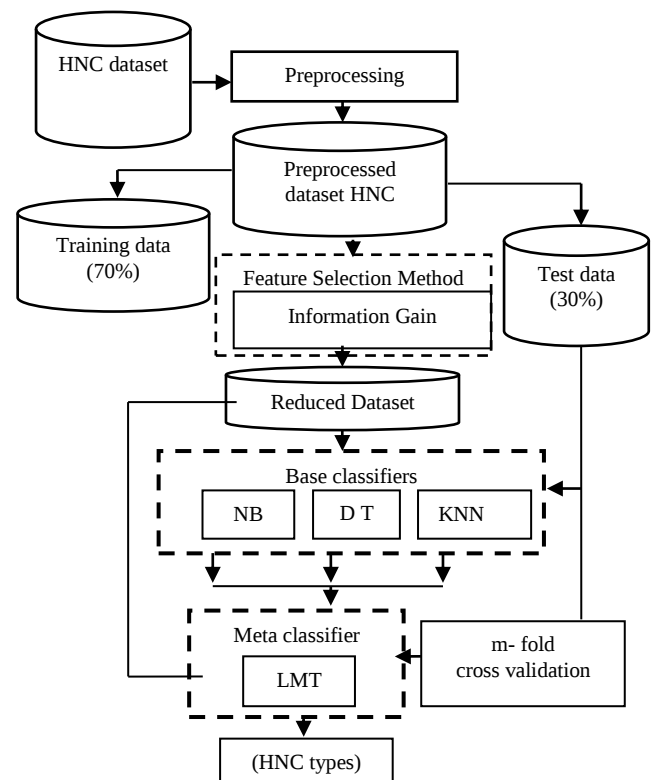


Fig 1. Architecture of a proposed model for HNC diagnosis

NB represents Naïve Bayes; DT means Decision Tree; KNN is the K-Nearest Neighbors and LMT is the Logistic Model Tree; HNC connotes Head and Neck Cancer

The major components of the architecture for HNC diagnosis model are collection of data, preprocessing, feature selection, base models and meta models/super learners.

The components of the proposed ensemble model for diagnosis of head and neck cancers are explained as follows:

A. Collection of Data

Data were collected from three Hospitals in Nigeria: Obafemi Awolowo University Teaching Hospital, Ile-Ife; Federal Medical Centre, Owo and University of Medical Sciences Teaching Hospital, Akure. The data were preprocessed. Information Gain was deployed for feature selection so that irrelevant feature could be removed from the dataset. The HNC dataset gathered consists of 18 features which are: bleeding, poor appetite, weight loss, snoring, swelling, nasal blockage, mouth breathing, hyponasal speech (poor speech), halitosis, facial asymmetry, fatigue, hoarseness, dyspnea, tinnitus, haemoptysis, proptosis, odynophagia and dysphagia. The

classes include: Sinonasal, nasopharyngeal, laryngeal and thyroid. The number of instances is 1473. The attribute values are converted to numeric.

B. Feature Selection

During data collection, problems like irrelevant features might occur. We introduced feature selection method where Information Gain algorithm was used to remove irrelevant features from the dataset (HNC), to improve accuracy and to reduce training time. This task was tackled in the prepared data before applying learning algorithms. The Information Gain as a feature selection method is described thus:

Information Gain is used to measure the dependence between features and labels and calculates the information gain between the feature and the class labels C . It is a method that ranks features based on a relevancy score which is based on each individual attribute [15].

It finds the entropy of the classes, features subsets and weighted average entropy in order to get the information gain of each feature. The weighted information gains of the features are ranked. In information gain, a feature is relevant if it has a high information gain.

To calculate the information gain, the entropy of the HNC type (class) is calculated using Equation 1.

$$E = - \sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

Where P_i is the proportion of examples in HNC that belongs to the i -th class, n is number of classes and E is the entropy. Information Gain is expressed as the difference between the prior entropy of classes and posterior entropy [16].

C. Construction of the base model

The base models deployed for this study were Decision Tree, K-Nearest Neighbor and Naïve Bayes. The dataset was divided into: training data and testing data. 1031 records were used for training data while 442 records were used for test data. The training data or cases were assumed to be represented as a pair $[x_1, x_2, x_3, \dots, x_n \rightarrow y]$ where $x_1, x_2, x_3 \dots x_n$ are vectors of attribute values describing some cases while y is the appropriate class or target. The first base model is explained as follows:

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences. Decision tree is a predictive model to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). Decision tree is constructed in a top-down recursive divide-and-conquer manner [17].

The goal of decision tree is to create a model that predicts the value of a target variable based on several input variables. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. It is a flow-chart-like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. These nodes in turn will continue to split until a final node or leaf node is grown. The leaf node determines the final classification of the variable being tested.

The second base model employed in this paper was K-

Nearest Neighbors (KNN).

The k-NN algorithm belongs to the family of instance-based and lazy learning algorithms [18]. Instance-based algorithms are those algorithms that model the problem using data instances (or rows) in order to make predictive decisions [19].

When new unlabeled data comes in, k-NN operates in two basic steps: Firstly, it looks at the k closest labeled training data points. Secondly, using the neighbors' classes, k-NN gets a better idea of how the new data should be classified. For continuous data, k-NN uses a distance metric like Euclidean distance to locate what is closer to k [20]. The Euclidean distance is given in Equation 2

$$Euclidean = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

where x represents the query-instance and y represents all the training samples;

$(x_i - y_i)$ represents distance between the query-instance and all the training samples.

The third base model was Naive Bayes classifier. It is a classification technique and frequently trumps a more sophisticated predictive analytic tool which is designed for use when predictors are independent of one another within each class [21].

Naive Bayes Classifier technique is based on Bayesian theorem. It is a conditional probability model that enables conditional predictions. Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$ using the Equation 3:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (3)$$

$P(c|x)$ is the posterior probability of class (target) given predictor (attribute) of class.

$P(c)$ is called the prior probability of class.

$P(x|c)$ is the likelihood which is the probability of predictor of given class.

$P(x)$ is the prior probability of predictor of class.

It combines this model (conditional probability) with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the Maximum a Posteriori (MAP) decision rule.

D. Stacking with Logistic Model Tree (LMT)

The predictions of the base models were combined, and at super level or meta level, Logistic Model Tree Algorithm was used for building the stacking.

Logistic Model Tree (LMT) is a supervised training model that combines Logistic Regression and Decision Tree [22].

Logistic model tree is based on a decision tree that has linear regression model at its leaves to provide a piecewise linear regression model. The algorithm makes use of cross validation to find a number of LogitBoost iterations that prevent overfitting.

IV. RESULT

When Information gain algorithm was applied, the threshold was set at <0.2 , the result is presented in Table 1 where features with higher values were used and passed for the training of base models and a meta model.

TABLE I: RESULT OF INFORMATION GAIN

Ranked Feature	Information Gain values
Nasal blockage	0.791
Dyspnoea	0.757
Haemoptysis	0.757
Snoring	0.749
Hoarseness	0.74
Facial Asymmetry	0.542
Fatigue	0.362
Hyponasal	0.335
Speech	
Proptosis	0.325
Tinnitus	0.317
Poor Appetite	0.30
Odynophagia	0.303
Mouth Breathing	0.302
Halithosis	0.269
Swelling	0.236
Bleeding	0.225
Dysphagia	0.206

A. Results of the Information Gain with Decision Tree Data

The selected features by the information gain were trained on the base models: Decision Tree (C45), KNN and Naïve Bayes in the ratio 70% of the dataset for training set and 30% of the data for test set. The performance metrics we considered for base level are accuracy, precision recall and F1 score. The results are presented in Table 2 and accuracies of the base models on Information Gain is shown in Figure 2

TABLE II: RESULTS OF THE PERFORMANCE EVALUATION FOR THE BASE MODELS WITH INFORMATION GAIN

Base models	Accuracy	Precision	Recall	F1 Score
Decision Tree	93.21%	0.943	0.932	0.931
KNN	94.80%	0.949	0.948	0.948
Naïve Bayes	94.12%	0.945	0.941	0.941

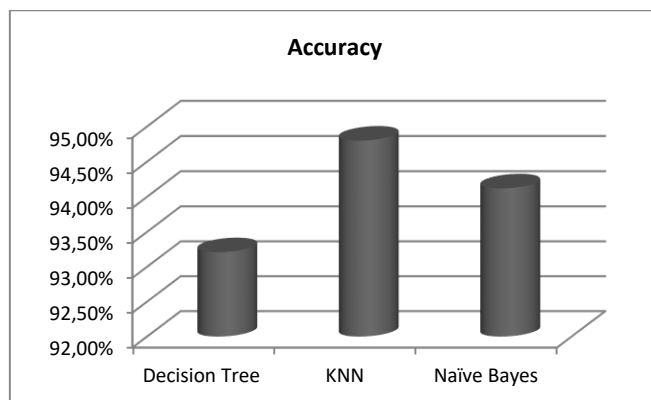


Fig 2. Accuracy of the base models with Information Gain

B. Stacking with Logistic Model Tree (LMT) with Information Gain Selection

The predictions of the three base models (Decision tree (C4.5), KNN and Naïve Bayes) were stacked with Logistic Model Tree. This was used on the selected features of Information Gain. The stacking was done on the basis of cross validation. The results is presented in Table 3

TABLE III: RESULT OF THE PERFORMANCE EVALUATION OF STACKED ENSEMBLE MODEL USING LMT

Metrics	Values
Accuracy	95.11%
Precision	0.952
Recall	0.951
F1 Score	0.951
Kappa	0.9336
Statistic	
Mean Absolute error	0.0304

V. DISCUSSION

From the results presented, the following facts are inferred:

The HNC dataset was trained on Information Gain method with base models namely Decision Tree, KNN and Naïve Bayes. The result showed that the accuracy of Decision Tree (C4.5) was 93.21%, KNN was with accuracy of 94.80% and Naïve Bayes with accuracy of 94.12%. Based on accuracy, it indicated that KNN had the highest accuracy of 94.80% followed by Naïve Bayes with accuracy of 94.12% while Decision Tree had the least accuracy of 93.21%. Using F1 score for evaluation, Decision Tree, KNN and Naive Bayes have 0.931, 0.948 and 0.941 respectively. From this metric, it showed that KNN had the highest F1 Score followed by Naïve Bayes and the least was Decision Tree. This implied that KNN was a better predictive model to predict or diagnose the type of HNC when Information Gain method was applied in the ratio 70% of the dataset for training set and 30% of the dataset for test set.

Following the predictions of the base learners, they (predictions) were combined. Logistic Model Tree was used at a stacked meta level to train the HNC dataset in a stratified cross validation. This was done on the selected features using Information Gain.

The results of the stacked model of Head and Neck Cancer dataset was 95.11% for Logistic Model Tree (LMT). This showed that LMT was higher than the values of the three base models. i.e. There was an improvement of the accuracy.

VI. CONCLUSION

HNC has become a global disease that its mortality and morbidity rate are on the increase because of late presentation and lack of access to the specialists in Ear, Nose, Throat/Head and Neck. Due to the common clinical features patients with HNC present with, it is of importance to deploy the use of stacked ensemble learning algorithms to predict the appropriate head and neck cancer type.

This study recommends that machine learning algorithms can be used to diagnose head and neck cancers early in developing countries especially as patients in these countries (developing) present to primary health care facilities as first port of call. The study helps primary health workers to diagnose early and refer to the tertiary health institutions where ENT specialists are found.

ACKNOWLEDGMENT

The following people are acknowledged: Prof Esan (OAUTH, Ife), Dr Pelemo (UNIMEDTH, Akure), Dr Ariyibi (FMC, Owo), Dr Ameye (OAUTH, Ife)

REFERENCES

- [1] Vipul, N., Punita, L., Mranalini, V. and Rajan, Yadav (2015). Evaluation of fatigue in Head and Neck Cancer Patients undergoing (intensity modulated radiation therapy) radiotherapy: A prospective study. *Asian Journal of Oncology*. 1(1). DOI: 10.4103/2454-6798.165111
- [2] GBD (2015). Mortality and Causes of Death, Collaborators. Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980-2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet*. 388 (10053): 1459–1544.
- [3] World Health Organization (2014). *World Cancer Report 2014*. Chapter 5.8.
- [4] National Cancer Institute. Available at <https://www.cancer.gov/about-cancer/understanding/statistics>
- [5] Zhi, C., Minoru, N., Chen, H., Scott, P. R., Xuan Hui, J. A. M., Michael, R. B., Ana, P. K., Brandi, R. P., Laura, B., Mariah, M., Amanda, Choflet, K. S., Shinya, S., Kazuki, U., John, W. W., Todd, R. M. and Harry, Q. (2018). Evaluation of Classification and Regression Tree (CART) Model in Weight Loss Prediction Following Head and Neck Cancer Radiation Therapy. *Advances in Radiation Oncology*.3(3): 346–355
- [6] Amanda, C., Kousuke, S., Shinyasugiyama, Kazuki, U., John, W. W., Todd, R. M. and Harry, Q. (2018). Evaluation of Classification and Regression Tree (CART) Model in Weight Loss Prediction Following Head and Neck Cancer Radiation Therapy. Elsevier Inc. on Behalf of the American Society for Radiation Oncology. 3(3): 346–355
- [7] Adoga, A. A., Silas, O. A., Nimkur, T. L. (2009). Open Cervical Lymph Node Biopsy for Head and Neck Cancers: Any Benefit? *Head Neck Oncol*.1:9.
- [8] Hongxun, W., Zhaohong, D., Bingjie, Z. and Qianyun L.(2016). Classifier Model Based on Machine Learning Algorithms: Application to Differential Diagnosis of Suspicious Thyroid Nodules Via Sonography. *American Journal of Roentgenology*. 207 (4): 859-864
- [9] Massa, S. T., Osasuwa-Peters, N, Christopher, K.M., Arnold, L. D., Schootman, M., Walker, R. J. and Varvares, M. A. (2016). *Competing Causes of Death in the Head and Neck Cancer Population*. National Center for Biotechnology Information, U.S. National Library of Medicine Elsevier Ltd.
- [10] Vidhu, R. and Kiruthika, S. (2016). A New Feature Selection Method for Oral Cancer using Data Mining Techniques. *International Journal Of Advanced Research In Computer and Communication Engineering (IJARCCE)* 5 (1)
- [11] Prerana, P. S. and Khushboo, T. (2015). Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network. *International Journal of Research in Management, Science & Technology*. 3(2)
- [12] Durairaj, M. and Deepika R. (2015). Prediction of Acute Myeloid Leukemia Cancer Using Data Mining- A Survey. *International Journal Of Emerging Technology and Innovative Engineering I* (2)
- [13] Abdelghani, B. and Erhan, G. (2005). Predicting Breast Cancer Survivability using Data Mining Techniques. Department of Computer Science, the George Washington University, Washington DC
- [14] Sato, F., Shimada, Y., Selaru, F. M., Shibata, D., Maeda, M., Watanaabe, G., Mori, Y. S. S., Imamura, M. and Meltzer, S. J. (2005). Prediction of Survival in Patients with Esophageal Carcinoma Using Artificial Neural Networks. *American Cancer Society*.
- [15] Osiris Villacampa (2015). *Feature Selection and Classification Methods for Decision Making: A Comparative Analysis*. Nova Southeastern University NSUworks. CEC Theses and Dissertations, College of Engineering and Computing
- [16] Jasmina, N., Perica, S. and Dusan, B. (2011). Toward Optimal Feature Selection Using Ranking Methods and Classification Algorithm. *Yugoslav Journal of Operations Research*. 21(1): 119-135
- [17] Jiawei H., Micheline, K. and Jian, P. (2011). *Data Mining: Concepts and Techniques* 3rd Edition
- [18] Altman, N. S. (1992). *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*. 46 (3): 175–185.
- [19] Wu, X., Kumar V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., Mclachlan, G. J., Ng, A., Philip S. B., Zhi-Hua Zhou, Y., Steinbach, M., Hand D. J., and Steinberg, D. (2007). *Top 10 Algorithms in Data Mining*. Knowledge Information System. Springer-Verlag London Limited
- [20] Sagar, S. N. (2012). *A Comparative Study of Classification Techniques in Data Mining Algorithms*. International Conference on Computer Science and Electronics Engineering
- [21] Han, J., Kamber, M. and Pei, J. (2012). *Data Mining: Concepts and Techniques*, 3rd Edition. Elsevier, Amsterdam
- [22] Landwehr, N., Hall, M. and Frank, E. (2005). *Logistic Model Trees*. *Machine Learning*.59: 161.



Folake Akinbohun works with Rufus Giwa Polytechnic, Owo, Ondo State, Nigeria. She bags Master of Technology in Computer Science from Federal University of Nigeria in 2012. She has rounded up her Ph.D in Computer Science at the same University. Her areas of specialization are Data Science and Machine Learning. She has many publications both at local and international levels. She is a member of Nigeria Computer Society (NCS) and Women in Artificial Intelligence (WAI).



Ambrose Akinbohun is Principal Consultant in Ear, Nose, Throat, Head and Neck Surgery of University of Medical Science Teaching Hospital, Akure, Nigeria. He possesses special interest in Head and Neck Surgery. He is a trainer of medical students and Residents (Post Graduate Medical Doctors in Training). He is the current Head of Department and has authored a number of publications. He is a public speaker at both radio and television programmes. He is a fellow of the

West African College of Surgeons and an examiner of the West African College of Surgeons.

Adekunle Daniel is a Senior Lecturer and Honorary Consultant, Department of Otorhinlaryngology, Head and Neck, University of Ibadan, Oyo State, Nigeria. He has a number of peer-reviewed publications to his credit.

Oyinloye Oghenerukevwe Elohor is a lecturer in Computer Science Department, Ekiti State University, Ado Ekiti, Ekiti State, Nigeria. She bags PhD in Computer Science from Federal University of Technology, Akure, Nigeria. She is the Head of Department. She has many publications to her credit.