



CLASSIFICATION OF HEAD AND NECK CANCER TYPES USING MACHINE LEARNING ALGORITHM

Prof O. Olabode¹

¹*Federal University of Technology, Akure, Ondo State, Nigeria*

Prof A. O. Adetunmbi²

²*Federal University of Technology, Akure, Ondo State, Nigeria*

Folake Akinbohun³

³*Rufus Giwa Polytechnic, Owo, Ondo State, Nigeria*

Dr Ambrose Akinbohun⁴

⁴*University Teaching Hospital, Akure, Ondo State, Nigeria*

Article DOI: <https://doi.org/10.36713/epra3289>

ABSTRACT

The worldwide incidence of head and neck cancer exceeds half a million cases annually. The morbidity and mortality of head and neck cancers considering thyroid, nasopharyngeal, sinonasal and laryngeal were reported high. The degree of facial disfigurement is unrivalled. Information Gain and Chi Square, Decision and Naïve Bayes were deployed for the study. The dataset was divided into training and test data. The results showed that the performance of Naïve Bayes outperformed Decision Trees. With the application of machine learning algorithms, head and neck cancer can be classified.

KEYWORDS: *Head and Neck, thyroid, Chi Square, Information Gain*

INTRODUCTION

Cancer occurs because of mutations in the genes responsible for cell multiplication and repair (National Cancer Institute, 2015). Head and neck cancer begin with the mutation of healthy cells and grows without control around head and neck regions in the human body. The prevalence of primary head and neck tumors in Port Harcourt had increased. Sinonasal was adjudged the commonest site of cancer of head and neck cancer (Onotai and Nwogbo, 2012).

Akinkugbe *et al.*, (2010) hinted that 8.2 million people die from cancer worldwide every year (Head and Neck Cancers (HNC) inclusive); 4 million people die prematurely. 12.5% of all deaths were attributed to

cancer. If this continues, 16 million new cases would be diagnosed yearly; 70% out of the 16 million would be in developing countries.

HNCs are on the increase in Africa because of late presentation. Like many other diseases in Africans, cancer does not form part of governments' priority list. This may be due to lack of awareness of the real burden of the disease because the prevalence still remains largely undetermined in this environment. The epidemiology of cancer increases every year especially in developing countries like Nigeria.

The epidemiological information of HNC calls for in-depth research where computing methods can be



adopted to diagnose without delay so that the morbidity and mortality associated with late presentation could be drastically reduced.

RELATED WORK

Researches had shown the prevalence of head and neck cancer, different approaches on the use of diseases are presented below:

It was reported that more than 70% of throat cancers are at an advanced stage when discovered due to late presentation (American Cancer Society, 2002). Head and neck cancer specifically, nasopharyngeal cancer was the commonest cause of death in young men in Southeast China and Taiwan (*Titcomb, 2001*). Laryngeal cancer incidence is higher in African Americans, Asian and Hispanic populations and there is a lower survival rate for similar tumor states in African Americans with head and neck cancer (*Gourin et al., 2006*).

Heikki (2002) averred that the effect of professional diagnostic delays was a strong predictor of survival of patients of laryngeal cancer (squamous cell carcinoma). The method used was a population-based sample of 66 patients with Laryngeal Cancer Squamous Cell Carcinoma (LSCC) in Northern Finland between 1990 and 1995 which were investigated using statistical analysis.

Jajroudi *et al.* (2014) predicted the survival of thyroid cancer where ANN (Artificial Neural Network) and Logistic Regression were used. Effective features in thyroid cancer had been selected based on supervision by radiation oncologists. After data pruning, 7706 samples were studied with 16 attributes.

Amanda *et al.* (2018) evaluated the prediction model for weight loss (WL) in head and neck cancer (HNC) patients who were treated with radiation therapy (RT). Two prediction modes at different time points were developed to predict weight loss ≥ 5 kg at 3 months post-radiation therapy which were: during radiation therapy planning and at the end of treatment (EOT) using additional on-treatment toxicities and quality of life data. CART algorithm was used on the two prediction models. The study deduced that the potential value of an informatics had facilitated insight into the prediction of weight loss using the CART algorithm.

Fabio *et al.* (2017) used a predictive model to distinguish between malignant and benign thyroid nodules. Chi-square test was for categorical variables and t-test for continuous variables. Logistic regression was used to identify features independently associated with malignancy (dependent variable).

Adisa *et al.* (2012) conducted a study on the implications for treatment planning in limited resource settings on pediatric head and neck cancers in Nigeria. A retrospective study using data from the University College Hospital (UCH) Ibadan between 1990 and 2008 was conducted. Statistical analysis was used on the categorical data which were expressed as percentages and Chi-square statistics. Continuous data were summarized using mean, standard deviation (SD) and confidence intervals. The data were further compared using t-test and/or one-way analysis of variance test as appropriate. The outcome was that head and neck cancer occurred in children.

Renske *et al.* (2010) reported the knowledge of general practitioners about nasopharyngeal cancer at the Puskesmas in Yogyakarta, Indonesia. The level of awareness of general practitioners in prompt diagnosis of nasopharyngeal cancer was shallow.

The study of Jyoti *et al.* (2011) tends to achieve development of a predictive data mining for heart disease diagnosis. Three different supervised machine learning algorithms i.e. Naïve Bayes, K-NN, Decision List algorithm were used for analyzing the dataset. The use of Naïve Bayes performed best among K-NN and Decision Tree to predict heart disease.

Anurag and Sudsanshu (2010) compared two data mining techniques on thyroid cancer dataset. C4.5 and C5.0 data mining techniques were compared such that tree size of C4.5 was very large compared to C5.0. After pruning C5.0 tree generated more accurate rule set than C4.5.

Dursun (2009) worked on analysis of prostate cancer data using a data mining approach. Three popular data mining techniques such as decision trees, artificial neural networks and support vector machines along with the most commonly used statistical analysis technique logistic regression to develop prediction models for prostate cancer survivability.

METHODS

Head and neck cancers types can be classified using machine learning algorithms which consists of different components such as data collection, data preprocessing, feature selection methods (Filter-based methods) and building of models and predictive result.



HNC Diagnosis Model

The features of HNC are considered to predict the type of cancer in head and neck region a patient has.

Hence diagnosis of HNC is directly related to the features (x) of HNC in patient where the features are represented by x_1, x_2, \dots, x_n .

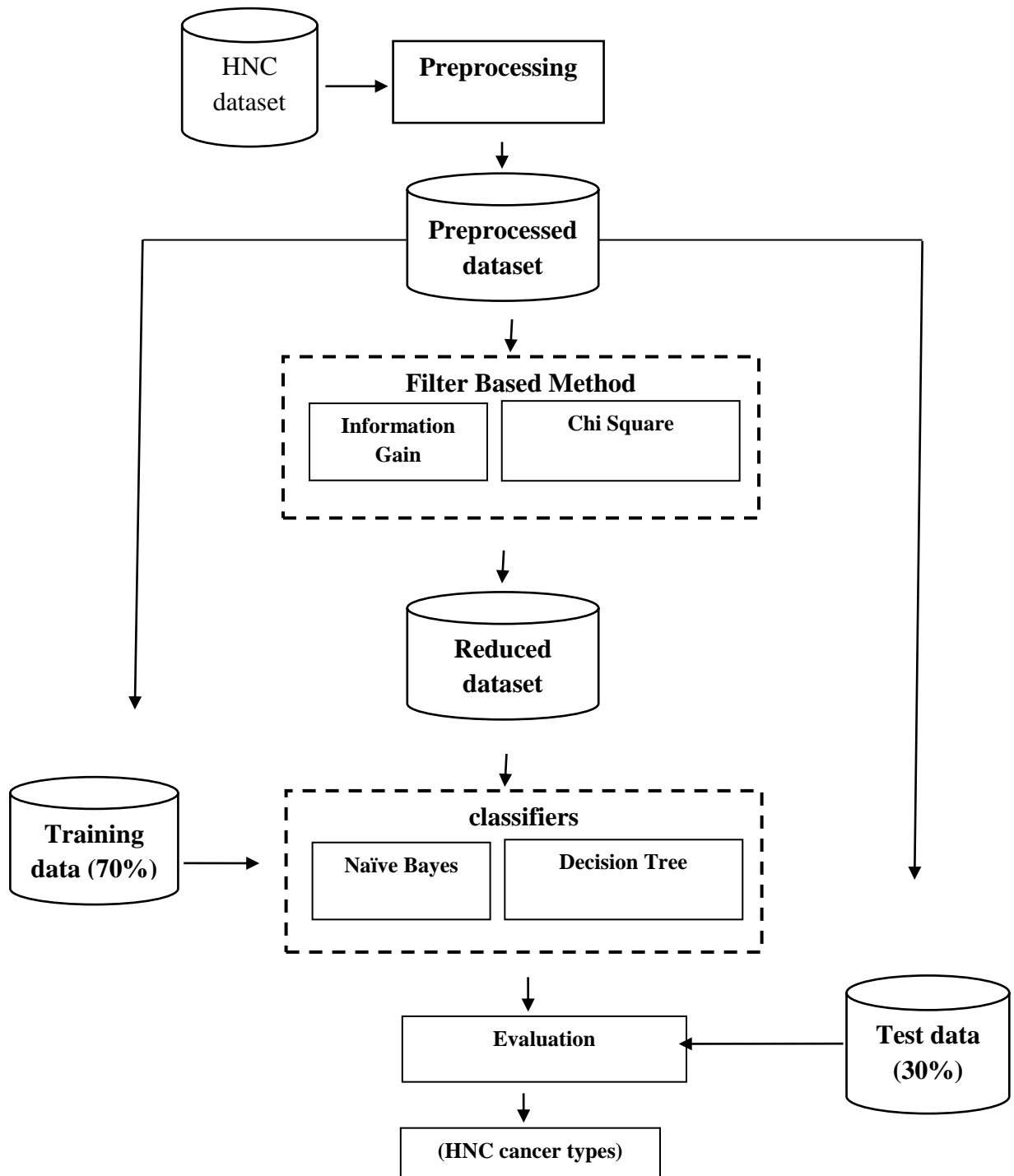


Figure 1: Architecture of HNC diagnosis model



HNC Data Set

The raw data were collected from three hospitals: University of Medical Sciences, Teaching Hospital, Akure, Federal Medical Centre, Owo, Ondo

State and Obafemi Awolowo University Teaching Hospital Complex, Ile-Ife, Nigeria.

The dataset consists of four classes which include Nasopharyngeal cancer, sinonasal cancer, laryngeal cancer and thyroid cancer, eighteen features

Table 1: Data Set

s/n	Feature	Attribute value
1	Bleeding	Mild/ moderate/severe/?
2	Poor appetite (Anorexia)	Yes/no
3	Weight loss	Yes/no
4	Snoring	Yes/no
5	Swelling	Small/medium/large
6	Nasal blockage	Yes/no
7	Mouth breathing	Yes/no
8	Hyponasal speech (poor speech)	Yes/no
9	Halithosis (mouth odour)	Yes/no
10	Facial Asymmetry	Yes/no
11	Fatigue	Yes/no
12	Hoarseness	Yes/no
13	Dyspnoea (difficult breathing)	Yes/no
14	Tinnitus (abnormal sound in the ear)	Yes/no
15	Haemoptosis (coughing out blood)	Yes/no
16	Proptosis (bulging eye)	Yes/no
17	Odynophagia (painful swallowing)	Yes/no
18	Dysphagia	Yes/no
	Class- Cancer type	Sinonasal/nasopharyngeal/ laryngeal/ thyroid

FEATURE SELECTION METHOD

After the data collection and preprocessing, filter -based methods followed:

Feature selection is the process of selecting a subset of relevant features for use in model construction (Jason, 2014). The reason for performing feature selection is to remove attributes that do not contribute to the efficacy and the efficiency of the models.

Two feature selection methods deployed were information Gain and chi square. They are described as follows:

I.Information Gain: Information Gain (IG) measures how much information a feature gives about the class. Features that perfectly partition should give maximal information. Unrelated features should give no information. The features with highest information gain are rated above the lowest features.

The Information Gain can be calculated using Equation 1 to find the entropy of the features and classes and use Equation 2 to calculate the Information Gain of each features in the HNC dataset.:

$$E = - \sum_{i=1}^n p_i \log_2 p_i \tag{1}$$

where P_i is the proportion of examples in HNC that belongs to the i -th class

n is number of classes, E is entropy

The gain of this test is

$$\text{Gain} = \text{Info (Class: cancer type)} - \text{Info (Attribute: for example Bleeding)} \tag{2}$$

II. Chi Square Test

Chi-square test was used to select relevant attributes in the dataset. It is a test to measure dependence between features, using this function “weeds out” the features that are the most likely to be independent of class.



Calculate Chi square (χ^2) statistic using Equation 4

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad 3$$

O is the observed frequency/count

E is the expected frequency/count

χ^2 is the Chi Square

Σ is the sum of above across all cells

SELECTING THE MODELS MODELS

After the filter-based method, the HNC data set consists of 1473 records and 18 features. 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 models considered for head and neck cancer prediction are: Naïve Bayes and Decision Tree.

A. Naïve Bayes

Naive Bayes is a probabilistic classification algorithm that is based on Bayes' Theorem. Naïve Bayes is a conditional probability model where the features are independent of each other.

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 5

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad 5$$

$P(c|x)$ 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.

Bayes Algorithm has three phases: learning phase, testing phase and estimate class phase for categorical distribution as presented below:

Algorithm

Learning phase

For each class value C_m

 Compute $P(C_m)$ //compute class /HNC types

For each attribute X_n

 // categorical distribution for C_m and x_n

 if $P(x_n|C_m) = 0$ then compute $P(x_j|C_m)$

$$P(x_j|C_m) = \frac{\text{count}(T, C_m, x_j) + 1}{\text{count}(T, C_m) + k}$$

 Else

$$P(x_n|C_m) = \frac{\text{count}(T, C_m, x_j)}{\text{count}(T, C_m)}$$

$P(C_m|x) = P(x_1, x_2, \dots, x_n|C_m)$ // compute distribution $D_{m,n}$

 // Using the chain Rule of the conditional probability

$$P(C_m|x) =$$

$$P(x_1|C_m)P(x_2|C_m, x_1)P(x_3|C_m, x_1, x_2)$$

$$\dots P(x_n|C_m, x_1, x_2, \dots, x_{n-1})$$

Testing phase

$$\text{Given unknown } x' = [x'_1, x'_2, \dots, x'_n]$$

Estimate class

$$\text{Return } C_m = \text{argmax}_{C_m} = P(x|C_m)$$

B. Decision Tree

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).

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. Decision tree is constructed in a top-down recursive divide-and-conquer manner. (Jiawei *et al*, 2011; Ian and Elibe, 2005).

RESULT AND DISCUSSION

Results of the Information Gain on Decision Tree (C45) and Naïve Bayes

The Information Gain of the features and Chi Square values are presented in Table 2.



Table 2: Information Gain and Chi Square results

Information Gain		Chi Square	
Features	Gain Values	Features	Values
Nasal Blockage	0.791	Dyspnea	0.959
Dyspnoea	0.757	Haemoptysis	0.959
Haemoptysis	0.757	Nasal Blockage	0.814
Snoring	0.749	Hoarseness	0.784
Hoarseness	0.74	Snoring	0.752
Facial Asymmetry	0.542	Fatigue	0.577
Fatigue	0.362	Poor Appetite	0.561
Hyponasal Speech	0.335	Odynophagia	0.559
Proptosis	0.325	Facial Asymmetry	0.557
Tinnitus	0.317	Tinnitus	0.494
Poor Appetite	0.305	Hyponasal Speech	0.364
Odynophagia	0.303	Mouth Breathing	0.338
Mouth Breathing	0.302	Proptosis	0.332
Halithosis	0.269	Halithosis	0.312
Swelling	0.236	Dysphagia	0.237
Bleeding	0.225		
Dysphagia	0.206		

After the methods had been set up experimentally, we considered the performance metrics of the models using accuracy, precision, recall and F1score. The selected features by the Information Gain were trained on Decision Tree (C45) and Naïve Bayes

models. The result is given in Table 3. Figures 2 and 3 show the graphical representation of accuracies of the classifiers with both Information Gain and Chi Square methods.

Table 3: Result of the performance evaluation for the models using Information Gain method

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree	93.21%	0.943	0.932	0.931
Naïve Bayes	94.12%	0.945	0.941	0.941

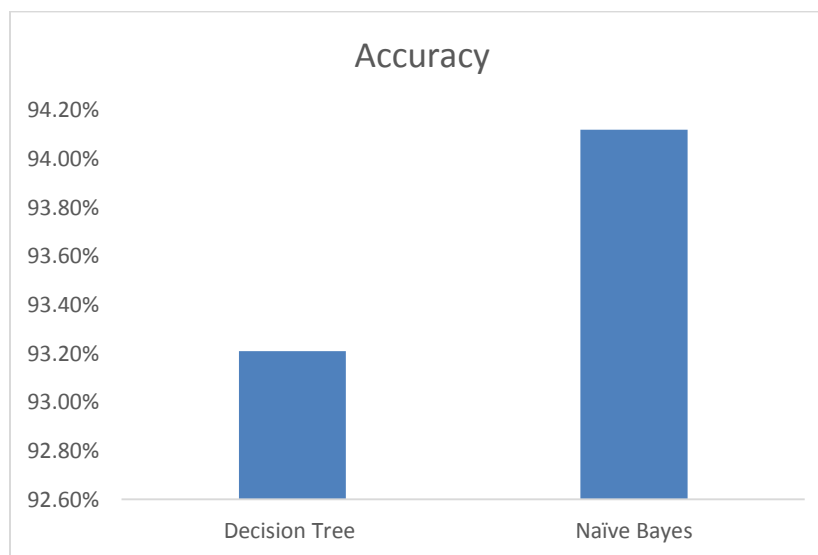


Figure 2: Accuracy of the learning algorithms (models) with information gain



Results of the Chi Square on Decision Tree (C45) and Naïve Bayes

Using Chi Square feature selection method, the features selected based on Chi Square were trained on Decision Tree and Naïve Bayes. Table 3 shows the performance metrics of the classifiers.

Table 3: Result of the performance evaluation for the models using Chi Square method

Models	Accuracy	Precision	Recall	F1 Score
Decision Tree	92.76%	0.932	0.928	0.927
Naïve Bayes	93.89%	0.942	0.939	0.939

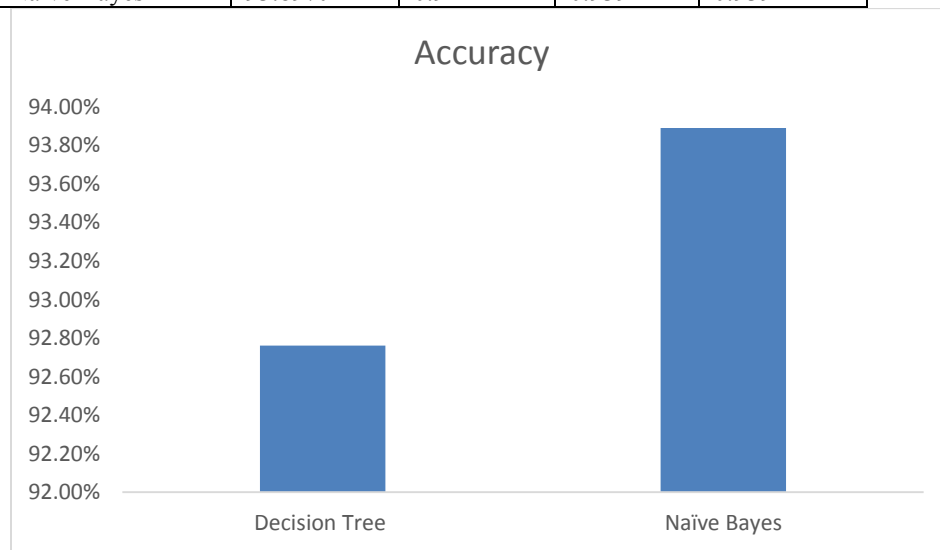


Figure 3: Accuracy of the learning algorithms (Models/Classifiers) with Chi Square

Discussion on the Performance Metrics of Head and Neck Dataset

The results of the models: Decision Tree (C4.5) and Naïve Bayes when feature selection methods such as Information Gain and Chi Square were performed, the inferences were stated below:

The HNC dataset was trained on Information Gain method with models namely Decision Tree, and Naïve Bayes. The result showed that the accuracy of Decision Tree (C4.5) was 93.21% and Naïve Bayes with accuracy of 94.12%. Naïve Bayes had higher Precision, Recall and F1 Score than Decision Tree. Based on the performance metrics, Naïve Bayes performed better than Decision Tree. This implied that Naïve Bayes was a better predictive model to predict or diagnose the type of HNC when Information Gain method was applied.

The HNC dataset was trained with Chi Square method with models namely Decision Tree and Naïve Bayes. The results showed that the accuracy of Decision Tree (C4.5) and Naïve Bayes were 92.76%

and 93.89% respectively. Naïve Bayes had higher Precision, Recall and F1 Score than Decision Tree. This means that Naïve Bayes had higher performance metrics than Decision Tree. The application of Chi Square method showed that Naïve Bayes model outperformed better than Decision Tree as depicted in Figures 2 and 3 respectively.

CONCLUSION

Cancer in head and neck regions is deadly. Late presentation and lack of access to the specialists cause a great challenge in early diagnosis in the healthcare system in developing countries. Two feature selection methods namely Information Gain and Chi Square and supervised learning algorithms namely Decision and Naïve Bayes were deployed for the study. The results showed that the model performance of Naïve Bayes outperformed Decision Trees when feature selection methods were applied. Naïve Bayes model could be employed to classify head and neck cancer types. This helps the non-specialist doctors and other primary health care workers who are not in the field of ENT to



diagnose and make referral early without delay to the appropriate specialists (ENT doctors).

ACKNOWLEDGEMENT

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

REFERENCES

1. Adisa, O. A., Agaku, T. I. and Akang, E. E. U (2012). Paediatric Head and Neck Cancers in Nigeria: Implications for Treatment Planning in Resource Limited Settings. 53 (4): 245-248
2. Akinkugbe, O. O., Lucas, A. O., Onyemelukwe, G. C., Yahaya, H. and Saka, M. J. (2010). Non-communicable diseases in Nigeria: The coming epidemics, Nigerian health review 2010, Health Reform Foundation of Nigeria (HERFON)
3. 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
4. Anurag, U. and Sudsanshu, K. (2010). Empirical Comparison by Data Mining Classification Algorithms (C 4.5 & C 5.0) for Thyroid Cancer Data Set. International Journal of Computer Science & Communication Networks. Volume 3(1): 64-68.
5. Dursun, D. (2009). Analysis of Cancer Data: A Data Mining Approach. Expert Systems: The Journal of Knowledge Engineering. 26(1). Blackwell Publishing Ltd
6. Fabio, M., Laura, M. D. and Cecilia D. (2017). A Predictive Model to Distinguish Malignant and Benign Thyroid Nodules Based on Age, Gender and Ultrasonographic Features. Brazilian Journal of Otorhinolaryngology.
7. Gourin, C. And Podolsky, R (2006). Racial Disparities in Patients with Head And Neck Squamous Cell Carcinoma. Laryngoscope. 116 (7): 1093–106.
8. Heikki, T., Petri, K., Kalevi, H. and Olli-Pekka, A. (2002). Diagnostic Delays In Laryngeal Carcinoma: Professional Diagnostic Delay is a Strong Independent Predictor of Survival. Journal of the Sciences and Specialties of the Head and Neck. 25(5): 389-394
9. Jajroudi, M., Baniyadi, T., Kamkar, L., Arbabi, F., Sanei, M., and Ahmadzade, M. (2014). Prediction of Survival in Thyroid Cancer Using Data Mining Technique. Technology in Cancer Research and Treatment.13(4)
10. Ian, H. W. and Eibe, F. (2005): Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition. San Fransisco: Morgan Kaufmann
11. Jason, B. (2014). An Introduction to Feature Selection. Machine Learning Process Jiawei H., Micheline, K. and Jian, P. (2011). Data Mining: Concepts and Techniques 3rd Edition
12. Jyoti, S., Ujma, A. and Dipesh, S. (2011): Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. International Journal of Computer Applications. 17(8) (0975 – 8887).
13. National Cancer Institute (2015). Definition of cancer. Available from <http://www.cancer.gov>dictionaries>
14. Onotai, L.O. and Nwogbo, A. C. (2012). Primary Head and Neck Malignant Tumours in Port Harcourt, Nigeria: A Revisit. Journal of Medicine and Medical Sciences. 3(2): 122-125
15. Renske, F., Marrten, A. Wilderman, B. S., Sofia, M. H. and Bing, T. I. (2010). Knowledge of General Practitioners about Nasopharyngeal Cancer at the Puskesmas In Yogyakarta, Indonesia. 10(1). BMC Medical Education
16. Titcomb, C. (2001). High Incidence of Nasopharyngeal Carcinoma in Asia. Journal in Surgery and Medicine. 33 (3): 235–238. https://en.wikipedia.org/wiki/Naive_Bayes_classifie