

# A Comparative Study of Predicting Customer Churn and Lifetime

O. Adesua  
Creative Research Networks  
Research Support Affiliates  
Accra Ghama  
[adesuaomolola@yahoo.com](mailto:adesuaomolola@yahoo.com)

P.A. Danquah  
Council for Sci & Ind. Research  
Accra, Ghana  
[danquahpaul@gmail.com](mailto:danquahpaul@gmail.com)

O.B. Longe  
American University of Nigeria  
Yola, Nigeria  
[olumide.longe@aun.edu.ng](mailto:olumide.longe@aun.edu.ng)

**Abstract:** *The problem to be investigated in this research is that of predicting customers who are at risk of leaving the company, a term called churn prediction in telecommunication. The aim of this research is to predict customer churn and further focus on creating customer lifetime profiles. These profiles will allow the company to fit their customer base into n categories and make a long estimation on when a customer is potentially going to terminate their service with the company. The research then proceeds to provide a comparative analysis of neural networks and survival analysis in their capabilities of predicting customer churn and lifetime.*

**Key words:** *GSM networks, Base station, Mobile station, Signal strength, GSM service provider*

## I. INTRODUCTION

Churn' is a word derived from change and turn [42]. It means the discontinuation of a contract. The biggest revenue leakages in the telecom industry are increasing customers churn behavior. Such customers create an undesired and unnecessary financial burden on the company. This financial burden results in to huge loss of the company and ultimately may lead to sickness of the company, detecting such customers well in advance is an objective of this research.

Major concern in customer relationship management in telecommunications companies is the ease with which customers can move to a competitor, a process called "churning". Churning is a costly process for the company, as it is much cheaper to

retain a customer than to acquire a new one [43]. It is stated that the costs of acquiring a new customer is five to ten times greater than that of retaining an existing one [1]. Also, the results of Dawkins and Reichheld's research shows that only five percent increase in retention rate yields in about 25 to 95 percent increase in the net present value of customers across a wide range of industries, such as credit cards, auto services, insurance brokerage, etc [37].

As a result of these and similar facts, enterprises are getting more and more interested in customer retention instead of acquiring new customers. Indeed, firms are concluding that the best core marketing strategy for the future is to retain existing customers and avoid customer churn ([35] and [36]).

This is exactly what churn management does. "Churn management consists of developing techniques that enable firms to keep their profitable customers and it aims at increasing customer loyalty." [34] That is, treating customers such that "they remain client of their original supplier even if a competitor proposes more advantageous conditions." [34]. Churn prediction models are developed by academics and practitioners to effectively manage and control customer churn in order to retain existing customers. As churn management is an important activity for companies to retain loyal customers, the ability to correctly predict customer churn is necessary.

Statistical and data mining techniques have been utilized to construct the churn prediction models. The data mining techniques can be used to discover interesting patterns or relationships in the data, and predict or classify the behavior by fitting a

model based on available data. In other words, it is an interdisciplinary area with a general objective of predicting outcomes and employing sophisticated data processing algorithms to discover mainly hidden patterns, associations, anomalies, and/or structure from extensive data stored in data warehouses or other information repositories [40]. There are several data mining techniques that are proposed to predict potential customers that are most likely to churn. Among the popular techniques to predict customer churn are: neural networks, support vector machines and logistic regression models [41].

The goal of most churn prediction model is to distinguish churners from non-churners as much as possible. Although in general this approach gives satisfying results, the time aspect as to when churning will take place often is neglected. Similarly, companies are increasingly viewing customers in terms of their lifetime value.

Customer lifetime profiling therefore helps to establish customer segmentation model which defines customer types in terms of lifetime value (LTV) or loyalty index. Not only do the telecommunications companies distinguish between which customers stay longer and which ones stay shorter, they also distinguish between which customers are highly profitable and which ones are low profitable or not profitable.

This research investigates a comparative analysis of neural network and survival analysis in predicting customer churn and lifetime.

## II. LITERATURE REVIEW

Building an effective customer churn prediction model using various techniques has become a decisive topic for business and academics in recent years [1]. Table 1 list relevant papers and the data mining model/technique applied in them.

Table 1: RELATED WORKS

Data Mining Technique	References
Neural Network	[4], [5], [6], [7], [12], [13], [14], [16], [17], [22], [23], [25], [26], [27], [28]
Decision Tree	[3], [4], [6], [9], [10], [11], [12], [13], [14], [16], [19], [25], [28]
Logistic Regression	[2], [6], [11], [13], [18], [19], [20], [23], [28], [29], [30], [31], [33]
Random Forests	[2], [16], [18], [20], [23], [31], [32]
Support Vector Machine	[2], [13], [14], [16], [18], [19], [20]
Survival Analysis	[22], [24], [32]
Bayesian Network	[4], [13], [15]
Self Organizing Maps	[3], [5]
AdaCost	[28]
Gradient Boosting Machine	[20]
Linear Discriminant Analysis	[14]
AdaBoost	[14]
Rough Set Theory	[8]
K-Nearest Neighbor	[6]
K Means	[12]
Time series	[33]
Taylor-Butina	[33]
ROCK (Robust Clustering using linKs)	[25]
Regression Forests	[31]
Linear Regression	[31]
Association Rules	[5]
Sequence Discovery	[21]

As it is seen in table 1, a list of all reviewed references that have exploited each method is identified. Classes are sorted based on two factors: number of papers in that field and the recency of publication. For each of the top-nine techniques, an example paper, which is indicated by bolded font in the table, is briefly explained. The rationale for selecting explained papers is based on recency and benchmarking results; that is, the latest published paper in each category, in which the related model has exhibited the best performance amongst other benchmarks, is summarized.

Reference [17] uses Genetic Algorithm to improve Neural Network by optimizing the network weights. This research is conducted in the context of a mobile telecommunications company and has developed two models with fitness functions based on cross entropy (log maximum likelihood) and model accuracy. For optimizing these two models, authors have exploited the crossvalidation method.

In [11], first input variables, which are selected according to receiver operating curve (ROC) of a single input predictor, are divided into several groups (e.g. demographic, bill information, etc.) based on the concepts they describe. Then each group is used as the input of a dependent Alternative Decision Trees (ADTrees) model, which is an improvement to decision trees. The outputs of ADTrees models constitute the inputs of a Logit model that will predict the churn. Recursive feature elimination (RFE) method is exploited for eliminating ineffective features in this stage. This model is benchmarked with TreeNet model (i.e. winner of Gold Prize) on a telecom company dataset and has recorded a very similar performance according to ROC.

Reference [6] assumes a complex relationship between the independent and target variables. It proposes using K-Nearest Neighbor (KNN) method as a solution for dealing with such challenge. It first exploits a single input KNN for changing the values of each independent feature. Then the new dataset is used for training and using Logit model. The proposed method has outperformed LR, C4.5, and RBF regarding a test on four distinct datasets and accuracy and ROC criteria.

By combining the cost-sensitive learning method of Weighted Random Forests with the over-sampling technique used in Balanced Random Forests, [16] has implemented Improved Balanced Random Forests. In the new approach, repeated sampling and higher penalties for wrong classifications of members of minority class helps in selecting the best subset of input features. The chief aim of this method is to deal with class imbalance in applications like churn management. The results of comparing the new method with Weighted Random Forests, Balanced Random Forests, Decision Tree, Neural Network and SVM illustrates its better accuracy, lift and lift curve.

In [19] a two step framework for using SVM in churn prediction is suggested. In the first step, the best C (i.e. a SVM model parameter) and subset of features are chosen by applying RFE and L2-SVM techniques on the primary C candidates and set of features. Also sampling is conducted in this step by Support Vector Sampling. Then the selected variable C, the subset of features, and the primary sample are used to train a nonlinear model called RBF-SVM in the next step. Also other features of RBF-SVM are determined in this stage. The final model is the best one among the two linear and nonlinear generated models. Authors has taken the help of auxiliary methods like 10-fold cross-validation, line and pattern search in order to improve their results which have outperformed that of Logit, C4.5, L1-SVM and CART regarding AUC measure.

Reference [22] has exploited the Neural Network method to predict the hazard and survival rate of customers. For doing so, first the hazard rate is calculated by Kaplan-Meier equation. Then the resulting values are set as the target value in training a Neural Network model. The new model is named BP ANN & Survival Analysis. The second contribution of this paper is establishing a customer segmentation model which defines six customer types based on Life Time Value (LTV) and Customer Survival Phase. The latter measure is computed regarding a proposed curve, which is derived from Product Life Cycle. The authors suggest appropriate retention strategies for each segment.

In [15] the simple Bayesian method is improved by applying Genetic Algorithm (GA) to it. In fact, GA is employed to optimize the structure of the network and the weights of its edges. Comparing this method with Bayesian Networks, TAN and APRI in the context of a credit company and regarding True Positives, True Negatives, True Positive Rate and True Negative Rate, the new method shows an outstanding performance.

Reference [3] constructs a two step model in which first the probability of attrition is calculated for a customer and then, if required, an appropriate retention strategy is proposed according to the cluster that includes the customer. The authors use C5.0 and Growing Hierarchical Self Organizing Map (GHSOM) respectively in the two sequential steps. They also amend the GHSOM method and also improve C5.0 results by exploiting cross-validation based on entropy and gain rate.

Reference [28] explains a research conducted in the context of financial service. It provides a new definition for churn, which is slumping LTV. It also defines a new loss function based on the profit and uses it to propose a new model evaluation criterion called area under profit curve (AUPROC). Another contribution of this research is improving the traditional AdaBoost (i.e. a model improvement method) technique to a newer one called AdaCost, which incorporates the new loss function in calculations. Taking the AUPROC, AUC, cumulative profit percentage and model accuracy into account for different models with time horizons of 3 and 6 months, AdaCost outperforms other models.

Reference [38] identifies the variables that affect churn in reverence of customer complaints data and provides a comparative analysis of neural networks, regression trees and regression in their capabilities of predicting customer churn. It showed significant accuracy for predicting customer churn using repairs and complaints data, proving that repairs and complaints influence customer's decisions to stay with their service providers.

Reference [39], proposed a neural network (NN) based approach to predict customer churn in subscription of cellular wireless services. The results of experiments indicate that neural network based approach can predict customer churn with accuracy more than 92%.

Further, it was observed that medium sized NNs perform best for the customer churn prediction when different neural network's top ologies were experimented.

### III. RESEARCH OBJECTIVES

The objectives of this reserach are to develop models to predict customer churn/survival using neural network and survival analysis. The second one is to demonstrate/test the 2 models. The final objective is compare the results from the 2 techniques.

### IV. RESEARCH METHODOLOGY

Records from a Telecommunications company will be used in this research. These pertinent records will be generated from its database like billing and customer service database. Required input fields or variables will be generated from these records which will be divided into the training and testing sets. This research will be done in two stages The first stage is to generate a prediction model for each of the chosen techniques (Survival analysis and Neural Network). Each of the models will constructed using the same training dataset. Testing dataset will then be used to evaluate the perfomance of these models. The goal of this stage is to determine which customers would churn or otherwise. The second stage will involve analysing and predicting customer lifetime using the same techniques and dataset. The goal at this stage is to determine the length of time that the customers will stay with the company. The performance of the techniques are then compared in terms of accuracy in their predictions.

## REFERENCES

- [1] KhakAbi, S., Gholamian, M.R. and Namvar, M. "Data mining applications in customer churn management," International conference on Intelligent Systems Modelling and Simulations, 2010, pp. 2592–2602.
- [2] K. Coussement and Dirk Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert Systems with Applications, vol. 34, 2008, pp. 313–327.
- [3] Bong-Horng Chu, Ming-Shian Tsai and Cheng-Seen Ho, "Toward a hybrid data mining model for customer retention," Knowledge-Based Systems, vol. 20, 2007, pp. 703–718.
- [4] X. Hu, "A Data Mining Approach for Retailing Bank Customer Attrition Analysis," Applied Intelligence, vol. 22, 2005, pp. 47–60, Springer.
- [5] H. S. Song, J. K. Kim, Y. B. Cho and S. H. Kim, "A Personalized Defection Detection and Prevention Procedure based on the Self-Organizing Map and Association Rule Mining: Applied to Online Game Site," Artificial Intelligence Review, vol. 21, 2004, pp. 161–184.
- [6] Y. M. Zhang, J. Y. Qi, H. Y. Shu, and J. T. Cao, "A Hybrid KNN-LR Classifier and its Application in Customer Churn Prediction," Proc. the IEEE International Conference on Systems, Man and Cybernetics, Oct. 2007, pp. 3265–3269.
- [7] G. Song, D. Yang, L. Wu, T. Wang, Sh. Tang, "A Mixed Process Neural Network and its Application to Churn Prediction in Mobile Communications," Proc. Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), 2006.
- [8] James J.H. Liou, "A novel decision rules approach for customer relationship management of the airline market," Expert Systems with Applications, vol. 36 (3), April 2009, pp. 4374–4381.
- [9] M. Zan, Z. Shan, L. Li, L. Ai-jun, "A Predictive Model of Churn in Telecommunications Based on Data Mining," Proc. IEEE International Conference on Control and Automation, IEEE Press, 2007.
- [10] Yi-Fan Wang, Ding-An Chiang, Mei-Hua Hsu, Cheng-Jung Lin, Ilong Lin, "A recommender system to avoid customer churn: A case study," Expert Systems with Applications, vol. 36, 2009, pp. 8071–8075.
- [11] J. Qi et al., "ADTreesLogit model for customer churn prediction," Annals of Operations Research, vol. 168, 2009, pp. 247–265, Springer.
- [12] Shin-Yuan Hung, David C. Yen and Hsiu-Yu Wang, "Applying data mining to telecom churn management," Expert Systems with Applications, vol. 31, 2006, pp. 515–524.
- [13] J. Zhao and Xing-Hua Dang, "Bank Customer Churn Prediction Based on Support Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example," Proc. 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008 (WiCOM'08), Oct. 2008, pp. 1–4.
- [14] Y. Xie and X. Li, "Churn Prediction with Linear Discriminant Boosting Algorithm," Proc. the Seventh International Conference on Machine Learning and Cybernetics, Kunming, July 2008.
- [15] Hongmei Shao, Gaofeng Zheng and Fengxian An, "Construction of Bayesian Classifiers with GA for Predicting Customer Retention," Proc. Fourth International Conference on Natural Computation, IEEE Computer Society Press, 2008.
- [16] Y. Xie, X. Li, E.W.T. Ngai and W. Ying, "Customer churn prediction using improved balanced random forests," Expert Systems with Applications, vol. 36, 2009, pp. 5445–5449.
- [17] P. C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services," Expert Systems with Applications, vol. 36, 2009, pp. 6714–6720.
- [18] K. Coussement, Dirk Van den Poel, "Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers," Expert Systems with Applications, vol. 36, 2009, pp. 6127–6134.
- [19] S. Lessmann and S. Voß, "A reference model for customer-centric data mining with support vector machines," European Journal of Operational Research, vol. 199 (2), Dec. 2009, pp. 520–530.
- [20] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," Expert Systems with Applications, vol. 36, 2009, pp. 4626–4636.
- [21] Ding-An Chiang, Yi-Fan Wang, Shao-Lun Lee and Cheng-Jung Lin, "Goal-oriented sequential pattern for network banking churn analysis," Expert Systems with Applications, vol. 25, 2003, pp. 293–302.



- [22] G. Zhang, "Customer Retention Based on BP ANN and Survival Analysis," Proc. International Conference on Wireless Communications, Networking and Mobile Computing, 2007 (WiCom), Sept. 2007, pp. 3406-3411.
- [23] W. Buckinx and D. Van den Poel, "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," European Journal of Operational Research, vol. 164, 2005, pp. 252-268.
- [24] B. Larivie`re, D. Van den Poel, "Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services," Expert Systems with Applications, vol. 27, 2004, pp. 277-285.
- [25] Lian Yan, Michael Fassino and Patrick Baldasare, "Predicting Customer Behavior via Calling Links," Proc. International Joint Conference on Neural Networks, Montreal, Canada, August 2005.
- [26] E Xu, S. Liangshan, G. Xuedong and Z. Baofeng, "An Algorithm for Predicting Customer Churn via BP Neural Network Based on Rough Set," Proc. the 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06).
- [27] "Predicting Customer Behavior in Telecommunications," L. Yan, R. H. Wolniewicz, R. Dodier, IEEE Intelligent Systems, IEEE Computer Society.
- [28] N. Glady, B. Baesens and C. Croux, "Modeling churn using customer lifetime value," European Journal of Operational Research, vol. 197, 2009, pp. 402-411.
- [29] Jae-Hyeon Ahna, Sang-Pil Hana and Yung-Seop Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," Telecommunications Policy, vol. 30, 2006, pp. 552-568.
- [30] K. Coussement and D. Van den Poel, "Integrating the voice of customers through call center emails into a decision support system for churn prediction," Information & Management, vol. 45, 2008, pp. 164-174.
- [31] B. Larivie`re and D. Van den Poel, "Predicting customer retention and profitability by using random forests and regression forests techniques," Expert Systems with Applications, vol. 29, 2005, pp. 472-484.
- [32] J. Burez and D. Van den Poel, "Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department," Expert Systems with Applications, vol. 35, 2008, pp. 497-14.
- [33] A. Prinzie T and D. Van den Poel, "Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM," Decision Support Systems, vol. 42, 2006, pp. 508-526.
- [34] M. Lejeune, "Measuring the Impact of Data Mining on Churn Management," Journal of Electronic Network Applications and Policy, vol. 11 (5), 2001, pp. 375-387.
- [35] M. Kim, M. Park and D. Jeong, The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services, *Telecommunications Policy* 28 (2004), pp. 145-159.
- [36] H.S. Kim and C.H. Yoon, "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market," Telecommunications Policy, vol. 28, 2004, pp. 751-765.
- [37] P. M. Dawkins and F. F. Reichheld, "Customer retention as a competitive weapon," Directors & Board, Summer 1990, pp. 42-7.
- [38] John Hadden, Ashutosh Twari, Rajkumar Roy and Dymtr Ruta, "Churn Prediction using Complaints Data" World Academy of Science, Engineering and Technology 19 2006.
- [39] Anuj Sharma and Kumar Pangrahi, "A Neural Network based Approach to Predicting Customer Churn in Cellular Network Services." International Journal of Computer Applications (0975 - 8887) Vol 27 - No 11, 2011.
- [40] Han, J., & Kamber, M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- [41] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecomm churn management. Expert Systems with Applications, 31(3), 515-524.
- [42] Lazarov, V. and Capota, M. "Churn Prediction".
- [43] Jadhav, R.J., and Pawar, U. T. Churn Prediction in Telecommunication using Data mining Technology, International Journal of Advanced Computer Science and Applications, Vol 2, No. 2, February 2011.

# Mobile 4G LTE Networks Mobility and Coverage for Some Locations in Ibadan Using Path Loss Analysis

O.E. Ogunsola  
Department of Physics  
University of Ibadan  
Ibadan, Nigeria  
E-mail:  
[seyiogunsola22@gmail.com](mailto:seyiogunsola22@gmail.com)  
Tel. No.: +2348050253416

O. Adekele  
Department of Computer Science  
University of Ibadan  
Ibadan, Nigeria

O.I. Olaluwoye  
Department of Physics  
University of Ibadan  
Ibadan, Nigeria

**Abstract:** The recent migration of most GSM service providers' networks in Ibadan from 3G to 4G, in preparation for the deployment of 5G technology in the nation necessitated the need to re-examine the GSM networks' mobility and coverage within the micro cells in-between a Base Station and a Mobile Station. This attempt is aimed at using existing Path Loss Propagation Models in proffering solutions to the negative consequences usually associated with call drops in the Urban and Suburban Areas of Ibadan due to inability of channels to handover as a result of path loss. The path loss (dB) analysis was carried out by measuring the Received Signal Strength RSS (dBm) at distances ranging from 0.05 km to 4 km in-between Base Stations and Mobile Stations using the factory fitted installed RSS software on Android phones. These measurements were taken for three weeks within the urban and suburban areas of the University of Ibadan campus, and its neighborhood community of Agbowo for ten selected existing Base Stations from four of the nationwide GSM Service Providers (SP1, SP2, SP3 and SP4) in Nigeria. The variation of path loss with the RSS for GSM Service Provider (SP1) propagating at 955MHz (reference distance of 0.05km), 1850MHz and 2120MHz, were 66.03 dB, 71.77 dB and 72.96 dB, respectively. However, at 4 km the path loss had risen to 101.59dB for 955MHz, 103.81dB for 1850MHz and 105dB for 2120MHz. Also, the path loss for the GSM service provider (SP2) propagating at 960MHz (reference distance of 0.05km), 1865MHz and 2150MHz were 66.07 dB, 71.84 dB and

73.08 dB, respectively. Moreover, in a similar manner to the SP1 service provider, at 4 km the path loss had risen to 104.14dB for 960MHz, 109.9dB for 1865MHz and 111.14dB for 2150MHz. Furthermore, the path loss for the GSM service provider (SP3) propagating at 950MHz (reference distance of 0.05km), 1835MHz and 2130MHz were 65.98 dB, 71.70 dB and 73.00 dB, respectively. Likewise, as was in the case of the SP1 and SP2 Service providers, the path loss at 4 km had risen to 104.05dB for 950MHz, 109.76dB for 1835MHz and 111.06dB for 2130MHz. Also, the path loss for the GSM service provider (SP4) propagating at 940MHz (reference distance of 0.05km), 1880MHz and 2140MHz, were 65.47 dB, 71.46 dB and 72.23 dB, respectively. Moreover, the path loss at 4 km had risen to 103.53dB for 940MHz, 109.52dB for 1880MHz and 110.29dB for 2140 MHz as was the case with the other GSM Service providers (SP1, SP2 and SP3) considered. Thus, the path loss increases with distance within the microcells of base stations. However, the path loss model with minimum path loss (dB) at a given distance enhances good coverage and handover postponement. Moreover, the mean square error values used in obtaining the accuracy between the measured and the Empirical models were 17.15dB, 59.69dB, 48.46dB, 60.52dB and 40.07dB for the C-model, Cost-OH, Sub-O, Lee-model and experimental model, respectively.

**Key words:** GSM networks, Base station, Mobile station, Signal strength, GSM service provider

## I. INTRODUCTION

Path loss or path attenuation is the reduction in the amplitude of electromagnetic waves travelling through space due to energy loss in the path between the transmitter and the receiver system. Signal path loss analysis plays a pivotal role in the design of radio communication system. The evaluation of path loss is very important from the Network Planning Level (NPL) to the link budgeting stage.

Thus, path loss measurement and monitoring provides helping tool in the determination of wireless communication (for voice and data) Quality of service (QoS) at NPL stage [1, 2]. The poor Received Signal Strength (RSS) at the receiver terminal is majorly caused by path loss [3, 4, 5]. However, mobile network mobility (traffic) can be varied with traffic parameters (bandwidth, delay, loss and availability).

In mobile network, the small area covered by a Base Station Transceiver (BST) is called a microcell. Moreover, there are several small overlapping areas within a BST. Hence, the closed distance cells require different frequencies in order for interference to be minimised, while for macro cell same frequency is required. However, it must be noted that all wireless communication operation requires the use of Key Performance Indicators (KPI) in accessing their network performance and QoS evaluation.

The coverage in a mobile communication is divided into a number of cells [6]. Hence, whenever a mobile unit is moving from one cell to an adjacent cell handover occurs. In essence, handover occurs by changing cells within the same BTS and the effect is more when movement is from one Mobile Station (MS) to another. However, a call is dropped when there is no channel to handover to within a MS. This happened when hand over initiation is not followed by handover execution.[1]. Also, path loss parameters such as transmitter power, transmitter location, antenna gain, antenna height, receiver sensitivity and the type of transmission analysis are helpful in radio communication system. In essence, path loss estimation is therefore a required parameter in determining both the Quality of Service (QoS) and signal strength at different locations before communication equipments are installed.

### A. Path Loss Propagation Models

Several propagation models already developed have been used over the years to predict signal coverage [7] which is very essential in the design of all 4G mobile networks. These propagation models are also important in predicting the path loss effect on radio system NPL which can be used as the controlling factor in system performance and coverage for better QoS at reception. Some of these path loss models include:

- (i) Free space Path Loss Model
- (ii) Clutter Factor Model
- (iii) Okumura's-Hata Model
- (iv) COST231- Hata
- (v) Lee Model
- (vi) Log-distance path loss Model

### B Free space Path Loss Model

This model takes into account a scenario where the propagated wave radiates freely to an infinite distance without being absorbed or reflected [8, 9]. This ideal case is not practicable in real life situations and is as presented in equation 1:

$$Pl(dB)_{FS} = 32.4 + 20\text{Log}(f_c) + 20\text{Log}D \quad (1)$$

### C. Clutter Factor Model

This proposed model consists of the plane earth loss plus an extra loss component called the clutter factor to the existing factors taken into consideration by various models [10]. This was done by assigning k and n for different frequencies and environments using the model [11] which considered large number of measurements taken from some American cities. The total loss defined by this model is obtained by using equation 2:

$$Pl(dB)_{CF} = 76.3 + 40\text{Log}D + 20\text{Log}f_c - 20\text{Log}h_b - 10\text{Log}h_m (h_m \leq 10m) \quad (2)$$

### D. COST231- Hata Model

The COST-HATA model was enhanced by the European COST 231 project (cooperation of scientific and Technical research) for the purpose of extending the limitations of the HATA models and to apply appropriate correction factors [12] as shown in equations 3 and 4:

$$Pl(dB)_U = 41.33 + 33.9\text{Log}f_c - 13.82\text{Log}h_b + x(\text{Log}D) - R + k_m \quad (3)$$

$$R = 3.2(\text{Log}(11.75h_m))^2 \text{ and } k_m = 3\text{dB}$$



$$Pl(dB)_{SU} = 41.33 + 33.9 \log f_c - 13.82 \log h_b + x(\log D) - R + k_m$$

$$R = 32(\log(11.75h_m))^2 \text{ and } k_m = 0 \text{ dB}$$

#### E. Okumura's-Hata Model

This is a simple mathematical model obtained as a function of frequency ( $f_c$ ) antenna heights, path loss exponent ( $n$ ) and distance in equations 5, 6 and 7, respectively by Okumura, from empirical data in Tokyo from his observations of what was designed [1].

$$Pl(dB)_U = 74.52 + 26.16 \log f_c - 13.82 \log h_b + x \log D - y \quad (5)$$

$$f_c \geq 300 \text{ MHz}$$

$$Pl(dB)_{SU} = 54.15 + 26.16 \log f_c - 13.82 \log h_b + x \log D - 2(\log \left( \frac{f_c}{28} \right))^2 \quad (6)$$

$$f_c \geq 300 \text{ MHz}$$

$$Pl(dB)_{op} = 28.61 + 26.16 \log f_c - 13.82 \log h_b + x \log D - z \quad (7)$$

where

$$y = 3.2 \left( \log(11.75h_m) \right)^2$$

$$x = 44.9 - 6.55 \log h_b$$

$$z = 4.78 (\log f_c)^2 + 18.38 \log f_c$$

$$D = \text{distance (km)}$$

$$f_c = \text{Frequency of transmission}$$

$$h_m = 1.5 \text{ m}$$

$$h_b = 36 \text{ m}$$

#### F. Lee Model

Lee model is a power law model, derived from measurements in different locations, using a procedure for estimating an effective base station antenna height for variations in terrain as shown in equation 8 [1].

$$Pl(dB)_L = 29 - 20 \log h_{b(\text{eff})} + n \log D - P_o$$

Where,  $n$  and  $P_o$  are given by measurements, and  $h_{b(\text{eff})}$  is the effective base station antenna height

#### G. Log-distance path loss Model (i.e. Experimental Model)

The log-distance path loss propagation models shows that the coverage received signal power

(4) This Model is the commonly used path loss empirical model and it divides the prediction area into open ( $Pl_{op}$ ), suburban ( $Pl_{SU}$ ) and urban ( $Pl_U$ ) [6]. Okumura's-Hata is very suitable for frequency operation of 150-1500 MHz, at a distance of 1 km to 20km. In addition, the base station antenna height ( $h_b$ ) of above 30m and mobile station height ( $h_m$ ) of 3m are also utilised as correction factors.

decreases logarithmically with distance for outdoor or indoor radio channels [10]. The mean path loss for an arbitrary Transmitter-Receiver (T-R) distance is expressed as a function of distance, using path loss exponent ( $n$ ) as shown in equations 9 and 10

$$Pl(dB) = Pl(d_0) + 10n \log(d/d_0) \quad (8)$$

where

$n = \text{the path loss exponent}$

$d_0 = \text{the closed - in reference distance}$

$d = \text{the Tx - Rx distance}$

$$n = \frac{\sum_{i=1}^k (Pl(d_i) - Pl(d_0))}{\sum_{i=1}^k \left[ 10 \log_{10} \left( \frac{d_i}{d_0} \right) \right]}$$

However, it must be noted that path loss parameters have been modelled by researchers, but report has shown that the existing models are not appropriate when deployed in areas outside the limit of which it was originally designed [8]. Moreover, in migrating from 3G to 4GLTE, field measurements are required in determining the mobility, handover and coverage of signals in mobile network and also in planning for the 5G mobility [13, 14].

Thus, this research is aimed at improving the quality of wireless services by carrying out site dependent measurements with a view to determining the suitable and acceptable Path loss model required.

However, measurements would be carried out for mobile communication network providers in Nigeria operating at GSM mobile frequencies of 900 MHz, 1800 MHz and 2100MHz. Furthermore, in this work, the comparative analysis of path loss for 4GLTE networks in the urban and rural environmental area shall be carried out within the University of Ibadan campus and its neighborhood Agbowo community, Ibadan. Moreover, the results obtained shall be compared with the already existing path loss propagating models such as Clutter Factor, Okumura-Hata, COST-231 and Lee, amongst others.

## II. METHODOLOGY

Data measurements were taken at various locations within the University of Ibadan campus and its neighborhood of Agbowo community. The base stations were classified into two: the Urban and Suburban areas. The parameter for determining the sub-urban area was based strictly on the low traffic density of the considered locations. Also, the buildings in this vicinity were mainly bungalows with some sparsely located storey buildings of about two to three floors. It was observed that these sub-urban locations are mainly on U.I. campus namely: Faculty of Science, CBN, Amina way, Carver way1, Carver way 2, Libra way including Alabanla, located at Agbowo. However, the choice of the urban locations was based on the high density of both human and vehicular traffic, including compactly built storey-buildings and offices with communication towers. These urban locations are mainly at Agbowo namely: Transformer, Major Salawu, Ogunyemi including Jaja clinic at UI.

The Received Strength Signal (RSS) measurements (dBm) were taken from Nigeria's GSM mobile spectrum [15] for four GSM service providers (SP1, SP2, SP3 and SP4) transmitting at: frequencies of 940MHz, 950MHz, 955MHz, 960MHz, 1835MHz, 1850MHz, 1865MHz, 1880MHz, 2120MHz, 2130MHz, 2140MHz and 2150MHz. These RSS measurements were taken using the signal strength software installed on Android phone for each of these frequencies within the microcell of Base station and Mobile station at 50m intervals each from 0.05km to 4km, for three consecutive weeks on a daily basis at 08Hr, 12Hr, 15Hr and 17Hr. These RSS (dBm) measurements were further utilised in determining the path loss.

The technical wireless parameters of an Israeli based communication company [9] were adopted and modified in order for the path loss values to be correctly analysed for the test locations.

## IV. RESULTS

The four GSM service providers' path loss and RSS were analysed for coverage strength and call handover within the microcells over the test distance. The path loss was determined using equation 1.

### *A. Variation of Path Loss (PL) and Received Signal Strength (RSS) of GSM provider (SP1)*

The variation of path loss with the RSS for GSM service provider (SP1) propagating at 955MHz, 1850MHz and 2120MHz for 4G signal with distance was presented in Table I. The path loss was 66.03 dB at the reference distance 0.05km for the 955 MHz, 71.77 dB for 1850MHz and 72.96 dB for 2120MHz,. Also, the corresponding RSS values were -38.03 dBm, -43.77dBm and -44.96dBm respectively. At 4 km the path loss had risen to 101.59dB for 955MHz, 103.81dB for 1850MHz and 105dB for 2120MHz with the RSS valves -70.07 dBm, -81.83 dBm and- 83.02 dBm respectively. It showed that the value of path loss increases with distance within the microcells of the base station. Call drop due to hand over problem can occur due to increase path loss despite that the RSS values look quite good.

TABLE I: PATH LOSS (dB) AND RSS (dBm) FOR VARIOUS FREQUENCY SPECTRUM  
FOR GSM SERVICE PROVIDER ONE (SP1)

SP1						
	955mHz		1850mHz		2120mHz	
D(km)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)
0.05	66.03	-38.03	71.77	-43.77	72.96	-44.96
0.1	72.05	-44.05	77.79	-49.79	78.98	-50.98
0.15	75.57	-47.57	81.32	-53.32	82.5	-54.5
0.2	78.07	-50.07	83.81	-55.81	85	-57
0.25	80.01	-52.01	85.75	-57.75	86.94	-58.94
0.3	81.59	-53.59	87.34	-59.34	88.52	-60.52
0.35	82.93	-54.93	88.67	-60.67	89.86	-61.86
0.4	84.09	-56.09	89.83	-61.83	91.02	-63.02
0.45	85.11	-57.11	90.86	-62.86	92.04	-64.04
0.5	86.03	-58.03	91.77	-63.77	92.96	-64.96
1	92.05	-64.05	97.79	-69.79	98.98	-70.98
2	98.07	-70.07	103.81	-75.81	105	-77
3	101.59	-73.59	107.34	-79.34	108.52	-80.52
4	104.09	-76.09	109.83	-81.83	111.02	-83.02

#### B. Variation of Path Loss (PL) and Received Signal Strength (RSS) of GSM provider (SP2)

The variation of path loss with the RSS for GSM service provider (SP2) propagating at 960MHz, 1865MHz and 2150MHz for 4G signal with distance up to 4 km was presented in Table II. The path loss was 66.07 dB at the reference distance 0.05km and for 960 MHz, 71.84 dB for 1865MHz and 73.08 dB for 2150MHz.

Also, the corresponding RSS values were -38.03 dBm, -38.07dBm and -43.84dBm respectively. At 4 km the path loss had risen to 104.14dB for 960MHz, 109.9dB for 1865MHz and 111.14dB for 2150 MHz with the RSS values -76.14 dBm, -81.90 dBm and -83.14 dBm respectively. The path loss increases with distance within the microcells of the base station also. Call drop due to hand over problem can occur due to increase path loss as high as 111.14dB though the RSS values look quite good.

TABLE II: PATH LOSS (dB) AND RSS (dBm) FOR VARIOUS FREQUENCY SPECTRUM  
FOR GSM SERVICE PROVIDER TWO (SP2)

SP2						
	960MHz		1865MHz		2150MHz	
D(km)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)
0.05	66.07	-38.07	71.84	-43.84	73.08	-45.08
0.1	72.1	-44.1	77.86	-49.86	79.1	-51.1
0.15	75.62	-47.62	81.39	-53.39	82.62	-54.62
0.2	78.12	-50.12	83.88	-55.88	85.12	-57.12
0.25	80.05	-52.05	85.82	-57.82	87.06	-59.06
0.3	81.64	-53.64	87.41	-59.41	88.64	-60.64
0.35	82.98	-54.98	88.74	-60.74	89.98	-61.98
0.4	84.14	-56.14	89.9	-61.9	91.14	-63.14
0.45	85.15	-57.15	90.93	-62.93	92.16	-64.16
0.5	86.07	-58.07	91.84	-63.84	93.08	-65.08
1	92.1	-64.1	97.86	-69.86	99.1	-71.1
2	98.12	-70.12	103.88	-75.88	105.12	-77.12
3	101.64	-73.64	107.41	-79.41	108.64	-80.64
4	104.14	-76.14	109.9	-81.9	111.14	-83.14

### C. Variation of Path Loss (PL) and Received Signal Strength (RSS) of GSM provider (SP3)

The variation of path loss with the RSS for GSM service provider (SP3) propagating at 950MHz, 1835MHz and 2130MHz for 4G signal with distance up to 4 km was presented in Table III. The path loss for the 950 MHz was 65.98 dB at the reference distance of 0.05km, 71.70 dB for 1835MHz and 73.00 dB for 2130MHz.

Also, the corresponding RSS values were -38.03 dBm, -37.98dBm and -43.70dBm respectively. At 4 km the path loss had risen to 104.05dB for 950MHz, 109.76dB for 1835MHz and 111.06dB for 2130 MHz with the RSS values -76.05 dBm, -81.76 dBm and -83.06 dBm respectively. The path loss increases with distance within the microcells of the base station also. Call drop due to hand over problem can occur due to increase in path loss as high as 111.06dB though the RSS values look quite good.

TABLE III: PATH LOSS (dB) AND RSS (dBm) FOR VARIOUS FREQUENCY SPECTRUM  
FOR GSM SERVICE PROVIDER THREE (SP3)

SP3						
	950mHz		1835mHz		2130mHz	
D(km)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)
0.05	65.98	-37.98	71.70	-43.70	73.00	-45.00
0.1	72.00	-44.00	77.72	-49.72	79.02	-51.02
0.15	75.53	-47.53	81.24	-53.24	82.54	-54.54
0.2	78.03	-50.03	83.74	-55.74	85.04	-57.04
0.25	79.96	-51.96	85.68	-57.68	86.98	-58.98
0.3	81.55	-53.55	87.27	-59.27	88.60	-60.60
0.35	82.89	-54.89	88.60	-60.60	89.90	-61.90
0.4	84.05	-56.05	89.76	-61.76	91.06	-63.06
0.45	85.07	-57.07	90.79	-62.79	92.08	-64.08
0.5	85.98	-57.98	91.70	-63.70	93.00	-65.00
1	92.00	-64.00	97.72	-69.72	99.02	-71.02
2	98.03	-70.03	103.74	-75.74	105.04	-77.04
3	101.55	-73.55	107.27	-79.27	108.56	-80.56
4	104.05	-76.05	109.76	-81.76	111.06	-83.06

#### D. Variation of Path Loss (PL) and Received Signal Strength (RSS) of GSM provider (SP4)

The variation of path loss with the RSS for GSM service provider (SP4) propagating at 940MHz, 1880MHz and 2140MHz for 4G signal with distance up to 4 km was presented in Table IV. The path loss for the 940 MHz at the reference distance of 0.05km was 65.47 dB, 71.46 dB for 1880MHz and 72.23 dB for 2140MHz. Also, the corresponding RSS values were -37.47 dBm, -43.46dBm and -44.23dBm respectively. At 4 km the path loss had risen to 103.53dB for 940MHz, 109.52dB for 1880MHz and 110.29dB for 2140 MHz with the RSS valves -75.53 dBm, -81.52 dBm and- 82.29 dBm respectively.

The path loss increases with distance within the microcells of the base station also. Call drop due to hand over problem can occur due to increase in path loss.



TABLE IV: PATH LOSS (dB) AND RSS (dBm) FOR VARIOUS FREQUENCY SPECTRUM FOR GSM SERVICE PROVIDER FOUR (SP4)

SP4						
	940mHz		1880mHz		2140mHz	
D(km)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)	PL(dB)	RSS(dBm)
0.05	65.47	-37.47	71.46	-43.46	72.23	-44.23
0.1	71.49	-43.49	77.48	-49.48	78.25	-50.25
0.15	75.01	-47.01	81.11	-53.11	81.77	-53.77
0.2	77.51	-49.51	83.5	-55.5	84.27	-56.27
0.25	79.45	-51.45	85.44	-57.44	86.21	-58.21
0.3	81.03	-53.03	87.03	-59.03	87.79	-59.79
0.35	82.37	-54.37	88.36	-60.36	89.13	-61.13
0.4	83.53	-55.53	89.52	-61.52	90.29	-62.29
0.45	84.55	-56.55	90.55	-62.55	91.31	-63.31
0.5	85.47	-57.47	91.46	-63.46	92.23	-64.23
1	91.49	-63.49	97.48	-69.48	98.25	-70.25
2	97.51	-69.51	103.5	-75.5	104.27	-76.27
3	101.03	-73.03	107.03	-79.03	107.79	-79.79
4	103.53	-75.53	109.52	-81.52	110.29	-82.29

#### E. RSS (dBm) Hourly Fluctuation of the GSM Service Providers

The fluctuations of the average signal strength of the various GSM Service Providers with distance for specific hours of the day (08hrs, 12hrs and 17hrs) for all the locations (Urban and suburban) considered is as presented in Figures 1 – 6.

The path loss effect on RSS(dBm) enhance the fluctuation as distance increases in both the urban and sub urban locations during the specified hours of the day . However, the value of RSS reduction was very significant within the 3-4km distance for all the service providers. Moreover, the effects of microcell inter hand over challenge within channels is also visible.

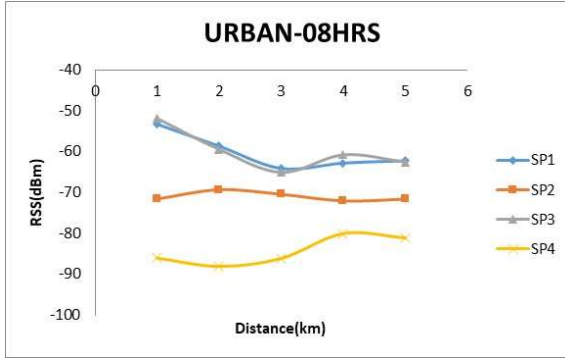


Figure 1: GSM Service providers' RSS(dBm) temporal fluctuations for urban 08hrs

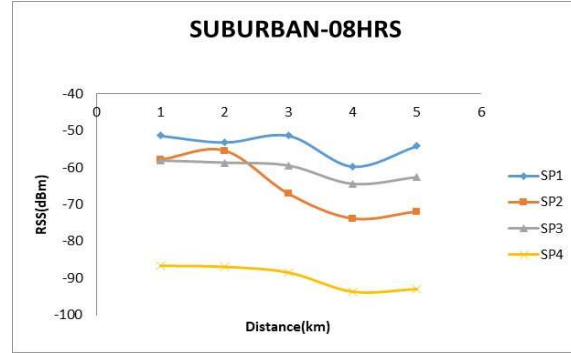


Figure 4: GSM Service providers' RSS(dBm) temporal fluctuations for suburban 08hrs

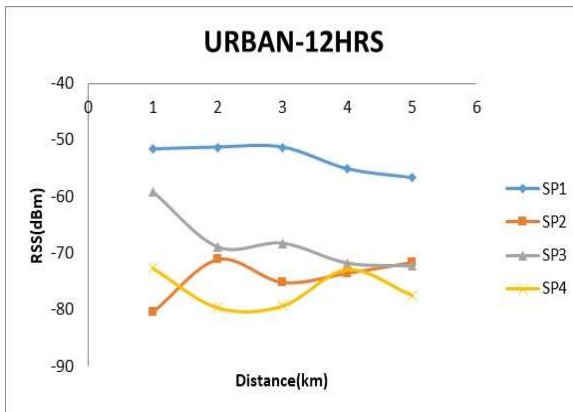


Figure 2: GSM Service providers' RSS(dBm) temporal fluctuations for urban 12hrs

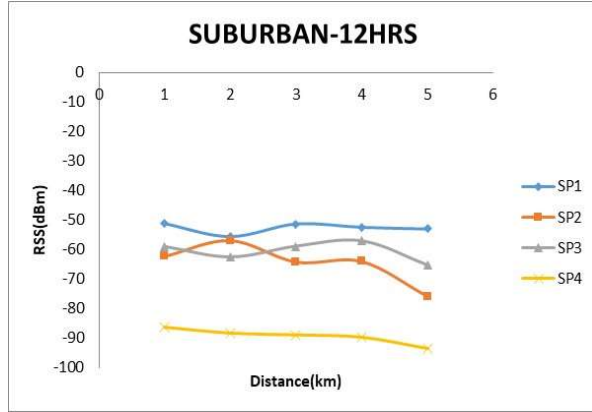


Figure 5: GSM Service providers' RSS(dBm) temporal fluctuations for suburban 12hrs

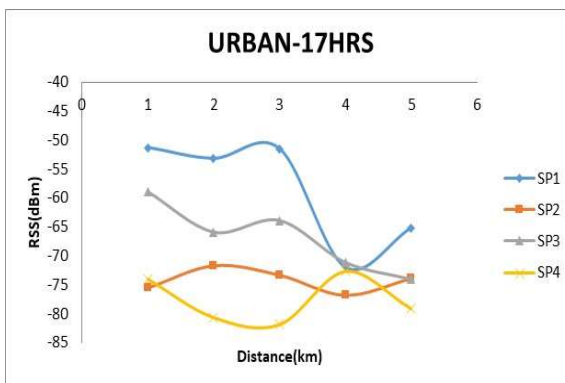


Figure 3: GSM Service providers' RSS(dBm) temporal fluctuations for urban 17hrs

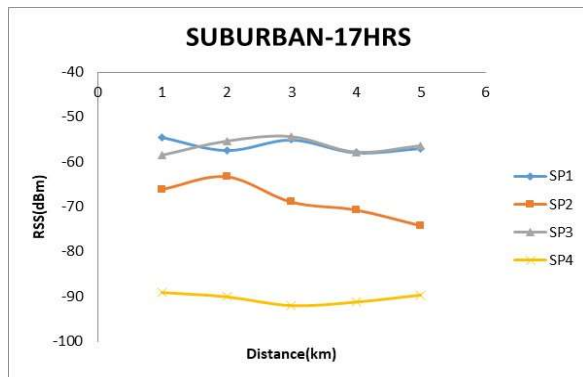


Figure 6: GSM Service providers' RSS(dBm) temporal fluctuations for suburban 17hrs

#### F. RSS(dBm) and GSM frequencies (900 MHz, 1800 MHz and 2100 MHz)

The average Received Signal Strength, RSS(dBm) information that can be useful in the GSM signal budget link estimation for all SP's and for

communication system design and planning in Ibadan, Nigeria is as presented in Table V

TABLE V: THE DAILY VARIATION IN AVERAGE RECEIVED SIGNAL STRENGTH ACROSS THE GSM SERVICE PROVIDER'S FREQUENCY SPECTRUM (SP) FOR THE TESTED URBAN AND SUBURBAN AREAS AT UI AND AGBOWO BETWEEN 8AM AND 5PM.

.Agbowo and UI	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)	RSS (dBm)
D(km)	940 (MHz)	950 (MHz)	955 (MHz)	960 (MHz)	1835 (MHz)	1850 (MHz)	1865 (MHz)	1880 (MHz)	2120 (MHz)	2130 (MHz)	2140 (MHz)	2150 (MHz)
0.05	-37.89	-37.98	-38.03	-38.07	-43.7	-43.77	-43.84	-43.91	-45.08	-45	-45.04	-44.96
0.1	-43.91	-44	-44.05	-44.1	-49.72	-49.79	-49.86	-49.93	-51.1	-51.02	-51.06	-50.98
0.15	-47.43	-47.53	-47.57	-47.62	-53.24	-53.32	-53.39	-53.45	-54.62	-54.54	-54.58	-54.5
0.2	-49.93	-50.03	-50.07	-50.12	-55.74	-55.81	-55.88	-55.95	-57.12	-57.04	-57.08	-57
0.25	-51.87	-51.96	-52.01	-52.05	-57.68	-57.75	-57.82	-57.89	-59.06	-58.98	-59.02	-58.94
0.3	-53.45	-53.55	-53.59	-53.64	-59.27	-59.34	-59.41	-59.48	-60.64	-60.6	-60.6	-60.52
0.35	-54.79	-54.89	-54.93	-54.98	-60.6	-60.67	-60.74	-60.81	-61.98	-61.9	-61.94	-61.86
0.4	-55.95	-56.05	-56.09	-56.14	-61.76	-61.83	-61.9	-61.97	-63.14	-63.06	-63.1	-63.02
0.45	-56.98	-57.07	-57.11	-57.15	-62.79	-62.86	-62.93	-63	-64.16	-64.08	-64.12	-64.04
0.5	-57.89	-57.98	-58.03	-58.07	-63.7	-63.77	-63.84	-63.91	-65.08	-65	-65.04	-64.96
1	-63.91	-64	-64.05	-64.1	-69.72	-69.79	-69.86	-69.93	-71.1	-71.02	-71.06	-70.98
2	-69.93	-70.03	-70.07	-70.12	-75.74	-75.81	-75.88	-75.95	-77.12	-77.04	-77.08	-77
3	-73.45	-73.55	-73.59	-73.64	-79.27	-79.34	-79.41	-79.48	-80.64	-80.56	-80.6	-80.52
4	-75.95	-76.05	-76.09	-76.14	-81.76	-81.83	-81.9	-81.97	-83.14	-83.06	-83.1	-83.02

#### G. Path Loss Exponent (n) for GSM providers for Experimental Model

The experimental Path loss model for the location was determined by using the location parameters. These values of n (Table VI) were obtained by using equations 9 and 10 and the test information for the locations. These test information includes distance from the base station, the whole frequencies of SP1, SP2, SP3, SP4 providers, transmission powers of 44dBm,  $h_m$  of 1.5 meters,  $h_b$  of 36meter, distance D (0.05 – 4.0)km and frequencies,  $f_c$ (Sp's).

TABLE VI: MAXIMUM PATH LOSS EXPONENT (n) FOR ALL GSM PROVIDERS FREQUENCIES

GSM Providers	Path loss Exponent (n)
SPI	2.117529
SP2	2.548445
SP3	2.000518
SP4	2.152145

#### H. Experimental and Empirical Path Loss Models for Urban and suburban Areas

The received signal strength (RSS) and Path loss (PL) from the base stations' micro and macro cells using various path loss models were calculated

against distance from Base Station to Mobile Station at three different frequencies 900, 1800 and 2100 MHz for all the considered GSM service providers (SP1, SP2, SP3 and SP4) and were presented in figures 7 to 10. The Free Space Model was obtained by using equation 1, Clutter Factor model was obtained from equation 2, COST231- Hata model was obtained using equations 3 and 4, and Lee model was obtained using equation 8. All these models were compared with the Log-distance path loss model (i.e. Experimental Model) obtained from equations 9 and 10 and the substitution of the corresponding Path loss exponent (n) obtained in Table 10.

### I. Comparison between Experimental and Empirical Path Loss Models for Urban and suburban Areas of SP1 and SP2

The location test parameters for GSM-SP1 ( $f_c = 955, 1850, 2120$  MHz,  $h_m = 1.5$  m and  $h_b = 36$  m) and for GSM-SP2 ( $f_c = 960, 1865, 2150$  MHz,  $h_m = 1.5$  m and  $h_b = 36$  m) were substituted into the appropriate equations (1-10) as applicable. The results obtained were as presented in Table VII, and Figures 7 and 8. All these models overestimated the Experimental model except Lee model which under estimated it.

TABLE VII: PATH LOSS MODELS FOR GSM-SP1 AND GSM-SP2 FREQUENCY BANDS IN IBADAN

Path loss Model	f(MHz)	GSM-SP1	Path loss Model	f(MHz)	GSM-SP2
C-Model	2120	$109.92 + 40\text{LogD}$	C-Model	2150	$110.06 + 40\text{LogD}$
Cost_H		$155.78 + 34.71\text{LogD}$	Cost_H		$155.99 + 34.71\text{LogD}$
Sub-OH		$144.16 + 34.71\text{logD}$	Sub-OH		$144.37 + 34.71\text{logD}$
Urban-OH		$156.63 + 34.71\text{LogD}$	Urban-OH		$156.83 + 34.71\text{LogD}$
Experimental		$89.57 + 21.2\text{logD}$	Experimental		$89.57 + 21.2\text{logD}$
Lee-Model		$58.81 + 38.4\text{logD}$	Lee-Model		$58.81 + 38.4\text{logD}$
Path loss Model	f(MHz)	SP1	Path loss Model	f(MHz)	SP2
C-Model	1850	$108.75 + 40\text{LogD}$	C-Model	1865	$108.89 + 40\text{LogD}$
Cost_H		$153.78 + 34.71\text{LogD}$	Cost_H		$153.92 + 34.71\text{LogD}$
Sub-OH		$142.59 + 34.71\text{logD}$	Sub-OH		$142.69 + 34.71\text{logD}$
Urban-OH		$154.62 + 34.71\text{LogD}$	Urban-OH		$154.74 + 34.71\text{LogD}$
Experimental		$89.57 + 21.2\text{logD}$	Experimental		$89.57 + 21.2\text{logD}$
Lee-Model		$58.81 + 38.4\text{logD}$	Lee-Model		$58.81 + 38.4\text{logD}$
Path loss Model	f(MHz)	SP1	Path loss Model	f(MHz)	SP2
C-Model	955	$103.01 + 40\text{LogD}$	C-Model	960	$103.05 + 40\text{LogD}$
Cost_H		$144.07 + 34.71\text{LogD}$	Cost_H		$144.15 + 34.71\text{LogD}$
Sub-OH		$134.78 + 34.71\text{logD}$	Sub-OH		$134.78 + 34.71\text{logD}$
Urban-OH		$144.88 + 34.71\text{LogD}$	Urban-OH		$144.88 + 34.71\text{LogD}$
Experimental		$89.57 + 21.2\text{logD}$	Experimental		$89.57 + 21.2\text{logD}$
Lee-Model		$58.81 + 38.4\text{logD}$	Lee-Model		$58.81 + 38.4\text{logD}$

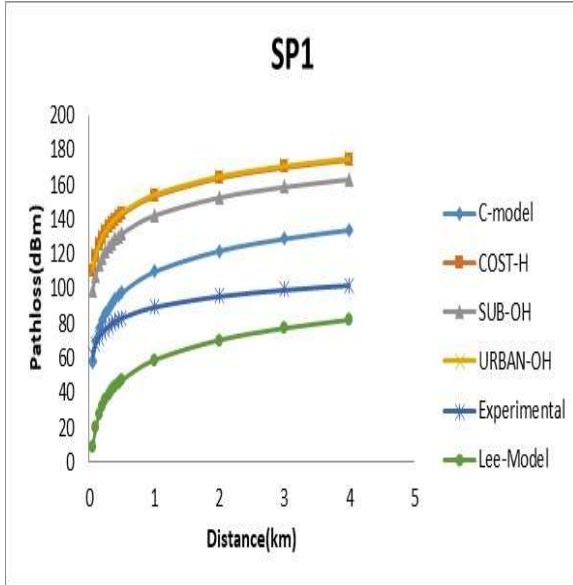


Figure 7: Path Loss Comparison for SP1 at a constant distance from Base station

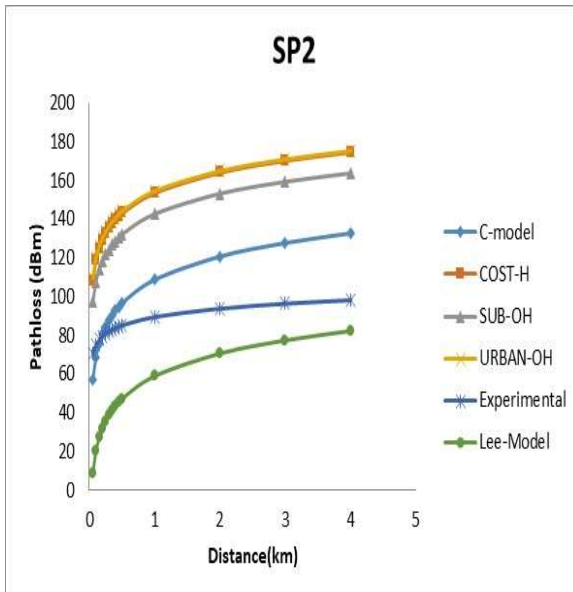


Figure 8: Path Loss Comparison for SP2 at a constant distance from Base station

#### J. Comparison between Experimental and Empirical Path Loss Models for Urban and suburban Areas of SP3 and SP4

The location test parameters for GSM-SP3 ( $f_c = 950, 1835, 2130$  MHz,  $h_m = 1.5m$  and  $h_b = 36m$ ) and for GSM-SP4 ( $f_c = 940, 1880, 2140$  MHz,  $h_m = 1.5m$  and  $h_b = 36m$ ) were substituted into the appropriate equations (1-10) as applicable.

The results obtained were as presented in Table VIII, and Figures 9 and 10. All these models overestimated the Experimental model except Lee model which under estimated it.



TABLE VIII: PATH LOSS MODELS FOR GSM-SP3 AND GSM-SP4 FREQUENCY BANDS IN IBADAN

Path loss Model	f(MHz)	GSM-SP3	Path loss Model	f(MHz)	GSM-SP4
C-Model	2130	109.98+ 40LogD	C-Model	2140	110.02+ 40LogD
Cost _H		155.85+34.71LogD	Cost _H		155.92+34.71LogD
Sub-OH		144.22+ 34.71logD	Sub-OH		144.27+ 34.71logD
Urban-OH		156.69+34.71LogD	Urban-OH		156.77+34.71LogD
Experimental		89.57+21.2logD	Experimental		89.57+21.2logD
Lee-Model		58.81+38.4logD	Lee-Model		58.81+38.4logD
Path loss Model	f(MHz)	SP3	Path loss Model	f(MHz)	SP4
C-Model	1835	108.60+ 40LogD	C-Model	1880	108.89+ 40LogD
Cost _H		153.66+34.71LogD	Cost _H		154.92+34.71LogD
Sub-OH		142.50+ 34.71logD	Sub-OH		142.7+ 34.71logD
Urban-OH		154.50+34.71LogD	Urban-OH		154.86+34.71LogD
Experimental		89.57+21.2logD	Experimental		89.57+21.2logD
Lee-Model		58.81+38.4logD	Lee-Model		58.81+38.4logD
Path loss Model	f(MHz)	SP3	Path loss Model	f(MHz)	SP4
C-Model	950	102.96+ 40LogD	C-Model	940	102.87+ 40LogD
Cost _H		143.99+34.71LogD	Cost _H		143.84+34.71LogD
Sub-OH		134.59+ 34.71logD	Sub-OH		134.59+ 34.71logD
Urban-OH		144.72+34.71LogD	Urban-OH		144.65+34.71LogD
Experimental		89.57+21.2logD	Experimental		89.57+21.2logD
Lee-Model		58.81+38.4logD	Lee-Model		58.81+38.4logD

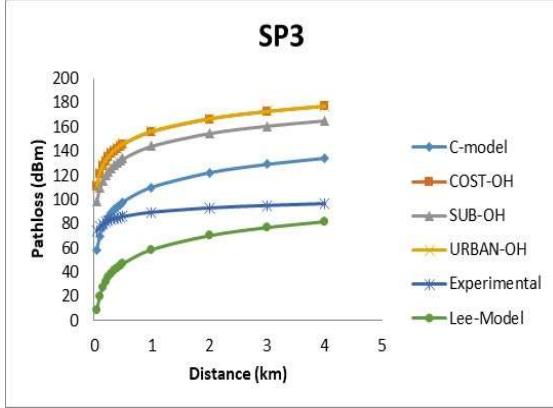


Figure 9: Path Loss Comparison for SP3 at a constant distance from Base station

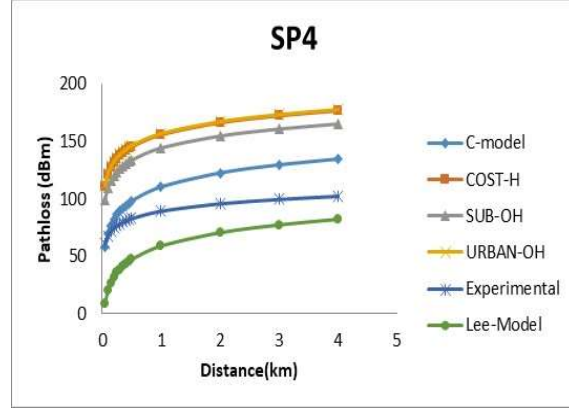


Figure 10: Path Loss Comparison for SP4 at a constant distance from Base station

Table IX: Mean square Error (MSE) for all the GMS service Providers considered at the Location in Ibadan.

	C-model	Cost-OH	Sub-O	Lee-model	Experimental model
SP1	16.85	59.76	48.65	60.58	38.09
SP2	16.76	57.72	46.80	58.54	41.49
SP3	18.07	59.24	47.91	60.07	42.77
SP4	16.91	62.03	50.46	62.87	37.95
Mean	17.15	59.68	48.45	60.52	40.07

$$MSE = \sqrt{\sum_{i=1}^K \left( \frac{(P_E - P_r)^2}{N} \right)} \quad (11)$$

Where,  $P_r$  is the Empirical model used predicted values and  $P_E$  is the measured values for each instance of distance,  $d$ .

The overestimation and underestimation was due to the differences in the physical development of Tokyo and Europe from where the models were originally developed, unlike Ibadan Nigeria where this present work is carried out. Thus, in order to improve on the accuracy of the utilised Empirical models, the mean square error (MSE) between the measured and the Empirical models (Table IX) were calculated using equation 11.

This MSE values were 17.15dB, 59.69dB, 48.46dB, 60.52dB and 40.07dB for the C-model, Cost-OH, Sub-O, Lee-model and experimental model, respectively.

#### IV. CONCLUSION

The path loss calculations for radio and wireless survey tools determined by signal strength at different locations were presented. This type of wireless survey tool is essential in determining radio signal strengths before installing any communication equipment. The parameters necessary to determine handover measurements are the received signal strength, the signal to noise ratio and the bit error rate. However, it is important to note that RSS measurements are affected by path loss which is dependent on distance. The RSS and Path loss from test locations were calculated at three frequencies: 900, 1800 and 2100 MHz for GSM Service providers SP1, SP2, SP3 and SP4 using different path loss models. However, the Model with minimum path loss value reduces the

number of handoffs and also enhances good coverage and handover postponement. Moreover, handover measurements provide information on mobility in GSM network, coverage and the quality of service. Also, the mean square error values used in obtaining the accuracy between the measured and the empirical models were 17.15dB, 59.69dB, 48.46dB, 60.52dB and 40.07dB for the C-model, Cost-OH, Sub-O, Lee-model and experimental model, respectively.

#### REFERENCES

- [1] Archana, Sharma, J.P., Sharma, D. & Sharma, P. K. (2014). Outdoor Propagation Path Loss Models: A Review. *International Journal of Scientific & Engineering Research*. 5(4), 1109.
- [2] Kumar, K.A.M. (2011). Significance of Empirical and Physical Propagation Models to Calculate the Excess Path Loss. *Journal of Engineering Research and Studies*. 3(5), 223-228.
- [3] Adeleke, O., Ojo, J.S., & Ajewole, M.O. (2014). Determination of the calibration factor for the rain rate derived from TRMM satellite using MRR data over Akure, Nigeria. *International Journal of Emerging Trends in Science and Technology*. 1(7), 1151-1165.
- [4] Prajesh, P. & Singh, R. K. (2011). A survey on various propagation model for wireless communication. 5<sup>TH</sup> IEEE International Conference on Advanced Computing and Communication Technology, 62.
- [5] Khan, J. (2010). Handover management in GSM cellular system. *International Journal of Computer Application*. 8(12), 14-24.
- [6] Azubogu, A.C.O., Okoli, G.N., Idigo, V.E. & Ohaneme, C.O. (2010). Empirical-Statistical Propagation Path loss Model for Suburban environment of Nigeria at 800MHz band. *The IUP Journal of Science and Technology*, 7(2), 45-54.
- [7] Purnima, K.S. & Singh, R.K. (2012). Cell Coverage Area and Link Budget Calculation in GSM System. *International Journal of Modern Engineering Research*. 2 (2), 170-176
- [8] Freeman R. L. (2007). *Radio System Design for Telecommunications*, New Jersey: John Wiley. <http://www.breezecom.com>. Retrieved October 22nd, 2020.

- [10] Ifeagwu E.N. & Edeko, A. . (2015). Characterization of Mobile Radio Propagation Channel using Empirically based Path loss Model for Suburban Environments in Nigeria. *International Journal on Recent and Innovation Trends in Computing and Communication*. 3(5), 3300 – 3304.
- [11] Egli, J.J. (1957). Radio Propagation above 40MC over Irregular Terrain. *Proceedings of the IRE*. 45(10), 1383-1391.
- [12] Singh, Y. (2012). Comparison of Okumura, Hata and Cost-231 Models on the Basis of Path Loss and Signal Strength. *International Journals of Computer Applications*. 59(11), 37-41..
- [13] Erceg, V., Greenstein, L.J., Ijandra, S.Y., Pakoff, S.R., Gupta, A., Kulic, B., Julius, A.A. & Bianchi, r. (1999). An empirically path loss model for wire-less channels in suburban environment. *IEEE journals on selected areas of communications*. 17, 1205-1211.
- [14] Sharma, D. & Singh, R. K. (2012). Analysis of Handover Initiation Using Path Loss to Sustain QoS. *International Journal of Scientific and Engineering Research*. 3(5),1-6.
- [15] <http://www.ncc.ng.or>. Retrieved October 22nd, 2020.

# Development of a Risk-Factor System for Predicting Occurrence of Breast Cancer Using Fuzzy Logic

A. O. Fabiyi  
Federal Polytechnic Ede  
Ede, Nigeria  
fabiyiademolaebony@ya  
hoo.com, 08025600073

G.O. Ogunleye  
Federal University Oye-  
Ekiti, Nigeria  
07066358913

N.C. Onyeka  
Federal Polytechnic Ede  
Ede, Nigeria  
camiliaobi@live.com  
07033310638

B.M.G. Amosa  
Allbytes Consults  
Gbagada Lagos, Nigeria  
\*amosabmg@gmail.com  
08034719314

**Abstract:** *The high mortality rate associated with cancer and the inability to detect the disease early has led to a catastrophic reduction in the rate of survivability of the disease in women. We have attempted to improve on the rate of survival by using Fuzzy logic to develop a risk factor system for detection of breast cancer. The system has been implemented using the Mamdani fuzzy logic approach in MATLAB. A graphical user interface (GUI) has been developed using Microsoft Visual Studio 2012. The GUI was powered by a Fuzzy Logic Library and Visual C# was used as its programming language.*

**Keywords:** *Breast Cancer, Disease, Risk factor, Fuzzy Logic.*

## 1. INTRODUCTION

Mortality due to breast cancer can be reduced by the early diagnosis of disease, as well as by early treatment initiation. Many symptomatic breast cancer patients experience long delays in obtaining diagnosis and treatment which can negatively affect their prognosis. Patients with breast cancer suffer distress due to the disease and its treatment. Furthermore, advances in medicine have prolonged life expectancy in these patients. Breast cancer is the second most common cancer overall, and by far the most common cancer in women. In 2012, 1.67 million new cases of breast cancer (25% of all incident cancer cases) were estimated worldwide. It is the most common cancer in women in both more and less developed regions,

with slightly more cases estimated to have occurred in LMICs (883 000 cases) than in more economically developed regions (794000 cases) in 2012 [1]. Breast Cancer is the most frequent cancer among women, impacting 2.1 million women each year, and also causes the greatest number of cancer-related deaths among women. In 2018 it is estimated that 627,000 women died from breast cancer approximately 15% of all cancer deaths among women [2]. Cancer is a potentially fatal disease which is majorly caused by environmental factors that mutate genes encoding critical cell-regulatory proteins [3].

Agents that cause cancer also known as carcinogens can be seen in the air, food and water, chemicals and also sunlight that people are exposed to. According to [4], 650,000 people out of the estimated 965 million are diagnosed with cancer annually and the lifetime risk of an African woman dying from cancer is two times higher than the risk of a woman in a developed country. Cancer is known to comprise of a set of more than 200 different diseases. It can be in a widely accepted sense described as an uncontrolled growth and spread of abnormal cells in the body. A cell is the basic unit of life. Cells in living thing split to produce more cells when the living thing harboring them requires it. At some point, cells keep on splitting or dividing without control thus creating excess cells which are not needed. This in turn causes the formation of mass tissue also known as a tumor. Tumors can be found in tissues of different types. There are currently many reasons that make cancer prevail.



These reasons can vary from increased ability to take care of diseases which will cause a delay in their progression to inability to diagnose and treat a disease which leads to death and disability [5]. Predicting the future occurrence of cancer is a very much wanted technique which will be a very great help in providing early detection of cancer and in like manner increase the rate of survivability.

The rise in the incidence of breast cancer in recent years should have prompted women with an awareness and knowledge to seek medical advice with minimal breast symptoms. However, only a small number of women are still aware of this internationally. Due to the conservative nature of many societies, many women refrain from seeking medical advice out of shyness or stigma until their disease becomes advanced. Research conducted has shown that there is high mortality rate and there has been a great lapse in the early detection of cancer hence, a catastrophic reduction in the rate of survivability.

This is a major problem since cancer is a kind of disease that is painless at its early stages and might not be noticed until it has become a big problem to contend or deal with and sometimes ends with the death of the individual hosting it. This research is aimed developing a risk factor system for detection of breast cancer using fuzzy logic. This will be used to enlighten people if they are at the risk of getting cancer hence, encouraging early detection which will in turn increase the rate of survivability. There are different types of cancers in which this study can be applied to but the study would be limited to only breast cancer.

## II. RELATED LITERATURES

In 2009, breast cancer became the most frequently diagnosed form of neoplastic disease in women in India and is now the most common cause of cancer death in the country, accounting for more than a fifth of all female cancer mortality. The main risk factors for breast cancer include individual factors like race and ethnicity, overweight and obesity, physical inactivity, alcohol use, and smoking. Breast cancer risk can also significantly increase for women with first-degree relatives with breast cancer and women with increased breast density.

Reproductive risk factors include early age at menarche, nulliparity, late age at first birth, lack of breast feeding, oral contraceptive use, menopausal status, and menopausal hormone therapy. In many LMICs, changes in reproductive factors, lifestyle and increased life expectancy have led to a sharp rise in the incidence of breast cancer. Presently, national and international health authorities like the U.S. Preventive Services Task Force (USPSTF), the Canadian Task Force on Preventive Health Care (CTFPHC) or the World Health Organisation share the opinion that not enough scientific evidence exists to show that breast self-examination (BSE) can save lives or made women to detect breast cancer at an earlier stage and therefore, should not be taught on population-wide level. They also recommend that BSE should not be promoted since there is evidence suggesting that such public health interventions may actually cause harm. However, in countries where mean tumour sizes are above [1].

Computing, in its regular ideology, is situated on manipulation of numbers and symbols [6]. To explain better, computing with linguistic variables is a methodology in which the object of computation are words and propositions drawn from a natural language, e.g., significant increase in water consumption, small, large, far from recommendations, etc.

Soft Computing is a branch or study under computing that involves the building of machines that are termed wise and intelligent [7]. Intelligence gives the ability or provides the power to derive an answer and not simply arrive at the answer. Intelligence can also be more generally and widely defined as the ability to perceive, identify and retain knowledge or information and apply it to itself or other instances of knowledge or information providing referable understanding models of any size, complexity, or density, due to any conscious or subconscious imposed will or instruction to do so ([www.wikipedia.com](http://www.wikipedia.com)).

Soft computing gives a highly noticeable paradigm shift in the field of computing, which mirrors the fact that human minds, unlike computers in the present day, embody an outstanding ability to store and process information which is pervasively imprecise, uncertain and lacking [8].

Soft computing has been regarded as a fast coming collection of methodologies, which shoots at exploiting tolerance for impression, partial truth to achieve robustness, total low cost and uncertainty. It mimics consciousness and cognition in several important respects [7]. Soft computing aims at developing a computer or machine which will work in the same way or manner as humans do. The knowledge and wisdom of humans can be re-enacted in computers and systems in some artificial manner.

The combination of the principal constituents of soft computing which are FL, NC, EC, ML and PR has been found to be very effective in solving problems. An example of this can be found in Neuro-fuzzy Systems. Neuro-fuzzy systems are on a daily basis becoming known as consumer products such as washing machines, air conditioners etc. and are also used in industrial applications. The incorporation of soft computing in consumer products and industrial systems leads to the development of systems which have high Machine Intelligence Quotient (MIQ) [8]. The high MIQ of systems which have soft computing incorporated in them is responsible for the fast growth in the number variety of applications of soft computing. Therefore, these successful applications of soft computing implies that the impact of soft computing will be highly felt the next couple of years.

#### A. Expert Systems (Es)

An expert system is a computer program that represents and reasons with knowledge of some specialists' subject with a view to solving problems or giving advice [9]. Expert Systems are computer programs that are derived from a branch of computer science research called Artificial Intelligence (AI) in [10]. The scientific goal of an AI is to understand intelligence which in turn leads to the building or development of computers that exhibit intelligent behaviour.

AI is deals with the concepts and methods of symbolic inference, or reasoning, by a computer, and how the knowledge used to make those inferences will be represented inside the machine. Expert systems are also known as knowledge based systems. The usage of ES has been largely successful because ES limits the field of interest to a narrowly defined specific domain area that can be naturally described by explicit verbal rules.

An ES's knowledge base is embedded with the human expert's knowledge (e.g. skilled doctor or lawyer). This human expert knowledge is what is used by the ES in its computerized consulting service. An ES gives advice to its users by combining facts in its Working Memory (WM) (WM contains facts about a problem that are discovered during a consultation) with domain knowledge in its knowledge base to draw conclusions about a problem.

#### B. Fuzzy Logic

Fuzzy logic is generally an expansion of deterministic logic, i.e. the truth value is ranged from 0 to 1 other than a binary value. The supposition of fuzzy logic is mainly intended at turning a black and white problem into a grey problem [11]. In the context of set theory, deterministic logic is matching to crisp sets, this means that each element in a set has a full membership to the set, i.e. the element totally feel right to the set. In contrast, fuzzy logic is equivalent to fuzzy sets, which means that each element in a set only has a half-done membership to the set, i.e. the element belongs to the set to a certain degree. The degree of a fuzzy membership is determined by a particular membership function such as trapezoidal membership function, triangular membership function and Gaussian membership function [12].

#### C. Fuzzy Logic as a Prediction Tool

Prediction according to the Encarta dictionary is defined as the statement about the future i.e. a statement of what somebody thinks will happen in the future. Predicting the occurrence of an event requires imprecise and uncertain knowledge. Predicting a system is normally done by getting knowledge from past events for which historical data is gotten and analyzed to get to know the resulting sequence or pattern in the market [13].

Knowledge from the past is applied mainly to acquire the patterns that formally existed. Getting knowledge about the past can provide to some extent knowledge about the future. Fuzzy logic has emerged overtime as a valuable technique in predicting the occurrence of future events [14]. It has been used and still being used as a prediction tool for many systems since its inception in 1965. Fuzzy logic is almost perfect for the prediction of an event because prediction requires knowledge that is rather imperfect and uncertain.

The accuracy of a fuzzy logic prediction system is highly dependent on the variable which includes the system, expert rules and the membership functions [15]. A fuzzy model operates on the level of linguistic terms and it also represents and processes uncertainty. Various application of fuzzy logic for prediction is presented in [16], [17], [18], [19], [20]. Literatures in Artificial Intelligence in Breast Cancer Research are discussed in [21], [22], [23]. Benefits of Fuzzy logic as a prediction system are presented in [24].

### III. DEVELOPMENT OF A RISK-FACTOR SYSTEM FOR PREDICTING OCCURENCE OF BREAST CANCER USING FUZZY LOGIC

Breast cancer is often known as the most commonly diagnosed cancer in women today in the world. Most women at various stages undergo different tests to determine if they will develop breast cancer. These tests are commonly used to determine if these women have the BRCA1 and BRCA2 cancer genes. BRCA1 and BRCA2 are human genes that produce tumor suppressor proteins. These proteins help repair damaged DNA and therefore, play a vital role in ensuring the stability of the cell's genetic material. When either of these genes is mutated, such protein product is either not made or does not function correctly.

However, since mutations in these genes account for just a small fraction of all breast cancers, a negative test result doesn't mean the women who go for these tests will be in the clear. Likewise, testing positive doesn't mean the woman might have breast cancer in the future. Early detection of breast cancer improves the chances of a curative treatment and a lot of progress has been made in this field during the last decade. Screening and diagnosis techniques have enhanced early detection of breast cancer over the years. Some of these screening and diagnostics techniques include:

- i. Breast self-examination: The American Cancer Society (ACS) stated that every woman should start monthly breast self-examination at the age of 20. An analysis carried out on 12 studies involving 8,118 patients with breast cancer correlated the performance of breast self-examination with tumor size and regional lymph node status. Women who performed breast self-examination were more likely to have smaller tumors and less likely to have axillary node

metastases than those who did not. This technique has been flawed due to the fact that it is rarely performed well.

- ii. Clinical breast examination: Women are recommended to begin clinical breast examination at age 20. This was stated by the American Cancer Society. They are also recommended to have an examination every 3 years between ages 20 and 39.
- iii. Mammography: The American Cancer Society, the American College of Radiology, and the American Medical Association each have since updated their guidelines since 1997 and recommend annual mammography beginning at age 40. The National Cancer Institute (NCI) also updated their guidelines in 1997. They recommended that women should undergo screening mammography every 1-2 years beginning in their 40s.

#### A. Fuzzy Logic Model

The proposed fuzzy logic model will use the Mamdani fuzzy logic approach. The Mamdani approach was chosen because it is widely accepted for capturing expert knowledge, flexible, uses defuzzification technique, and has output membership functions.

The Mamdani model in this research was designed based on various potential breast cancer risk factors. Some potential breast cancer risk factors include:

- i. Age: The risk of developing breast cancer increases with age. It has been found that rates are generally low in women younger than 40. This rate increases after age 40 and are highest in those 70 and above.
- ii. Sex: Majority of breast cancer cases are normally associated with females. Breast cancer also occurs in males but it is very rare.
- iii. Family History: Women who have relations like a mother or sister with breast cancer have a greater risk of getting breast cancer. The risk is multiplied if the woman has multiple relatives who have the disease.
- iv. Alcohol: Research consistently shows that drinking alcoholic beverages -- beer, wine, and liquor -- increases a woman's risk of hormone-receptor-positive breast cancer. Alcohol can increase levels of estrogen and

- other hormones associated with hormone-receptor-positive breast cancer. Alcohol also may increase breast cancer risk by damaging DNA in cells.
- v. Menarche Age: Early menstruation in a woman can increase the risk of breast cancer. This is possible because an early period exposes a woman's body to greater amounts of estrogen (reproductive hormone in a woman) over her lifetime. Increased levels of estrogen in a woman over a long period of time can increase the risk that cells in the breast will become cancerous.
- vi. Age at Menopause: Women who have menopause at a later age have higher risk of being diagnosed with breast cancer. This is possible because an early period exposes a woman's body to greater amounts of estrogen (reproductive hormone in a woman) over her lifetime. Increased levels of estrogen in a woman over a long period of time can increase the risk that cells in the breast will become cancerous.
- vii. Number of Births: Women who have less than 2 children have a higher risk of breast cancer. The reason is that pregnancy changes hormone levels and breast tissue in a way helps to protect breast cells from becoming cancerous. The more pregnancy a woman has, the more her breast tissue changes and the more protection she gets.
- viii. Birth Weight: A woman who weighs more when she gives birth has a higher risk of breast cancer before menopause. Researchers are not sure yet why this occurs but studies are being conducted on different influences like pregnancy hormones and other prenatal factors to learn more about breast cancer development.
- ix. Tumor Size: Tumors occur as a result of formation of mass tissue which is caused by the continuous splitting and dividing of cells without control thus creating excess cells which are not needed. Tumor size measured in centimeters (cm) ranges from 0.1 – 5.0cm. Women with tumor size < 1.0cm have lower risk than women with tumor size  $\geq 5.0$ cm

After reviewing the above potential risk factors, a few factors were chosen to be incorporated in the model. These factors are: age, menarche age, family history, alcohol intake, and tumor size. In accordance to these chosen factors, membership functions were created using the Triangular Membership Function. The architecture of the breast cancer system which involves the input variables, membership functions, rules and the defuzzification process which produces the crisp value is presented in Figure 1

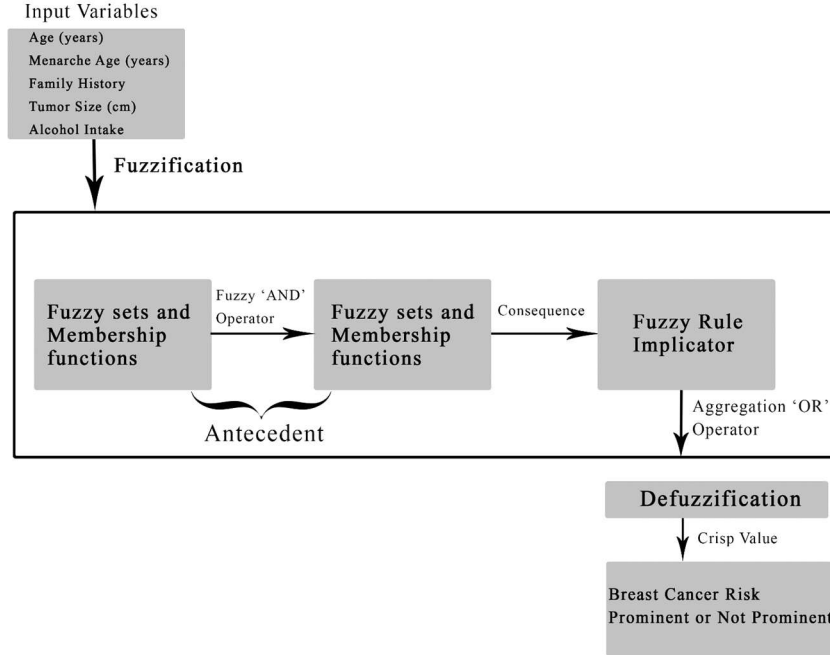


Figure 1: Architecture of the Breast Cancer Fuzzy Inference System

### B Linguistic Variables

Tables showing the various linguistics variables for the design are in Tables 1 – 4.

TABLE 1: LINGUISTIC LABELS FOR FUZZY VARIABLES

S/NO	PARAMETERS	LINGUISTIC LABELS
1	Age	Range one(R1), range two(R2), range three(R3), range four(R4)
2	Menarche Age	Early(EL), normal(NM), late(LT)
3	Family History	YES, NO
4	Tumor Size	Small(SM), medium(MD), large(LG)
5	Alcohol Intake	Less(LS), more(MR)
6	Breast Cancer Risk	Low risk(LR), risky(RS), high risk(HR)

TABLE 2: NUMERICAL VARIATION INTERVAL FOR INPUT VARIABLES

INPUT VARIABLES	MIN VALUE	MAX VALUE
Age	40	100
Menarche Age	10	15
Family History	0.1	1.0
Tumor size	0.1	5.0
Alcohol intake	0.1	1.0

TABLE 3: NUMERICAL VARIATION INTERVAL FOR OUTPUT VARIABLES

OUTPUT VARIABLES	MIN VALUE	MAX VALUE
Breast cancer risk	0%	100%

TABLE 4: LINGUISTIC LABELS AND THEIR RANGES

S/NO	PARAMETERS	RANGE
1	Age	Range one: 40 – 49 Range two: 50 - 59 Range three: 60 - 100
2	Menarche age	Early: 10 – 11 Normal: 12 - 13 Late: 14 - 15
3	Family History	Yes: 0 - 0.5 No: 0.6 - 1
4	Tumor size	Small: 0.1 - 1.0 Medium: 1.1 - 2.0 Large: 2.1 - 5.0
5	Alcohol intake	Less: 0.0 - 0.5 More: 0.6 - 1
6	Breast cancer risk	Low risk: 1 – 33.3% Risky: 33.4 – 66.6% High risk: 66.7 – 100%

#### C. Rule Base

The defined rules for the risk-factor system for predicting the occurrence of breast cancer using fuzzy logic is presented in Table 5

TABLE 5: DEFINED RULES FOR THE FUZZY MODEL

Rule number	Age	Tumor size	Menarche age	Alcohol intake	Family History	Output: breast cancer risk
1	R1	SM	EL	LS	YES	LR
2	R1	SM	EL	LS	NO	LR
3	R1	SM	EL	MR	YES	RS
4	R1	SM	EL	MR	NO	LR
5	R1	SM	NM	LS	YES	LR
6	R1	SM	NM	LS	NO	LR
7	R1	SM	NM	MR	YES	RS
8	R1	SM	NM	MR	NO	LR
9	R1	SM	LT	LS	YES	LR
10	R1	SM	LT	LS	NO	LR
11	R1	SM	LT	MR	YES	RS
12	R1	SM	LT	MR	NO	LR
13	R1	MD	EL	LS	YES	HR
14	R1	MD	EL	LS	NO	RS
15	R1	MD	EL	MR	YES	HR
16	R1	MD	EL	MR	NO	HR
17	R1	MD	NM	LS	YES	RS
18	R1	MD	NM	LS	NO	RS
19	R1	MD	NM	MR	YES	HR
20	R1	MD	NM	MR	NO	RS
...	...	...	...	...	...	...
60	R2	MD	LT	MR	NO	RS
61	R2	LG	EL	LS	YES	HR
62	R2	LG	EL	LS	NO	HR
63	R2	LG	EL	MR	YES	HR
64	R2	LG	EL	MR	NO	HR
65	R2	LG	NM	LS	YES	HR
66	R2	LG	NM	LS	NO	RS
67	R2	LG	NM	MR	YES	HR
68	R2	LG	NM	MR	NO	HR
69	R2	LG	LT	LS	YES	HR
70	R2	LG	LT	LS	NO	RS
71	R2	LG	LT	MR	YES	HR
72	R2	LG	LT	MR	NO	HR
73	R3	SM	EL	LS	YES	RS
74	R3	SM	EL	LS	NO	RS
75	R3	SM	EL	MR	YES	HR
76	R3	SM	EL	MR	NO	RS
77	R3	SM	NM	LS	YES	RS
78	R3	SM	NM	LS	NO	RS
79	R3	SM	NM	MR	YES	HR
80	R3	SM	NM	MR	NO	RS

#### IV IMPLEMENTATION OF A RISK-FACTOR SYSTEM FOR PREDICTING OCCURRENCE OF BREAST CANCER USING FUZZY LOGIC

We presented the necessary development and implementation requirements for actualizing the

Breast Cancer Prediction System, it also describes and explains the necessary procedures to be followed. These procedures are to be followed for the development, testing, and implementation of the proposed system and other development and/or implementation related issues.



The development tools used for the design and implementation of the system are; Matrix Laboratory (MATLAB), Microsoft Visual Studio and C# based fuzzy logic library.

#### A. Implementation of the Model with Matlab

The Mamdani Fuzzy Logic for the Model is in Figure 2

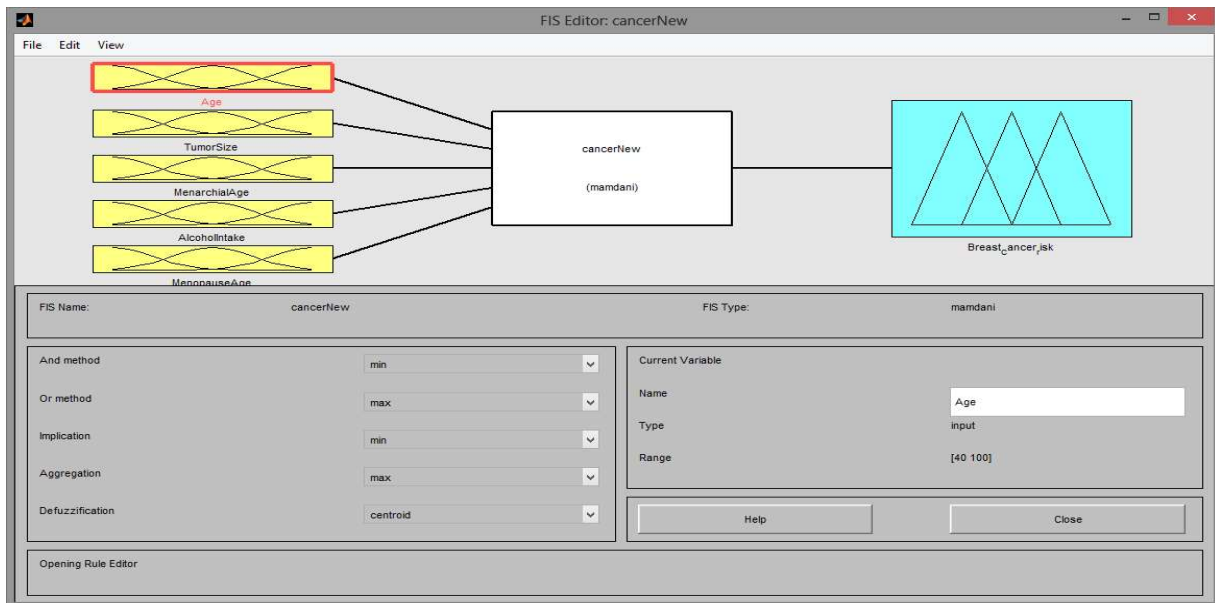


Figure 2: Mamdani Fuzzy Logic for the Model

#### B. Creating Membership Functions

Membership functions are created using the membership function editor. The following steps are used for the specification of the input/output membership functions:

- i. Select each of the input/output membership variables
- ii. Set both the range and display to the vector
- iii. Select add MF's from the edit menu. (Hint: this adds MF's to the chosen input variable)
- iv. Select membership function type for the added MF's e.g. trimf, which is the triangular membership function
- v. Pick each MF and give appropriate linguistic variables with their corresponding ranges
- vi. Pick each MF and give appropriate linguistic variables with their corresponding ranges

### C. Membership Functions

The input membership functions used for the design are presented in figures 3 – 7

#### i. Age

The age input variable which comprises of three triangular membership functions. The display range ranges from [40-100] is as presented in Figure 3.

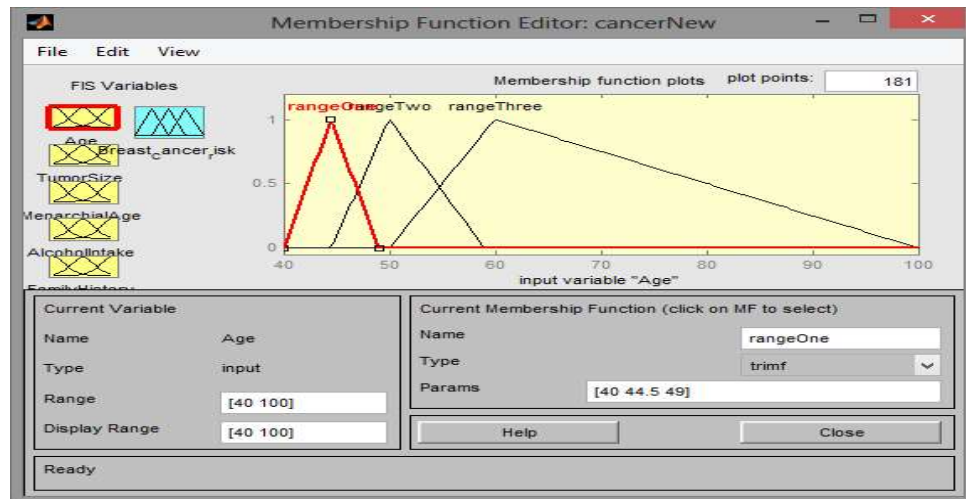


Figure 3: Age input variable

#### ii. Menarche Age

Figure 4 shows the menarche age input variable which comprises of three triangular membership functions and the display range ranges from [10-15]

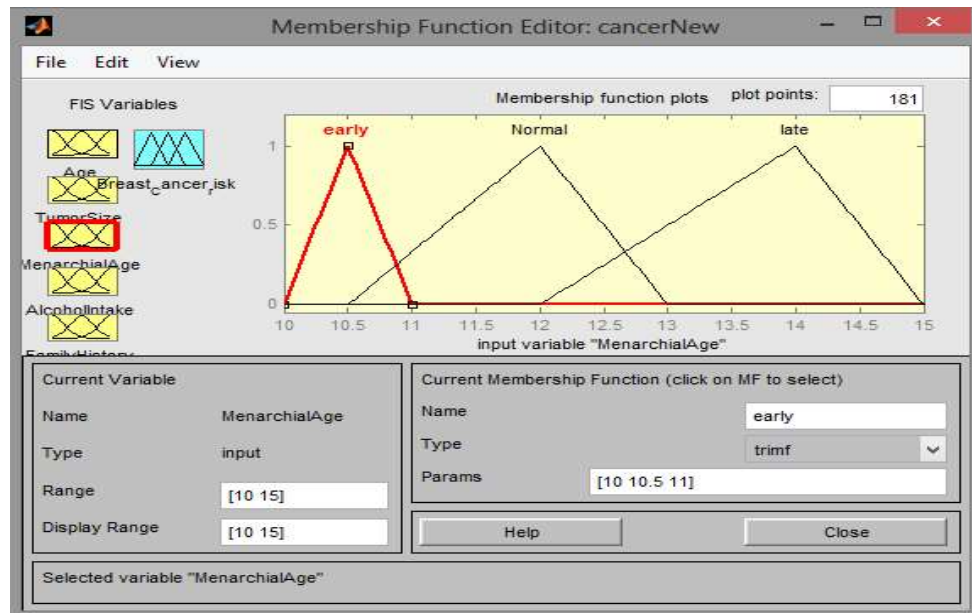


Figure 4: Menarche age input variable

### iii. Family History

The family history input variable which comprises of two triangular membership functions. The display range ranges from [0-1] is in Figure 5

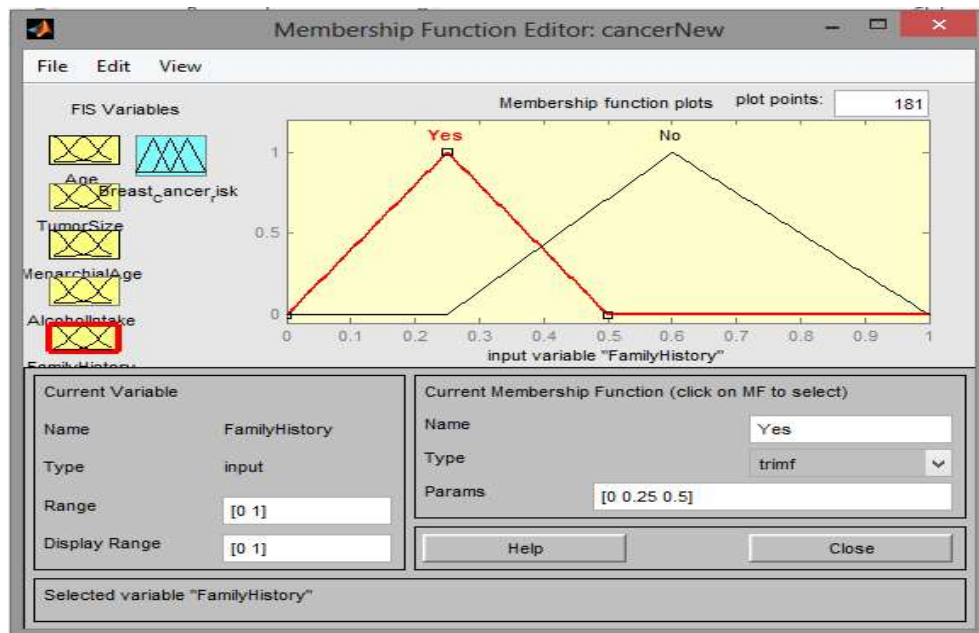


Figure 5: Family history input variable

### iv. Tumor Size

The tumor size input variable which comprises of three triangular membership functions and the display ranges from [0-5] is as shown in Figure 6.

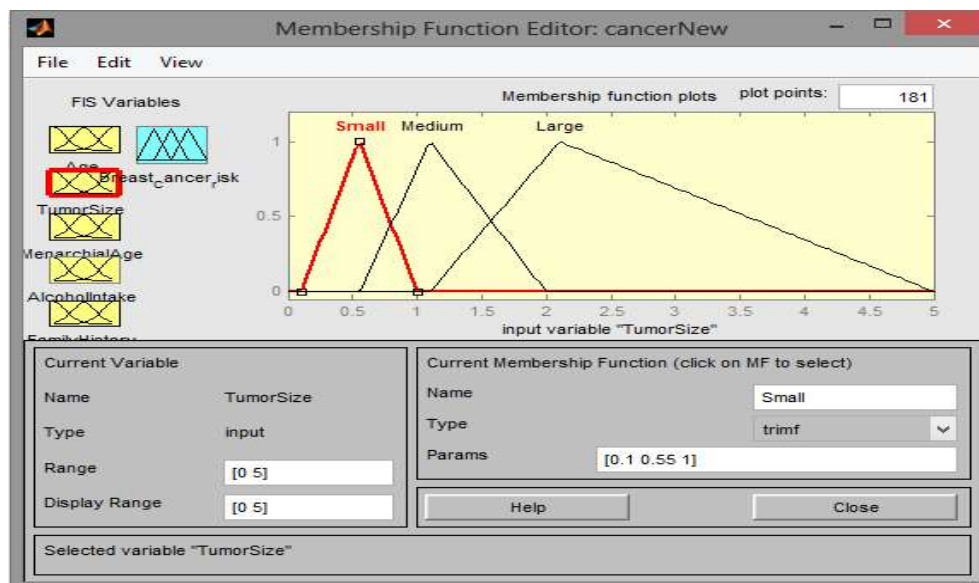


Figure 6: Tumor size input variable

#### v. Alcohol Intake

The alcohol input variable which comprises of two triangular membership functions and the display ranges from [0-1] is in Figure 7.

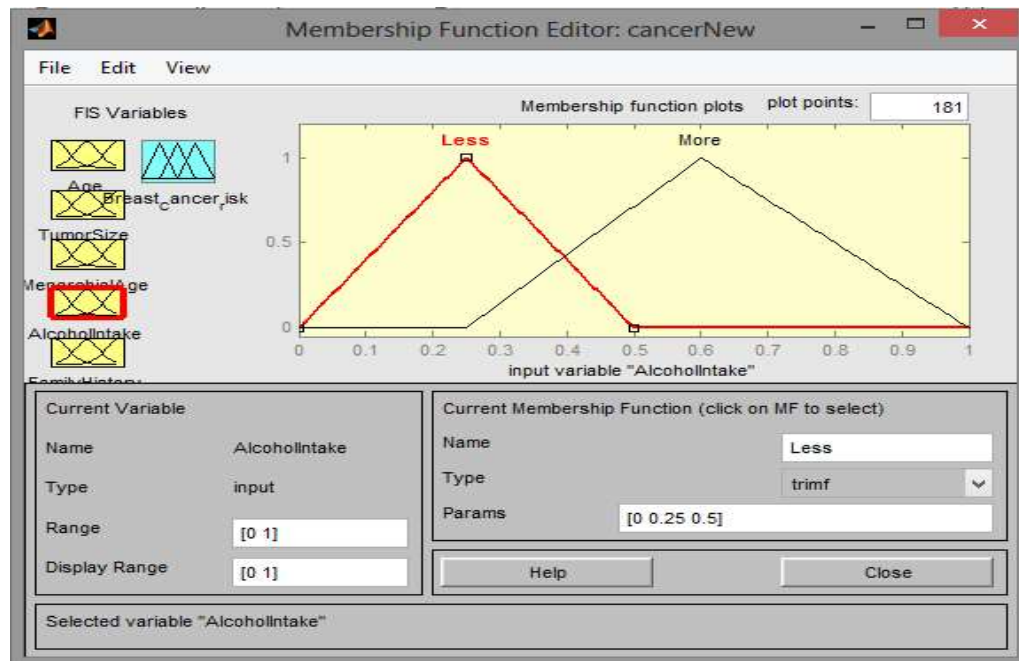


Figure 7: Alcohol intake input variable

#### d. Output Membership Function

The breast cancer risk-output variable which comprises of three triangular membership functions and the display ranges from [0-100] is presented in Figure 8.

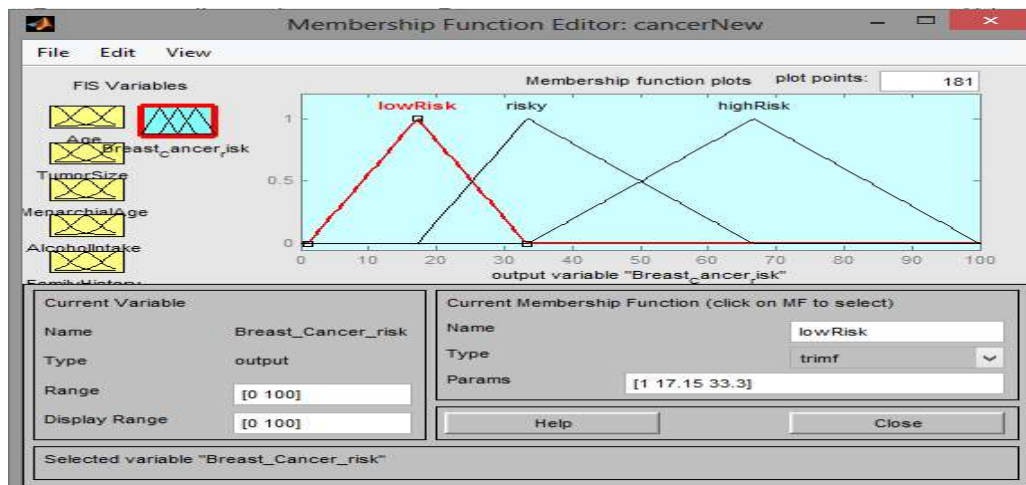


Figure 8: Breast cancer risk output variable



#### e. Rule-Base

After the determination of membership functions and membership degrees for the system, fuzzy logic model formed. Rules of the fuzzy logic model were determined using acquired data and expert opinions. A total of 108 rules were generated. The Fuzzy model rule editor and Fuzzy model rule viewer are presented in Figures 9 and 10 respectively.

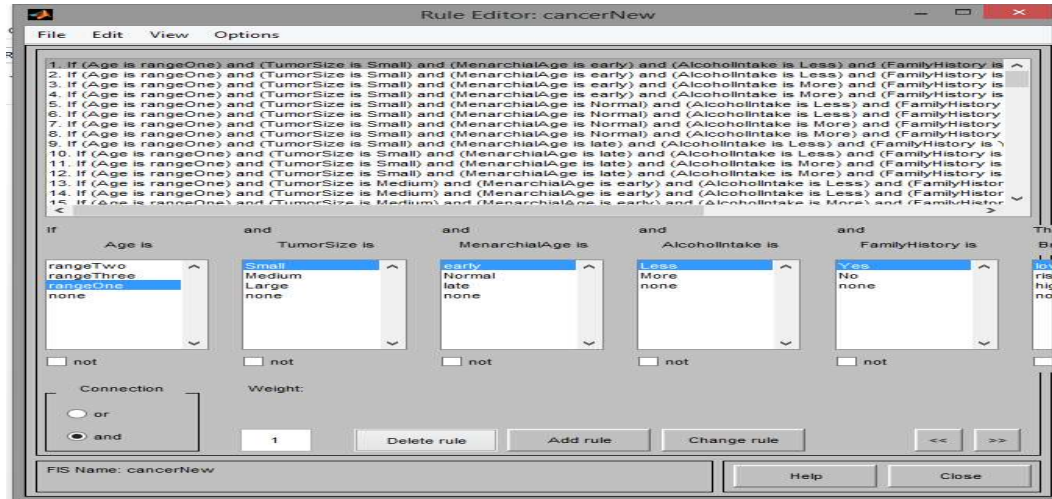


Figure 9: Fuzzy model rule editor

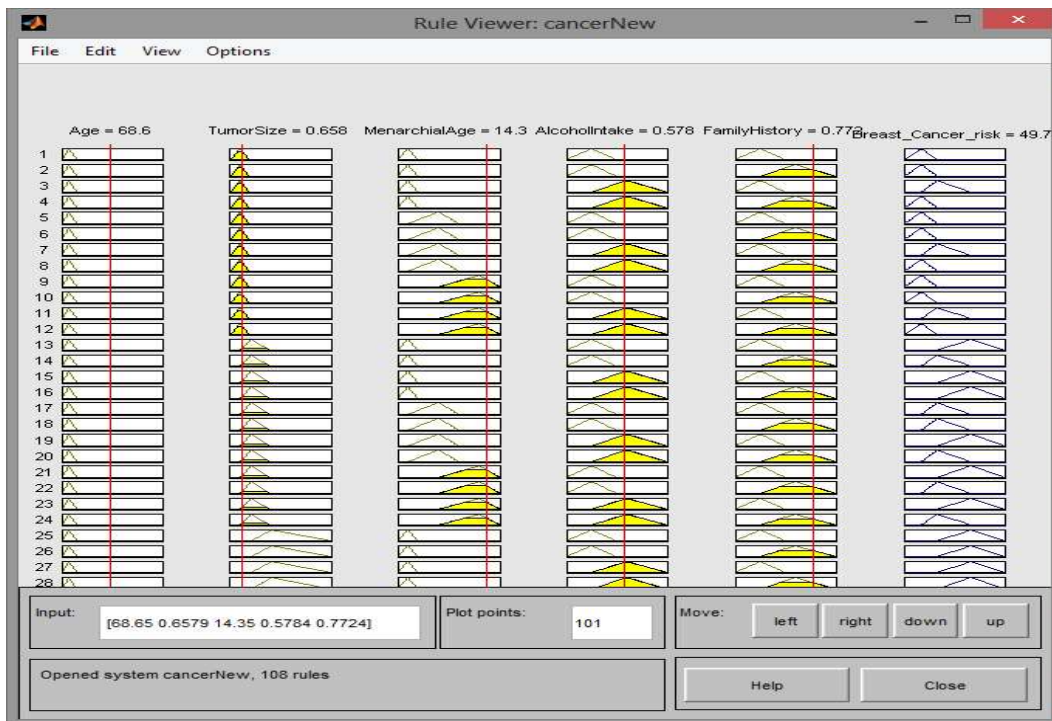


Figure 9: Fuzzy model rule editor

#### f. Graphical User Interface Documentation

The interfaces of the Breast Cancer Risk Expert System application are in Figure 11. The Risk Analysis Form is the form that is used to determine the breast cancer risk of a patient. The form is powered using a fuzzy logic library. Users are required to fill in all fields appropriately. After all fields have been filled appropriately, the user is then required to click the 'Check Risk' button which shows the user the risk value of the patient being analyzed.

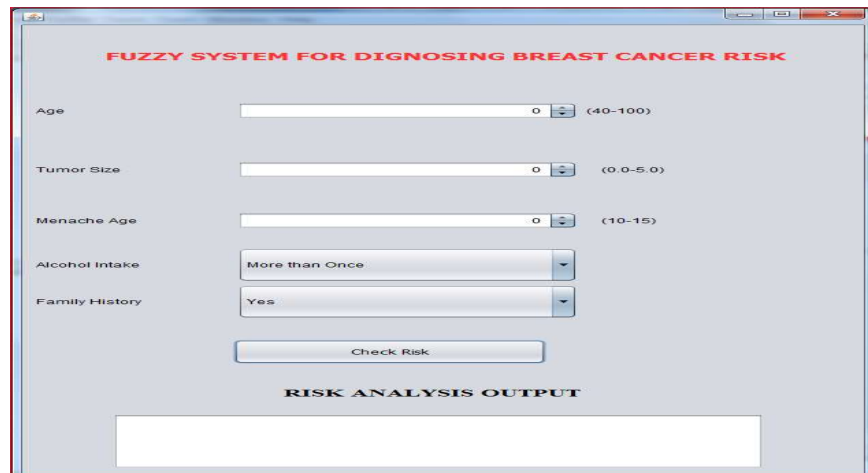


Figure 11: Risk Analysis form

#### g. System Test

The validity of the presence of breast cancer in a patient is done through the risk analysis dialog box (Figure 12).

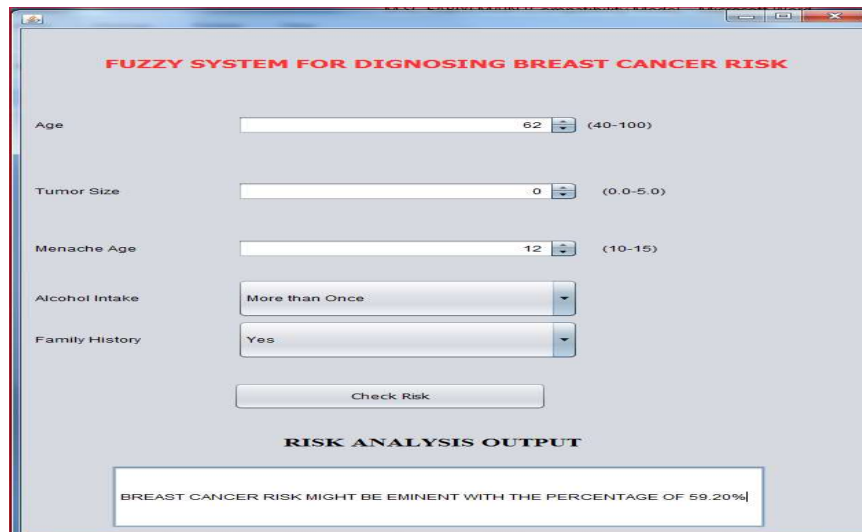


Figure 12: Risk analysis form test



## V. CONCLUSION

The dramatic rise in the incidence of breast cancer in the recent time has necessitated the government to initiate specific measures to increase awareness and educate women to seek early medical advice when they experience breast symptoms. Given that most women display positive attitudes towards breast cancer screening, therefore there is an opportunity to develop a system for early detection of the ailment. This simple, inexpensive technique would be more acceptable to diagnose patients early, rather than the expensive and resource-intensive mammography procedures recommended in high-income countries in addition to breast self-examination and clinical breast examination. Early detection of breast cancer is one of the most important means or tools for successful treatments. In this research work, a desktop based medical expert system using the Mamdani fuzzy logic approach for standardized prediction has been developed and implemented to predict the presence of breast cancer in a woman. The system alongside other medical examinations and tests would help medical practitioners predict the presence of breast cancer.

This research work has been designed to help its users to predict the future occurrence of breast cancer in women but it's still dependent on some other medical examinations. The fuzzy logic model used in this research work utilizes only a few risk factors for its prediction. The system can be improved further by considering other relevant risk factors which can be beneficial to the improvement of the system. Other models for prediction can also be used to enhance this research work. One of these models is known as the Artificial Neural Fuzzy Inference System (ANFIS).

## REFERENCES

- [1] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. Lyon, France: International Agency for Research on Cancer, 2013. <http://globocan.iarc.fr>.
- [2] WHO (2020) Breast Cancer. [www.who.int/diagnosis-screening/breast](http://www.who.int/diagnosis-screening/breast).
- [3] Malcom, R. (2001). Cancer. *Encyclopedia of Life Sciences*. Retrieved from [www.els.net](http://www.els.net)
- [4] Fatimah, A. (2009). Epidemiology and Incidence of Common Cancers in Nigeria, Cancer Reg & Epid wkshop, College of Medicine, University of Lagos, Nigeria.
- [5] Crimmins, E., Hayward, M., & Saito, Y. (1994). Changing Mortality and Morbidity Rates and Health Status and Life Expectancy of the Older Population. *Demography*, 31(1), pp 159-175.
- [6] Jasenka, G., Ivana, R., & Zelimir, K. (2012). Application of Fuzzy Logic in Diet –Therapy Advantages of Application. Emerging Technologies and Applications. IntechOpen Limited, London, SW7 2QJ, United Kingdom.
- [7] Devendra, K. (2008). Soft Computing: Techniques and its Applications in Electrical Engineering. doi:10.1007/978-3-540-77481-5\_1
- [8] Bullinaria A., 2005a. IAI: Production systems, <http://www.cs.bham.ac.uk/~jxb/IAI/w7.pdf>. Accessed 29<sup>th</sup> June, 2019.
- [9] Ramik, J and Imanek, J. (1985) "Inequality relation between fuzzy numbers and its use in fuzzy optimization," Fuzzy Sets and Systems, Vol.16, No.2, pp.123–138
- [10] Akinyokun, O.C., Iwasokun, G., Arekete, S.A., Samuel, O.W. (2015). Fuzzy logic-driven expert system for the diagnosis of heart failure disease. *Artificial Intelligence Research* 4(1), pp 12-21
- [11] Zadeh, L. (2004). Fuzzy Logic Systems: Origin, Concepts and Trends, *Information Sciences*, 172, pp 1-40.
- [12] Lee, C. (1990). Fuzzy Logic in Control Systems: Fuzzy Logic Controller. *IEEE Transaction of System, Man and Cybernetics*, pp 404-418.
- [13] Justin, W., & Eric, Z. (2006). Prediction Markets in Theory and Practice. *National Bureau of Economic Research*, pp 11-12.
- [14] Vaidehi, V., Monica, S., Mohamed, S., Deepika, M., & Sangeetha, S. (2008). A Prediction System Based on Fuzzy Logic, Proceedings of the World Congress on Engineering and Computer Science, October 22 -24, 2008, San Francisco, USA
- [15] Michael, H. (2010). Hotel Booking Prediction by Means of Fuzzy Logic Prediction. *BMI Paper*.
- [16] Abdusselam, A., Mehmet, O., & Mehmet, C. (2005). Water Consumption Prediction of Istanbul City by Using Fuzzy logic Approach. *Water Resources Management*, 19, 641-645. doi:10.1007/s11269-005-7371-1.
- [17] Amosa, B.M.G., Ekuewa, J.B., Onyeka, N.C and Fabiyi, O.A. (2018). Fuzzy Logic Expert System for Diagnosis of Chronic Liver Diseases. International Conference on new Trends in Engineering and Technology. GRT Institute of Engineering and Technology, Chennai, Tamil Nadu, India
- [18] Agboola, A., Gabriel, A., Aliyu, E., & Alese, B. (2013). Development of a fuzzy logic based rainfall prediction model. *International Journal of Engineering and Technology*, pp 427 - 435.
- [19] Kailan, S., & Zakir, H. (2013). Applying Fuzzy Logic to Risk Assessment and Decision Making. Retrieved from [www.casact.org/research/Applying-Fuzzy-Logic.pdf](http://www.casact.org/research/Applying-Fuzzy-Logic.pdf)
- [20] Amosa, B.M.G., Adisa, S.P., Ateko, B.C. Ugwu, J.I. (2018). Adaptive Neuro Fuzzy Expert System for Diagnosing HIV. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. Volume 3, Issue 5, pp. 254 – 265
- [21] Navjot, K., Navleen, S., & Anand, N. (2013). Review of Expert Systems Based on Fuzzy Logic. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(3), pp 1334-1339.

- [22] Reus N (1994) Assessment of benefits and drawbacks of using fuzzy logic, especially in fire control systems. Fysisch en Elektronisch Lab TNO the Hague (Netherlands).
- [23] Rosa, L., & Cuauhtemoc, L. (2013). Fuzzy Logic Model for Predicting the Number of Online Courses Needed from Number of Students Enrolled in Higher Education. *International Conference Frontiers in Education: CS and CE*.
- [24] Yilmaz, A., & Ayan, K. (2011). Cancer Risk Analysis by Fuzzy Logic Approach and Performance Status of the Model, *Turk J Elec Eng & Comp Sci*, 21: pp 897 –912.



# Smart Phone App to Prevent Sharing Microphones in The Covid-19 Era

O. E. Ekong, V.A. Eyo, A.P. Ekong  
Department of Computer Science,  
Akwa Ibom State University  
Akwa Ibom State, Nigeria

OtuekongEkong@aksu.edu.ng<sup>1</sup>, VictoWills@gmail.com<sup>2</sup>, AnietieEkong@aksu.edu.ng<sup>3</sup>  
+2348135406037, +2347064559585, +2348035467965

**Abstract:** *Any smart phone can now be turned into a sound capturing device because of its built-in microphone. The ubiquity of smart phones expands the number of people able to participate in meetings. This paper presents methods and results of a sound capturing system capable of functioning as an ad-hoc microphone that feeds into a public address system. The Recursive Waterfall approach guided the design and implementation. The system comprises of Wi-Fi-Mic mobile app and Wi-Fi-Mic server app, both developed using Java. The Wi-Fi protocol was adopted to link the two apps so as to accommodate many participants. The traditional cables can be used to channel sound from the server app into a public address system for additional sound amplification. Results show that, when used with an external microphone (mouth-piece), this system can achieve performance nearing that of a typical wireless microphone.*

**Key words:** *Acoustic sensing, Audio systems, Bring Your Own Device, Mobile computing, Meeting support system.*

## I. INTRODUCTION

Meetings are a crucial component of human existence. Meetings provide avenues for man to communicate verbally and nurture social relationships with other meeting attendees (Xu, Yu, Wang & Ni, 2014). In a large gathering, a public address (PA) system is usually needed to obtain input sound from microphones, musical instruments, acoustic sound source, or recorded sound.

Thereafter to amplify the sound; and finally project the sound through loudspeakers. A microphone is the sensing unit that captures human voice or other acoustic sound source. Without amplifying the input sound other meeting attendees would have difficulty following what is going on. In some large meetings, the microphones provided are:

- Not portable – thus requiring meeting attendees wishing to make contributions to manoeuvre to the nearest microphone stand.
- Fewer than the number of speakers/participants – thus requiring microphones to be passed around to whosoever wishes to speak per time.

Such practices can lead to: a lot of time being wasted trying to access a microphone so as to make a contribution; participants jousting for attention thereby disturbing the audience; loss of audience control; and possible disruption of the underlying meeting. In this era of COVID-19 pandemic, such practices can lead to an increase in the risk of exposure.

In several countries, COVID-19 outbreaks have been linked to large gatherings, such as religious events, sporting events, social events, youth camps, conferences, and exhibitions. In the absence of a vaccine, event organizers must take reasonable steps to reduce the risk of infection. As part of the measures necessary to reduce the risk of infection, the government and public health authority in several countries demanded that event organizers of large gatherings comply with the following (NCDC, 2020 Jun 13):

- Attendance should not exceed 1/3 of sitting capacity.
- Attendees should be made to sit and maintain two metres distance from each other.
- Practices that require direct contact (e.g., handshakes, hugs, and kisses) should be discouraged.
- Practices that require sharing of materials (e.g., microphones, communal cups, ablution kettles, and offering trays) should be limited.

There are two established media through which audio signals are channelled: wired (corded) and wireless (cordless). In the wired media, sound waves are transmitted as pulses through cables (wires); whereas in the wireless media, sound waves are transmitted as radio waves through air. All the processes that enable audio signals to be sent or routed over a media to a receiver over a given distance is referred to as sound channelling (Abubakar, 2009; Zhu, Liu, Yu, Pfeffer, Dave & Nam, 2004). Basically, sound channelling refers to the uni-directional transmission of sound from transmitter to receiver.

A typical wireless microphone uses radio waves that operate in the Very High Frequency (VHF) or Ultra High Frequency (UHF) band. One common problem with wireless transmission, especially in the VHF band, is noise. The audio quality is susceptible from interferences from nearby electronic devices that have transmitters such as other wireless microphones and mobile phones. This can be overcome by switching to relatively newer options for wireless communication over a short range, such as Wireless Fidelity (Wi-Fi) and Bluetooth.

There is an abundance of personal smart mobile communication devices in the market at the present time. The term 'smart' is used here to denote that an end user can install his/her preferred mobile applications in addition to the bundled mobile applications. Examples of such smart mobile communication devices include: smart phones, personal digital assistants (PDAs), tablets, notebook computers, and laptop computers. Many of these devices come with Wi-Fi and Bluetooth communication capability as a standard feature.

For several years now, the practice referred to as Bring Your Own Device (BYOD) has witnessed an increase in adoption in several organizations. Employees are provided with avenues to use their own personal smart mobile communication devices in formal work environments to carry out several tasks including syncing with their work emails, syncing with their work calendars, editing Word documents, compiling Excel spread sheets, preparing PowerPoint presentations, and assessing corporate data (Ogie, 2016). This enables the organization to derive various benefits, such as: increased productivity, improved employee satisfaction, cost savings, and increase in revenue (Pelino, 2012).

Although smart mobile communication devices are ubiquitous, they are largely under-utilized. Many users focus on the primary functions and ignore or are ignorant of the derivative or ancillary capabilities. A typical smart phone is packed with several sensors that can be re-purposed to meet diverse needs with the use of clever software applications (apps). This research work explores sound channelling from a smart phone's built-in microphone sensor to a Wi-Fi connected personal computer (PC) for onward amplification through a public address system.

## II. STATEMENT OF PROBLEM

Many of us walk around with sophisticated mobile devices which are equipped with diverse sensors. However, we hardly exploit the capabilities of these sensors to tackle some of the common challenges that confront us. One challenge we frequently face is the need to capture, channel, amplify and project sound in large gatherings. This work seeks to tackle this challenge by developing an app that extends the functionality of our smart phone's in-built microphone beyond merely enabling telephone conversations.

## III. OBJECTIVE

The aim of this research is to develop a system for capturing sound through a smart phone's microphone sensor and channelling of the sound through a wireless connection to a computer acting as the server for onward transmission to the amplifying and loudspeaker units of a public address system.

The proposed system comprises of a mobile app developed on Android Studio and native app acting as the server developed using Java.

#### IV. RELATED WORK

This section details some of the efforts by researchers to re-purpose the built-in microphone in smart phones for derivative or ancillary capabilities. Xu, Yu, Wang and Ni (2014) proposed a smart phone-based meeting support system, named *SmartMic* that connects to a wireless speaker via Bluetooth connection or directly to an amplifier using the 3.5mm jack port. Meeting attendees were to speak through their smart phone instead of the traditional public address microphone. The system was designed to readily integrate with common facilities available meeting venues. Their preliminary experiment results showed that the system supported easy participation of meeting attendees, enhanced user experiences, facilitated social interactions, and increased the meeting productivity.

As a consequence of its use of the Bluetooth communication protocol, only 10 meeting attendees can connect per time; and the coverage area is restricted to 100m. This work attempts to resolve these challenges by proposing a system that utilizes the Wi-Fi communication protocol. Celestina, Hrovata and Kardous (2018) observed that there are many sound measuring apps for various mobile platforms in the market but only a fraction of these apps achieve results near enough to that obtained from professional sound level measuring instruments.

In their paper, they developed a sound level meter app and also presented methods and results of calibrating the app according to relevant ANSI (American National Standards Institute) and IEC (International Electrotechnical Commission) sound level meter standards.

They reported that, when used with an external microphone, their sound level meter app can achieve compliance with most of the requirements for Class 2 of IEC 61672/ANSI S1.4-2014 standard. Picaut, Fortin, Bocher, Petit, Aumond and Guillaume (2019) proposed an open crowd-sourcing method for the assessment of noise in an environment.

This solution comprised of a smart phone application and a spatial data infrastructure specifically developed to collect physical data (noise indicators and GPS positions) and perceptual data (pleasantness) from the environment. Although any number of contributors could use the proposed system, results of higher relevance were achieved when the approach was restricted to a limited number of supervised contributors.

#### V. METHODOLOGY

The Recursive Waterfall Model guided the sequence of processes and procedures relied upon for the design and development of this work. The Recursive Waterfall Model was adopted because it allows a loop to the previous levels of operations so that changes can be effected whenever the need arises.

#### VI. CONCEPT DESIGN

The proposed system, called *Wi-Fi-Mic*, consists of two components as shown in Fig. 1: the Android-based mobile app that captures sound using the built-in microphone or a mouth-piece attached to it through the 3.5mm jack port; and the native app acting as the server that resides on a laptop/PC to receive the audio signal transmitted wirelessly to it from the smart phone for onward transmission to the amplifying and loudspeaker units of a public address system.



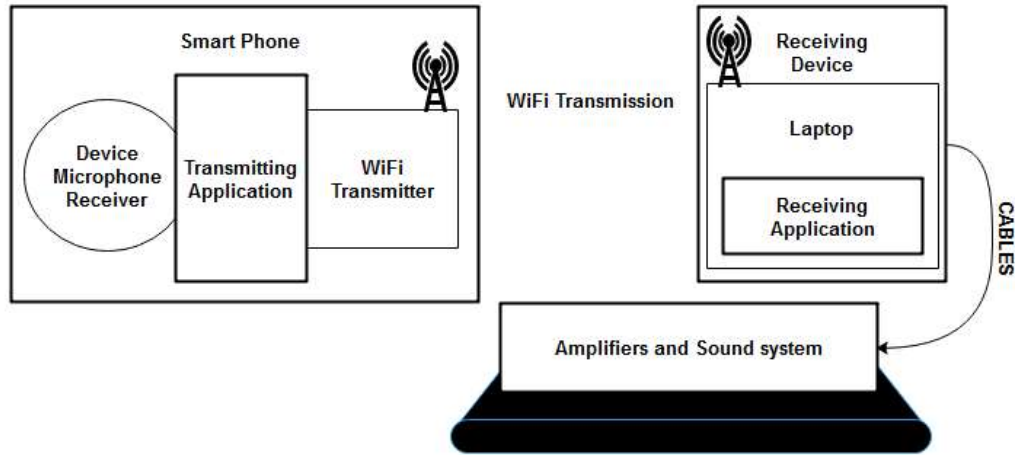


Fig. 1: Architecture of *Wi-Fi-Mic* System

Although Unified Modelling Language (UML) is not a method or procedure for system development, it can be used with any software development

methodology to simplify the overall comprehension of a proposed system. Fig. 2 details the flowchart that guided the development of the *Wi-Fi-Mic* system.

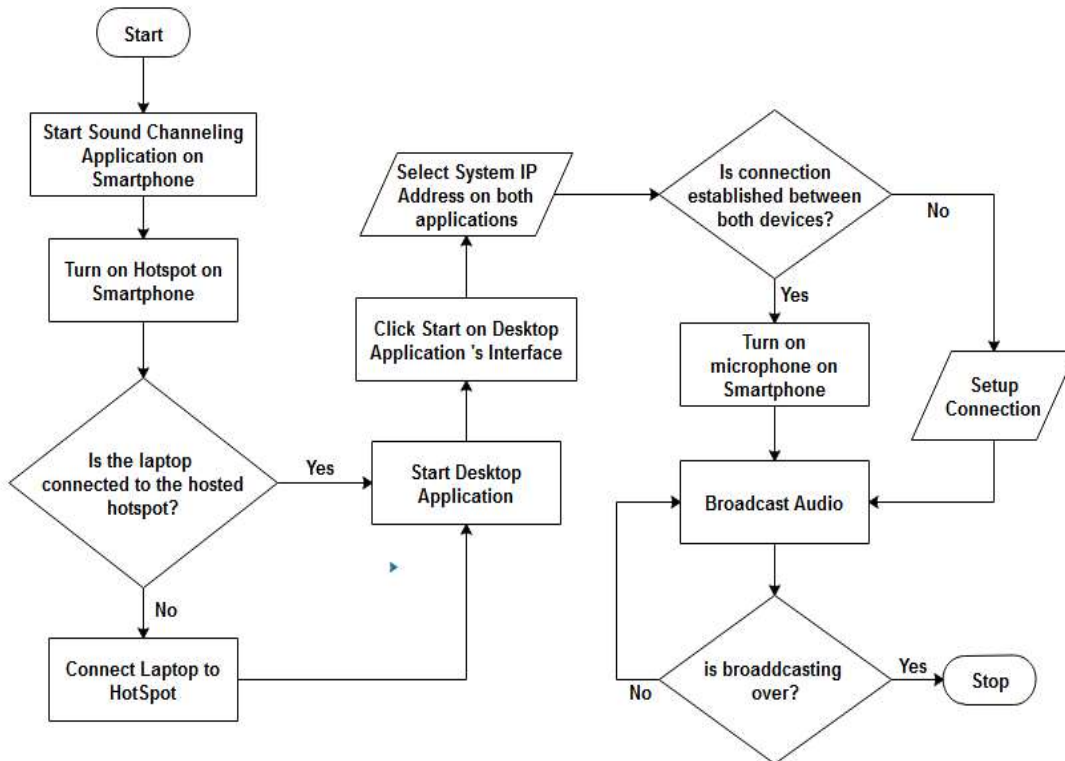


Fig. 2: Flowchart of *Wi-Fi-Mic* System Operations

The *Wi-Fi-Mic* server app (see Fig. 3) was developed using Java on the NetBeans Integrated Development environment (IDE). The server app will run on a computer which has the following minimum requirements: 32-bit or 64-bit Windows Operating System, 1.0 GHz of processor speed, 1 GB of random access memory (RAM), and screen resolution of 800 x 600 pixels. The *Wi-Fi-Mic* mobile app (see Fig. 4) was developed using Java on the Android Studio (version 3.3). The smartphone app will run on Android phones with a minimum requirement of: Operating System – Ice Cream Sandwich, Android Version – 4.0, Memory (RAM) – 500MB, CPU Frequency – 1.3 GHz, Internal Storage – 1 GB, Screen Resolution – 854 x 480.

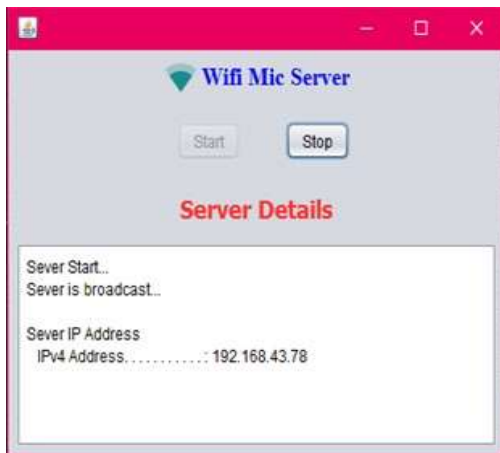


Fig. 3: Wi-Fi-Mic Server Set-up



Fig. 4: Wi-Fi-Mic Smartphone Set-up

## VII. DISCUSSION OF FINDINGS

Upon launching the *Wi-Fi-Mic* server app, the user is taken to the server set-up screen. Similarly, upon launching the *Wi-Fi-Mic* mobile app, the user is taken to the client set-up screen which prompts the user to establish a connection with the server by either connecting with available servers that have been automatically discovered or by providing the Internet Protocol (IP) Address of a hidden server. User authentication was not incorporated in this prototype. The system is designed to allow any meeting attendee with access to the Wi-Fi network to connect.

Experimental findings showed that more than 10 participants were able to connect simultaneously to the Wi-Fi-Mic system than would have been possible if the Bluetooth communication protocol was used. In addition, the audio quality was comparable to that of a standard wireless microphone despite some of the meeting attendees being more than 20 meters from the Wi-Fi router.

## VIII. CONCLUDING REMARKS

Most people today carry around their own personal smart mobile communication devices. These devices can be made to serve as *ad-hoc* wireless microphones that channel acoustic sound into a public address system so as to encourage and enable active participation of meeting attendees in small, medium or large gatherings, either through speech or singing. By choosing to use Wi-Fi as the communication protocol, the system was able to accommodate a larger number of meeting attendees simultaneously; a wider coverage area; and greater flexibility in the freedom of movement of meeting attendees.

## IX. CONTRIBUTIONS TO KNOWLEDGE

This study has demonstrated that utilizing a smart phone's built-in microphone as an *ad-hoc* wireless microphone for channelling sound into a public address system is readily achievable. The study has also highlighted the possibility of scaling up the number of active meeting attendees by deploying a Wi-Fi connection rather being restricted by the limitations of Bluetooth connection.

## X. RECOMMENDATION AND FUTURE WORK

In compliance with the measures necessary to reduce the risk of exposure to infectious diseases such as COVID-19, the use of a smart phone's microphone as an *ad-hoc* microphone can serve as a means of limiting the sharing of microphones in large gatherings. In this prototype, user authentication was not incorporated. In future work, it should be incorporated to ensure that only authenticated meeting attendees are allowed to connect to the server app. This will prevent the meeting from being usurped by detractors.

## REFERENCES

1. Abubakar S. B. (2009). *Wireless Microphone System* (BSc dissertation, Universiti Teknikal Malaysia, Melaka).
2. Celestina, M., Hrovata, J., and Kardous, C. A. (2018). Smartphone-based sound level measurement apps: Evaluation of compliance with international sound level meter standards. *Applied Acoustics*, 139, 119-128. <https://doi.org/10.1016/j.apacoust.2018.04.011>
3. Nigeria Centre for Disease Control (2020 Jun 13). Nigeria Releases COVID-19 Guide to Re-Open Places of Worship. Retrieved from <https://www.ncdc.gov.ng/news/255/nigeria--releases-covid-19-guide-to-re-open-places-of-worship>
4. Ogie, R., (2015). Bring your own device: an overview of risk. *IEEE Consumer Electronics Magazine*, 5(1), 114-119. <https://doi.org/10.1109/MCE.2015.2484858>
5. Pelino, M. (2012). *Building the case for a Bring-Your-Own-Device (BYOD) program*. Cambridge, MA: Forrester
6. Picaut, J., Fortin, N., Bocher, E., Petit, G., Aumond, P., and Guillaume, G. (2019). An open-science crowdsourcing approach for producing community noise maps using smartphones. *Building and Environment*, 148, 20-33. <https://doi.org/10.1016/j.buildenv.2018.10.049>
7. Xu, H., Yu, Z., Wang, Z., and Ni, H. (2014). SmartMic: a smartphone-based meeting support system. *The Journal of Supercomputing*, 70, 1318-1330.
8. Zhu, C., Liu, G., Yu, Q., Pfeffer, R., Dave, R. N., & Nam, C. H. (2004). Sound assisted fluidization of nanoparticle agglomerates. *Powder Technology*, 141 (1-2), 119-123.

# Security Enhancement of an Automated Teller Machine Using Fingerprint and Password

C. Nwankwo  
Augustine University  
Ilara Epe, Nigeria  
[Chidera.Nwankwo@augustinuniversity.edu.ng](mailto:Chidera.Nwankwo@augustinuniversity.edu.ng)  
+2348100936083

G. Oletu  
American University of  
Nigeria  
Yola, Nigeria  
[grace.Oletu@aun.edu.ng](mailto:grace.Oletu@aun.edu.ng)  
+2348057202980

O.B. Longe  
American University of  
Nigeria  
Yola, Nigeria  
[olumide.Longe@aun.edu.ng](mailto:olumide.Longe@aun.edu.ng)  
+2348160900893

**Abstract:** *This paper focuses on how to enhanced security of Transactions in Automatic Teller Machine system using a multi-factor authentication system (Password and Fingerprint). In order to achieve security and to overcome illegal activities, shortcoming of piracy in money transactions, we propose the idea of using fingerprints of customers and password instead of traditional use of PIN number. After authorized verification, the customer will be able to proceed for transaction else after four successive wrong attempts, the ATM card will be ejected automatically and a message will be sent to the registered mobile number. The proposed system is developed to provide better security to the ATMs.*

**Keywords—** ATM, Fingerprint, PIN, Biometric.

## I. INTRODUCTION

Fast development of banking technology has changed the way banking activities are dealt with. One banking technology that has impacted positively and negatively to banking activities and transactions is the advent of automated teller machine (ATM). With an ATM, a customer is able to conduct several banking activities such as cash withdrawal, money transfer, paying phone and electricity bills beyond official hours and physical interaction with bank staff. In a nutshell,

ATM provides customers a quick and convenient way to access their bank accounts and to conduct financial transactions.

Personal identification number (PIN) or password is one important aspect in ATM security system. PIN or password is commonly used to secure and protect financial information of customers from unauthorized access [1]. An ATM (known by other names such as automated banking machine, cashpoint, cash machine or a hole in the wall) is a mechanical system that has its roots embedded in the accounts and records of a banking institution [1]-[2].

It is a computerized machine designed to dispense cash to bank customers without need of human interaction; it can transfer money between bank accounts and provide other basic financial services such as balance enquiries, mini statement, withdrawal and fast cash among others [3]

The paper is arranged as follows. Section II provided the background of ATM security and the need for biometrics. Section III Proposed System Model, Section IV Software Output for the Proposed System and Section V presented the results obtained from the survey. Section VI concluded the paper.

## II. RESEARCH BACKGROUND

Crime at ATMs has become a nationwide issue that faces not only customers, but also bank operators and this financial crime case rises repeatedly in recent years [4]. A lot of criminals tamper with the ATM terminal and steal customers' card details by illegal means. Once users' bank card is lost and the password is stolen, the users' account is vulnerable to attack.

Traditional ATM systems authenticate generally by using a card (credit, debit, or smart) and a password or PIN which no doubt has some defects [5]. The prevailing techniques of user authentication, which involves the use of either passwords and user IDs (identifiers), or identification cards and PINs (personal identification numbers), suffer from several limitations [6]. Passwords and PINs can be illicitly acquired by direct covert observation. When credit and ATM cards are lost or stolen, an unauthorized user can often come up with the correct personal codes.

Despite warning, many people continue to choose easily guessed PIN's and passwords - birthdays, phone numbers and social security numbers. Recent cases of identity theft have heightened the need for methods to prove that someone is truly who he/she claims to be. Biometric authentication technology may solve this problem since a person's biometric data is undeniably connected to its owner, is non-transferable and unique for every individual. The system can compare scans to records stored in a central or local database or even on a smart card.

Biometrics can be defined as a measurable physiological and behavioural characteristic that can be captured and subsequently compared with another instance at the time of verification. It is automated methods of recognizing a person based on a physiological or behavioural characteristic [7]. It is a measure of an individual's unique physical or behavioural characteristics to recognize or authenticate its identity [8]. Common physical biometrics characteristics include fingerprint, hand or palm geometry, retina, iris and face while popular behavioural characteristics are signature and voice. Biometrics technologies are a secure means of authentication because biometrics data are unique, cannot be shared, cannot be copied and cannot be lost. Shaikh and Rabaiotti

[9] analyzed the United Kingdom identity card scheme. Their analysis approached the scheme from the perspective of high volume public deployment and described a trade-off triangle model. They found that there is a trade-off between several characteristics, i.e., accuracy, privacy and scalability in biometric based identity management system, where emphasis on one undermines the other.

Amurthy and Redddy [6] developed an embedded fingerprint system, which is used for ATM security applications. In their system, bankers collect customers' finger prints and mobile numbers while opening accounts, then customer only access ATM machine.

The working of the ATM machine is such that when a customer place a finger on the finger print module it automatically generates every time different 4-digit code as a message to the mobile of the authorized customer through GSM modem connected to the microcontroller. The code received by the customer is entered into the ATM machine by pressing the keys on the touch screen. After entering it checks whether it is a valid one or not and allows the customer further access. Das and Schouten and Jacobs [8] presented an evaluation of the Netherlands' proposed implementation of a biometric passport, largely focusing on technical aspects of specific biometric technologies (such as face and fingerprint recognition) but also making reference to international agreements and standards (such as ICAO and the EU's "Extended Access Control") and discussed the privacy issue in terms of traditional security concepts such as confidentiality.

The use of fingerprints as a biometric is both the oldest mode of computer-aided, personal identification and the most prevalent in use today [10]. In the world today, fingerprint is one of the essential variables used for enforcing security and maintaining a reliable identification of any individual. Fingerprints are used as variables of security during voting, examination, operation of bank accounts among others. They are also used for controlling access to highly secured places like offices, equipment rooms, control centers and so on [11].



### III. PROPOSED SYSTEM

Figure 1 is the conceptual model that defines the structure, behavior, and more views of a way that supports the ATM system. It represents the actions, processes, and operations within the modeled system. The user has to insert his/her bank card to log in to the system. The user will be prompted for Biometric verification to be provided by the user.

If the fingerprint is matched, the user will be prompted for Password, if matched the user will be allowed to carry out the following transactions such as: Find Transfer, Mini Statement, Bill Payment, Cash Withdraw, and Balance Enquiry. After finishing a successful transaction, the user can either continue or Card ejected and exit the ATM. If not the user will be allowed with four more attempts to log in. In the case of four successive wrong attempts, the account will be blocked, Card ejected and a message sent to the bank and cardholder. The activity diagram of the proposed system is displayed in as flowchart diagram in Figure 2.

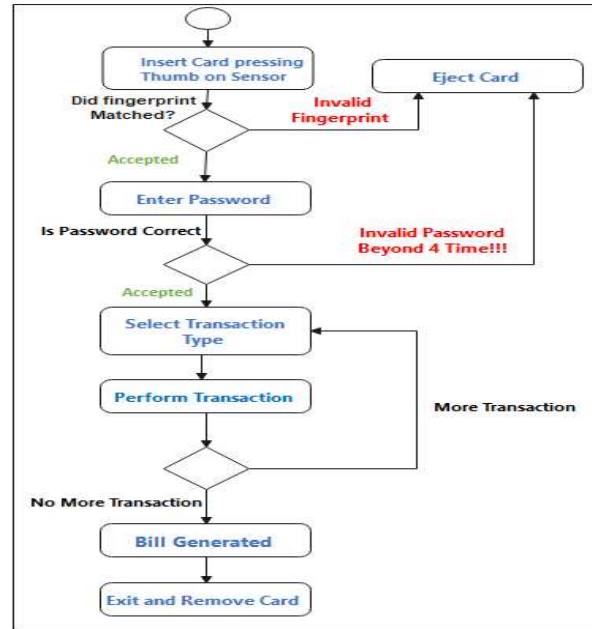


Figure 2 Activity Flowchart Diagram for the Proposed System

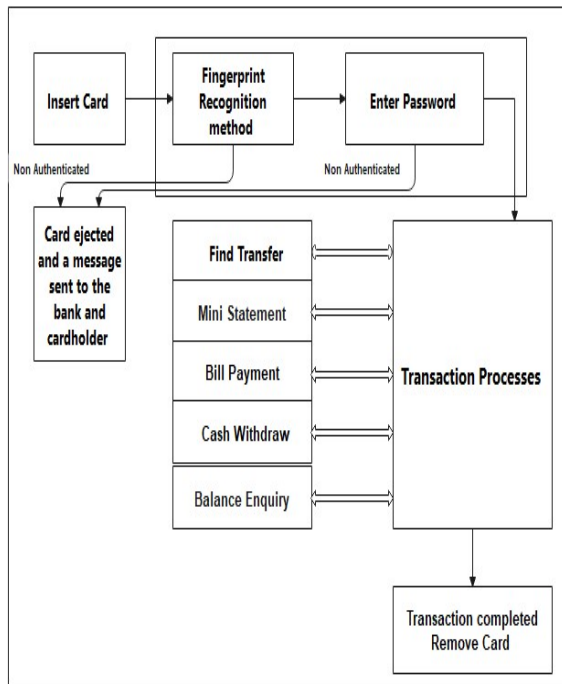


Figure 1 Proposed System Model

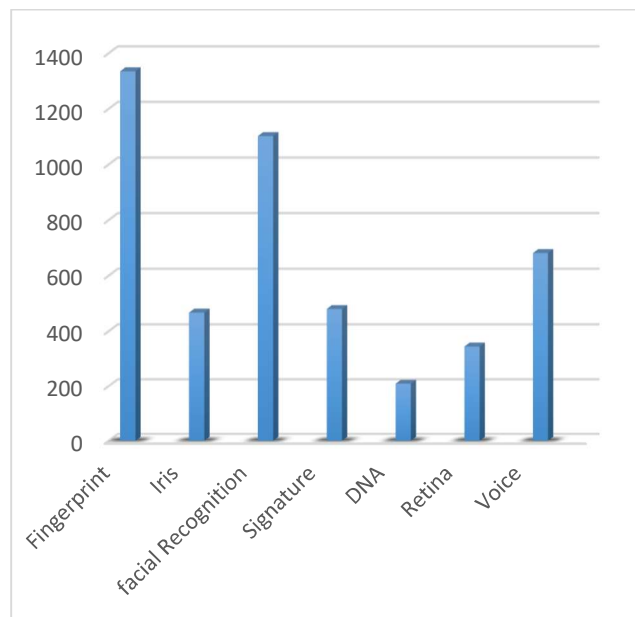


Figure 3 Combination of Biometrics Characteristics with Password



**Table 1: Respondent Information**

Age	Frequency	Percentage
18-24	262	18.0
25-34	450	30.0
35-44	367	24.0
45-54	177	12.0
54-64	150	10.0
Above 65	94	6.0
Total	1500	100%

#### IV. SOFTWARE OUTPUT

*//This displays what happens when a customer carries out a single transaction*

Insert your card and press your thumb on the sensor:  
Please input your password:

Welcome!

Select a Transaction Type:

1. Fund Transfer
2. Mini Statement
3. Bill Payment
4. Cash Withdraw
5. Balance Enquiry

Performing Fund Transfer....

Operation done!

Do you want to perform another transaction! (Y/N):

n

Generating Bill...

Please collect your receipt

Ejecting card...

Your card has been ejected

*//This displays what happens when a customer carries out multiple transactions*

Insert your card and press your thumb on the sensor:

Please input your password:

Wrong password, please try again!

Please input your password:

Wrong password, please try again!

Please input your password:

Welcome!

Select a Transaction Type:

1. Fund Transfer
2. Mini Statement
3. Bill Payment
4. Cash Withdraw
5. Balance Enquiry

Performing Cash Withdraw....

Operation done!

Do you want to perform another transaction! (Y/N):

y

Select a Transaction Type:

1. Fund Transfer
2. Mini Statement
3. Bill Payment
4. Cash Withdraw
5. Balance Enquiry

Displaying Mini Statement....

Operation done!

Do you want to perform another transaction! (Y/N):

n

Generating Bill...

Please collect your receipt

Ejecting card...

Your card has been ejected

*//This displays what happens when a customer inputs wrong password more than 4 times*

Insert your card and press your thumb on the sensor:

Please input your password:

Wrong password, please try again!

Please input your password:

Wrong password, please try again!

Please input your password:

Wrong password, please try again!

Please input your password:

Wrong password, please try again!

You have entered an incorrect password 4 times...

Your card has been ejected!

#### V. RESULTS AND DISCUSSION

A total of 1500 respondents were obtained from the survey carried out and the analysis of all the data obtained from the respondents shows that 18% of the respondents are between 18-24, 30% of respondents are between the ages of 25-34, and 24% of the respondents are between the ages of 35-44, 12% of the respondents are between 45-55, 10% of the respondents are between 54-64 years and finally 6% of the of the respondents were 65 and above in Table 1. In Figure 3, 89% respondents preferred Fingerprint and Password authentication enhance security in ATM as when compared to other biometric methods.

## VI. CONCLUSION

Automatic Teller Machines is the most used technology in the increasing financial transaction of the current world. Biometrics, particularly fingerprint scanning, continues to gain acceptance as a reliable form of securing access through identification and verification processes. Using the fingerprint and Password authentication method at the ATM will reduce various fraudulent activities which in turn increases security and provides the user with a safe and reliable method of transaction.

## REFERENCES

- [1] S.S, Das and J. Debbarma, "Designing a Biometric Strategy (Fingerprint) Measure for Enhancing ATM Security in Indian e-banking System", *International Journal of Information and Communication Technology Research*, vol.1, no. 5, pp.197-203, 2011.
- [2] W.W.N. Wan, C.L. Luk, and C.W.C. Chow, "Customers Adoption of Banking Channels in Hong Kong", *International Journal of Bank Marketing*, vol. 23, no. 3, pp. 255-272, 2005.
- [3] Wikipedia the free encyclopedia, "Biometrics", Downloaded March 20, 2012 from <http://en.wikipedia.org/wiki/Biometrics>.
- [4] B. Richard and M. Alemayehu, "Developing E-banking Capabilities in a Ghanaian Bank: Preliminary Lessons. *Journal of Internet Banking and Commerce*, vol. 11, no. 2, 2006. Downloaded March 15, 2012 from <http://www.arraydev.com/commerce/jibc/>
- [5] P.K. Amurthy and M.S. Redddy, "Implementation of ATM Security by Using Fingerprint recognition and GSM", *International Journal of Electronics Communication and Computer Engineering* vol.3, no. 1, pp. 83-86, 2012.
- [6] N.K. Ratha, J.H. Connell, and R.M. Bolle, "Enhancing Security and Privacy in Biometrics-based Authentication Systems, *IBM Systems Journal*, vol. 40, no. 3, pp. 614-634, 2001.
- [7] N.K. Ratha, S. Chikkerur, J.H. Connell and R.M. Bolle. "Generating Cancelable Fingerprint Templates", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, 2007.
- [8] B. Schouten and B. Jacobs, "Biometrics and their use in e-passport", *Image and Vision Computing* vol. 27, pp. 305-312. 2009,
- [9] S.A. Shaikh and J.R. Rabaiotti,. "Characteristic trade-offs in designing large-scale biometric-based identity management systems". *Journal of Network and Computer Applications* vol. 33, pp. 342-351, 2010.

- [10] L. O’Gorman, “Overview of fingerprint verification technologies”, Elsevier Information Security Technical Report, vol. 3, No. 1, 1998.
- [11] G.B. Iwasokun, O.C. Akinyokun, B.K. Alese, and O. Olabode. “Fingerprint Image enhancement: Segmentation to thinning”, International Journal of Advanced Computer Science and Applications, vol. 3, no. 1, pp. 15-24, 2012.

# Towards the Analysis and Prediction of Early Childhood Development Using Data Mining Classification Techniques

<sup>1</sup>I. Aimufua, <sup>2</sup>S. Rakshit, <sup>3</sup>N.R. Vajjhala & <sup>4</sup>O.B. Longe

School of IT & Computing

American University of Nigeria, Yola, Nigeria

[ikponmwosa.aimufua@aun.edu.ng](mailto:ikponmwosa.aimufua@aun.edu.ng); [sandip.rakshit@aun.edu.ng](mailto:sandip.rakshit@aun.edu.ng) [narasimha.vajjhala@aun.edu.ng](mailto:narasimha.vajjhala@aun.edu.ng);

[olumide.longe@aun.edu.ng](mailto:olumide.longe@aun.edu.ng)

**Abstract** - *In recent years, the use of technology in almost every sector has been on the increase, which has led to massive growth in the generation and usage of data. Medical, Business and other leading industries have surveyed and noted that data repositories will be a useful tool in the designing of business strategies, analyzing of unstructured and structure data in order to gain useful knowledge.*

**Keywords:** *Early Childhood Development Analysis and Prediction Using Data Mining Classification Techniques*

## I. INTRODUCTION

In the educational domain, especially the early childhood education, there are very rich information available that we could leverage on to determine the development stage of a child and hidden patterns of a child learning ability or disability, but we are faced with weak knowledge and tools which are needed to harness this potentials. This study focuses on investigating which data mining classification technique the will be most suitable in building a predictive model that can identify the social, cognitive and emotional stage of a child. Through massive literature review, it was found that EDM has started gaining traction as researcher are now integrating learning management system(LMS) with teaching module in order to gain useful insight. As the education data mining get larger into early childhood education by so doing having different characteristics, different data mining technique will have their own predictive efficiency [1].

In this study, we will comprehensively study and compare different data classification technique and their prediction accuracy for early childhood dataset. Authors compared J48, Naïve Bayes, Random Forest, SVM and k-NN classifiers using performance measure like kappa statistics, ROC, RMSE and MAE using a data mining analytical tool called WEKA. Author also compared their accuracy measure like TP rate, FP rate, precision, recall and F measure. The result show from all analysis indicate that J48 classifier has a better classification accuracy and prediction rating over other tested algorithm using early childhood data set.

### A. Background of Study

Early childhood can be classified as the stage where major transition in infant occur which is a very significant influence on the development of the child and the transition to middle childhood stage, late childhood stage and then adolescence. According to [3] early childhood education should aim at holistic development of a child's Social, emotional, cognitive and physical need in order to build a foundation for lifelong learning and wellbeing.

In low and middle income countries, there are more than 200 million children under the age of 5 that have not gotten to their full developmental potential due to certain factors like poverty, poor nutrition and other factors [2].

Research has proven that the experiences of a child in their early childhood have a long term effect on the psychological, cognitive and general brain function.

Also, the exposure of a child to an environment that is not so conducive for upbringing during the first few years of their life lowers the child intelligent quotient (IQ), can also lead to poor academic achievement, increased anti-social behaviours and lowers economic productivity in their adulthood [4].

A child's characteristic of behaviour and self control in his developmental stage (age 1-3) predicts psychopathology [5]. Hence, the proper development of early childhood is the physical, mental, behavioural, cognitive and spiritual foundation of a well productive adulthood.

In [6] factors which are associated with resilience in early childhood development were listed listed to be the connection to competent and caring adult, cognitive and self regulation skill, a self positive image and finally motivation. The presence of these factors in early child development can influence several aspect of an overall child's functioning, which include the social behaviour, the academic performance and mental health

Data Mining is a multi disciplinary field that is gaining a lot of focus as it is used to know or identify hidden information or pattern from a large dataset. The field of data mining has gotten a whole lot of traction because of its analytical capabilities as it is been applied to several research domain which includes Medical prediction and Health care analysis, Social Network Mining, Business analysis/Intelligence and Education [7]

In the educational domain (early childhood education) there are several rich information available but we are faced with a very weak knowledge and tools which are needed to identify hidden pattern for early childhood development [9] [9]. Data Mining technique can be applied to the field of Education especially the early childhood education to help build predictive models that will be able to identify the current developmental stage such as the social, cognitive and emotional stage of a child and the support the child needs and be able to build systems for early prediction and correction of behavioural pattern of a child

Data Mining is a field in computer science that involves several computational processes, artificial intelligence, statistical machineries and tools which is used for extracting useful information and for pattern discovering. Mining data from early childhood education can lead to better understanding on the teaching method a child need, Data mining could be used in developing a predictive model for the educational ecosystem that collects, process and review hidden information or pattern of a child which can help in predictions and classification's of childhood challenges to better improve the early childhood developmental process.

## II. PROBLEM STATEMENT

### A. Problem Statement

Most study on data mining has always focused on higher level of education, medical analysis and prediction and lastly business prediction model development and analysis, but little or no study have been done in developing and analyzing a predictive model for early childhood development dataset.

This study will focus on evaluating different classifiers technique with different key performance indicators targeted at early childhood development dataset to determine which classifier has a better prediction and a more accurate performance in developing a data mining model for early childhood development [10].

## III. RESEARCH QUESTION

In trying to understudy the educational data mining of early childhood, we came up with this approach on

- Which classification technique (J48, Random Forest, Naïve Bayes, SVM, and KNN) will be most appropriate in developing a predictive model for early childhood education system in Nigeria?
- What are the relevant factors that influence early childhood development in Nigeria?

#### IV. RESEARCH THRUST

The research aims to analyze and predict the early childhood development using data mining classification technique. The objective is to investigate which data mining algorithm is more accurate in developing a data mining model for early childhood education. The specific objective of this proposal includes

- To investigate how close the classification technique predictions is to possible actualization
- To investigate the technique which are suitable to predict classes of needed attention in Early childhood education dataset
- Implementation of the predicted model algorithm using WEKA to analyze Early childhood education dataset

##### *A. Scope of Study*

The emphasis of this research will be to identify the classification technique and analyze the algorithm best suited for model development given a specific dataset of early childhood education. We will be using some indicators as performance measurement like Accuracy for value prediction, MAE (Mean Absolute Error) to check how close the prediction is to possible outcome, RMSE (Root Mean Squared Error/Deviation) to differentiate the predicted value and observed value, K-statistics to compare the accuracy of the system to random numbers and ROC (Receiver Operating Characteristics) for graphic representation of performance.

##### *B Justification of Research*

The Justification of this research is to exploit classification algorithm available in data mining to provide an accurate and precise performance rating for each algorithm given a dataset for early childhood development. Investigating algorithm with best performance will provide insightful knowledge for data mining model development for early childhood education in Nigeria

#### V. RESEARCH AGENDA

We will review literature on data mining analysis on two different research domain; medical domain and educational domain. Application of data mining classification technique for the virtualization of hidden pattern and the comparison of different classifiers to indicate which have a better accuracy will be discussed in this chapter. Furthermore, the research gap, research framework underpinning data mining in the educational domain will also be examined.

A detail of the research method to be used will be outlined. Also, the research approach, the theoretical background of the thesis will be extensively discussed. The method in which data is to be collected, the attribute needed and the tool that is needed to analyze the data will be discussed. We will then proceed to the descriptive analysis of the data collected. The experimental result from the analysis is to be extensively discussed. We will then come out with the result obtained. Also, the chapter concludes and gives recommendation for further research.

#### REFERENCES

- [1] Agresti, A., & Finlay, B. (2009). Statistical Methods for the Social Sciences. Pearson Prentice Hall.
- [2] Alaa, T. (n.d.). Classification assessment methods | Elsevier Enhanced Reader. <https://doi.org/10.1016/j.aci.2018.08.003>
- [3] Ambili, & Afsar. (2016). A Framework for Child Development analysis and Learning Disability Prediction using a Hybrid Naive Bayes and Decision Tree Fusion Technique – NB Tree. <https://www.semanticscholar.org/paper/A-Framework-for-Child-Development-analysis-and-a-%E2%80%93-Ambili-Afsar/13b46ecaf0fb49ecc19cb31ddbe2228053739e3b>
- [4] Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, 60(3), 383–398. <https://doi.org/10.1007/s11423-012-9235-8>



- [6] Ankita, N., & Anjali, R. (2017). Analysis of Student Performance Using Data Mining Technique. [www.ijircce.com](http://www.ijircce.com)
- [7] Aqra, I., Herawan, T., Ghani, N. A., Akhunzada, A., Ali, A., Razali, R. B., Ilahi, M., & Choo, K.-K. R. (2018). A novel association rule mining approach using TID intermediate itemset. PLOS ONE, 13(1), e0179703.  
<https://doi.org/10.1371/journal.pone.0179703>
- [8] Baker-Henningham, H. (2014). The role of early childhood education programmes in the promotion of child and adolescent mental health in low- and middle-income countries. International Journal of Epidemiology, 43(2), 407–433.  
<https://doi.org/10.1093/ije/dyt226>
- [8] Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. Archives of General Psychiatry, 53(11), 1033–1039.  
<https://doi.org/10.1001/archpsyc.1996.01830110071009>
- [9] Chakrabarty, A., Mannan, S., & Cagin, T. (2016). Chapter 8—Inherently Safer Design. In A. Chakrabarty, S. Mannan, & T. Cagin (Eds.), Multiscale Modeling for Process Safety Applications (pp. 339–396). Butterworth-Heinemann.  
<https://doi.org/10.1016/B978-0-12-396975-0.00008-5>
- [10] Chen, M.-S., Han, J., & Yu, P. S. (1996). Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 8(6), 866–883. <https://doi.org/10.1109/69.553155>

# Development of One-Time Password (OTP) for ATM/POS : A Review of Related Works

S. Jurgene & O.B. Longe  
Department of Information Technology  
National Open University of Nigeria  
Yola, Nigeria  
[sarahjuergenjr85@gmail.com](mailto:sarahjuergenjr85@gmail.com)

**Abstract:** *This study is aimed at designing a One-Time Password (OTP) for ATM/POS. The emergence of the Internet Banking leads to the Introduction of Plastic payment cards that provide a suitable and secure medium which people conduct a variety of financial transactions. But with this exciting innovation, it has also led to crime opportunities called “PLASTIC FRAUDS”. Notwithstanding, this crime must be controlled and that is why the banks must adopt a preventive and proactive method to fight the crimes. This calls for a One-Time Password (OTP) for ATM/POS. In this paper, we reviewed Related Literature and set the agenda for the research.*

**Abstract- One-Time Password (OTP), ATM/POS, Security, Online Banking, Protection**

## I. INTRODUCTION

Plastic fraud is defined as the use of plastic payment cards such as Debit Card, Credit Card or ATM Card information to perform a transaction without the knowledge or the permission of the Owner, or the issuer. [1]. To prevent Plastic Payment Card Fraud, attention must be paid to different areas and these include the technical and non-technical methods that can be adopted. Varieties of researches have been conducted in the field of detection and prevention of Plastic Payment Card Fraud. Other security techniques have been proposed to preventing Plastic Fraud; also, the One Time Password Technology has been implemented on other areas to ensure security.

## II. RELATED WORKS

[2] Carried out a study on Nigeria and their study revealed that (E-banking is still at the infant level in the country with most of the banks having mainly information sites and providing little Internet transactional services. However, most studies in these areas revealed that there has been a very steady move away from cash as transactions are now being automated. Crime and corruption represent a major concern for business executives not only in Nigeria but also in other parts of Africa. In Nigeria, the most serious problem to economic activities and business are crime and corruption which averages 75% and 71% respectively [3]

By definition; cybercrime may be referred to as any form of misconduct in cyber space. It is simply defined as the criminal use of the Internet. Cybercrime is believed to have started in the 1960's in the form of hacking. This was followed by privacy violations, telephone tapping, trespassing and distribution of illegal materials in the 1970s. The 1980s witnessed the introduction of viruses. The fast pace of development of ICT from the 1990s till today has added to the list of criminal exploits in cyber space. Today, the Internet is used for espionage and as a medium to commit terrorism and transnational crimes) [4]. [8] [15]

With (E-banking gaining ground in Nigeria and other parts of Sub-Sahara Africa, customers and online buyers are facing great risk of unknowingly passing on their information to fraudsters. "Hackers" get information of those who have made purchases through websites and then make fake cards, which they use with less detection. Absence of a law specifically dealing with card-related crimes in Nigeria may be giving thieves a loophole to operate freely. Finally, they provide insight into how cybercrime impacts on E-banking from a Nigerian perspective using social theories to explain causation with a view of guiding policy makers on behavioral issues that should be considered when formulating policies to address cyber-criminal activities in Nigeria [2] [6] [7] [14]

[5] defined computer crime as "any criminal activity involving an information technology infrastructure: including illegal access or unauthorized access, illegal interception that involves technical means of non-public transmissions of computer data to, from or within a computer system; data interference that include unauthorized damaging, deletion, deterioration, alteration or suppression of computer data; systems interference that is interfering with the functioning of a computer system by inputting, transmitting, damaging, deleting, deteriorating, altering or suppressing computer data; misuse of devices, forgery (ID theft), and electronic fraud)".

The authors categorized cybercrime in Nigeria into two, Crimes that target computer network or devices directly and Crimes facilitated by computer network. conducted an investigation and review criminal laws in Nigeria also they investigated on cybercrime and its socio-economic consequences and the damages on the image of Nigeria, they also analyze the activities of internet fraudsters also known as "Yahoo Yahoo boys" in Nigeria who uses false pretense to extort money from unsuspected victims.

Furthermore, the authors highlighted the point that (as the time of writing the paper) there is no specific law in the country to combat cybercrime, which makes the country a safe haven for the criminal to operate freely. However, they examine already existing laws that though they are not directly related to cybercrime, but can be used to limit it to some extent.

Some of these laws are examined including the section 418 and 419 of the Nigeria constitution. The authors concluded that Cyber Crime is a threat to the economy and they also made recommendations which include:

The introduction of Cyber Police who will be trained to handle cybercrime related issues in the country. The introduction of a central computer crime response wing to serve as an agency to hint the government, organizations and other investigative agencies and guide in investigations.

Lastly the authors recommend the introduction of a resource center which will comprise of expert and professional on cyber-crime and related issues to establish rules and standards and also guide procedures.

Bhasin (2007) identified that business is subjected to (crime with the advent of computers especially the internet. He noted that cybercriminal uses information technology to perpetrate the crimes. Furthermore, he highlighted that the banking sector comprises of both public and private sector and also foreign banks not to forget small or regional and co-operative banks. All these banks use various Information Technology resources e.g., ATM, phone banking etc. with these cybercrimes present a high risk to the financial institutions. He further discussed the commonly high-tech crimes perpetrated against banks as he outlined Phishing, Identity Theft, Worms and Trojan horses, Spyware, Internet search engines, Blackmail and Denial of Service (DOS) or Distributed Denial of Service.

He further explained Phishing as (masquerading an illegitimate website to make it look like the website the victim is banking with to collect or steal customer details. He identifies Identity Theft as another major problem which is also related to phishing giving the definition as "Manipulating or improper accessing another person's identity information" in other to fraudulently establish a claim over the account benefit.[8]



AMERICAN  
UNIVERSITY  
OF NIGERIA



IEEE  
COMPUTER  
SOCIETY

IEEE Computer Society Technical Paper Series

in Collaboration with

The 26<sup>th</sup> iSTEAMS Bespoke Multidisciplinary Conference &  
The School of IT & Computing  
American University of Nigeria, Yola

According to the author, (Worms and Trojan are significant threats to banks, he defines worms as “a program (or algorithm) that replicates itself over a computer network and usually performs a malicious action, such as using up the computer’s resources and possibly shutting the system down”. He relates the activities of worm to that of a computer virus, but Trojan on the other end doesn’t replicate but can be destructive also, Trojan conceal virus and also spywares like key logger.[8]

The author describes spyware as “Any software that covertly gathers user information through the user’s Internet connection without his or her knowledge, usually for advertising purposes” and they range from harmless pop-up to ability to record any activity on a computer and transmit it remotely to the hacker. Internet search engines such as Google can be used to pull out sensitive information out a website e.g., credit card details, admin login pages etc.

Finally, he talked about the risk of (Denial of service attack against online banking, Denial of service does harm by bringing down computer or the network. Distributed Denial of service occurs when the attack is launched simultaneously with various innocent client computers against a computer system or network. He emphasized that there are no other option banks can do to curb cybercrime than to be proactive; more than thirty (30) percent of successful hacks are committed by employees or in-house worker.)

He recommended that the first line of defense should start with senior management not the Information Technology because implement policies would not be a cure at all. Then categorized a comprehensive approach to physical, technical and administrative security control as follows; preventive, detective, deterrent, recovery.

The author identified Risk assessment as to direct the rest of action and lead to effectiveness if properly carried out, implement prevention techniques, policies and tool. Also have a sound business recovery plan in policies and procedures in case of successful attack.

And lastly, he also recommended that banks should develop incident response plan as part of policies then educate employees through training and seminars. Lastly bank should use specific Information Technology system to countermeasure and help mitigate crimes e.g., Fraud Detection System. [9] in their paper titled “Towards Ameliorating Cybercrime and Cyber Security” The methodology employed by the authors in performing the research was collection of data which include the use of questionnaire, personal interviews, Observation and so on. They analyzed the gathered information and make some recommendation towards making the cyberspace a safer place.

The authors describe “cybercrime as wreaking havoc on computer data or networks through interception, or destruction of such data. It also involves committing crime with the use of computer system or against them. They categorized cybercrime into three; Cybercrime against person/individual, Cybercrime against property and Cybercrime against government. They outline the causes of these crimes which include the sake of been recognized, another reason is the zeal to make quick money by the hacker and also using cybercrime to fight for a cause in which the hacker believes in”

They suggest “cybercrime can be eradicated by first identifying the challenges of already existing system; they suggest investment in education and harmonization of international cooperation and law and encourage coordination and cooperation between national law enforcement agencies”

The authors further identify the type of people who are involved in cybercrime as Idealist who are young people between the ages of 13-21 and their motivation is just to be in the spotlight of the media. The other type of people involve in cybercrime are the Greed Motivated hackers, who are very dangerous because they are ready to commit any crime so far it will mean making money; and the last set as the Cyber terrorist, the most dangerous and their aim is not just money but to stand for what they believe is just. Their main target is mainly government.

They concluded by alerting that cybercrime and cyber security must be a great concern to all government in the world and countries who neglect or fail to tackle it swiftly will suffer great consequences.

[11] in his paper entitled “Automated Teller Machine (ATM) Frauds in Nigeria: The Way Out” emphasize the objectives of the paper as to (scrutinize various ATM frauds in the country and to provide solutions to mitigate the fraud in the banking industry, the methodology that is been used to carry out the research was sampling 5 banks randomly from the 25 banks. The sampled banks are First Bank, UBA, Union Bank, Guarantee Trust Bank, and Zenith Bank. Questionnaires were distributed to 50 customers per sampled banks in Abuja.

A Scale of 5-points was used to measure the level of agreement or disagreement by the respondents. Frequency distribution was used to analyze the data collected and examined the pattern of response to each variable under investigation the study seeks to investigate the dimensions of ATM frauds in Nigeria, the frequency counts and percentages were used to capture the responses of the respondents. From the above gender distribution of respondents, 52% of the respondents were males, while 48% were females. From the respondents' age classification, 40.8% of the respondents were within the age bracket of 31-40 years while 31.2% of the respondents were within the age bracket of 41-50 years. In other words, 72% of the respondents were youths whose ages range between 31 and 50 years.

This is an indication of the level of literacy of the respondents. 56% of the respondents were married, 28% were single and 10% and 6% were divorcee and widows / widowers respectively. Students, civil servants and self-employed business men and women fall into the categories of the singles and the married which constitutes about 84%. The level of education of the customer in all the 5 sampled banks are as follows: 46% of the sampled customers have tertiary education, 36% have secondary education while 10% had primary education. 8% of the sampled customers were illiterate three dimensions featured prominently in the level of agreement and disagreement on dimensions of ATM frauds in banks.

The three prominent dimensions are ranked in the Dimensions that are 20% and above are card jamming, shoulder surfing and stolen ATM cards. The three constitute about 65.2% of ATMs fraud cases in Nigeria, 80 respondents (32%) favored video surveillance as a method of checkmating the ATM frauds, 50 respondents (20%) supported setting withdrawal limit while 40 respondents amounting to 16% supported remote monitoring. 14% of the respondents believed that customers' awareness is very central to checkmating ATM frauds. Many customers have received text messages from hackers asking them to send their pin codes. He concluded by noting that every nation has a peculiar ATM fraud that is common to it.

The e-banking has great possibilities but that would be dependent on the extent to which the ATM frauds are controlled. There are many other products that are ATM related that have been developed in developed countries. For such products to have a hold in Nigeria, the ATM fraud-related problems must be solved. Such products are electronic fund transfer at the point of sale and electronic card products. Recommendations were made to both banks and customers to curb this crime”

[11] in his paper entitled “A Five Way Fuzzy Authentication for secured banking” proposed to combine the use of Pin Number along Keypad ID, RFID Tag, Fingerprint. Image, One Time Password generated to users' phone. and another One Time Password given by the user to the server for authentication to secure Banking, this was related to the security issues associated with the existing three factor authentication protocols, which makes use of the RFID, Pin number and Biometrics. This proposed system makes use of RFID, Radio Frequency Identification which is a wireless non-contact radio system to transfer data from a tag attached to an object for the purpose of automatic tracking, it makes use of a five-factor authentication which includes the OTP and keypad ID as additional authentication factor. It makes use of fingerprint recognition technologies to analyze global pattern schemata on the fingerprint along with small unique marks, (minutiae).



The system works : the client insert card into a card reader, then input the pin and his/her fingerprint, if the inputted pin and fingerprint matches the transaction is allowed, if the pin doesn't match, the client cannot proceed, if only the pin matches and the fingerprint didn't not match up to 60% the fuzzy logic will be applied and the system will generate an OTP automatically and sent to the real user's mobile number using RSA algorithm, the generated OTP is then inputted by the user with the keypad I'd, if it matches perfectly transaction is allowed, else rejected

The analysis shows that the work satisfies all security requirements on five factor authentication and has several other practice-friendly features. [12] in their paper entitled " two factor Authentication Using Mobile Phones" examine the problem associated with static password, as user tends to write them down on paper or store them, some users uses the same password for multiple accounts and some password are easy to guess, in addition to hackers' techniques to steal password like, sniffing, snooping, dictionary attack, etc.

They therefore propose the use of two factor authentication which is a mechanism which uses a mobile based software token system that will supposedly replace the existing hardware and computer-based software tokens, the proposed system according to them is secure and it's made up of 3 parts: software installed on client's phone, a server software and a GSM modem connected to the computer.

The system work on two (2) mode of operation which is the "Connectionless Authentication System' which generates a One Time Password without connecting to the client server. The mobile phone acts as a token and uses certain unique factor to generate the OTP locally. The client may use the password online or on ATM, the Second Mode which is the "SMS Based Authentication" works in case of failure in the first mode or the password is rejected, in this mode the mobile phone request for the OTP from the server by sending via SMS, a unique information to the client and the server verifies this message content, if correct it generate the OTP and sends it back to the originating phone number, with a time limit, all these messages are subjected to charges.

In generating the OTP, the algorithm makes use of the IMEI number of the phone, the phone number, the Pin and the timestamp all concatenated and hashing the result with SHA-256 which returns a 256bits message. It's then XOR-ed with 256character, with a Base64 encoded that yields a 28-character password. They have design for the client, database and server for implementation and when tested with different method, they got 100% accuracy in the randomly generated number.

[13] in their paper entitled "Enhancing ATM security using fingerprint and GSM Technology" highlighted that there is need to improve security in ATM transactions due to the increase in criminal activities. They proposed a system which will add to the already existing method of using PIN a fingerprint enrollment, and a GSM technology connected to the microcontroller which sends a 4-digit code to the user.

The system consists of three validation functions, it first validates the pin number then the fingerprint, before the sending the 4-digit GSM modem to the phone number of the user. They did a survey collecting data on the effectiveness of the system with the result being recorded at twenty (20) people reported it has normal, fifty (50) people reported it good and seventy-five (75) people reported it as the best. This survey was carried out among thirty-five (35) professors, twenty-five (25) students, thirty-five (35) bank employees and twenty-five (25) government workers.

### III CONCLUSION

With all this reviewed literature, it is a certainty that Plastic Payment Card has come to stay and its being generally accepted worldwide including Nigeria, with the Implementation of the Cashless economy policy, it is obvious that the usage plastic payment card will definitely increase which will lead to more crime rate and the need of these crimes to be curb, thus leading to the proposal of this software which serves as prevention for Plastic fraud.



All the methods reviewed (both proposed and already in use) are all very exciting, the reason to opt for using One Time Password to improve security is because of the following reasons:

- i) Using the Biometric Technology (such as Fingerprint, Retina) will come with extra cost in purchasing devices for enrollment of the biometric trait, while OTP will make use of the already existing GSM technology in banks.
- ii) Also, the Biometric Technology will limit some Nigerians that are not used to going to ATM or other terminals to perform transaction but instead send people they trust (child) to perform the transaction. This may discourage them from using plastic payment card which will hinder the cashless economy policy. But with the OTP, it is still possible for them to send people they trust, all they need to do is to give them their mobile phone.
- iii) Implementing the biometrics together will consume a lot of time which will not go down well with the Nigerians because of the long queues that are evident on the ATM terminals across the country

#### REFERENCES

- [1] Moon, D., Flatley, J., Green, B. & Murphy, R. (2010): Acquisitive Crime and Plastic Card Fraud: British Crime Survey, Home Office Statistical Bulletin.
- [2] Wada, F. and Odulaja, G.O. (2012); Assessing Cybercrime and its Impact on E-Banking in Nigeria using Social Theories. African Journal of Comp & ICTs. Vol 5. No. 1. pp 69-82.
- [3] National Fraud Authority (2012); Annual Fraud Indicator, United Kingdom, March 2012
- [4] Olasunkanmi, O. O (2010): "Computer Crimes and Countermeasures in the Nigeria Banking Sector" journal of internet banking and commerce. 15(1) pp 1-10
- [5] Ehimen, O. R. and Bola, A. (2009): Cybercrime in Nigeria. Business Intelligence Journal- January 2010 Vol3 No.1 pp 93-98.
- [6] Longe, O.B & Danquah, P. (2012): Mitigating Socially Engineered Cyber Deception and Theft: An Ethnographic Approach. Information Technology in Developing Countries, IFIP Publication - Vol.22, No. 3. [www.iimahd.ernet.in/egov/ifip/nov2012/longe.htm](http://www.iimahd.ernet.in/egov/ifip/nov2012/longe.htm)
- [7] Longe, O.B. & Wada, F. (2012). Action Speaks Louder than Words - Understanding Cyber Criminal Behavior Using Criminological Theories. Journal of Internet Banking and Commerce. Vol. 17, No. 1.
- [8] Longe, O.B, Lawal, B.O & A. Ibitola (2014): Strategic Sensor Placement for Intrusion Detection in Network-Based IDS. I.J. Intelligent Systems and Applications, 2014, 02, 61-68. Published Online January 2014 in MECS (<http://www.mecspress.org/>). DOI: 10.5815/ijisa.2014.02.08 (Indexed in Scopus, IJIS)
- [9] Ayofe, A. N. and Oluwaseyifunmitan, O. (2009): Towards Ameliorating Cybercrime and Cybersecurity: International Journal of Computer Science and Information Security. Vol.3, No.1, 2009.
- [10] Adeoti, J. O. (2011): Automated Teller Machine (ATM) Frauds in Nigeria: The Way out: Journal of Social Science, 27(1): pp 53-58.
- [11] Jenifer Raja Shermila: A Five Way Fuzzy Authentication for Secured Banking: International Journal of Engineering Research and Application, Vol.2 Issue 4. July 2012, IJERA, pp 375-379.
- [12] Aloul, F. S., Zahidi and W. El-Hajj (2012) "Two Factor Authentication Using Mobile Phones".
- [13] Padmapriya, V. and Prakasam, S. Ph.D. (2013): Enhancing ATM security using Fingerprint and GSM Technology: International journal of Computer Applications (0975-8887), Vol. 80 No.16. pp 234-238
- [14] Nigeria Deposit Insurance corporation 2012; "Annual Reports", Nigeria, December 2012
- [15] Wikipedia, the free encyclopedia, on One-Time password.

# Design and Implementation of an Automated Train Transportation Ticketing System for Kaduna to Abuja Train Stations

A.B. Mustapha & Abdullahi A.B.  
School of IT & Computing  
American University of Nigeria  
Yola, Nigeria

[ahmad.mustapha@aun.edu.ng](mailto:ahmad.mustapha@aun.edu.ng); [abubakar.abdullahi@aun.edu.ng](mailto:abubakar.abdullahi@aun.edu.ng)

***Abstract - Transportation is an easy way of moving goods, people or animals from one place to another. There are different types of means to which things can be transported but the area we are going to concentrate on is transportation by train. Rail transport is a means of transporting people and goods from one place to another via a train. Trains are vehicles that runs on rails or tracks. Ever since humans came to existence, the drive we have towards knowledge, inventions and innovation is on an unending loop. Thanks to that, we have profoundly moved forward in terms of our technological developments and thus created things that added more value and efficiency to our day to day activities. In this work, we lay out a research intention directed at designing and implementing an automated train transportation ticketing system for the Kaduna-Abuja Train Stations.***

## I. BACKGROUND OF THE STUDY

“The Train Ticket Reservation System is an online application that allow guests to access transport ticket, purchase and pay the ticket on the internet” [1] E-ticketing is a medium that allows passengers to get their tickets electronically without stress of being in a long queue. “Electronic ticketing for urban or rail public transport is usually referred to as travel card or transit pass”[4] The use of an online train ticketing system brings efficiency in the process of ticket acquisition, managing train schedule and records.

“The Nigerian Railway Corporation traces its history to the year 1898, when the first railroad in Nigeria was constructed by the British colonial government” [2] One of the first standard gauge railway modernization projects (SGRMP) is the Abuja-Kaduna rail. It has a total length of 186km and was contracted by China Civil and Engineering Construction Company (CCECC). “The construction of the Abuja-Kaduna rail line was started in February 2011 and completed in December 2014 with China’s support. Nigerian President Muhammadu Buhari officially inaugurated the line for commercial services in July 2016” [3]

The fact that train transport is the cheapest means of transportation and would be able to transport a large percentage of the population when fully utilized is the main factor of motivation. Also, the recent interest of the government in the railway sector plays a vital role in being the main motivation for this project. Our interest is to solve the problem that passengers face when going through the manual process.

## II. PROBLEM STATEMENT

Being a Nigerian, we can say that one thing is certain when it comes to buying or getting tickets manually and that is long queues. Long queues at ticket counters is almost inevitable. Traveling Stations do not usually provide adequate number of ticket counters in order to serve with absolute efficiency and effectiveness.

Another reason is the long duration of time that takes the passengers to get their ticket or time the person in charge at the counter takes to settle the passenger's change. Automated ticketing systems will resolve this issue.

**Data redundancy (repetition):** This implies that data repetitions can occur in different ways, files or the same data will be repeated in different formats. The problem faced in manual processes is that maintaining such large data is a big problem and no adequate security measures are put in place. The implementation of an automated system will help manage these problems and would require log in details in order to access the system and access would be denied when a false information is typed in.

### III. AIMS

The aim of this project is to design and implement an automated train ticketing system which will fit in with the transport system and will be cost friendly.

### IV. OBJECTIVES

The successful implementation of the system means that time taken to get tickets at the ticket counter will significantly reduce. The project objectives are as follows;

- 1 To identify the significance of the project in providing information that is useful to railway passengers and use the web based application to overcome some of the problems the current manual system brings.
- 2 To help keep customer records because running this kind of services often comes with a huge amount of data which would be really hard to sort out manually.
- 3 To build a web based application on train reservation to take care of online booking and know the train schedules.

### V. SCOPE AND LIMITATION

The system to be developed will include online ticket booking; keeping of customer records, train schedules and prices for various destinations and will have a page dedicated for customer feedback on their trips or complains they might have.

The system will not have the feature to cancel reservation; specific seat reservation, and other management issues.

### REFERENCES

- [1] Kumar & A.Gopikrishna. (2015). "Robust and Secure Online Bust Ticket Reservation System." *International Journal of Computer Science and Engineering* 157.
- [2] Nigeria, Media. (2018). Media Nigeria. April 26. Accessed November 28, 2020. <https://www.medianigeria.com/history-of-nigeria-railway-corporation/>.
- [3] Railway Technology. Accessed 12 05, (2020). <https://www.railway-technology.com/projects/abuja-kaduna-rail-line/>.
- [4] Ainin & Mohezar. (2008). "E-Ticketing as a New Way of Buying Tickets: Malaysian Perceptions." *Journal of Social Sciences* 150.
- [5] Asaad, Ayad & Hayder. (2012). "Design and Developing Online Iraqi Bus Reservation System Using Unified Modeling Language." *International Journal of Scientific Knowledge* 9.
- [6] Athukorala & Dissanyake. (2008). "Automated Train Ticket System." University of Moratuwa, Sri Lanka, Sri Lanka.
- [7] Roy Dholakia, R; Uusitalo,O. (2002). "Switching to electronic stores: consumer characteristics and the perception of shopping benefits"." , *International Journal of Retail & Distribution Management* 459-469.

