

Predicting Breastfeeding Practice of Nigerian Child Using Machine Learning and Deep Learning Algorithms

¹Jelilat Olusola Yusuf*, ²Abubakar Muhammad Miyim

¹ICT Directorate,
Federal University Dutse

²Department of Computer Science,
Faculty of Computing,
Federal University Dutse

Email: jelilatjoy@fud.edu.ng

Abstract

The accessibility and rapid growth of data with the conspicuous rise in hardware technologies have led to the development of new studies in distributed and deep learning. This work reports the prediction of breastfeeding practice of a child using machine and deep learning algorithms. Despite that exclusive breastfeeding reduces the child mortality rate caused by pneumonia and diarrhea and other benefits attached to it the percentage of mothers who practice it in Nigeria still fall short to 29% in 2018 according to Nigeria Demographic Health Survey (NDHS) compare to the global target of 70% by 2030. The aim of the study is to develop a model for the prediction of breastfeeding practice in Nigeria. This study adopted Sample, Explore, Modify, Model and Access (SEMMA) data mining methodology using 2018 NDHS dataset. Experiments were performed with ML algorithms (RF, J48, JRIP, and SVM) built with WEKA software as well as with DNN algorithm using python. Some control parameters were applied to configure the DNN model while considering the number of layers, neurons within each layer, activation function for each layer, ADAM algorithm, epochs (iterations), learning rate and percentage split of dataset between training and testing subsets. The performance of the DNN model in predicting breastfeeding practice was evaluated and compared with previous study that used ML models only. It was found that DL has a better child prediction of breastfeeding practice than ML models with accuracy of 97.9%. This could serve as a supporting tool for healthcare practitioners to support breastfeeding practice in Nigeria.

Keywords: Machine Learning, Deep Neural Network, Random Forest, Dataset, Breastfeeding.

INTRODUCTION

Breastfeeding is one of the desirable infant feed in global health to save lives, improves the development, health of individuals and also promote economic development. Every \$1 (₦315.25) invested in breastfeeding in low and middle income countries can generate as much as \$35 (₦11,033.75), in economic returns (Alive and Thrive, 2017). Furthermore only 17 percent of babies enjoy exclusive breastfeeding between zero to five months in Nigeria which is well below the 50 percent global target (Alive and Thrive, 2017). The low awareness of exclusive breastfeeding has real effects on human life, the quality of life, and national economic outcomes (Alive and Thrive, 2017).

*Author for Correspondence

In order to reduce the global childhood mortality caused by pneumonia, WHO and UNICEF made efforts to significantly reduce pneumonia-related incidence and mortality by 2025 (WHO and UNICEF, 2013). They both conducted a series of five multi-country workshops which resulted in the publication of a Global Action Plan for Pneumonia and Diarrhea (GAPPD) in 2013 so as to develop target interventions for diarrhea and pneumonia (Qazi *et al.*, 2015). As part of its strategy to protect children from pneumonia, GAPPD encouraged exclusive breastfeeding for the first six months of a child's life. However, the International Vaccine Access Center (IVAC) reported that as of 2015, only an approximate 17% of Nigerian mothers breastfed their children exclusively for the first six months of life (IVAC, 2015). There is also a very low awareness among Nigerian mothers on the ability of breastfeeding to prevent pneumonia. A study revealed that only 16.7% of mothers in Nigeria were aware that an early childhood pneumonia can be prevented by breastfeeding (Ekure *et al.*, 2013).

In recent years, machine learning has become well known in research and has been incorporated in a large number of applications, including multimedia concept, retrieval, image classification, video recommendation, social network, analysis, text mining, and so on (Pouyanfar *et al.*, 2018). Among various machine-learning algorithms, "deep learning," also known as representation learning (Deng, 2014) is widely used in these applications. The explosive growth and availability of data and the remarkable advancement in hardware technologies have led to the emergence of new studies in distributed and deep learning. Deep learning gives the healthcare institutions, policy experts and physicians the opportunity make data driven judgments that will promote patient treatment, disease control, and equally provide health care decisions (Tobore, 2019).

There have been numerous empirical studies on the various applications of a machine learning in hospital settings for prognostication (Weng *et al.*, 2017, Taylor *et al.*, 2016), the prediction of mortality in the hospital setting (Arya *et al.*, 2013), the prediction of under-five mortality (Adeyinka *et al.*, 2019) and the prediction of breastfeeding practice (Taddele *et al.*, 2020). Meanwhile, the application of deep learning is yet to be demonstrated in the prediction of breastfeeding practice. The objective of this study is to fill the gap of evaluating ML and DL in prediction and compare the result to see which performs better in predicting breastfeeding practice of a Nigerian Child and major determinants of breastfeeding practice.

The related literature reviewed shows that gaps still exist in the use of DNN in predicting the breastfeeding practice of a child. From the literature, other studies that used ML did not achieve up to 98% prediction accuracy as achieved by DNN and attributes like the number of antennal visits and mother's age which were ranked above other attributes by RF algorithms as important determinant of breastfeeding practice were not included in the previous study.

DL models are capable of solving complex problems that require discovering hidden patterns in the data or deep understanding of intricate relationships between large numbers of interdependent variables. DL unlike ML requires less preprocessing of data, the network itself takes care of feature extraction that needed to be completed by domain expert when using ML techniques.

SYSTEM MODELLING

This section discusses the stages followed while modeling, evaluating and comparing the prediction performances of ML algorithms in breastfeeding practice of a child and is also summarized in Fig. 2.1.

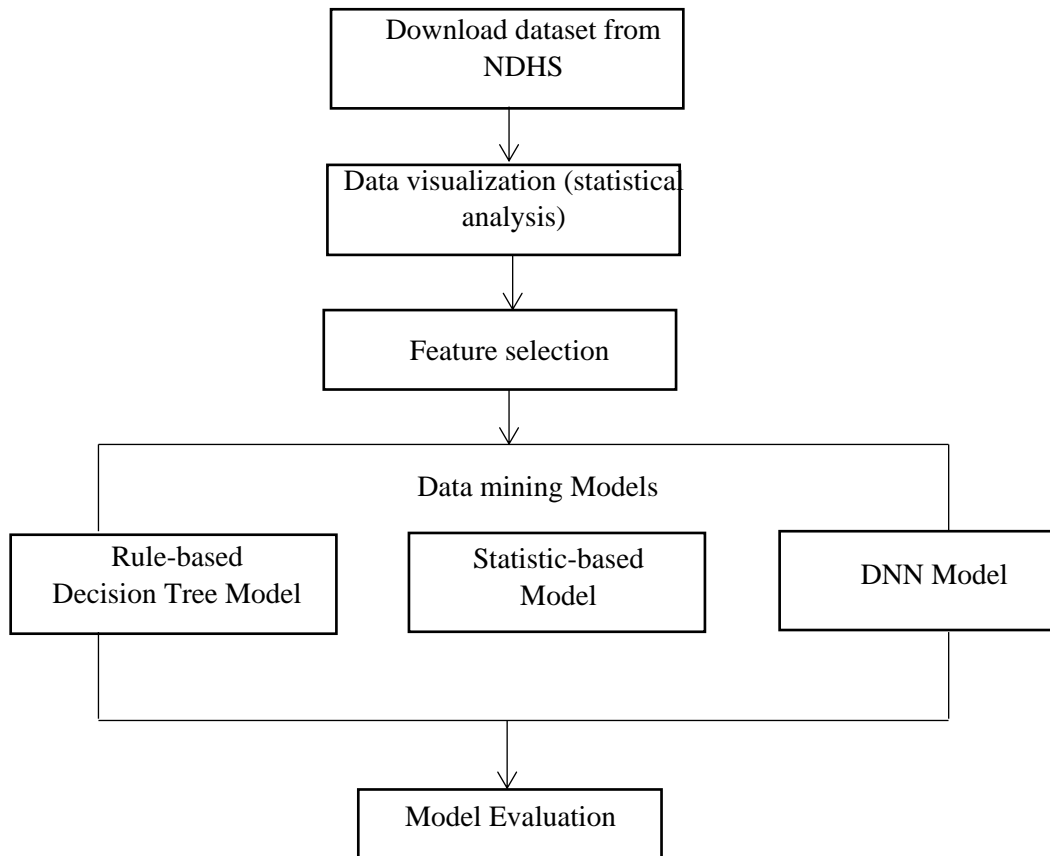


Figure 2.1: Workflow Diagram for Data Mining

There are several methodologies for carrying out data mining tasks like classifications and predictions such as KDD, SEMMA and CRISP DM. The methodology adopted for this study is Sample, Explore, Modify, Model and Access (SEMMA). It is a methodology that is easier to apply to general Data Science tasks.

Sample

The process sequence begins with data of mothers within child bearing age obtained from the 2018 Nigeria Demographic Health Survey (NDHS) repository through an official application letter to request for permission to download and use the dataset for this research and approval was granted.

Explore

The dataset was visualized with SPSS to examine the attributes and values of the dataset so as to understand the dataset during experimentation. Statistical analysis was done to obtain the relevant, high level information that is useful for the proposed modeling and also checked for the quality of the dataset in terms of error and missing values. Therefore, the attributes selected for the model building are statistically described as summarized in Table 2.1.

Table 2.1: Descriptive Statistical Summary of Socio-Demographic Characteristics

Attributes	Values	Percentage	Attributes	Values	Percentage
Place of delivery	Home	58.8	Region	North Central	17.3
	Public Hospital	27.7		North East	21.3
	Private Hospital	12.2		North West	30.4
birth weight	Below average	15.1		South East	11.2
	Average	51.7		South South	9.4
	Above average	33.2		South West	10.4
Duration of breastfeeding	Ever breastfed	67.1	Educational Status	Illiterate	45.4
	Never breastfed	3.5		Primary	15.5
	Still breastfeeding	29.4		Secondary	31.3
Wealth Status	Poor	46.6		Higher	7.8
	Middle	21.1	Currently amenorrhea	Yes	62.4
	Rich	32.3		No	37.6
Frequency of watching television	Yes	41.8	Use of internet	Yes	9.3
	No	58.2		No	90.7
Had diarrhea recently	Yes	13.2	Type of place of residence	Urban	34.5
	No	86.8		Rural	65.5
Currently pregnant	Yes	13.2	Had fever in last two weeks	Yes	22.2
	No	86.8		No	68.3
Child is alive	Yes	90.5	Births in the last five years	1 & 2 births	85.7
	No	9.5		3& 4 births	14.3

Modification

The dataset which contains 587 attributes was subjected to simple statistical analysis to certify the quality of the dataset so as to view the missing and error values and also to capture useful information about the dataset. Data on children less than six months of age were obtained and feature selection was performed using Information Gain Attribute Evaluator algorithm with a ranker search method in WEKA software to determine the relevant attributes for this study. The relevant attributes selected are Number of antenatal visits, currently amenorrhea, currently pregnant, child is alive, Respondent’s current age, had diarrhea recently, Region, Wealth status, Birth weight, the Highest educational level, place of delivery, had fever in the last two weeks, the frequency of watching television, use of the internet, the type of place of residence, and also births in the last five years accordingly. The attributes selected by the algorithm, their rank and information gain value is attached to the appendices. The harmonized data were thereafter assigned for training and testing.

Model

A short description of all the algorithms used in this study is provided in this section. These algorithms belong to the category of supervised learning methods and can be further categorized into a rule-based decision tree, statistical learning algorithms and deep neural network, as described in the following sections.

Machine Learning (ML) approach

In this study, the first set of experimentation was performed with ML of two sub-categories that is a rule based decision tree and statistical learning. The algorithms considered for decision tress are J48, RF, PART and JRip. J48 decision tree algorithm builds decision tree from a set of predefined dataset using the concept of information entropy and attribute ordering. Random forest is a great algorithm for both classification and regression problems to produce a predictive model. Its default hyperparameters return great results and it is great at avoiding

over-fitting. Decision tree algorithms and SVM algorithm for statistical learning were all evaluated for the best model for predicting breastfeeding practice in Nigeria. All the ML models were built using WEKA software with holdout (70/30) test option and compared for better performance.

Deep Learning Approach

Deep neural network (DNN) algorithms are a further development of Artificial Neural Network (ANN) which contains several hidden layers and these layers are configured with a large number of neurons. The hidden layers in ANN contain a specified number of neurons, with weights, biases as well as activation functions applied. In the DL algorithm, the output of each layer serves as input to the next layer by applying Eq. (1) starting with the input layer of the network where network is $h^0 = x$ (Lee et al., 2018).

$$h^k = (W^k h^{k-1} + b^k) \sigma^k \text{-----(1)}$$

where k = number of DNN network layers; W^k = neuron weights for the k layer; h^k = the array of output for the k layer; b^k = bias for the k layer; and, σ^k = activation function for the k layer. The final layer’s output, y , is the dependent variable’s prediction. The activation functions are used to accent the nonlinearity of relationships between input and output variables (Taqi et al., 2018). Key characteristics of activation factors are that they should be differentiable making them easier to apply particularly with gradient-based optimizers (Bengio 2009; Asghari et al., 2020). The two activation functions used in this study are as shown in Table 2.2.

Table 2.2: Activation Functions

Activation function	Formula
Sigmoid	$s(x) = \frac{1}{1 + e^{-x}}$
Relu (Rectified linear unit)	$y(x) = \max(0, x)$

There are certain control parameters that could be applied to configure DNN models which are the number of layers, the number of neurons within each layer, activation function applied to each layer, optimization algorithm, epochs (iterations), learning rate as well as percentage splits of dataset between training and testing subsets. Most of these control parameters are established by trial and error to suit specific datasets (Ng 2016). The control parameters considered while developing the DNN model in this study are listed in Table 2.3.

Table 2.3: Control parameters for the DNN model used to predict breastfeeding practice of a child

Control Parameters	Value
Number of variables in input layer	17
Number of hidden layers	2
Number of neurons in each hidden layers, 1 and 2	12, 8
Number of neurons in the output layer	1
Activation function used in the input and hidden layers	ReLU
Activation function applied to output layers	Sigmoid
Optimization algorithm	Adam
Number of iterations (Epochs)	200
Learning rate	0.01

A sequential model was created and layers added one at a time until network architecture was satisfactory as shown in figure 2.4. The input layer, which is the number of input features was 17 input variables which are all numerical values and thus makes it easy to use directly with neural networks which expect numerical input and output values. The problem is a

binary classification as the Normal breastfeeding practice was represented as 1 and Poor as 0. The model maps rows of input variables (X) to output variable (y), which is represented as $y = f(X)$.

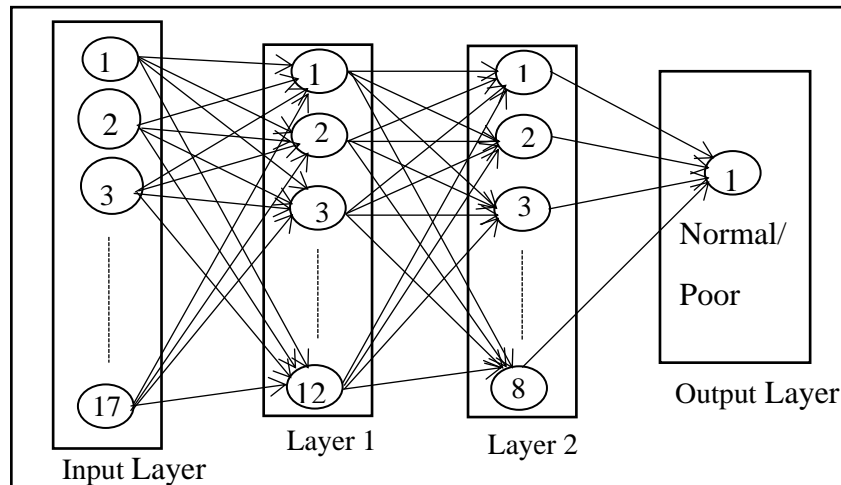


Figure: 2.4: Architecture of the DNN model for breastfeeding practice

The input variables (X) used are Place of delivery, Birth weight, Region, Frequency of watching television, Use of the internet, Wealth Status, Births in the last five years, Currently amenorrhoea, Currently pregnant, Respondents current age, Type of place of residence, Educational Status, Child is alive, Number of antenatal visits during pregnancy, Had fever in the last two weeks, Had diarrhea recently and Duration of breastfeeding while the output variables (y) is Class variable (0 or 1).

Neurons or nodes are a computational unit that has one or more weighted input connections, it connects directly to an input variable and contributes to an output variable. In this study the first hidden layer has 12 nodes and uses the ReLU activation function, the second hidden layer has 8 nodes and also uses the ReLU activation function while the output layer has only one node and uses the sigmoid activation function which is known for squashing values between ranges of 0 to 1.

The model was compiled using the numerical libraries with backend TensorFlow which chooses the best way to represent the network for training and making predictions to run on the hardware. Since training a network means finding the best set of weights to map inputs to output in our dataset, we specified binary cross entropy as the loss argument because it's a classification problem and loss function is used to evaluate a set of weights. The DNN was trained using the efficient stochastic gradient descent algorithm, ADAM optimizer which automatically tunes itself and gives good results. The training process ran for a specified different number of iterations through the dataset called epochs and will be discussed in the next chapter. The holdout approach which is based on percentage split for training and testing was adopted for the DNN algorithm, thus 70% for training and 30% for testing.

Modelling Tools

The modelling tools used for this study are:

- i. **WEKA** (Waikato Environment for Knowledge Analysis): This is free software developed at University of Waikato, New Zealand and is available under the GNU General Public License. It is a popular suite of machine learning software written in Java. WEKA

contains several visualization tools and algorithms for data analysis and predictive modelling. This software contains collection of machine learning algorithms for solving real-world data mining problems and runs on almost any platform (Sudhir & Kogbe, 2013). Weka supports several standard data mining tasks such as data preprocessing, clustering, classification, regression, visualization and feature selection (Sunita & Lobo, 2011). This package is easy to use because of its graphical user interfaces and also has the advantage of running on almost any modern computing platform since it is fully implemented in the Java programming language.

ii. **KERAS:** This is an open source neural network library written in python and capable of running on top of TensorFlow, Theano, Microsoft Cognitive Toolkit and R. It is user friendly, modular, extensible and designed to enable fast experimentation with deep neural networks (Chollet, 2019). The modelling activities carried out by Keras are Data preparation, Activation functions, Data sequence reading, Deep Neural Network implementation, Prediction and Evaluation.

This study made use of some important features available in Keras which are commonly used in neural network building blocks such as layers, activation functions, optimizers and also tools to make working with text data easier by simplifying the coding necessary for writing DNN code.

Access

This last stage focuses on the evaluation and the most common measure which is accuracy was considered for the performance evaluation of each model using precision, recall and f-measure. Precision quantifies the number of positive class predictions that actually belong to the positive class. Recall quantifies the number of positive class predictions made out of all positive examples in the dataset. F-Measure provides a single score that balances both the concerns of precision and recall in one number. It is the weighted average of precision and recall, takes both false positives and false negatives into consideration.

The formulas for calculating the metrics are as given in equations 1-3 as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{Trues Positives} + \text{False Positives}} \quad (2)$$

Where, True Positives means the number of Normal breastfeeding practice classified as Normal while they were actually Normal. False Positives means the number of Poor breastfeeding practice classified as Normal while they actually were Poor.

$$\text{Recall} = \frac{\text{True Positives}}{\text{Trues Positives} + \text{False Negatives}} \quad (3)$$

False Negatives means the number of Normal breastfeeding practice classified as Poor while they were actually Poor.

$$\text{F-measure} = \frac{2 * \text{Precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

RESULT AND DISCUSSION

In this section, the result of the performances of different ML models and DL model used for the prediction of breastfeeding practice is evaluated and discussed. The models were evaluated based on their Precision, Recall and F-measure accuracies.

Experimental Results

In experimentation conducted, holdout test option (70/30) was considered for all the algorithms, DNN with 3 layers at epoch 200 achieved the highest accuracy of 97.9%, Random

forest had the highest accuracy of 87.8% and performed better than J48 with 87.5%, JRip with 87.6% and PART had 85.3% for rule based decision trees while SVM had the lowest accuracy of 81.3%. Random forest performed better for ML model prediction of breastfeeding practice. The classifier performance for Each ML model in terms of precision, recall and f-measure are summarized in Table 3.1 and figure 3.1 below.

Table 3.1: Classifier Performance based on Accuracy Measures

Model	Accuracy	70/30 Test Options	Model	Accuracy	70/30 Test Options
J48	Precision	0.866	JRip	Precision	0.877
	Recall	0.868		Recall	0.879
	F-measure	0.866		F-measure	0.876
Random Forest	Precision	0.879	SVM	Precision	0.817
	Recall	0.881		Recall	0.818
	F-measure	0.878		F-measure	0.803
PART	Precision	0.850	DNN	Precision	0.997
	Recall	0.852		Recall	0.962
	F-measure	0.851		F-measure	0.979

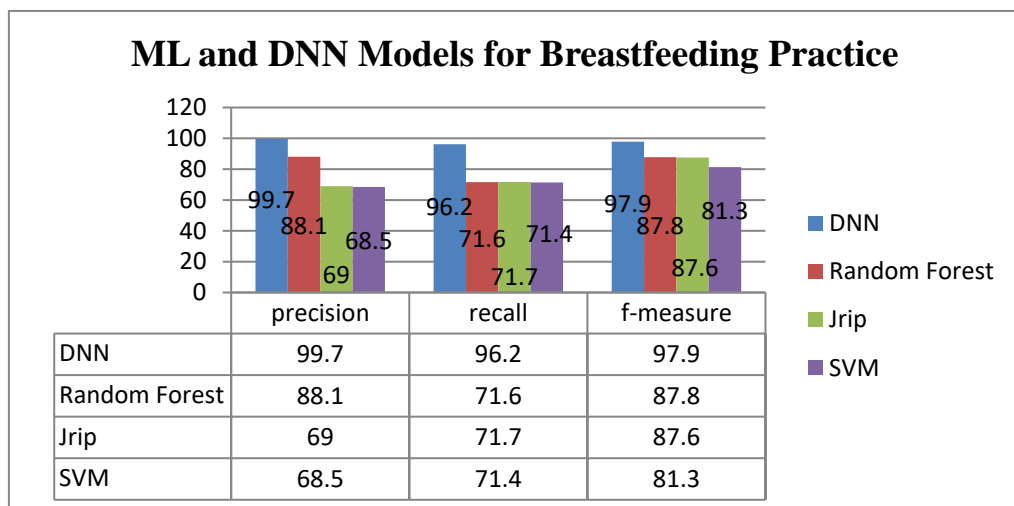


Figure 3.1: A Comparison of the ML Models

Result Evaluation

This section describes and compares the performances of various experiments carried out on ML and DNN models for predicting breastfeeding practice of a Nigerian child. Table 3.3 and figure 3.3 below shows the ML model and DNN model with the best accuracy.

Table 3.3: A Comparison of the ML and DNN models

Model	Precision	Recall	F-measure
Random Forest	0.879	0.881	0.878
DNN	0.997	0.962	0.979

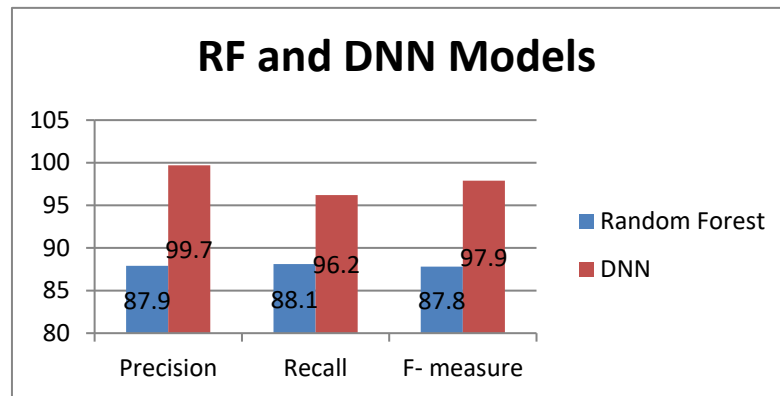


Figure 3.3: Random Forest

Compared and DNN

Models

DISCUSSION OF RESULT

To substantiate findings from this study there is need to compare it with previous study. This work has been able to predict breastfeeding practice of a child with better accuracy of 97.9% with the DNN model better than the accuracy of 96.86% achieved by previous work. Furthermore, in term of attributes that determine the breastfeeding practice of a child this study was able to detect from previous patterns in the dataset with the use of RF that mother’s age and number of antenatal visits are very important attributes that determines the breastfeeding practice of a child and this attributes are missing from previous work. Random Forest also evaluates attribute importance based on the average impurity decrease and number of nodes using that attribute, the algorithm ranked the attributes as shown in Table 3.4.

Table 3.4: Attribute Importance

Average Impurity Decrease	No of nodes	Attributes	Average Impurity Decrease	No of nodes	Attributes
0.37	13302	Respondents current age	0.29	628	Currently amenorrhea
0.35	8434	Place of delivery	0.29	2420	Use of the internet
0.35	8406	birth weight	0.28	2448	Type of place of residence
0.34	10154	Number of antenatal visits during pregnancy	0.27	2644	Had fever in the last two weeks
0.34	6794	Wealth Status	0.25	2910	Births in the last five years
0.33	5046	Frequency of watching television	0.22	1589	Had diarrhea recently
0.31	6775	Region	0.16	552	Currently pregnant
0.3	3238	Educational Status	0.09	421	Child is alive

From the above ranking this study has been able to establish the determinants of breastfeeding practice with the use of Random Forest algorithm. The result clearly ranked respondents age, place of delivery, birth weight, the number of antenatal visit during pregnancy, wealth status higher than other attributes. The two attributes added to this study which is respondent’s age and number of antenatal visits during pregnancy which was missing from the attributes used in the previous study are part of the major determinant in predicting the breastfeeding practice of a child.

This could be because mothers who attend antenatal are more aware from the sensitization by healthcare practitioners than mothers who do not. The age of the breastfeeding mother according to health care practitioners also affects milk production has said that milk yield is dependent on the amount of functional tissue in the breasts and may decrease with age due to what they called disuse atrophy.

CONCLUSION

This study has been able to prove that DL can predict breastfeeding practice of a child better than ML models with 3 layers DNN model using a different number of epochs which tuned the model to achieved accuracy of 97.9% at epoch 200. The two attributes added to this study which is respondent's age and number of antenatal visits during pregnancy which was missing from the attributes used in the previous study were ranked higher than other attributes as major determinant in predicting the breastfeeding practice of a child. Since DNN model has proven to be better in predicting the breastfeeding practice of a child, then it can be deployed as decision support model of breastfeeding practice for the healthcare practitioners. Further research on this work should implement a user interface by deploying this DNN model to mobile devices using TensorFlow mobile which will serve as user interface that can be easily used at the hospital level to predict breastfeeding practice of a Nigerian child.

REFERENCES

- Adeyinka, E. A., Bela, S., Jing, S. (2019). Predicting Under-five Mortality Across 21 Low and Middle Income Countries using Deep Learning Methods. *Level 8, G40, Griffith University School of Medicine, Parkland Drive, Gold Coast, 4222, Australia*. Pp.1-3
- Alive & Thrive, UNICEF (2017). The Economic Cost of not Breastfeeding on Human Capital Development and Health Systems in Nigeria. Bill & Melinda Gates Foundation and the government of Canada and Ireland. p1
- Arya R, Wei G, McCoy, J.V., Crane, J., Ohman-Strickland, P. *et.al.* (2013). Decreasing Length of Stay in the Emergency Department With a Split Emergency Severity Index 3 Patient Flow Model. Gerson L, editor. *Acad Emerg Med*. 2013 Nov;20(11):1171-9.
- Chollet, F. (2018). Keras: The Python Deep Learning Library. *Astrophysics Source Code Library, record ascl-1806.022*
- Ekure, E. N., Esezobor, C. I., Balogun, M. R., Mukhtar, M.Y., Ojo, O. O. *et.al.* (2013). Mothers and Childhood Pneumonia: What should the Focus of Public Campaigns be? *Nigerian Journal of Paediatrics*, vol. 40, no. 1, pp. 24-29, 2013.
- International Vaccine Access Center (IVAC), (2015). Johns Hopkins Bloomberg School of Public Health, "Pneumonia and Diarrhoea Progress Report 2015: Sustainable Progress in the Post-2015 era," 2015. <http://www.jhsph.edu/research/centers-and-institutes/ivac/resources/IVAC-2015-Pneumonia-Diarrhea-Progress-Report>. (Accessed 13, June 2021)
- Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y. *et.al.* (2018). "A survey on deep learning: Algorithms, techniques and applications" *ACM Computing Surveys*, Vol51, issue 5, pp1-36, <https://doi.org/10.1145/3234150>
- Qazi, S., Aboubaker, S., MacLean, R., Fontaine, O., Mantel, C. *et.al.* (2015). Ending Preventable Child Deaths from Pneumonia and Diarrhoea by 2025. Development of the Integrated Global Action Plan for the Prevention and Control of Pneumonia and Diarrhoea, *Archives of Disease in Childhood*, vol. 100, no. 1, pp. S23 - S28, 2015.
- Sudhir, B. J. and Kodge, B. G. (2013). Census Data Mining and Data Analysis using WEKA.

- (ICETSTM – 2013) *International Conference in “Emerging Trends in Science, Technology and Management-2013, Singapore.*
- Taddele, W.A. (2020). Prediction Model for Breastfeeding Practice among Ethiopian Children using Decision Tree and Rule Induction Algorithms: Data mining” *Research Square.* DOI: <https://doi.org/10.21203/rs.3.rs-58065/v1>
- Taylor, R.A., Pare, J.R., Venkatesh, A.K., Mowafi, H., Melnick, E.R. *et.al.* (2016). Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Jones A, editor. *Acad Emerg Med.* 2016 Mar 1; 23(3): pp. 69-78.
- Tobore, I. (2019). “Deep learning intervention for health care challenges: some biomedical domain considerations” National Library of Medicine, National Centre for Biotechnology Information, 8600 Rockville Pike Bethesda, MD 20894.
- Weng, S. F., Warren, S. M. and Walter, P. (1943). “A logical calculus of the ideas immanent in nervous activity”. *Bulletin of Mathematical Biophysics* 5, 4 (1943), pp. 115-133
- WHO and UNICEF, (2013). Executive Summary on Ending Preventable Child Deaths from Pneumonia and Diarrhea by 2025: The integrated Global Action Plan for Pneumonia and Diarrhea (GAPPD). 2013. https://apps.who.int/iris/bitstream/handle/10665/7920/9789241505239_eng.pdf (Accessed 15 July, 2021)