



An Artificial Neural Network Model for Predicting Children at Risk of Defaulting from Routine Immunization in Nigeria

Abraham Eseoghene Evwiekpaefe¹, Valerie Plangan Lawi²

^{1,2} Department of Computer Science, Nigerian Defence Academy, Nigeria
E-mail: aeevwiekpaefe@nda.edu.ng¹, lawivalerie@gmail.com²

Abstract

It has been widely recognized that immunization remains one of the most successful for decreasing child mortality rates and preventing several serious childhood diseases globally. This study proposed a prediction model for accurate identification of routine immunization defaulters in Nigeria. The proposed framework classified defaulters at five different risk stages: insignificant risk, minor risk, moderate risk, major risk and severe risk to reinforce targeted interventions by accurately predicting children at risk of defaulting from the immunization schedule. Data from Nigerian Demographic and Health Survey 2018 was obtained for this study and thirty-four (34) demographic and socio-economic factors were used to predict children at risk of defaulting from routine immunization in Nigeria by using Artificial Neural Network (ANN) to train the dataset. The results indicated that ANN model produced an accuracy of 99.16% for correctly identifying children who are likely to default from immunization series at different risk stages. Other performance measures include Precision of 99%, Recall of 99% and F1 Score of 99%. The model was further validated using one thousand (1000) dataset, out of which nine hundred and seventy four (974) were correctly predicted.

Keywords: Artificial Neural Network (ANN), Immunization, Vaccination, Immunization Defaulters.

1. INTRODUCTION

Immunization and vaccination are two of the most essential public health interventions, and they represent a cost-effective strategy for reducing both morbidity and mortality associated with infectious illnesses [1]. Immunization is a global success story in terms of health and development, saving millions of lives each year. Immunization is an essential component of primary health care and a basic human right. It's also one of the best health investments you can make with your money. According to the World Health Organization (WHO), vaccinations prevent 2 to 3 million deaths annually, but 1.5 million more might be prevented with increased vaccine uptake.



Despite the free routine vaccinations available in low- and middle-income countries (LMICs), many children fail to receive all of the recommended vaccinations, receive them too late for their age, or stop receiving them altogether [2]. With an annual population growth rate of 2.83 percent, Nigeria has the highest density of population in Africa and ranks second globally in terms of under-five mortality. 12.2 of the 20 million under- vaccinated and unvaccinated children worldwide are from only 10 countries (62 percent). This list consists of Nigeria as one of the most economical public health treatments. Vaccination coverage is one of the metrics used to track progress toward reductions in child morbidity and mortality, according to the 2018 Nigeria Demographic Health Survey (NDHS). One of the most crucial ways to reduce childhood morbidity and mortality is through vaccination against diseases like diphtheria, pertussis, tetanus, polio, and measles. Therefore, all health systems must make achieving and maintaining high levels of vaccination coverage a top priority. Data on immunization coverage can be used as an indicator of a health system's ability to provide crucial services to the most vulnerable group of a population in order to track progress toward this goal.

In Nigeria, vaccination rates have improved over the past 10 years. While the trends show improvement, they still fall short of Sustainable Development Goal 3, for which the target is achieving more than 90% coverage of all basic vaccinations among children aged 12-23 months [3]. This study aims to develop a predictive model that can identify children at risk of defaulting from routine immunization to achieve maximum immunization coverage in Nigeria.

Several studies have been conducted in recent years to assist and improve child vaccination coverage across the country. This section describes in detail the research works related to this study, such as socioeconomic factors associated with child immunization, determinants of incomplete childhood immunization, and using predictive analytics to identify children at risk of defaulting from routine immunization.

One of Nigeria's major public health issues is the partial immunization rate against diseases that can be prevented by vaccination. This is especially true in rural areas. It is poorly understood what causes incomplete immunization and what causes missed opportunities. [4] sought to identify the causes of incomplete immunization and the elements leading to missed opportunities for immunization in infants younger than one year of age. The study was conducted in Nassarawa state's Awe LGA. Vaccination status and missed opportunities for immunization were calculated by proportion. The Chi-square test with a 5% threshold of significance was used to calculate differences in proportions. To compare mean values amongst subclasses, the ANOVA test was performed. The chi-square test was used to examine associations between variables and missed opportunities or incomplete vaccination status. A binary logistic regression model was used to fit the variables together in order to evaluate their relative relevance in relation to the

dependent factors and any potential confounding. Through the use of a cross-sectional study design, mothers of children within one year of age served as the study's subjects. The immunization card was used to determine the optimal utilization of all available chances for immunization as well as to verify the accuracy and completeness of the immunization schedule. About two third (62.8%) of the children were not fully immunized by one year of age, 33.4% had experienced a missed opportunity for immunization and 36.4% were partially and incorrectly immunized. The most frequent reasons for incomplete immunization given by parents are long distance walking (17.5%), long waits at the health care facility (15.2%), and parental resistance, disagreement, or concern about immunization safety (38.8%). To increase the percentage of children who arrive at the health facility fully immunized, missed chances for immunization and incomplete immunization must be avoided, especially in rural areas where immunization coverage is lower than the anticipated national coverage (minimum 80 percent).

Numerous studies have focused on socio-economic issues to identify the causes of low vaccination coverage. To assess the effects of educational intervention for women with low literacy rates on their children, [5] conducted a study on a low-income community in Karachi, Pakistan. A Poisson regression model was applied to estimate the intervention's impact. The multivariable Poisson regression model includes the child's immunization status at enrolment, the mother's assessment of the influence of immunization on child's health, maternal education, paternal occupation, ownership of home, cooking fuel used at home, and place of residence. The intended educational intervention for less literate mothers increased the immunization completion rate of the DPT-3/Hepatitis B vaccine by 39%, as was clear from the results, and this had important ramifications for obtaining higher immunization ratios. The value of maternal education in Kenya, which is still a country with low immunization rates, was the subject of a similar study by [6]. Retrospective cross-sectional data from the Kenya Demographic and Health Survey conducted in 2008–2009 were used in the study. STATA version 13.1 was used to conduct data analysis for descriptive, bivariate, univariate, and multivariate logistic regression analyses. It was discovered that women with knowledge above the primary level are substantially more likely to complete their immunizations. Even after accounting for individual and community-level factors in model development, it was discovered that maternal literacy is statistically necessary for a child's well-being.

Aside from literacy rate, several other socioeconomic factors influence immunization coverage. [7] assessed the childhood immunization rate at the district level by considering health facilities in India as well as other contributing socioeconomic factors. The study made use of data from the 2008 Indian District Level Household Survey (DLHS-3). The findings revealed that an individual's income is strongly related to immunization coverage rate and plays an important

role in prediction. Their research found that maternal education is statistically significant and has a positive impact on immunization outcomes.

[8] conducted research to examine the factors that influence a child in the USA to be fully vaccinated. An immunization prediction model was created using information from the National Immunization Survey as well as a variety of socioeconomic and demographic factors. Based on several demographic and socioeconomic factors, logistic regression was used to determine whether a preschool-aged child in the USA today is likely to receive vaccinations. Categories of educational attainment, firstborn child, race and ethnicity, age of mother, and census region are significant variables in the model. This model confirms the importance of geography in immunization outcomes, but it does not conclusively show that children born later are less likely to get fully vaccinated than children born first. All census regions and educational levels were determined to be significant. Overall, these models show that socioeconomic and demographic factors influence children immunization rates and, if used effectively, can help policy makers and public health authorities better understand immunization rates and develop policies to improve them.

Study by [9] recognized the importance of vaccination timeliness and used routine data to assess the efficacy of immunization programs. The data includes 1782 Welsh-residing children born between 2000 and 2001 who were enrolled in the Millennium Cohort Study and whose parents approved the linking of their records to the National Community Child Health Database at the age of seven. StataSE 13 was used for all analyses. This study brought to light the fact that postponing vaccinations has a significant negative impact on children's immunity, placing them at a higher risk of contracting infectious diseases.

Study by [10] used a secondary dataset analysis of the Nigeria Demographic and Health Survey (NDHS), 2013, to identify individual and socioeconomic characteristics linked to childhood immunization coverage in Nigeria. Univariate, bivariate, and multivariate statistics were used to evaluate the dataset after it had been downloaded, checked for completeness, and validated. According to the findings, 22.1 percent of the 27,571 kids aged 0 to 59 months had completed their vaccinations, while 29.1 percent had never had any vaccinations. Immunization coverage was significantly associated with childbirth order, delivery place, child number, and presence or absence of a child health card. Immunization coverage was substantially correlated with maternal age, geographical location, education, religion, literacy, wealth index, marital status, and occupation. Paternal education, age and occupation were all strongly related to coverage. In a multivariate analysis, the respondent's age, educational level, and wealth index remained significantly related to immunization coverage at a 95% confidence level. The study reveals socioeconomic, parental, and child challenges to Nigeria's successful immunization programs. In a related study, [11] used socioeconomic features and immunization status from a survey carried out in Punjab, Pakistan, in 2011.

Secondary data from the 2011 Multiple Indicators Cluster Survey served as the study's foundation (MICS). Binary logistic regression and descriptive statistics were used in the data analysis. The availability of immunization cards, the mother's literacy level, and the location of the child's birthplace were found to be the key factors influencing vaccine coverage variation.

There were no studies that focused on the risk of partial or no vaccination affecting the entire immunization program. Furthermore, none of the previous studies used this information to improve immunization coverage or to avoid future delays [12]. The initial recognition of the application of predictive analytics and its potential to increase immunization coverage was made by [2]. A predictive analytics algorithm was created to determine which children are most likely to miss their follow-up vaccination appointments for any vaccine on the routine immunization schedule. 47,554 longitudinal immunization records that were divided into training and validation sets were used to construct the algorithm. The method that forecasts the risk of each child defaulting from the follow-up immunization visit was created using four machine learning models: random forest, recursive partitioning, support vector machines, SVMs, and C-forest. The gender of the child, the language that is spoken at home, the location of the child (town or city), the enrollment vaccine, the timeliness of the vaccination, the enrolling staff (the vaccine provider or others), the date of birth (exact or estimated), and the child's age group were all taken into account in the models as predictors of defaulting. Accuracy, precision (positive predictive value), sensitivity, specificity, negative predictive value, and area under the curve (AUC) were all evaluated for each model.

Study by [13] designed an automated platform using R Markdown to examine content from Twitter and Google Trends linked to immunization. In order to predict whether a child will receive immunizations or not, the study developed data mining models using patient survey data. It also used time series forecasts to predict trends in immunization coverage. In this study, SVM, KNN, SuperLearners, Random Forest, C-Forest, Neural Networks, Boosting, Decision Trees, Bagging, and Naïve Bayes were used to predict the immunization status of children. The study generated a visual dashboard of the online content, and the SuperLearners data mining algorithm was the most effective in predicting a child's immunization status, with 76%, sensitivity of 30.01%, and specificity of 80%. Time series forecasts have a mean absolute error of 6.83.

Study by [14] developed a hybrid model to predict and target vaccination rates in less immunized regions in India. The hybrid deep learning framework Rank-Based Multi-Layer Perceptron (R-MLP) used data from the updated District Level Household Survey-4 (DLHS). The R-MLP model predicted and classified vaccination rates for partially immunized people into extreme, low, and medium ranges. The model identified the loss values in order to identify the target regions where health care programs are needed to increase the level of immunization

among children. The proposed hybrid deep learning models were trained and validated using Python-based deep learning libraries Keras and TensorFlow. The proposed hybrid deep learning model's performance was compared to that of other variant machine learning techniques such as Decision Tree C5.0, Naive Bayes, and Linear Regression. With an accuracy of 95.58%, recall of 78.17%, and precision of 85%, the hybrid deep learning system clearly outperformed any other alternative approach.

Study by [12] proposed a predictive framework for accurately identifying children who are at risk of missing any of the vaccines on the immunization schedule. A sample dataset extracted from the Pakistan Demographic and Health Survey (PDHS, 2017-2018) with 7153 data records containing 19 demographic and socioeconomic attributes was used for predicting defaulters and the identification of association rules to understand the relationship between the child's demographics and vaccination status. Using a multilayer perceptron (MLP) classifier, the proposed model correctly identified the children who are likely to default from immunization series at various risk stages, achieving 98% accuracy and 0.994 for the area under the curve (AUC).

Study by [15] used a random forest classification algorithm to develop a model that predicts infant immunization completion rates in order to improve immunization service delivery and utilization. Using DPT3 as an identifier classified into three categories, this model predicts who is likely to complete the recommended immunization vaccines according to schedule. The study made use of existing secondary electronic immunization records data from the MyChild System implemented at the Mukono district health facility. The data used was gathered between 2015 and 2020. This model's predictors include the child's date of vaccine administration, tetanus exposure, HIV exposure of child, date of birth, and whether the caregiver was counseled. The accuracy of the model was 76%.

Study by [16] developed a predictive model that predicts defaulting and non-defaulting children in upcoming immunization visits using nine (9) ML algorithms. Random forest achieved 81.9%, 83.6% sensitivity, and 80.3% specificity. The model made use of 3113 records from the Paigham-e-Sehat study. The study found that vaccination coverage at birth, parental education, and the socioeconomic conditions of the defaulting group were the main determinants of vaccination coverage.

Predictive modeling is a novel concept in the field of immunization, and its potential to revolutionize immunization service delivery is still to be identified [12]. [13] were the first to recognize the application of predictive analytics to improve immunization coverage. However, even if a child misses only one dose, the model places that child at a high risk of defaulting. This prompted [12] to create a model that will accurately predict children at risk of missing routine immunizations by categorizing them into five (5) categories rather than the three (3)

categories proposed by [13] based on their vaccination status. Also identifying the demographic and socioeconomic factors that contribute to children missing routine immunizations is critical for public health because they are at a higher risk of contracting serious childhood infections such as pneumonia, meningitis, or measles. Identifying these factors will also allow for more focused and targeted interventions. Despite the fact that there is literature on demographic and socioeconomic factors that identify children at risk of not receiving complete routine immunization in Nigeria, to the best of our knowledge, little attempt has been made to use predictive models to support the findings above. The goal of this study is to predict children at risk of defaulting from routine immunization in Nigeria by using multiple demographic and socioeconomic factors identified in the literature.

2. METHODS

In this study, a predictive model was built using neural network techniques as shown in Figure 1. The model shows the different steps taken to develop a model that will accurately predict children that are at risk of defaulting from routine immunization.

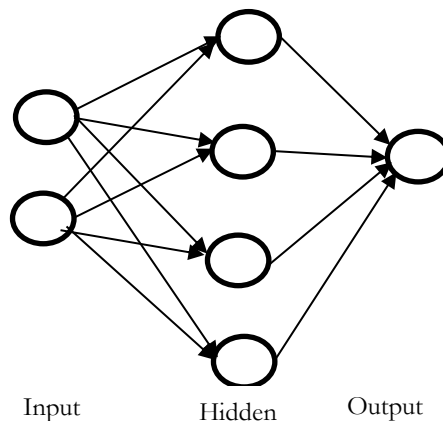


Figure 1. The ANN Architecture

2.1 The Proposed Framework

The goal of the model is to enhance classification accuracy in predicting and identifying the defaulters at different levels of risk of missing their next scheduled immunization based on demographic and socio-economic factors. This model as shown in figure 2 describes an approach where an artificial neural network was used to train a dataset which was used for prediction. The prediction is based on multiple factors (socio-economic factors and immunization data provided).

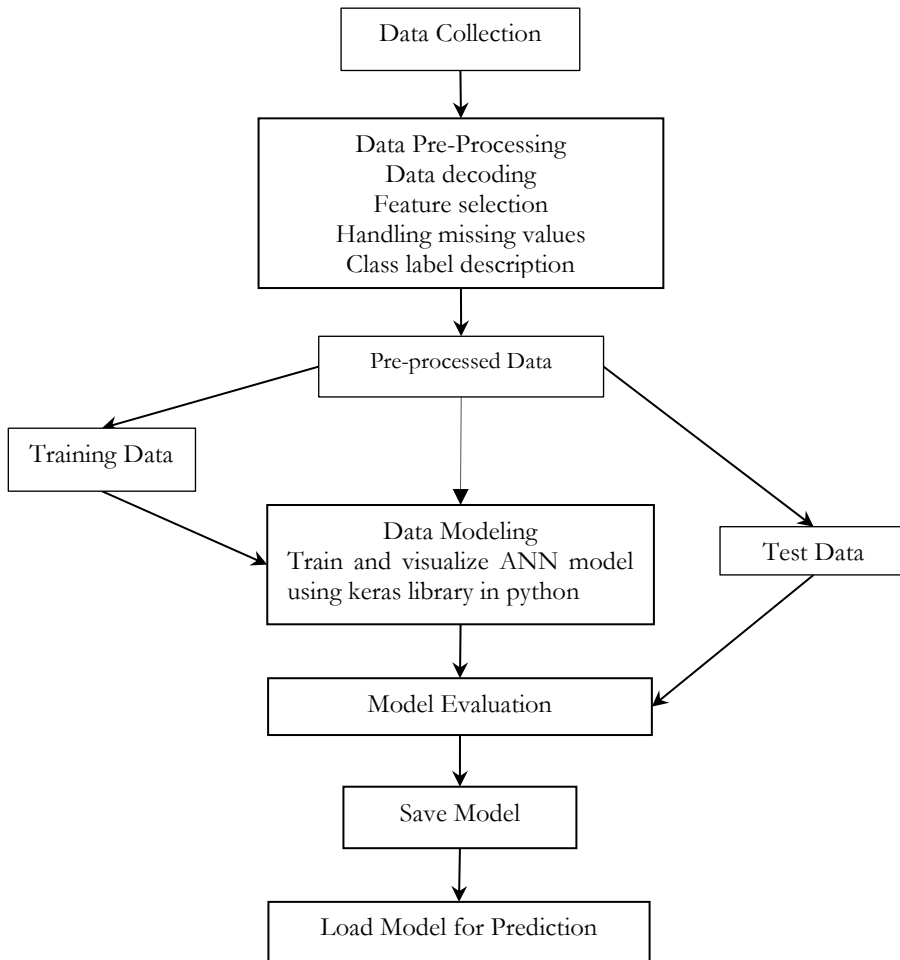


Figure 2. Proposed Framework for Predicting Children at Risk of Defaulting from Routine Immunization

2.2.1 Data Source and Research Area

Dataset was extracted from Nigerian Demographic Health Survey (2018). After submitting the study's abstract, the dataset was made accessible on the www.dhsprogram.com website. Due to the fact that the survey is conducted every five (5) years with assistance from the United States Agency for International Development, the 2018 dataset was chosen because it was the most recent survey (USAID). Nigeria was the area where the study was carried out based on the report of Nigerian Demographic report on child health.

2.2.2 Data Pre-processing

Data preprocessing is critical to preparing the data for efficient and effective model building, as unclean data has a negative impact on model performance. The following are the main steps taken in preprocessing: attribute selection, handling missing values, class label description, and data normalization.

2.2.3 Data Decoding

In this step, the numerical demographic values were used to generate nominal values. The survey dataset contains the metadata required to decode the numerical value of demographic variables. Data from the demographic and metadata variables were extracted, decoded to nominal values, and then used to interpret the variables

2.2.4 Attribute Selection

When selecting the qualities or features for the study, consideration was given to how relevant they were to the goals of neural network modelling. As a result, thirty-four (34) attributes were chosen as shown in Table 1.

Table 1. Tabular View Showing the Attributes, Contributing factor, Description and Values

No	Attribute	Contributing Factor	Description	Attribute Values
1.	Maternal age	Mother	The age of the mother at the time of delivery	15,16,17.....49
2.	Region	Mother	The different geo-political zones in Nigeria	North central, Northeast, North West, South West, South South, South East
3.	Religion	Mother	A particular system of faith and worship	Catholic, Other Christian, Islam, Traditionalist, No Religion
4.	Wealth index	Mother & Father	This is a composite measure of a household's cumulative living standard.	Poorest, Poorer, Middle, Richer, Richest
5.	Birth order	Child	The order in which a child is born	1,2,3.....21
6.	Health card	Child	The health record card that contains information about vaccination	Yes, no

No	Attribute	Contributing Factor	Description	Attribute Values
			dates and records.	
7.	Maternal Highest Education	Mother	The highest level of education of the mother	No education, Primary, Secondary, Higher
8.	Literacy	Mother	The ability of the mother to read and write	Cannot read at all, able to read only parts of a sentence, able to read whole sentence, no card with required language, blind/visually impaired
9.	Native language	Mother		English, Hausa, Yoruba, Igbo, Other
10.	Health card and/or vaccination document	Child	The health record card and/or documents that contains information about vaccination and dates and records	Does not have health card, has only health card and was seen, has only health card and wasn't seen, has only other document and was seen, has only other document and wasn't seen, has card/card/another document but only card was seen, has card/other document but only other document was seen, has card/other document and both were seen, has card/other document and none were seen
11.	Total number of children	Child	The total number of children born in a family	1,2,3.....20
12.	Ethnicity	Mother		Ekoi, Fulani, Hausa, Ibibio, Igala, Igbo, Ijaw/Izorn, Kanuri/Berberi, Tiv, Yoruba, Other, don't know
13.	Age of child	Child		0, 1, 2
14.	Place of delivery	Child	The place where the child was born	Home, health facility
15.	Attended ANC	Mother	The mothers who paid at least one visit to the clinic during pregnancy.	Never attended, attended
16.	Paternal age	Father	The age of the father	15,16,17.....59
17.	Paternal educational attainment	Father	The level of educational attainment of the father	No education, Incomplete primary, Complete primary, Incomplete secondary, Complete secondary, Higher, don't know

No	Attribute	Contributing Factor	Description	Attribute Values
18.	Paternal Highest Education	Father	The highest level of education of the father	No education, Primary, Secondary, Higher
19.	Received BCG	Child	The BCG vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
20.	Received PENTA1	Child	The Penta 1 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
21.	Received POLIO 1	Child	The Polio 1 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
22.	Received PENTA 2	Child	The Penta 2 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
23.	Received POLIO 2	Child	The Polio 2 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
24.	Received PENTA3	Child	The Penta 3 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
25.	Received POLIO3	Child	The Polio 3 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
26.	Received MEASLES 1	Child	The Measles 1 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
27.	Received MEASLES 2	Child	The Measles 2 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
28.	Received POLIO 0	Child	The Polio 0 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
29.	Received HEPATTIS B	Child	The Hepatitis B vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
30.	Received Pneumococcal 1	Child	The Pneumococcal 1 vaccination	No, Vaccination date on card, reported by mother, Vaccination marked on card

No	Attribute	Contributing Factor	Description	Attribute Values
			status of the child	
31.	Received Pneumococcal 2	Child	The Pneumococcal 2 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
32.	Received Pneumococcal 3	Child	The Pneumococcal 3 vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
33.	Received IPV	Child	The IPV vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card
34.	Received VITAMIN A	Child	The Vitamin A vaccination status of the child	No, Vaccination date on card, reported by mother, Vaccination marked on card

2.2.5 Handling Missing Values

Different missing values, such as spaces, "Null," special characters like "?" and unknown values, were eliminated in order to get good results. This is because there are only two possible options in immunization i.e., either a child vaccinated or not vaccinated for a specific disease. As a result, the missing values removed from the original set reduced the number of records from 33,924 to 15,294. The dataset was split into two parts; the first was used for training and testing the model, and it contained 14,294 records and 35 fields.

2.2.6 Class Label Description

Before passing the training set to model building, training data was labeled into five categories by classifying the children according to their vaccination status:

- 1) Children who have not taken even a single vaccination dose were labeled as unvaccinated.
- 2) Children who have taken 5 or less than 5 doses out of a total of 16 were labeled as partially low.
- 3) Children who have taken 6 to 10 out of the 16 doses were labeled as medium.
- 4) Children who have taken 11 or more out of 16 were labeled as partially high.
- 5) Children who have taken all doses of vaccinations and those who have taken the basic vaccinations are labeled as fully immunized. A "fully

immunized child" is a child who has received one dose of BCG, three doses of OPV (excluding OPV given at birth), three doses of DPT vaccine, and one dose of measles vaccine by 12 months of age.

The output/target class which is named Defaulter was coded as shown in Table 2 before the data was trained and tested using an artificial neural network.

Table 2. Defaulter Class Label

Class	Vaccination Status	Risk Rate
0	Partially Low	Major Risk
1	Fully Immunized	Insignificant Risk
2	Partially High	Minor Risk
3	Unvaccinated	Severe Risk
4	Medium	Moderate Risk

2.2.7 Data Normalization

Data was normalized using the Min Max scaler which transforms features by scaling each feature to a range of 0 and 1 thereby making the input values appear easy for the algorithm. Eighteen (18) attribute values were normalized because they have large numbers. They are maternal age, religion, wealth index, total number of children, paternal educational attainment, paternal age, birth order number, place of delivery, number of ANC visits, native language, ethnicity, paternal educational level, region, maternal highest educational level, literacy, health card, health card and/or vaccination document and age of child

2.2.8 Data Modeling

Once the data was completely preprocessed, a prediction model was built using Artificial Neural Network (ANN). The socioeconomic factors were inputs in the multilayer FNN's input-output training set, and the defaulter class was the output. To train the model, only a training set was used. The model was evaluated, saved and used to make predictions on new data. Out of the 15,294 datasets, 14,294 were used for the training and testing of the models while 1000 datasets were kept for prediction.

2.2.9 Model Evaluation

Model Evaluation is the process of assessing how well a trained model performed against real data. In this study, the model was evaluated based on the following performance measures: classification accuracy, sensitivity, specificity, f1 score and Confusion matrix. To measure the performance of the model on unseen data, a testing and validation set was utilized.

3. RESULTS AND DISCUSSION

3.1 Performance Evaluation

The distribution as shown in Figure 3 illustrates that the highest number of defaulters are children with major risk of defaulting; this means that 3199 (22.4%) children out of 14,294 children took less than 5 out of the 16 vaccination doses, while the least number of defaulters are children with moderate risk of defaulting; this means that 2595 (18.2%) children have taken 6 to 10 doses out of the total 16 doses. The middle class, which is children with minor risk of defaulting, has 2840 (19.9%) children who have taken more than 11 doses. It also shows that only 2930 (20.5%) of children with insignificant risk of defaulting have taken all the basic vaccinations, while 2730 (19%) of children have not taken even a single vaccination dose. This means that only about 20.5% of the population has received all 8 basic vaccinations labeled as insignificant risk.

Class 0: Major Risk: 3199

Class 1: Insignificant Risk: 2930

Class 2: Minor Risk: 2840

Class 3: Severe Risk: 2730

Class 4: Moderate Risk: 2595

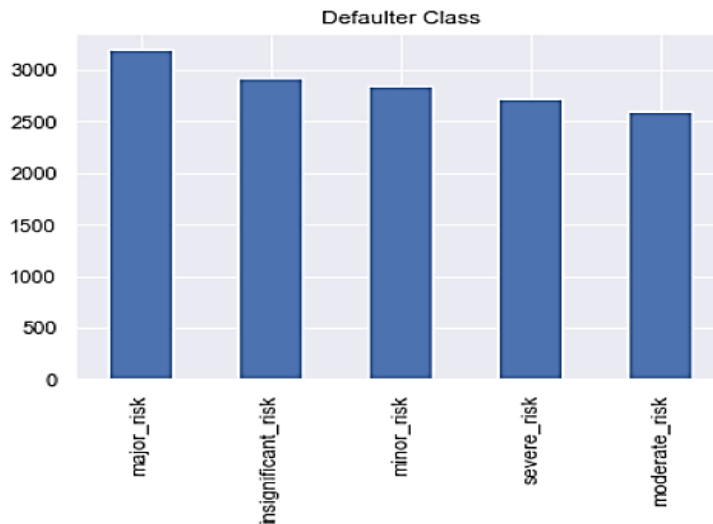


Figure 3. Defaulter Class Distribution

The model designed to predict children at risk of defaulting on routine immunization was the ANN algorithm. The data used for training was 80% while the data used for testing was 20%, which produced an accuracy of 99.16% and a loss of 0.0591 as shown in figure 4.

90/90 [=====] - 0s 1ms/step - loss: 0.0591 - accuracy: 0.9916

Test loss; 0.0590994618833065

Test Accuracy; 0.9916054606437683

Figure 4. Test loss and Test Accuracy for ANN

The precision, recall, f1 score are as shown in the classification report in Figure 5.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	566
1	0.98	0.98	0.98	543
2	0.99	0.98	0.99	570
3	1.00	0.99	0.99	631
4	1.00	1.00	1.00	549
accuracy			0.99	2859
macro avg	0.99	0.99	0.99	2859
weighted avg	0.99	0.99	0.99	2859

Figure 5. ANN Classification Report

The confusion matrix in figure 6 shows that from the 20% used for testing, the true positives for fully immunized, partially high, medium, partially low, and unvaccinated are 565, 534, 561, 626, and 549, respectively, i.e., they were correctly predicted.

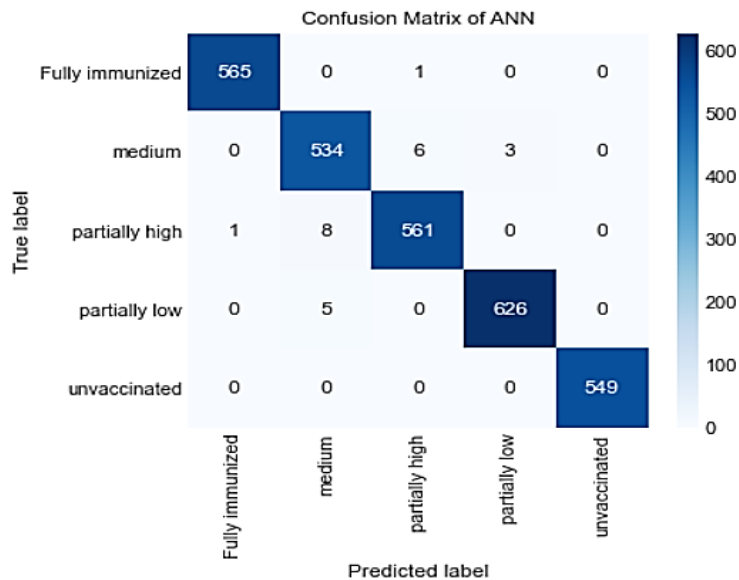


Figure 6. Confusion Matrix for ANN

Figure 7 displays the results of the model. The Artificial Neural Network produced an accuracy of 99.16%.

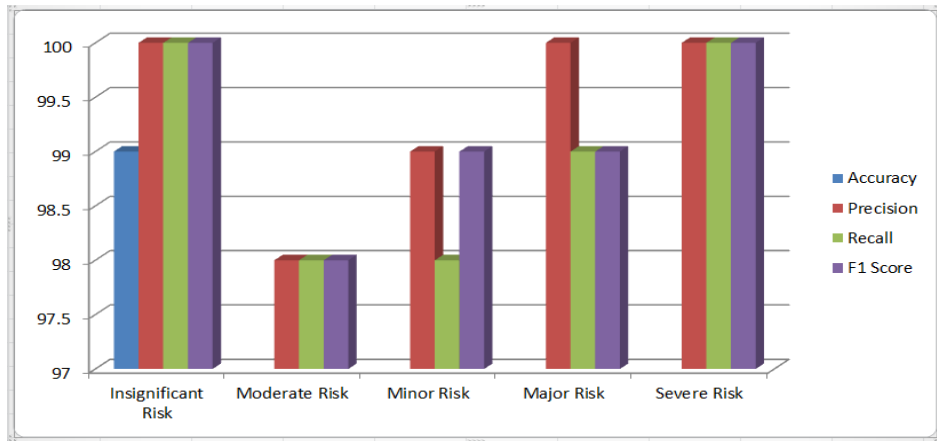


Figure 7: Chart Showing the Performance Results of the ANN

Model Accuracy: Figure 8 as shown below depicts an accuracy plot against 100 epochs. An epoch is defined as one complete pass of the training dataset through the algorithm. The figure also shows how the model's accuracy improved as the number of epochs increased during training and testing.

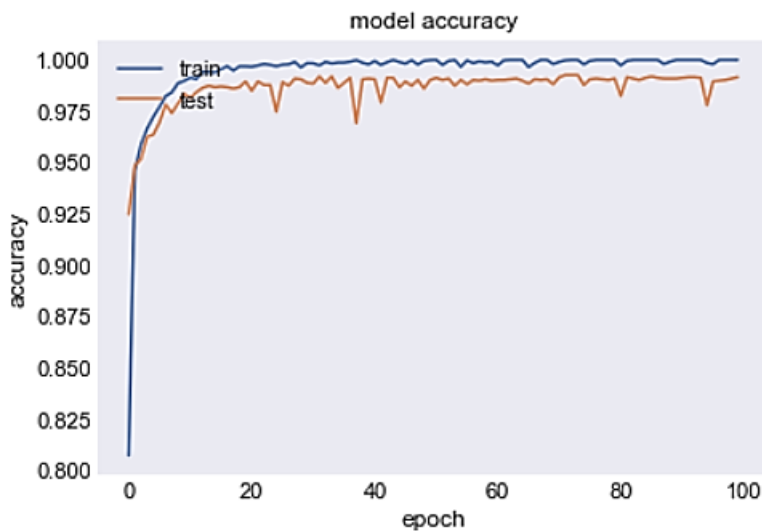


Figure 8: Illustrating ANN Model Accuracy against Epoch during Training and Testing

Loss: Figure 9 shows a plot of loss against epoch where 100 were considered in this study. It further illustrates how the loss of the model reduced with the increased epochs both during training and testing.

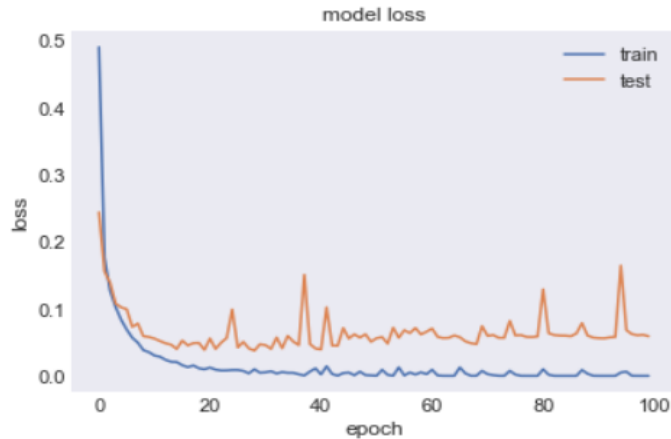


Figure 9: Illustrating ANN Model Loss against Epoch during Training and Testing

Visualization of Neural Network: The diagram in Figure 10 is the final visualization of the neural network. The figure illustrates the Routine Immunization neural network which has one input layer with 34 neurons, two hidden layers (the first hidden layer has 128 neurons while the second hidden layer has 8 neurons) and the output layer has 5 neurons for the 5 classes.

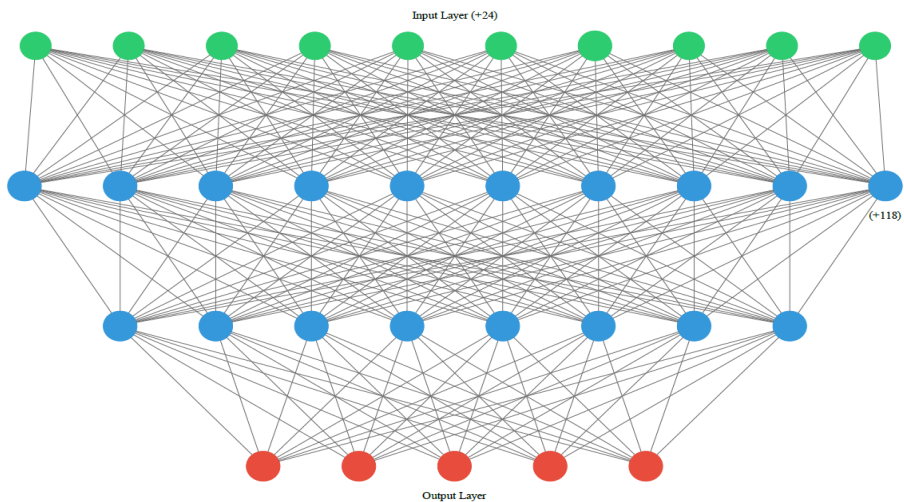


Figure 10. Routine Immunization Defaulter Neural Network

3.2 Discussion

This research aimed to develop a neural network model to predict children at risk of defaulting from routine immunization in Nigeria using the NDHS 2018 dataset. After rigorous data cleaning and preprocessing, 14,924 records were utilized, incorporating thirty-four independent attributes. These attributes included maternal age, region, religion, wealth index, birth order, health card, maternal highest education, literacy, native language, vaccination documents, total number of children, ethnicity, child's age, place of delivery, antenatal care attendance, paternal age, paternal education, BCG, Polio doses (0, 1, 2, 3), Measles doses (1, 2), Vitamin A, Hepatitis (birth), Penta doses (1, 2, 3), PCV doses (1, 2, 3), and IPV. The output classified the risk of defaulting into five categories: insignificant risk, minor risk, moderate risk, major risk, and severe risk. The dataset was then trained and tested using an Artificial Neural Network (ANN) classifier implemented in a Jupyter notebook via the Anaconda platform.

The results revealed that out of the 14,924 children analyzed, 3,199 (22.4%) received fewer than five of the sixteen recommended vaccination doses, indicating a high likelihood of defaulting. This group represented the largest among the five risk categories. Conversely, 2,595 children (20.5%) had received all eight basic vaccinations, classified as an insignificant risk, highlighting a relatively low fully immunized rate within the sample. For the testing phase, which utilized 20% of the dataset, the true positive rates for fully immunized, partially high, medium, partially low, and unvaccinated groups were 565, 534, 561, 626, and 549, respectively. This demonstrated that the model accurately predicted the immunization status with minimal misclassifications.

To further validate the model's accuracy, it was tested on an additional 1,000 datasets. Remarkably, 974 out of these 1,000 cases were accurately predicted, yielding a model accuracy of 99.16%. This high accuracy rate underscores the model's effectiveness in predicting the immunization status of children, suggesting its potential utility in identifying and mitigating the risk of defaulting in routine immunization programs in Nigeria. The study's findings could significantly contribute to public health strategies aimed at improving vaccination coverage and preventing disease outbreaks.

4. CONCLUSION

As immunization has remained the most effective and efficient public health intervention to date, the Nigerian Government should restructure the process to improve vaccination uptake and reduce childhood morbidity and mortality. Neural networks possess the ability to identify children at risk of defaulting from RI. Identifying such children is a major key to reducing child morbidity and mortality.

This study identified different socio-economic factors that can make a child default from RI from various literatures in Nigeria and beyond. The attributes were obtained from the NDHS 2018 dataset. The study developed an Artificial Neural Network which produced accuracy of 99.16%, recall of 0.99, precision of 0.99 and f1 score of 0.99 with a good confusion matrix. The model further used one thousand (1000) validation dataset out of which nine hundred and seventy-four (976) results were accurately predicted. This ANN model would assist the health care centers and NGOs targeted at RI intervention to identify children at risk of defaulting. Policy makers and the government can benefit greatly by performing analysis on RI datasets to reduce the frequency of dropouts.

REFERENCES

- [1] E. A. Ophori, M. Y. Tula, A. V. Azih, R. Okojie and P. E. Ikpo, "Current trends of immunization in Nigeria: prospect and challenges," *Tropical medicine and health*, vol. 42, no. 2, pp 67–75, 2014, doi:10.2149/tmh.2013-13.
- [2] S. Chandir, D. A. Siddiqi, O. A. Hussain, T. Niazi, M. T. Shah, V. K. Dharma, A. Habib, and A. J. Khan, "Using Predictive Analytics to Identify Children at High Risk of Defaulting from a Routine Immunization Program: Feasibility Study," *JMIR Public Health and Surveillance*, vol. 4, no. 3, pp 1-12, 2018, doi: 10.2196/publichealth.9681.
- [3] National Population Commission (NPC) [Nigeria] and ICF, Nigeria Demographic and Health Survey 2018, Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF, 2018
- [4] I. Abdurraheem, A. Onajole, A. Jimoh and A. Oladipo, "Reasons for incomplete vaccination and factors for missed opportunities among rural Nigerian children," *Journal of Public Health and Epidemiology*, vol. 3, no. 4, pp. 194-203, 2011.
- [5] A. Owais, B. Hanif, A. R. Siddiqui, A. Agha, and A. K. Zaidi, "Does improving maternal knowledge of vaccines impact infant immunization rates? A community-based randomized-controlled trial in Karachi, Pakistan," *BMC Public Health*, vol. 11, no. 1, pp. 1-8, 2011, doi: 10.1186/1471-2458-11-239
- [6] E. O. Onsomu, B. A. Abuya, I. N. Okech, D. Moore, and J. Collins-McNeil, "Maternal Education and Immunization Status Among Children in Kenya," *Maternal and Child Health Journal*, vol. 19, no. 8, pp. 1724–1733, 2015, doi: 10.1007/s10995-015-1686-1.
- [7] A. Rammohan, and N. Awofeso, "District-level variations in childhood immunizations in India: The role of socio-economic factors and health infrastructure," *Social Science & Medicine*, vol. 145, pp. 163–172, 2015, doi: 10.1016/j.socscimed.2015.05.004

- [8] E. Crouch, and L. A. Dickes, “A Prediction Model of Childhood Immunization Rates,” *Applied Health Economics and Health Policy*, vol. 13, no. 2, pp. 243–251, 2015, doi:10.1007/s40258-015-0157-6
- [9] S. Walton, M. Cortina-Borja, C. Dezateux, L. J. Griffiths, K. Tingay, A. Akbari, A. Bandyopadhyay, R. A. Lyons, and H. Bedford, “Measuring the timeliness of childhood vaccinations: Using cohort data and routine health records to evaluate quality of immunisation services,” *Vaccine*, vol. 35, no. 51, pp. 7166–7173, 2017, doi: 10.1016/j.vaccine.2017.10.085
- [10] O. Oleribe, V. Kumar, A. Awosika-Olumo, and S. D. Taylor, “Individual and socioeconomic factors associated with childhood immunization coverage in Nigeria,” *Pan African Medical Journal*, vol. 26, pp. 1–14, 2017, doi: 10.11604/pamj.2017.26.220.11453.
- [11] M. Ameen, S. Rasul, M. A. ul Haq, and Q. K. Mahmood, “Determinants of factors associated with childhood immunization in Punjab, Pakistan: evidence from the multiple indicator cluster survey,” *Journal of Public Health*, vol. 26, no. 5, pp. 495–499, 2018, doi: 10.1007/s10389-018-0902-z
- [12] S. Qazi, M. Usman, A. Mahmood, A. Afzaal Abbasi, M. Attique, and Y. Nam, “Smart Healthcare Using Data-Driven Prediction of Immunization Defaulters in Expanded Program on Immunization (EPI);” *Computers, Materials and amp; Continua*, vol. 66, no. 1, pp. 589–602., 2020, doi: 10.32604/cmc.2020.012507
- [13] N. Mannion, (2020). Predictions of Changes in Child Immunization Rates Using an Automated Approach: USA. School of Computing National College of Ireland. [MSc thesis, School of Computing National College of Ireland], 2020.
- [14] G. Mohanraj, V. Mohanraj, J. Senthilkumar, and Y. Suresh, “A hybrid deep learning model for predicting and targeting the less immunized area to improve children vaccination rate,” *Intelligent Data Analysis*, vol. 24, no. 6, pp. 1385–1402, 2020, doi:10.3233/ida-194820
- [15] B. Kembabazi, A Classification model leveraging Electronic Immunization Records to predict child immunization completion: Case study - Mukono Health facility (Thesis, Strathmore University), 2021, <http://hdl.handle.net/11071/12749>
- [16] F. Sameen, A. Momin Kazi, M. Kazmi, M. A. Abbasi, S. Ahmed Qazi, and K. L. Stergioulas, “Improving Routine Immunization Coverage Through Optimally Designed Predictive Models,” *Computers, Materials & Continua*, vol. 70, no. 1, pp. 375–395, 2022. doi:10.32604/cmc.2022.019167