

A Comparative Study of Selected Supervised Machine Learning Algorithms for Detecting Child Mortality

Muhammad Haladu¹, Olawoyin O. Ishaq², Musa Uba Muhammad², Abubakar Salisu³

¹Department of Statistics, Federal College of Agricultural Produce Technology, Kano

²Department of Statistics, Aliko Dangote University of Science and Technology, Wudil

³Department of Mathematics, Sa'adatu Rimi College of education Kumbotso, Kano

muhammadhaladu41@gmail.com¹, babington4u@gmail.com², musaubamuhammad@gmail.com²,
abubakarkiru17@gmail.com³

Abstract

The World Health Organization (WHO) defines under-five mortality rate as the probability that a child born in a specified year will die before reaching the age of five. It is a key metric for determining the quality of a country's healthcare system. In many parts of the world, under-five mortality is still a big concern, according to the World Health Organization (WHO). Under-five mortality has been a major issue in emerging nations. Even though the number of childhood deaths worldwide has significantly decreased, under-five mortality is still a significant public health concern in Nigeria. Nigeria now has the second-highest number of under-five deaths globally, after India, and the highest recorded number of under-five deaths in Africa (WHO, 2022). In Nigeria, around one million children under the age of five pass away each year, with over 60% of these deaths taking place between the ages of one and fifty-nine. According to data from the Nigeria Demographic and Health Surveys (NDHS), the under-five mortality rate increased from 157 per 1000 live births in 2008 to 138.8 per 1000 live births in 2020. This showed that the rate is still high, and the Millennium Development Goal (MDG) calls for reducing under-five mortality to 25 per 1,000 live births by 2030. The study's precise goals are to use supervised machine learning classification algorithms to predict mortality among children under five using nationally representative data on children under five and the specific objectives are, to predict under five mortality using machine learning classification models, to compare the performance of the supervised machine learning Classification models and lastly to apply the study's conclusions to improve and enhance relevant government policies and intervention plans. The study utilized secondary data from the 2018 Nigeria demographic and health survey data (NDHS). The 2018 NDHS is a nationally probability sample survey with approximately 42,000 households (NDHS 2018). The unit of analysis include children under-five years with total sample size of 33924 selected from 1389 clusters across Nigeria. The (2018 NDHS) dataset was downloaded from www.dhsprogram.com website, SPSS version 23 was used to open the data set and R programming Language is used to perform data analysis and processing. In this study, the outcome variable was under-five mortality measured as a binary outcome, under-five mortality measured as being alive (coded as 0) or dead (coded as 1). Twenty seven(27) predictors (features) used in this study are mother's age at birth, highest educational level, mother's body mass index (BMI), Currently breastfeeding, Child alive, whether the child was wanted, sex of the child, birth order, child is twin, age in five years, given child anything other than breast, Birth weight in kg, Current age of child, Size of child at birth, source of drinking water, toilet facility, household wealth index, marital status, place of residence type, region, Native language ,religion, antenatal care visit, Delivery by caesarean, place of delivery, covered by health insurance and Number of antenatal care visit. The study used 2018 demographic and health survey data to investigate the factors linked to under-five mortality in Nigeria. Four distinct classification techniques for supervised machine learning were compare. Support Vector Machine (SVM) and Naïve Bayes emerged as the top performers among the four evaluated models based on their accuracy followed by Artificial Neural Network and lastly Linear Discriminant Analysis(LDA) models with

accuracy values reaching 62.48%. This indicates their strong ability to correctly identify cases where a child is not alive while maintaining a moderate level of sensitivity 62.74% for detecting cases where a child is alive and Lastly Support vector machine model outperformed all other models by achieving perfect scores across all metrics, suggesting that it can accurately classify every instance in the dataset.

Keywords: Supervised Machine Learning Algorithm, Under-five mortality, Comparative Model, Confusion Matrix.

1. Introduction

The World Health Organization (WHO) define the under-five mortality rate as the probability that a child born in a specified year will die before reaching the age of five, if subjected to age-specific mortality rates of that era (WHO, 2018). It is commonly represented as the number of deaths per thousand live births. It is a key metric for determining the quality of a country's healthcare system. The world has acknowledged that under-five mortality is a critical threat to mankind that must be addressed, therefore its inclusion in the Sustainable Development Goals (SDGs). Health target 2 under SDG 3 aims to reduce under-five mortality to 25 per 1,000 live births by 2030. The death rates for children under five in underdeveloped nations are significantly lower than those in developed nations. Over the past three decades, there has been a notable improvement in the survival rate of children worldwide. According to WHO (2018), 1 in 26 children died before turning five in 2021, compared to 1 in 11 children in 1990. This means that millions of children now have a greater chance of surviving than they had in 1990. Furthermore, as stated by the United Nations International Children's Emergency Fund (UNICEF), "when compared to the 1990s, progress in lowering child mortality rates has accelerated in the 2000s period, with the annual rate of reduction in the global under-five mortality rate rising from 1.8% in the 1990s to 4.0% for 2000-2009 and 2.7% for 2010-2021 (UNICEF). From 93 deaths per 1,000 live births in 1990 to 38% in 2021, the global under-five mortality rate decreased by 59% (UNICEF, 2021). Even with these significant advancements, there is still a pressing need to improve child survival. About 13,800 children under five died every day in 2021 alone, which is an unacceptably high number of mostly avoidable child fatalities. One in eight Nigerian children never reach the age of five, with the country's under-five death rate of 132 per 1000 live births, according to the Nigeria Demographic Health Survey (NDHS, 2018). From 1990 to 2018, the number of infant deaths, which make up half of all child deaths, decreased from 87 per 1000 live births to 67. Every 10 minutes, a Nigerian woman passes away during childbirth, and every minute, a Nigerian child under five passes away. Lack of access to high-quality delivery services is a major problem in Nigeria, as seen by the high mortality rates. Caretakers describe a variety of challenges when it comes to obtaining healthcare, including issues like treatment costs, the poor condition of medical facilities, travel time, and a lack of awareness and expertise to make informed decisions and referrals (WHO, 2019). The Nigerian government is addressing the unacceptable high under-five morbidity death indices by implementing two critical child survival initiatives: Integrated Community Case Management (ICCM) and Integrated Management of Childhood Illness (IMCI).

2. Related Works

The review was basically base on aim, data, methods, findings and conclusion of the related studies.

Aayush et al(2016) Applied machine learning in the analysis of infant mortality and its factors using the united states dataset .The study used the birth files of the year 2013 and its important factors were identifies using Kendall rank correlation coefficient. A randomized split of 80% to 20% was used to build a training dataset of 32432 cases and a testing dataset of 8081 cases out of 40541cases. Three different classification models (Logistics Regression, Naïve Bayes and Lagragian Support Vector machine) were use for binary classification problem and three model were fitted with training data and cross validated in 10 folds using the test data. The evaluation metrics were used were accuracy, precision, recall and F1-score. The result determine that the performance of the logistic regression model was best with highest precision score follow by Naïve Bayes and Lagrangian support vector machine. The study also indicated that identifying a mother at high risk is more important than misclassification of a low risk mother given by recall. The result suggested that best model with high precision can be used to predict key factors and high risk mothers and thus they can be given proper medical care to mitigate the risk, hence reducing the infant mortality rate .

Abdulwahaba.& Bahaa (2023) Use an Artificial Neural Network Prediction Model for maternity Health care Quality in Nigeria, and the study showed that it is possible to predict the quality of maternity health care in Nigeria using artificial neural networks (ANNs) and binary classification. The study sought to address the difficulty of accurately assessing the quality of maternal health services, which is essential for guaranteeing good birth experiences and increasing mother outcomes, by utilizing a thorough approach and utilizing the power of artificial neural networks (ANNs). Maternal health services were accurately classified into "High_Quality" and "Low_Quality" groups using the binary classification model that was designed for this study. The artificial neural network (ANN) helped healthcare providers discover areas that needed improvement and optimize resource allocation for pregnant women by making data-driven predictions based on its capacity to comprehend complicated patterns within large datasets. The research's conclusions may have a big impact on legislators, medical professionals, and advocates for maternal health. They may direct initiatives to maximize healthcare resources and improve maternal health outcomes in Adamawa and throughout Nigeria. A number of criteria, including F1 score, recall, accuracy, and precision, are used to assess the overall performance. The suggested model obtains 88% accuracy, 88% precision, 89% recall, and an 88% F1 score on the testing dataset, according to the results. Additionally, the model offers a feature set that, in this instance, yields results with a high accuracy of 87.17%. The demands of the health facility are the best indicators of the delivery of maternal health care, according to data research on maternal health services facilities.

Adeyinka et al.(2020) Performed multilevel multinomial logistic regression analysis on data of a nationally representative sample of 29,786 (weighted=30,960) live births delivered 5 years before the survey to 18,497 women aged 15–49years and nested within 16,151 households and 2227 communities. The findings showed that Determinants of under-five mortality differ across the neonatal, post-neonatal and toddler/pre-school stages in Nigeria. Unexpectedly, attendance of skilled health providers during delivery was associated with an increased neonatal mortality risk, although its effect disappeared during post-neonatal and toddler/pre-school stages. Also the study indicates that maternal-level factors such as maternal education, contraceptive use, maternal wealth index, parity, death of previous children, and quality of prenatal care accounted for high variation (39%) in childhood mortalities across the communities. The inclusion of other compositional

and contextual factors had no significant additional effect on childhood mortality risks across the communities.

Ayele et al (2017) Study the Survival analysis of under-five mortality using Cox and frailty models with 2011 Ethiopian Demographic and Health survey data. From the results, it was found that under-five children who live in Addis Ababa had a lower hazard (risk) of death (p value=0.048). This could be as a result of higher health facilities and living standards in Addis Ababa, compared to other regions. Under-five children who lived in rural areas had a higher hazard (risk) of death compared to those living in urban areas. In addition, under-five children who lived in rural areas had 18% (p value=0.01) more hazard (risk) of death than those living in urban areas. Furthermore, with older mothers, the chance of a child dying before reaching the age of 5 is lower. The research that the chances of a child dying before reaching the age of 5 are less if the mother does not become pregnant again before the child reaches the age of 5. Therefore, giving birth when older and not becoming pregnant again before the child reaches the age of 5 is one means of reducing under-five mortality

Azuike et al (2019) applied univariate binary logistic regression to examine the association between the individual explanatory variables and the outcome variable finding from the result following multivariate analysis a number of determinants of under-five mortality were identified. They include: Geo-political zones of the children, Gender, Mother's educational level, Mother's number of births in the preceding 5 years, Mother's marital status, Maternal age, Mother's type of family planning method, Child's birth order, Family size, Family wealth index and Place of residence (urban/rural). The study finally conclude that Multiple logistic regression has enabled us to identify important determinants of under-five mortality in Nigeria and recommend that policy makers should consider the determinants as they design interventions to combat under-five mortality .

Brnabic&Lisa (2021) Investigated systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making, this review found a wide variety of approaches, methods, statistical software and validation strategies that were employed in the application of machine learning methods to inform patient-provider decision making. Based on these findings, there is a need to ensure that multiple modeling approaches are employed in the development of machine learning-based models for patient care, which requires the highest research standards to reliably support shared evidence-based decision making. Models should be evaluated with clear criteria for model selection, and both internal and external validation is needed prior to applying these models to inform patient care. Few studies have yet to reach that bar of evidence to inform patient-provider decision making .

Chen. et al. (2021) proposed that the discrimination analysis accurate rate may be significantly increased from 49.5 to 92% using the enhanced naive Bayes classification algorithm. The enhanced naive Bayes classification algorithm has a greater accuracy thanks to robustness analysis. The work achieves an improved naive Bayes classification algorithm by applying feature weighting and Laplace calibration. It has been discovered that the enhanced naive Bayes classification algorithm has an accuracy of more than 99% and stability when sample size is large; the algorithm's accuracy is more than 95% when sample attribute and number of categories are less than 400 and 24, respectively.

Deprez et al. (2017) showed that machine learning algorithms are useful to assess the goodness of fit of the mortality estimates provided by standard stochastic mortality models (they considered Lee-Carter and Renshaw-Haberman models). They applied a regression tree boosting machine to analyze how the

modeling should be improved based on feature components of an individual, such as its age or its birth cohort. This (non-parametric) regression approach then allows us to detect the weaknesses of different mortality models. In addition, they investigated cause-of-death mortality.

Dominic & Azuh (2019) studied factors influencing the survival of under-five children among women visiting government health care facility in semi-urban communities in Nigeria. Descriptive statistics and regression analyses were applied and result indicate that Knowledge of child preventable diseases, birth spacing, and duration of breast feeding and immunization status showed significant influence on child survival and conclude that knowledge of child preventable diseases and factors related to maternal child care practices were found to influence child survival in this communities.

Desalegn & Mosisa (2021) Conduct Descriptive statistics and Multivariable logistic regression analysis for the determinants of Diarrhea in Under-Five Children among health Extension Model and Non-Model families. The two-week childhood diarrhea prevalence in under-five children was 7.8% in model and 27.8% in non-model families. Childhood diarrhea in under-five children was 2.4 times more likely among health extension non-model families than model families. Gobir A.A et al (2018) employed multivariate analysis and suggested that children of uneducated fathers were more likely to have been ill compared to those of educated fathers (RR = 1.20; 95% C.I = 0.78 – 1.91). Children who were weaned abruptly and transferred to their aunties or grandmothers (geographic weaning) were more likely to have been ill compared to those weaned normally (RR = 1.71; 95% C. I= 0.97 – 2.03). The result indicated that several factors like paternal education and geographic weaning are social determinants of health of under-fives in the study area. Improved access to formal education, Health education on effects of harmful cultural practices; community-based health insurance scheme and rural development are recommended.

Guidingo et al. (2024) Take a look at the Overview on the Advancements of Support Vector Machine Models in Healthcare Applications and see how SVM's strong theoretical underpinnings and benefits have made it a very effective machine learning tool, particularly in the medical domain. Applying specialized SVMs to healthcare data has general benefits, as demonstrated by our investigation. Notably, compared to other machine learning techniques, SVM-based models are frequently more successful. Because support vector machines (SVMs) can automatically regularize higher-dimensional data, they can effectively handle overfitting problems. Different SVM extensions have been developed to address different problems. TWSVM, together with other extensions, is one type of SVM.

Huda M. Haroun et al (2014) Applied chi-square to determine the level of infant and under-five mortality rates and to examine the effect of socioeconomic, demographic and environmental factors on the health status of the children under five years. The outcome of the study showed that Infant mortality rate was 77 per 100 and child mortality rate was 67 per 100. The results revealed that immunization, child order, child birth weight, birth interval and contraceptive use had a significant influence on the mortality of children under the age of five. The mother's level of education is highly significant on the mortality of children under five years old.

Janardhanan et al (2015) used the RBF network, SVM classifier, and Naïve Bayes classifier in medical data mining. A variety of medical datasets are used to analyze and compare the prediction model's performance in forecasting diseases. Every dataset had a distinct amount of attributes and belonged to the binary class. Diabetes, cancer, and heart-related datasets are among the collections. It has been noted that SVM classifiers

yield higher classification accuracy percentages. The work has been carried out in the WEKA environment, and the outcomes demonstrate that SVM is the most reliable and efficient classifier for medical data sets.

Justice & Aheto (2019) employed Single level binary logistic and multilevel logistic regression model to investigate determinants of under-five mortality in Ghana. The fit of the model was checked using Variance Inflation Factor and Likelihood Ratio tests. The Receiver Operating Characteristic curve was used to assess the predictive ability of the models. A p -value < 0.05 was used to declare statistical significance. The study produced a good predictive model and identified increase in number of total children ever born, number of births in last 5 years, and mothers who did not intend to use contraceptive as critical risk factors that increase the odds of under-five mortality. Also, children who were born multiple and residing in certain geographical regions of Ghana is associated with increased odds of under-five mortality. Maternal education and being a female child decreased the odds of under-five mortality. No significant unobserved household-level variations in under-five mortality were found. The spatial map revealed regional differences in crude under-five mortality rate in the country. The study was concluded by identified critical risk factors for under-five mortality and strongly highlights the need for family planning, improvement in maternal education and addressing regional disparities in child health which could help inform health policy and intervention strategies aimed at improving child survival.

Khan et al (2021) examined trends and puts forward projections of under-5 mortality (U5M) in Bangladesh and identified the effects of maternal high-risk fertility behaviors and use of healthcare services. Chi-square (χ^2) test was used to identify if there was any association with maternal high-risk fertility behaviors and the use of healthcare services. A multivariate logistic regression model was used to determine the effects of fertility behaviors and healthcare usage on the occurrence of U5M adjusting with confounders. The study showed that U5M declined from 82.5 to 41.0 per 1000 live births during 1994–2014 and is projected to further reduce to 17.6 per 1000 live births by 2030. The study identified a noticeable regional variation in U5M with maternal high-risk fertility behaviors including age at birth < 18 years (aOR: 1.84, 95% CI: 1.23–2.76) and birth interval < 24 months (aOR: 1.56, 95% CI: 1.02–2.37) found to be significant determinants. There was a 39–53% decline in this rate of mortality among women that had used antenatal care services at least four times (aOR, 0.51, 95% CI: 0.27–0.97), delivery care (aOR, 0.47, 95% CI: 0.24–0.95), and had received postnatal care (aOR, 0.61, 95% CI: 0.41–0.91) in their last birth. Cesarean section was found to be associated with a 51% reduction in U5M (aOR, 0.49, 95% CI: 0.29–0.82) compared to its non-use.

The research concluded that with the current trend of reduction, Bangladesh will achieve the MDG target before the deadline. The study also found that maternal high-risk fertility behaviors and non-use of maternal healthcare services are very prevalent in some regions of Bangladesh and have increased the occurrence of U5M in those areas. This suggests therefore, that policies and programmes designed to reduce the pregnancy rates of women that are at risk and to encourage an increase in the use of maternal healthcare services are needed.

Las Johansen & Caluza (2018) applied descriptive-correlation design using data mining techniques specifically the decision tree algorithm, the annual population growth shows a very significant role in achieving a model in this study. Also, the generated model revealed very high acceptability of the model with a weighted average of 97.4% value of all classes of mortality rate under five years old. Finally, the generated model is an excellent baseline in predicting child mortality under five years old in the Philippines. However, further studies may be conducted to enhance this model.

Hussein et al. (2021) employed descriptive techniques to examine the levels and differentials of under-five mortality both in Kenya and in Egypt. In addition, background characteristics of women who had a last birth within the five-year period before the surveys were also investigated on the backdrop of under-five mortality. In descriptive analysis, Chi-square tests were used to test for existence of association between categorical variables, i.e. the dependent variable (child death) and some socio-demographic and behavioral indicators (independent) that determine childhood mortality .

Maind.& Wankar (2014) mentioned that artificial neural networks (ANNs) provide a number of benefits over traditional methods, and that you can often anticipate a network to train quite effectively based on the type of application and the strength of the internal data patterns. This holds true for situations where there may be a lot of dynamic or non-linear relationships. Discussions on neural networks are commonplace these days. Since nature itself provides evidence that such things are effective, their promise appears to be quite bright. Hardware development, however, is the technology's future and, in fact, the key to it all. The majority of neural network research being done today just verifies the theory's viability.

Masrie et al (2020) use the Kaplan Meir method and cox proportional hazards regression model to identify the survival status and the proximate determinants associated with infant mortality. The study used the 2016 Ethiopian Demographic and Health Survey records for all 10641 live births and survival data of all 2826 infants born five years before the survey were reviewed. The result of Kaplan-Meier estimation determined that about 65% infant death occurs early in the month of life immediately after births and reduced in the later months of follow-up time and the result of Cox proportional hazards model indicates that mothers' level of education, preceding birth interval, plurality, size of child at birth and sex of child were significant predictors of infant mortality. The study showed that risk of dying in infancy was lower for infants whose mothers had secondary education level. The study also recommended that close monitoring and supporting reproductive age mothers to increase the uptakes of family planning and antenatal care follow-up will increase the survival of infants.

Mohsin & Mustapha (2021) Upon reviewing a research paper titled "Machine Learning Applications based on SVM Classification," it became apparent that the study aims to address data processing by utilizing the support vector machine (SVM) algorithm across various domains, given its dependability and effectiveness as a classification technique in the machine learning domain. As a result, this paper covers the usage of SVM classifier by exploring numerous works. It was ultimately determined that SVM classifier has obtained high accuracy results in most of the applications, specifically for face recognition and disease identification applications. SVM classification has been used in many fields, including face recognition, plant disease identification, text recognition, sentiment analysis, and intrusion detection system for network security application.

Osita et al (2021) investigated Under-5 Mortality and its Associated Factors in Northern Nigeria, Findings from this study showed that children younger than 5 years old whose mothers did not use contraception had significantly greater odds of dying in the U5M period than those who had access to contraceptives.

Pienaar et al. (2022). Study an Artificial Neural Network Model for Predicting Pediatric Mortality in Two Tertiary Pediatric Intensive Care Units in South Africa, When the models' performances were compared, the ANN's AUROC (0.89) was greater than that of the recalibrated PIM3 model (0.86) and the logistic regression model (LR) (0.87). However, the ANN outperforms logistic regression and recalibrated PIM3 in the precision recall curve (AUPRC = 0.6 vs. 0.53 and 0.58, respectively). The ANN model had a slope

of 1.12 (intercept 0.01), the logistic regression model had a slope of 1.09 (intercept 0.05), and the recalibrated version of PIM3 had a slope of 1.02 (intercept 0.01). However, for the ANN model, the calibration curve was closer to the diagonal. The study comes to the conclusion that, although there are many obstacles, artificial neural network models are a practical approach for predicting mortality in lower- and middle-income nations.

Rathod et al (2023) Investigate Naïve Bayes algorithms for news classification and draw the conclusion that a well-liked machine learning technique for news story classification is the Naïve Bayes classifier. It offers many advantages, including ease of use, good performance, and strong performance on a variety of categorization tasks. Nevertheless, the researchers go on to say that an assessment of its performance on a specific dataset is necessary. This region can be further enhanced to increase the classifier's performance, handle more complicated problems, and apply the classifier to new domains. The Naïve Bayes classifier has a lot to offer information retrieval and natural language processing.

Rustam et al. (2020) Research is being done in an effort to create a machine learning-based system for predicting in the future how many COVID-19 cases will be affected. The work shows how machine learning algorithms can predict the number of future COVID-19 cases, which is currently seen as a possible threat to humankind. Specifically, the study used four conventional forecasting models to predict the COVID-19 hazardous factors: exponential smoothing (ES), support vector machine (SVM), least absolute shrinkage and selection operator (LASSO), and linear regression (LR). Each of the models predicts three different things, including the number of newly infected cases, the number of deaths, and the number of recoveries in the next

Victorseelan & Jaisankar (2019) In order to predict breast cancer, compare Statistical Discriminant Analysis with Artificial Neural Network Model. The correct prediction rate serves as a proxy for the model's power. According to the study, Neural Network analysis predicts more accurately than Discriminant analysis.

Sreehari (2020) Study Supervised machine learning algorithms SVM, Random Forest, KNN, Naïve Bayes Algorithm, and Softmax Algorithm have been applied for the stock price prediction. The results reveal that for large dataset, Random Forest Algorithm outperforms all the other algorithms in terms of accuracy and when the size of the dataset is reduced to almost half of the original, and then Naïve Bayes Algorithm shows the best results in terms of accuracy. Also, reduction in the number of technical indicators reduces the accuracy of each algorithm in predicting the stock market trends.

3. Methodology

The data source for this study is secondary data from the 2018 Nigeria demographic and health survey data (NDHS). The 2018 NDHS is a nationally probability sample survey with approximately 42,000 households (NDHS 2018).

The unit of analysis include children under-five years with total sample size of 33924 selected from 1389 clusters across Nigeria. The (2018 NDHS) dataset was downloaded from www.dhsprogram.com website, SPSS version 23 was used to open the data set and R programming Language is used to perform data analysis and processing .

Study Variables and Measurement

In this study, the outcome variable was under-five mortality measured as a binary outcome, under-five mortality measured as being alive (coded as 0) or dead (coded as 1). Twenty seven(27)predictors (features) used in this study are mother’s age at birth, highest educational level, mother’s body mass index (BMI), Currently breastfeeding, Child alive, whether the child was wanted, sex of the child, birth order, child is twin, age in five years, given child anything other than breast, Birth weight in kg, Current age of child, Size of child at birth, source of drinking water, toilet facility, household wealth index, marital status, place of residence type, region, Native language ,religion, antenatal care visit, Delivery by caesarean, place of delivery, covered by health insurance and Number of antenatal care visit.

3.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a technique based on the neural Structure of the brain that mimics the learning capability from experiences. It means that if a neural network is trained from past data, it will be able to generate outputs based on the knowledge extracted from the data (Punitha D.C & Vigneswari.T2019).

A neural network is called a mapping network, if it is able to compute some functional relationship between its input and output. Function approximation from a set of input-output pairs has numerous scientific and engineering applications.

Thus a neural network is extremely useful when you do not have any idea of the functional relationship between the dependent and independent variables. The network formed has to be trained so that it will learn an approximation. Learning in a neural network means, finding an approximate set of weights. ANN is a universal function approximated that can approximate any function with arbitrary accuracy. It is to be noted that ANN is a nonlinear model that can be implemented for most complex real world applications. There are two main categories of network architectures depending on the type of the connections between the neurons which include “feed-forward neural networks” and “recurrent neural networks”. If there is no “feedback” from the outputs of the neurons towards the inputs throughout the network, then the network is referred as a “feed-forward neural network”. Otherwise, if there exists such a feedback, i.e. a synaptic connection from the outputs towards the inputs (either their own inputs or the inputs of other neurons), then the network is called a “recurrent neural network”. Usually, neural networks are arranged in the form of “layer”s. Feed-forward neural networks fall into two categories depending on the number of the layers, either “single layer” or “multi-layer”.

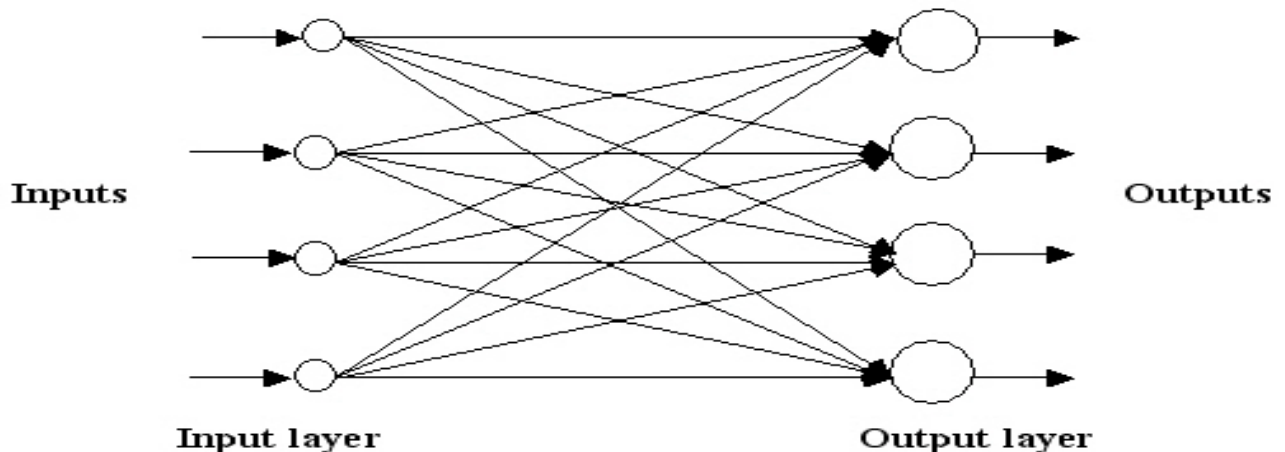


Figure 1: A single layer feed-forward neural network

In single layer feed-forward neural network (fully connected) is shown. Including the input layer, there are two layers in this structure. However, input layer does not count because there is no computation performed in that layer. Input signals are passed on to the output layer via the weights and the neurons in the output layer compute the output signals.

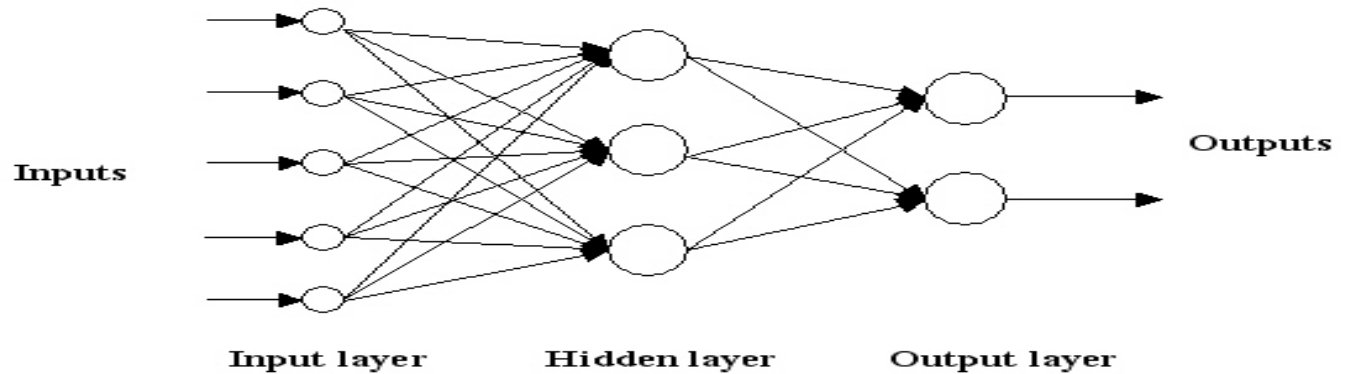


Figure 2: Multi-layer feed-forward neural network

A multi-layer feed-forward neural network with one “**hidden layer**” is depicted. As opposed to a single-layer network, there is (at least) one layer of “**hidden neurons**” between the input and output layers. According to Haykin.S(1999), the function of hidden neurons is to intervene between the external input and the network output in some useful manner. Existence of one or more hidden layers enable the network to extract higher-order statistics.

For the example, consider a multi-layer feed-forward neural network given in above diagram there is only one hidden layer and the network is referred as a 5-3-2 network because there are 5 input neurons, 3 hidden neurons, and 2 output neurons. From the above diagram 1 and 2, networks are “fully connected” because every neuron in each layer is connected to every other neuron in the next forward layer. If some of the synaptic connections were missing, the network would be called as “partially connected”. The most important feature of a neural network that distinguishes it from a conventional computer is its “learning” capability. A neural network can learn from its environment and improve its performance through learning. Haykin .S(1999) Defined learning in the context of neural networks in the literature as follows: “Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place”.

3.2 Back-Propagation Algorithm

Among many other learning algorithms, “**back-propagation algorithm**” is the most popular and the mostly used one for the training of feed-forward neural networks. It is, in essence, a means of updating networks synaptic weights by back propagating a gradient vector in which each element is defined as the derivative of an error measure with respect to a parameter. Error signals are usually defined as the difference of the actual network outputs and the desired outputs. Therefore, a set of desired outputs must be available for training. For that reason, back-propagation is a supervised learning rule.

3.3 Linear Discriminant Analysis (LDA)

In order to categorize a new member into one of these categories based on observations made by the member, discriminant analysis is employed to describe the differences between the groups. Many references refer to discriminant analysis as classification. Nonetheless, the term "classification" is used to refer to cluster analysis in a few places. Medical diagnosis, market research, anthropological specimen classification, business failure or success prediction, placing students or workers based on pretest results compared to previous student or worker placements, distinguishing between natural and artificial seismic activity, fingerprint analysis, image pattern recognition, and signal pattern classification are a few applications of discriminant analysis.

Discriminant analysis uses a vector of variables (X_1, X_2, \dots, X_n) to create a multivariate density function that describes a person or a phenomenon.

By creating a linear combination, the discriminant function maps the multidimensional properties of the density function of the population's variables onto a one-dimensional measure (Jeatrakul.P &Wong, K.W,2009).

The linear Discriminant function is as follows:

$$Z_i = XA = a_0 + a_1 X_{i1} + a_2 X_{i2} + \dots + a_n X_{in}$$

Where

Z= Discriminant Score

X= vector of n independents variables

A= Vector of Discriminant coefficients

Discriminant analysis minimizes the statistical distance between each observation and its own group means by computing the discriminant coefficients and choosing the proper weights to distinguish the average values of each group.

3.4 Naïve Bayes Model

This approach is based on the work of Thomas Bayes, who developed the classification-using Bayesian theorem. Applications with non-deterministic relationships between the attribute set and the class variable employ Bayesian classification. Even though the test record's attribute set is the same as that of some of the training instances, it is impossible to predict the test record's class label with precision. Noise or the existence of confounding variables that influence classification but are not considered in the study could be the cause of this.

Given the class label y , a naïve-bayes classifier assumes that the attributes are conditionally independent and estimates the class-conditional probability. We must estimate the conditional probability of each X_i given Y rather than calculating the class-conditional probability of every combination of X . Because it does not require a huge training set to generate a decent estimate of the likelihood, this method is more practical. The naïve bayes classifier classifies test records by calculating the posterior probability for each class Y as follows:

$$P(Y / X) = \frac{P(Y) \prod_{i=1}^d P(X_i / Y)}{P(X)}$$

The conditional probability for a categorical variable X_i is calculated based on the percentage of training cases in class y that have a given attribute value X_i . Since these points are averaged out when estimating the conditional probabilities of the data, the naïve bayes classifier is robust to separate noise points and irrelevant features. Their performance may suffer from linked traits. Domingos has confirmed that, in

reality, the Bayesian classifier performs fairly well, even in the presence of high attribute dependencies. It has also been demonstrated that attribute independence is not necessary for a Bayesian classifier to perform optimally in a zero-loss scenario. For the Bayesian classifier to be optimal, a number of required and sufficient requirements have been determined.

3.5 Support Vector Machine Model

Since Vapnik initially proposed the Support Vector Machine (SVM), the machine learning research community has shown a great deal of interest in this technique. A group of related supervised learning techniques used for regression and classification are called SVMs. Their classification is categorized as generalized linear. SVM has the unique ability to maximize the geometric margin and minimize the empirical classification error at the same time. SVM is hence known as Maximum Margin Classifiers. The foundation of SVM is Structural Risk Minimization (SRM). SVM creates a maximal separating hyperplane by mapping the input vector to a higher dimensional space. Each side of the hyperplane used to divide the data is built with two parallel hyperplanes. By maximizing the distance between the two parallel hyperplanes, a hyperplane is said to be separating. The classifier's generalization error is assumed to be better the greater the margin or separation between these parallel hyperplanes. Each side of the hyperplane used to divide the data is built with two parallel hyperplanes. By maximizing the distance between the two parallel hyperplanes, a hyperplane is said to be separating. The classifier's generalization error is assumed to be better the greater the margin or separation between these parallel hyperplanes.

$$\left\{ (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), \dots, (x_n, y_n) \right\}.$$

Where $y_n = 1 / -1$, a constant denoting the class to which that point x_n belongs. n = number of sample. Each x_n is p -dimensional real vector. The scaling is important to guard against variable (attributes) with larger variance. We can view this Training data, by means of the dividing (or separating) hyperplane, which takes

$$w \cdot x + b = 0$$

where w is a p -dimensional vector and b is a scalar. In relation to the dividing hyperplane, the vector w points perpendicularly. To expand the margin, add the offset parameter b . The solution is limited because the hyperplane is forced to pass through the origin in the absence of b . We are interested in SVM and parallel hyperplanes, just as we are in the greatest margin. The following equations can be used to describe parallel hyperplanes:

$$w \cdot x + b = 1, \quad w \cdot x + b = -1$$

If the training data are linearly separable, we can select these hyperplanes so that there are no points between them and then try to maximize their distance. By geometry, We find the distance between the hyperplane is $2 / |w|$. So we want to minimize $|w|$. To excite data points, we need to ensure that for all i either

$$w \cdot x_i - b \geq 1 \text{ or } w \cdot x_i - b \leq -1$$

This can be written as follows:

$$y_i (w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n$$

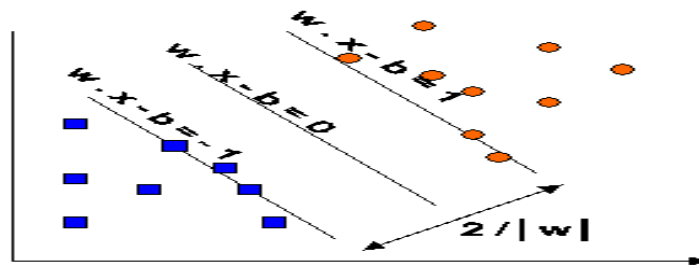


Figure 3: Support Vector Machine Model

Above Figure represent Maximum margin hyperplanes for a SVM trained with samples from two classes. Samples along the hyperplanes are called Support Vectors (SVs). A separating hyperplane with the largest margin defined by $M = 2 / |w|$ that is specifies support vectors means training data points closets to it. Which satisfy?

$$y_j [w^T \cdot x_j + b] = 1, \quad i=1$$

Optimal Canonical Hyperplane (OCH) is a canonical Hyperplane having a maximum margin. For all the data, OCH should satisfy the following constraints

$$y_i [w^T \cdot x_i + b] \geq 1; \quad i=1,2,\dots,1$$

Where 1 is Number of Training data point. In order to find the optimal separating hyperplane having a maximul margin, A learning macine should minimize $\|w\|_2$ subject to the inequality constraints

$$y_i [w^T \cdot x_i + b] \geq 1 \quad ; \quad i = 1, 2, \dots, 1$$

This optimization problem solved by the saddle points of the Lagrange's Function

$$L_p = L(w, b, \alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^1 \alpha_i (y_i (w^T \cdot x_i + b) - 1)$$

$$\frac{1}{2} w^T w - \sum_{i=1}^1 \alpha_i (y_i (w^T \cdot x_i + b) - 1)$$

Where α_i is a Lagranges multiplier .The search for an optimal saddle points (w_0, b_0, α_0) is necessary because Lagranges must be minimized with respect to w and b and has to be maximized with respect to nonnegative α_i ($\alpha_i \geq 0$). This problem can be solved either in primal form (which is the form of w & b) or in a dual form (which is the form of α_i).Equation number (4) and (5) are convex and KKT conditions, which are necessary and sufficient conditions for a maximum of equation (4). Partially differentiate equation (5) with respect to saddle points (w_0, b_0, α_0) .

$$\frac{dL}{dw_0} = 0$$

$$w_0 \sum_{i=1}^1 \alpha_i y_i x_i$$

And

$$\frac{dL}{db_0} = 0$$

$$\sum_{i=1}^1 \alpha_i y_i = 0$$

Substituting equation (6) and (7) in equation (5). We change the primal form into dual form.

$$L_d(\alpha) = \sum \alpha_i - \frac{1}{2} \sum_{i=1}^1 \alpha_i \alpha_j y_i y_j x_i^T x_j$$

In order to find the optimal hyperplane, a dual lagrangian (Ld) has to be maximized with respect to nonnegative α_i (i.e. α_i must be in the nonnegative quadrant) and with respect to the equality constraints as follow

$$\begin{aligned} \alpha_i &\geq 0 \\ i &= 1, 2, \dots, 1 \\ \sum_{i=1}^1 \alpha_i y_i &= 0 \end{aligned}$$

Note that the dual Lagrangian $L_d(\alpha)$ is expressed in terms of training data and depends only on the scalar products of input patterns $(x_i^T x_j)$. The goal of SVM is to produce a model which predicts target values of data instances in the testing set which are given only the attributes.

3.6 Kernel Selection of Support Vector Machine (SVM)

Training vectors x_i are mapped into a higher (may be infinite) dimensional space by the function Φ . Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimension space. $C > 0$ is the penalty parameter of the error term.

Furthermore, $K(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$ is called the kernel function. There are many kernel functions in SVM, so how to select a good kernel function is also a research issue. However, for general purposes, there are some popular kernel functions which include the following:

- i) Linear kernel: $K(x_i, x_j) = x_i^T x_j$.
- ii) Polynomial kernel: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- iii) Radial Basis Function (RBF) kernel: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$
 Where, $\gamma > 0$
 Sigmoid kernel: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Here, γ , r and d are kernel parameters. In these popular kernel functions, RBF is the main kernel function because of the following reasons.

- 1) The RBF kernel nonlinearly maps samples into a higher dimensional space unlike to linear kernel.
- 2) The RBF kernel has less hyperparameters than the polynomial kernel.
- 3) The RBF kernel has less numerical difficulties.

3.7 Training and Testing Data

Training data and test data are two important concepts in machine learning. Observations in the training set form the experience that the algorithms use to learn. In supervised machine learning problems, each observation consists of an observed output variable and one or more observed input variables.

The test set is use is a set of observations used to evaluate the performance metric. It's important that no observations from the training set are included in the test set. If the test set does contain examples from the training set, it will be difficult to assess whether the algorithms has learned to generalize from the training set or has simply memorized it. Randomize split of 70% to 30% was done where 70% of the total samples was used as trained data to prepare the models and the remaining 30% random sample was used as test data to predict the measures of the model performance.

3.8 Algorithms Performance Evaluation

The algorithm evaluation is mostly done by predicting accuracy and the most widely use techniques in machine learning approach model is confusion matrix. Confusion matrix is an N by N matrix used for evaluating the performance of classification model, where N is the number of target classes. The matrix compares the actual target values with those being predicted by the machine learning model .

It's a table with four different combinations of predicted and actual values in the case for a binary classifier. A true positive is an outcome where the model predicted correctly the positive class and true negative is an outcome where the model predicted correctly negative class.

3.8.1 Metrics Computed from a Confusion Matrix

The following metrics are calculated from the confusion matrix as follows:

ACCURACY: This measures the proportion of correct predictions made by a machine learning model and mathematically define by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

SPECIFICITY (TRUE NEGATIVE): Measures the proportion of actual negative instances that are correctly predicted by a machine learning model and its define as:

$$Specificity = \frac{TN}{TN + FP}$$

SENSITIVITY (TRUE POSITIVE) OR RECALL: Its measure the proportion of actual positive instances that are correctly predicted by a machine learning algorithms and is given by:

$$Sensitivity = \frac{TP}{TP + FN}$$

Where,;

TP= True Positive: Correctly Predicted Positives instances

TN= True Negative: Correctly Predicted Negatives instances

FP= False Positive: Incorrectly Predicted Positives instances

FN= False Negative: Incorrectly Predicted Negatives instances

4. Results and discussion

This goes into great detail about the findings and conversations derived from the research. The first part of the conversation looks at the connections between Nigerian child mortality and a number of other factors. then the modeling process and model performance comparison. A major public health concern in many developing nations, including Nigeria, is child mortality. Nigeria still has a high child mortality rate, especially in rural regions, despite efforts to lower it. Child mortality and other demographic, economic, and health-related factors were predicted and categorized in this study using techniques from Artificial Neural Networks (ANN), Discriminant Analysis (DA), Naïve Bayes (NB), and Support Vector Machines (SVM). It should be noted that the data utilized in this study was not balanced; instead, the oversampling approach was used to balance the data in order to get a non-bias result. Nevertheless, the results of this investigation can potentially offer important new information about the causes of child death in Nigeria.

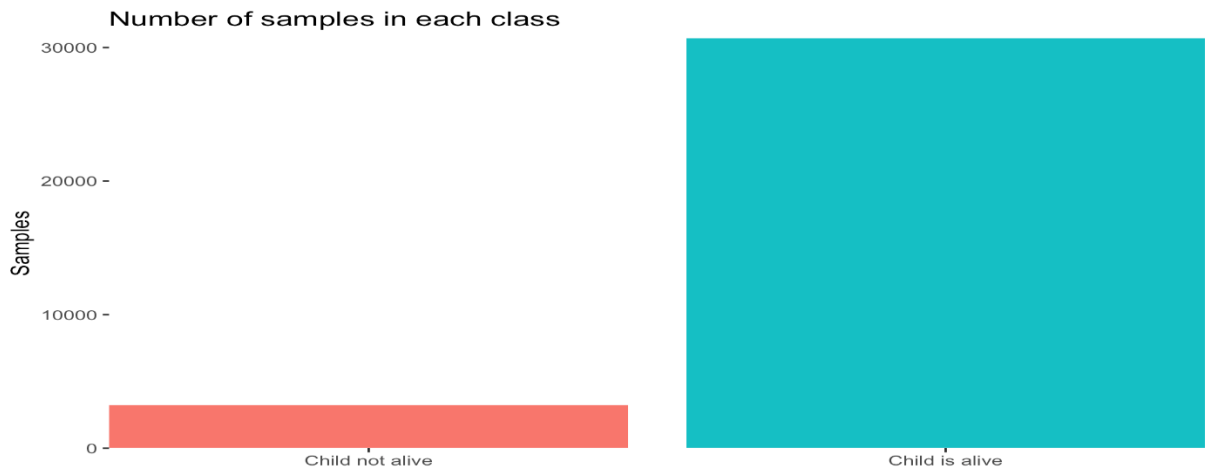


Figure 4: Presentation of Number of sample in each Class (NDHS 2018).

Considering that the "Child is alive" class has more samples than the "Child not alive" class, it would seem from the plot above that the former is the dominant class. The implication is that there are more cases of a child being alive than dead in the dataset under analysis. It might be crucial to take into account the disparity in sample sizes between the two groups when interpreting the findings of any further research. For instance, if a machine learning model is trained using the data to forecast child mortality, the model can be skewed to predict the dominant class that is, the child who is alive more accurately than the minority class that is, the child who is not alive.

As a result, when evaluating the findings of any further study, it is crucial to take the class distribution into account and to take action to correct any imbalances that can compromise the veracity of the findings. As a result, oversampling was used to balance the data.

4.1 Artificial Neural Network (ANN)

The Artificial neural network technique is applied to the same data used to check the performance against the other models.

Table 1: Artificial Neural Network (ANN) for under-five Mortality (NDHS 2018).

Metric	Specificity	Sensitivity	Accuracy	AUC
Score	0.9052	0.5223	0.7138	0.5806

The ANN model shows moderate performance, with an accuracy of 71.38%. It performs better in correctly identifying children who are not alive (specificity of 90.52%) than in identifying children who are alive (Sensitivity of 52.23%).

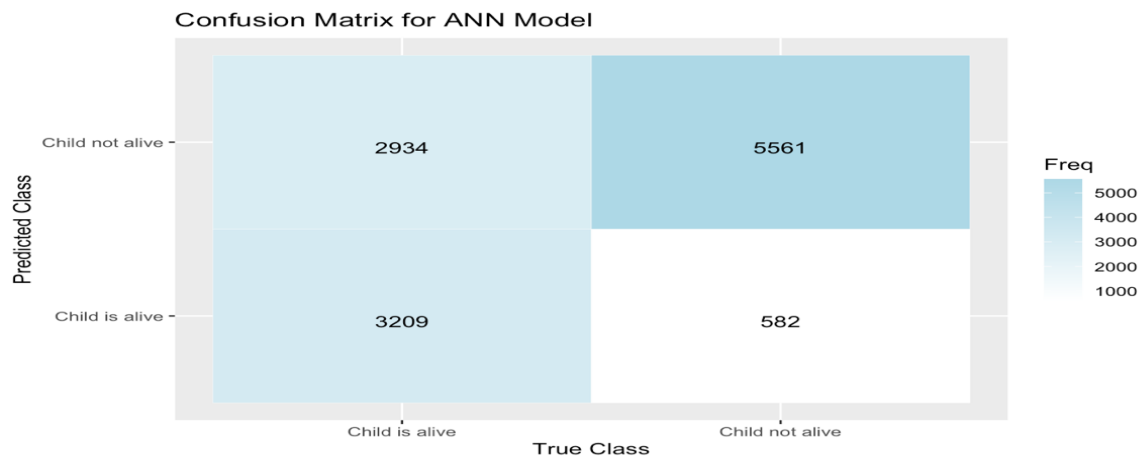


Figure 5: Confusion matrix for the ANN classifier for under-five Mortality (NDHS 2018).

From the above provided confusion matrix for ANN model that predicts whether a child is alive or not, we can see that the model correctly predicted 3209 children as being alive (true positives) and correctly identified 5561 children as not being alive (true negatives). However, the model also made 582 false positive predictions (predicted that a child is alive when they are not actually alive) and 2934 false negative predictions (predicted that a child is not alive when they are alive).

4.2 Linear Discriminant Analysis

The Linear Discriminant Analysis was also implemented to check the performance against the other models, table 2 shows the metrics score computed

Table 2: Linear Discriminant Analysis for under-five Mortality (NDHS 2018).

Metric	Specificity	Sensitivity	Accuracy	AUC
Score	0.6222	0.6274	0.6248	0.7056

The model has a moderate specificity (62.22%) and sensitivity (62.74%), indicating that it performs reasonably well in detecting when a child is alive and when a child is not alive. However, the performance is not particularly strong in either aspect. The accuracy of 62.48% suggests that the model has room for improvement. The relatively balanced but moderate sensitivity and specificity indicate that the model struggles with both types of errors (false positives and false negatives) to some extent.

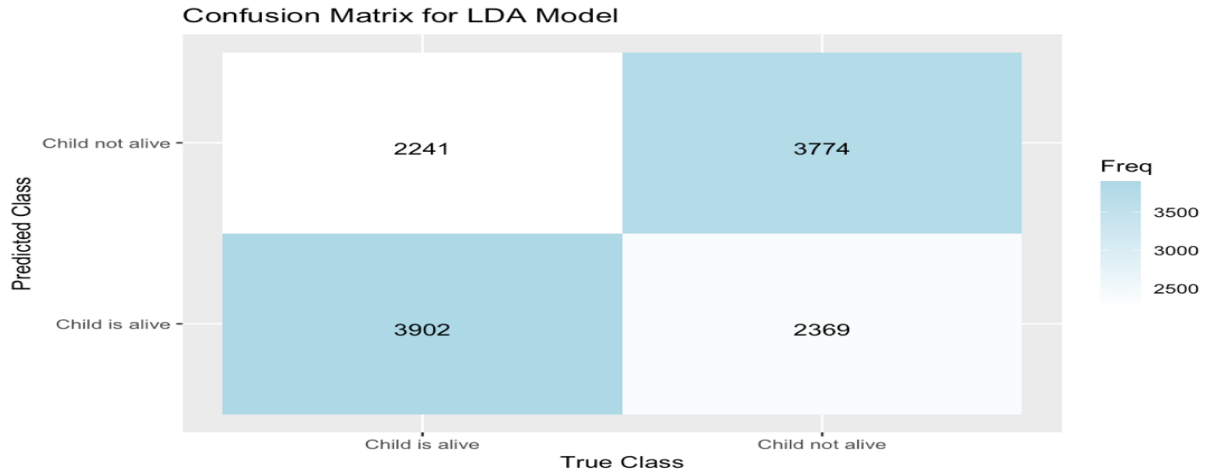


Figure 6: Confusion matrix for the LDA classifier for under-five Mortality (NDHS 2018).

The confusion matrix for Linear Discriminant model that predicts whether a child is alive or not, showed that the model correctly predicted 3902 children as being alive (true positives) and correctly identified 3774 children as not being alive (true negatives). However, the model also made 2241 false positive predictions (predicted that a child is alive when they are not actually alive) and 2369 false negative predictions (predicted that a child is not alive when they are alive).

4.3 Naïve Bayes

The Naïve Bayes was also implemented to check the performance against the other models.

Table 3: Naïve Bayes Metrics for under-five Mortality (NDHS 2018).

Metric	Specificity	Sensitivity	Accuracy	AUC
Score	0.9347	0.7296	0.8322	0.9108

The Naïve Bayes model has a very high specificity (93.47%), indicating that it is particularly strong in correctly identifying when a child is not alive. Although the sensitivity is relatively high at 72.96%, the model still misses nearly a quarter of the cases where the child not is alive. The high accuracy (83.22%) suggests that the Naïve Bayes model performs well overall, mainly when the goal is to minimise false positives.

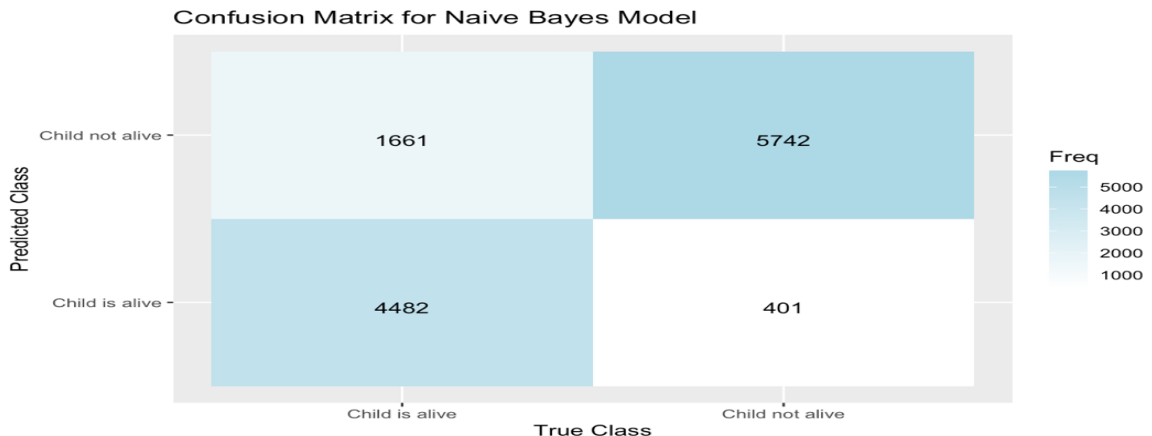


Figure 7: Confusion matrix for the Naïve Bayes classifier for under-five Mortality (NDHS 2018).

Naïve Bayes confusion matrix model correctly predicted 4482 children as being alive (true positives) and correctly identified 5742 children as not being alive (true negatives) and the model also made 401 false positive predictions (predicted that a child is alive when they are not actually alive) and 1661 false negative predictions (predicted that a child is not alive when they are alive).

4.4 Support Vector Machine

Support Vector Machine algorithms is applied to the same data used to check and compare the performance against the other models.

: Support Vector Machine Metrics for under-five Mortality (NDHS 2018).

Metric	Specificity	Sensitivity	Accuracy	AUC
Score	0.9992	0.9168	0.9543	0.9876

The SVM model has an exceptionally high specificity (99.92%), almost perfectly predicting true positives. This makes it highly reliable in correctly predicting when a child is not alive. The high sensitivity (91.68%) shows that the model is also very effective at correctly identifying when a child is alive. The model's accuracy of 95.43% is quite impressive, indicating that it performs very well overall across both classes (alive and not alive). The model is particularly strong at minimizing false positives and negatives, contributing to its high accuracy.

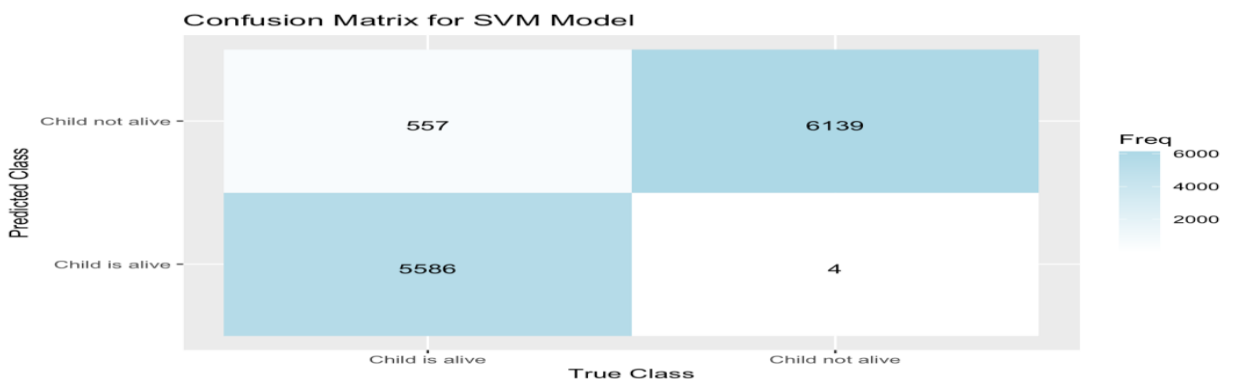


Figure 8: Confusion matrix for the support vector classifier for under-five Mortality (NDHS 2018).

Confusion matrix for above model that predicts whether a child is alive or not, we can see that the model correctly predicted 5586 children as being alive (true positives) and correctly identified 6139 children as not being alive (true negatives). However, the model also made 4 false positive predictions (predicted that a child is alive when they are not actually alive) and 557 false negative predictions (predicted that a child is not alive when they are alive).

Table 5: Models Comparison Table for under-five Mortality (NDHS 2018).

Model	Specificity	Sensitivity	Accuracy	AUC
Artificial Neural Network	0.9052	0.5223	0.7138	0.5806
Linear Discriminant Analysis	0.6222	0.6274	0.6248	0.7056
Naïve Bayes	0.9347	0.7296	0.8322	0.9109
Support Vector Machine	0.9992	0.9168	0.9543	0.9876

Support Vector Machine (SVM) and Naïve Bayes emerged as the top performers among the four evaluated models based on their accuracy follows by Artificial Neural Network and lastly Linear Discriminant analysis models with accuracy values reaching 62.48%. This indicates their strong ability to correctly identify cases where a child is not alive while maintaining a moderate level of sensitivity 62.74% for detecting cases where a child is alive and Lastly Support vector machine model outperformed all other models by achieving perfect scores across all metrics, suggesting that it can accurately classify every instance in the dataset.

5. Conclusion

The study used 2018 demographic and health survey data to compare the performances of some selected supervised learning algorithms. Four distinct classification techniques for supervised machine learning were compare. Artificial Neural Network, Naïve Bayes, Linear Discriminant Analysis and Support Vector Machine, the result showed that Support Vector Machine (SVM) and Naïve Bayes emerged as the top performers among the four evaluated models based on their accuracy followed by Artificial Neural Network and lastly Linear Discriminant Analysis (LDA) models with accuracy values reaching 62.48%. This indicates their strong ability to correctly identified cases where a child is not alive while maintaining a moderate level of sensitivity 62.74% for detecting cases where a child is alive and Lastly Support vector machine model outperformed all other models by achieving perfect scores across all metrics, suggesting that it can accurately classified every instance in the dataset.

The research recommends that elements that have been discovered to contributed to the increase in under-five mortality should be given more attention in order to bring it down to the lowest possible level.

Government should subsidize the cost of health care services at the grassroots to increase patronage. Government should create more awareness on the benefits of immunization, encourage longer period of breastfeeding and birth spacing, sensitize women more on the safety and importance of delivering their babies under the supervision of a trained healthcare professional, particularly for nursing mothers and

finally interventions are needed and should focus on children residing in low socio economic households with inadequate access to primary health care services.

References

- Aayush, K. S., Devaraj, P., Sneha, B., and Ramra, S. (2016). Application of machine learning in analysis of infant mortality and its factors. <https://www.researchgate.net/publication/303974365>, DOI: 10.13140/RG.2.1.3857.3687
- Abdulwahaba.S. and Bahaa.Y.B.(2023). Artificial Neural Network Prediction Model for Maternal Health Services Quality in Nigeria. *International journal of development mathematics*, ISSN: 3026-8656 (Print) | 3026-8699 (Online).
- Amanuel, K.A., Henonk, G., Woldu, M., E., Zhongha, Z. and Lingxia Z.(2022). Trends of under-five mortality and associated risk factors in Zambia: a multi survey analysis between 2007 and 2018. *Andegiorgish et al. BMC Pediatrics (2022) 22:341*
- Ayele, D., Temesgen, T., Zewotir. and Mwambi., H. (2017). Survival analysis of under-five mortality using Cox and frailty models in Ethiopia. *Journal of Health, Population and Nutrition (2017) 36:25 DOI 10.1186/s41043-017-0103-3*.
- Azuike, E.C., Onyemachi, P. E., Amah, C.C., Okafor., K.C., Anene., O., Enwonwu, K.G., Aniemen, R.C., Arua, N.E. and Ilika, A. L. (2019). Determinants of under-five mortality in south-eastern Nigeria. *MOJ Public Health*.
- Brnabic, A. and Lisa M., H. (2021). Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC Med Inform Decis Mak (2021) 21:54* <https://doi.org/10.1186/s12911-021-01403-2>
- Caroline., K. (2021). Supervise Learning approach to predict infant mortality in Kenya. [Masters thesis, University of Nairobi Kenya]
- Chen, R.C., Dewi, C., Huang, S. and Caraka, R., E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big data*. <https://doi.org/10.1186/s40537-020-00327-4>.
- Daniel, A., Nazeem, M., Pammla, P. and Elon, W., I. (2020). Inequities in child survival in Nigerian communities during the Sustainable Development Goal era: insights from analysis of 2016/2017 Multiple Indicator Cluster Survey. *BMC Public Health (2020) 20:1613*, <https://doi.org/10.1186/s12889-020-09672-8>.
- Deprez, P., Pavel., V., Shevchenko., Mario, V. and Wuthrich. (2021). Machine Learning Techniques for Mortality Modeling. *ArXiv: 1705.03396v1 [stat.AP] 7 May 2017*
- Desalegn, B. and Getu., M. (2021). Determinants of Diarrhea in Under-Five Children among health Extension Model and Non-Model families in Wama Hagelo District, Western Ethiopia. Community Base Comparative Cross-Sectional Study. *Journal of Multidisciplinary Healthcare. 2021;14:2803-2815*. doi:10.2147/JMDH.S324846.
- Dominic, E., Azuh., A., Chinedu, S. and Samuel., O. (2019). Factors influencing the survival of under-five children among women visiting government health care facility in semi-urban communities in Nigeria. *Article in Cogent Arts & Humanities · November 2019*. DOI: 10.1080/23311983.2019.1686800.
- Florkowski, C., (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev. 2008 Aug; 29 Suppl 1(Suppl 1): S83-7. PMID: 18852864; PMCID: PMC2556590*.
- Gobir, A. A., Joshua, I.A., Audu, O., Salaudeen, A.G., Bako, I.A. and Adeyemi, A.O. (2015). Social determinants of morbidity among under-fives in a rural community of north-western Nigeria. *JMBR: A Peer-review Journal of Biomedical Sciences, December 2015, Vol. 14 No. 2 pp 40-49*.

- Guido,R, Ferrisi,S, Lofaro,D. and Conforti,D (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: *Information* 2024, 15, 235. <https://doi.org/10.3390/info15040235>
- Huda,M., Haroun, M., Mahfouz,S. and Khalid, H., I.(2007).Level and determinants of infant and under-five mortality in wad-medani town, Sudan. *Journal of family community medicine*, 14(2): 65–69.
- Janardhanan,P., Heena,L. and Sabika, F.(2015).Effectiveness of Support Vector Machines in Medical Data mining: *journal of communications software and systems*, vol. 11, no. 1, march 2015.
- Johansen, B. C. (2018). Machine learning algorithms application in predicting Children mortality: A model development. *International Journal of information Sciences and Application*. ISSN 0974-2255 Volume 10, Number 1(2018). pp.1-6, <http://www.irphouse.com>
- Justice,M.A.K., (20219). Predictive model and determinants of under-five child mortality: evidence from the 2014 Ghana demographic and health survey. *BMC Public health* 19, Article number: 64 (2019).
- Khan, M.A., Khan, N., Rahman, O., Mustagir, G., Hossain, K.and Islam, R. (2021). Trends and Projections of under-5 mortalities in Bangladesh including the effects of maternal high-risk fertility behaviors and use of healthcare services. *PLoS ONE* 16(2): e0246210.
- Lucia,H., Lee, S., Liu, Y., Mishra, A., Sharrow, D. and You, D.(2020). *Levels & Trends in Child mortality Estimates developed by the UN Inter-agency Group for Child Mortality Estimation.Report (2020)*. United Nations Children's Fund ISBN: 978-92-806-5147-8.
- Masrie,G., Abate,D. and Abebaw, A.(2020).Proximate determinants of infant mortality in Ethiopia, 2016 Ethiopian demographic and health surveys. *Archives of Public Health* volume 78, Article number: 4 (2020).
- Mohamed, A., Hussein, M. M. and Doaa, H. (2021). Determinants of Under-Five Mortality: A Comparative Study of Egypt and Kenya. *Open Access Library Journal* Volume 8, e7889 ISSN Online: 2333-9721 ISSN Print: 2333-9705.
- Mustafa,D.A. and Mohsin,A..A.(2021).A review on Machine Learning Applications based on SVM Classification: *Qubahan aacademic journal* <https://doi.org/10.48161/qaj.v1n2a50>.
- National Population Commission (NPC) [Nigeria] and ICF. 2019. *Nigeria Demographic and Health Survey 2018*. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF.
- Osita,K.,Felix,A.,Anastasia,O.,Odumegwu,H.,Oforkansi,D.,Abada,CTanko, I. and .,Kingsley.(2021).Under-5 Mortality and its Associated Factors in Northern Nigeria.*Intertional journal of enviromntal research and public health*. *Pediatric Critical Care, a section of the journal Frontiers in Pediatrics*.doi: 10.3389/fped.2022.797080
- Pienaar,A.M.,Joseph,B. S., Luwes,N.and Lincoln, J. S.(2022) An Artificial Neural Network Model for Pediatric Mortality Prediction in Two Tertiary Pediatric Intensive Care Units in South Africa. A Development Study:
- Rustam,F., Reshi,A.A.,Mehmood,A.,Ullah,S.,Byungwonon, A..W. and Choi,G.S.(2020).COVID-19 Future Forecasting Using Supervised Machine Learning Models :*IEEE Access* 10.1109/ACCESS.2020.2997311,
- Sanni,Y.,Michel,E., Godson,T., Shah,V.,Ghose,B. and Bernard, K. (2017).Prevalence and determinants of childhood mortality in Nigeria. *BMC Public Health* (2017) 17:485.
- Sreehari, J. (2020). Stock Market Trend and Value Prediction, *International Research Journal of Engineering and Technology (IRJET)* e-ISSN: 2395-0056. Volume: 07.
- Srivastava,D., Bhambhu,L. (2019). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*.
- Vapnik. V (1995). The Nature of Statistical Learning Theory. NY: Springer-Verlag.