American Journal of Applied Mathematics and Statistics, 2021, Vol. 9, No. 2, 38-47
Available online at http://pubs.sciepub.com/ajams/9/2/1
Published by Science and Education Publishing
DOI:10.12691/ajams-9-2-1

Two-Stage Artificial Neural Network Regression Modelling for Wheezing Risk Factors Among Children - A Case Study of Gatundu Hospital, Kenya

Thomas Mageto*, John Onyango

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Kenya
*Corresponding author: tttmageto@gmail.com

Received February 19, 2021; Revised March 22, 2021; Accepted March 30, 2021

Abstract In Kenya wheezing that leads to asthma development in most cases remain under-diagnosed and under-treated. Currently there is no public supported wheezing and asthma care programmes to optimize care for patients with asthma which greatly compounds diagnosis and treatment of the disease. The aim of this study is therefore to consider and analyse the covariates of childhood wheezing among children below 10 years of age in Kenya, a case study of Gatundu hospital in order to improve the provision of wheezing and asthma care services in medical facilities. The possible risk factors in the study are selected from three major groups of demographic, socioeconomic and geographical location factors related to childhood wheezing. The longitudinal secondary data obtained from Gatundu hospital in Kenya were collected and a total of 584 complete cases were recorded. The predictor variables considered in the study include age of children in months, gender, exclusive breastfeeding, exposure to tobacco smoking, difficult living conditions, residence, atopy, maternal age and preterm births. Due to the binary nature of response variable in which data is recorded as presence or absence of wheezing, the risk factors were modelled using multiple logistic regression and Artificial Neural Network Models. Simple random samples of sizes n = 385 without replacement were selected and p-values at 5% level of significance for the variables were recorded. In multiple logistic regression, the five variables identified as possible risk factors for modelling with p-value less than or equal to 0.05 were selected that includes age of children, exclusive breastfeeding, exposure to tobacco smoking, difficult living conditions and residence that recorded p-values of 0.0151, 0.0000, 0.0071, 0.0274 and 0.0410. The best multiple logistic linear regression model selected was based on Akaike Information Criterion (AIC) criterion that recorded null deviance, residual deviance and AIC of 502.44, 179.57 and 191.57 respectively. The precision and accuracy of the multiple logistic regression model were recorded as 89.2% and 93.3% respectively. The Artificial Neural Network was considered for modelling as well, the model with one-hidden layer with four neurons in the hidden layer recorded precision of 97.1% and accuracy of 39.4% while the rest of the models with one hidden layer recorded precision and accuracy of 0.0% and 65.1% respectively. The Artificial Neural Network model with two-hidden layers were also considered and the Network with one neuron in both layers was selected as better performing model with precision and accuracy of 88.2% and 93.3%. The developed two-stage logistic Artificial Neural Network was found to have better performance compared to multiple linear logistic regression and Artificial Neural Networks since it recorded precision and accuracy of 97.1% and 99.0% respectively and hence recommended for consideration in modelling the risk factor of wheezing among children in Kenya.

Keywords: *logistic, wheezing, artificial neural network, two-stage regression, modelling*

Cite This Article: Thomas Mageto, and John Onyango, "Two-Stage Artificial Neural Network Regression Modelling for Wheezing Risk Factors Among Children - A Case Study of Gatundu Hospital, Kenya." *American Journal of Applied Mathematics and Statistics*, vol. 9, no. 2 (2021): 38-47. doi: 10.12691/ajams-9-2-1.

1. Introduction

1.1. Background

The contagious diseases continue to be the outstanding cause of morbidity and mortality for the population of Kenya but non-contagious diseases are also progressively enhancing to be a dominant element of ill health and death among children. The non-contagious diseases like asthma whose development is attributed to wheezing is the commonest chronic lung disease in the world affecting up to 18% of the world population and remains under-diagnosed and under-treated leading to poor quality of life and enormous social, family and economic costs thus a leading burden. Wheezing that develops to asthma and other obstructive airways diseases are non-contagious diseases that are increasingly becoming a treat among children in Kenya and therefore becoming a lifelong consequence for the sick, enormous health resource utilization and a substantial economic burden to the country at large [1]. The most common symptoms of asthma are wheezing, an expiratory breathing difficulty that is continuous, musical, whistling like and it is common during lower respiratory tract infection (LRTI) among children. Its heterogeneous nature renders it a complex predicament among children in which the patient experiences whistling (sound when breathing) breathlessness and tight chest that may feel like a band is tightening around it. Heterogeneity in the sense that it has a peculiar pathogenic mechanism variation in duration and severity. Infants whose wheezing condition warrants hospitalization are at the danger of developing asthma later in life and it has been established that of all children, at least a third of them often suffer from wheezing by three years of age and those whose wheezing warrant hospitalization, are about 2% [2]. Small airway size and or dysfunction in the regulation of airway tone seem to increase the risk of wheezing during lower respiratory tract infection (LRTI) in infancy and in childhood. The epidemiology of asthma, prevalence and risk factors have not been fully described in Kenya and therefore need to comprehensively carry out the study [3].

The relation between development of asthma and childhood wheezing will be a consequence of variations, system maturations and acquired modifications of the airway. Crucial determinants of the ultimate aftermath after early wheezing in childhood include host, viral and other factors of the environment [4]. Prospective studies have been conducted and several risk factors for early wheezing among children and subsequent development of asthma have been suggested that include history of allergy in family, maternal smoking during pregnancy or passive exposure after birth, personal history of rhinitis or eczema, gender, viral respiratory infections by respiratory syncytial virus (RSV) or rhinovirus and attending day care and use of medications such as paracetamol and antibiotics during infancy [5].

There is an increased incidents of asthma symptoms in overweight or obese female of school going ages. In the ages of 6 – 11 years of the female gender there is an increase of the risk of developing new asthma symptoms and increased bronchial responsiveness [6,7]. In most studies there is deficient in either determination of allergen-specific sensitization in infancy or the lie long follow up in revealing atopy becoming evident after infancy. This has been realized by most studies on early life risk factors for asthma, clinical or non-specific markers of atopy for example, atopic dermatitis or elevated entire serum immunoglobulin. Recent epidemiological evidence suggest that exposure to animals in early life cuts down subsequent allergy and asthma and accumulating evidence suggest that the environment microbiome contributes significantly to the development of asthma among children [8]. In line with this, being exposed to pets during the wheezing period in infancy has not raised the risk of asthma in later life. In addition, babies born pre-maturely are considered to be at a high risk of developing shortage in percentage forced expiratory volume in one second and have projected the number of admissions to hospital for respiratory related diseases when compared to children born after 37 weeks of pregnancy [9]. It is then evident that this study will explore the suggested risk factors and other emerging risk factors in the study to develop a foundation for the modelling of risk factors in wheezing among children in Kenya.

1.2. Statement of the Problem

In Kenya wheezing remain underdiagnosed and undertreated, since there is no public supported wheezing care programmes designed to optimize care for patients with asthma which greatly compounds the diagnosis and treatment. The children who are less than 10 years of age are more vulnerable due to high risk of wheezing and interventions targeting these children depend on prevalence of wheezing among children in the country as this contributes to the number of children and infants that need to be reached for testing and enrolled into care and retained on treatment programmes. This study will explore the risk factors associated with wheezing among children and develop a classification model for wheezing risk factors and suggest necessary interventions that will improve diagnosis and care among children to regulate the number of new wheezing cases which in turn lead to development of asthma which is a major cause of death among children in Kenya.

1.3. Hypothesis

i) H_0: There exist no relationship between demographic factors and wheezing among children.

H_1: There is a relationship between demographic factors and wheezing among children.

ii) H_0: There is no relationship between socio-economic factors and wheezing among children.

H_1: There is a relationship between socio-economic factors and wheezing among children.

iii) H_0: There is no association between geographical factors and wheezing among children.

H_1: There is a relationship between geographical factors and wheezing among children.

1.4. Objectives

1.4.1. General Objective

The general objective of the study is to model the risk factors of wheezing among children in Kenya: A case study of Gatundu hospital.

1.4.2. Specific Objectives

i) To determine the wheezing risk factors among children in Kenya: A case study of Gatundu hospital.

ii) To model the wheezing risk factors among children using multiple linear logistic regression models.

iii) To model the wheezing risk factors among children using Artificial Neural Network models.

iv) To model the wheezing risk factors among children using Two-Stage Artificial Neural Network regression models.

v) To determine the performance of the developed Two-Stage Artificial Neural Network regression model for classification of wheezing cases among children.

1.5. Justification of Study

The wheezing among children rates in Kenya remain high and poses a great treat in the lives of children and therefore this study aims at investigating and determining the wheezing risk factors in order to make a contribution in proposing the measures to be taken to reduce the risks of developing wheezing and its complications among children that leads to development of asthma and becomes a burden in stretching the medical resources and threat to the lives of children.

2. Literature Review

In the recent past several studies have been conducted in modelling the wheezing risk factors among children. A study has been carried considering multivariate analyses of wheezing among infants and recorded p value of more than 0.20 for the likelihood ratio test for inclusion. In the study, the risk factors from within the parental were considered then incorporated the allergic disposition and environmental exposure groups. The conclusion from the study was that the majority of children who wheeze in the first 6 months do not have the condition later after three years of age [10].

A cross-sectional study carried in Brazil to investigate wheezing in infants of 12 – 15 months used bivariate and multivariate analysis in which the results were expressed as odds ratio and 95% confidence intervals. The analysis revealed that history of asthma in family, history of previous pneumonia, having had more than six upper respiratory tract infections, living moderately polluted area, paracetamol use for URTI and antibiotic use for skin infection were possible risk factors of wheezing among infants [5].

A study carried in Canada examined the association of breastfeeding and wheezing in the first year of life for 2773 infants in which records of episodes were collected at 3, 6 and 12 months. In the study exclusive breastfeeding and partial breastfeeding were considered in which 21% of mothers recorded cases of asthma, the mothers who breastfed for at least 12 months were 46% and 21% of the children experienced wheezing. Among the mothers who reported cases of asthma, the breastfeeding factor was found to be inversely associated with infant wheezing independent of maternal smoking and education [11].

A longitudinal cohort study undertaken in China in which children of 6 – 72 months were considered and recruited in the study with at least three episodes of doctor-diagnosed wheezing who had history of congenital pulmonary airway malformation, bronchopulmonary dysplasia, trachea cannula, severe pneumonia, severe immunodeficiency disease and cardiovascular disease. The results indicated that the incidence of wheezing declined overall with age, rhinitis ever and aeroallergens sensitization and previous RV infection were found to be the most significant risk factors for persistent wheezing in children [12].

In USA investigation conducted to determine the prevalence and risk factors for asthma and wheezing among US adults were carried. A multiple logistic regression analysis were conducted and the results showed that low education level, female gender, current and past smoking status, pet ownership, lifetime diagnosis of physician diagnosed hay fever and obesity were all significantly associated with wheezing [13]. It was also observed that the use of household heating or cooking appliances were not significant on wheezing among adults. A similar study from Danish medical birth registry considered the maternal or paternal history of fever, sex, smoking during pregnancy occupation, income, season of birth, head circumference and gestation age. The effects of the risk factors were evaluated using the logistic regression models and independent effects of risk factors were estimated using the logistic regression in which the response was recorded as whether or not the child had wheeze ever or recurrent wheeze or atopic dermatitis, the data was collected in the fourth interview when child is 18 months, and mother asked whether the child had ever had an itchy rash and if so about localization and age when affected [14].

3. Methodology

3.1. Study Design

The cross-sectional study design was carried on data sets collected from Gatundu hospital in Kiambu County, Kenya. A total of 584 complete cases observations were taken from the medical records in which 204 were doctor-diagnosed wheeze cases whereas 380 were controls. The controls were at some point exposed to the risk factors under the study. Some of the variables were not categorical but were coerced to be categorical variables, for instance, maternal age. The sample size considered in the study was computed using the sample size determination relation given as [15]

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{d^2} = \frac{1.96^2 0.5(0.5)}{0.05^2} = 385 \tag{1}$$

Where n is the sample size, p is the optimal probability of success, d is the error of proportion of success and $z_{\alpha/2}^2$ is the variate from a standard normal distribution at α = 0.05 level of significance.

3.2. Multiple Logistic Regression Model

In conducting statistical evaluation on the determinants that immensely impact on wheezing among children, the multiple linear logistic regression model was considered that take the general equation of the form;

$$P(Y=1) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} \tag{2}$$

The distribution of our response variable Y now specified by the probabilities P(Y = 1) as success and P(Y = 0) as failure [16]

Such that

$$\frac{P(Y=1)}{1-P(Y=1)} = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}$$

Taking natural logarithms both sides we then have

$$\ln \frac{P(Y=1)}{1-P(Y=1)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \quad (3)$$

Where Y is the response variable, X_1, X_2… X_p are the predictors while β_0, β_1... β_p are the unknown constants.

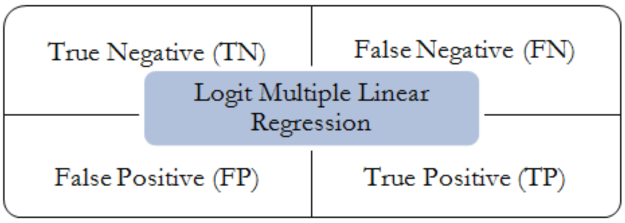

Figure 1. Binary Confusion Matrix

The unknown parameters are estimated using maximum likelihood method and the results are interpreted using the log-odds such that the log-odds greater than 0.5 implies there is an increased occurrence of wheezing while that of less than 0.5 implies decreased occurrence of wheezing in among children. The performance of the model is better assessed using the confusion matrix in which the parameters are classified in binary represented in a matrix format as in Figure 1 [17]. The accuracy and precision are commonly used to determine the performance of the model as given below.

$$Accuracy = \frac{TN+TP}{TN+FN+FP+TP} \quad (5)$$

$$Precision = \frac{TP}{FP+TP}. \quad 6)$$

3.3. Artificial Neural Network Model

Artificial Neural Network (ANN) is a machine learning approach that models similar to human brain consisting of a number of artificial neurons that tend to have fewer connections than biological neurons [18]. The neurons receives a number of inputs denoted by X_1, X_2, ..., X_p and the activation function is applied to these inputs and resulting output denoted by Y as shown in Figure 2 for one hidden layer with one neuron in the hidden layer and Figure 3 for one hidden layer with four neurons in the hidden layer.

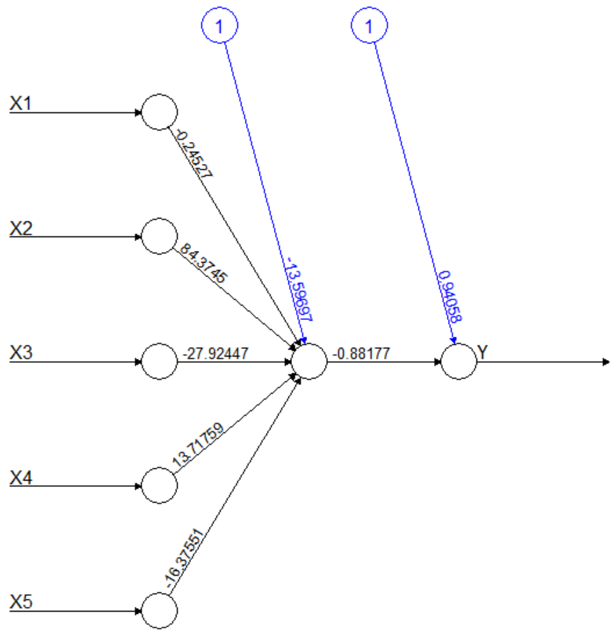

Figure 2. Artificial Neural Network (one hidden layer with one neuron)

The Artificial Neural Network can be represented using the model given as [19]

$$Y_0(t) = b_0 + \sum_{j=1}^{q} w_{0j}(t) \left\{ b_{1j} + \sum_{i}^{p} w_{1i}(t) x_i(t) \right\} \quad (7)$$

$$Z_0 = \Gamma(Y_0) \quad (8)$$

Where x_i ’s are the input, w_i are the weights of interconnections, b_0 is the bias on output layer, b_1 is bias on the single hidden layer, q is the number of neurons in single hidden layer, p is the number of input variables, Γ is the activation function that may be the linear, logistic sigmoid, bipolar sigmoid, tangent hyperbolic or radial basis functions [20].

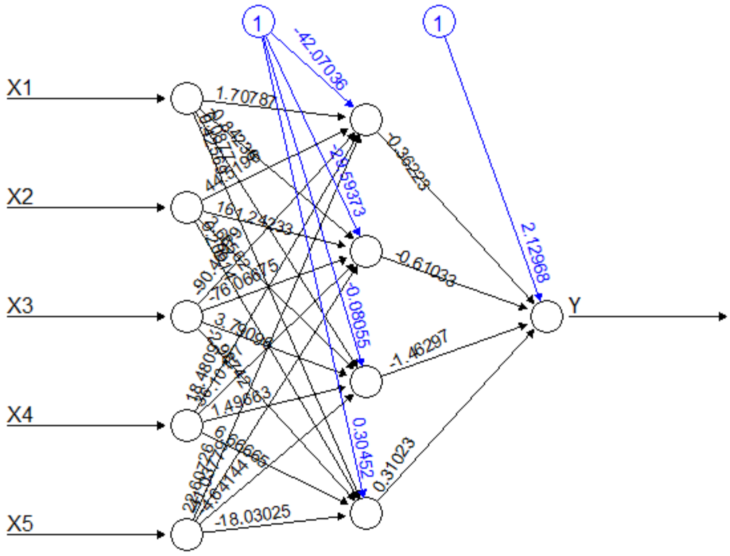

Figure 3. Artificial Neural Network (one hidden layer with four neurons)

3.4. Two-Stage Artificial Neural Network Model

In the two-stage Artificial Neural Network (ANN) model we have a model that takes into consideration the Artificial Neural Network model and incorporates with results that have already been obtained in previous model and thus improving the model that is now referred to as Two-Stage Artificial Neural Network modes. This is also a machine learning approach in which the neurons receives a number of inputs denoted by $X_1, X_2, \ldots, X_{p*}$ and the activation function is applied to these inputs and resulting output denoted by Y^*. The two-stage Artificial Neural Network model in machine learning can be represented using the model given as

$$Y_0^*(t) = \begin{cases} b_{00} + \sum_{j=1}^{q} w_{00j}(t)\left\{ b_{01j} + \sum_{i}^{p*} w_{01i}(t)\, x_i(t) \right\} \\ b_{10} + \sum_{j=1}^{q} w_{10j}(t)\left\{ b_{11j} + \sum_{i}^{p*} w_{11i}(t)\, x_i(t) \right\} \end{cases} \tag{9}$$

$$Z_0^* = \begin{cases} \Gamma\left(Y_0^*\right) \\ \Gamma\left(Y_1^*\right) \end{cases} \tag{10}$$

Where x_i's are the input p* variables that are significant upon fitting the multiple logistic regression analysis, w_i are the weights of interconnections on the input variables, w_j are the weights of interconnections for the neurons in the hidden layer, b_0 is the bias for the output, b_1 is the bias for the neurons in the hidden layer while Γ is the activation function [20].

3.4.1. Types of Neural Networks

There are three main types of neural networks that include the feed forward, radial basis function and recurrent neural networks [21]. In the feedforward neural network the input layer is connected to the hidden layer and this hidden layer is connected to another hidden layer if any otherwise it is connected to the output layer. The radial basis function are feed forward networks trained using supervised training algorithm typically configured with a single hidden layer that usually train much faster than back propagation networks. The recurrent neural network on the other hand contains at least one feedforward connection and possibly neurons with the self-feedback links. In this research the feed forward neural network has been selected due to its simple architecture that has minimum repetitions and therefore limits overfitting that saves in computation time and resources.

3.4.2. Number of Hidden Layers and Number of Neurons in Hidden Layer

In many artificial neural networks there are at least one hidden layer and it is rare to have more than two hidden layers [21]. The selection of the number of neurons in the hidden layer plays a critical role in determining the weights and predicted output. If there are too many neurons in the hidden layer, the model will experience overfitting while if there are fewer neurons there will be under fitting. It is then important to select the number of neurons in the hidden layer such that we don't experience overfitting or under fitting. The criterion adapted in this study is the rule of thumb method in which the number of hidden neurons is in the range between the size of the input layer and the size of the output layer [21].

3.4.3. Stopping Training Neural Network

The major challenge during the training of neural networks is how long one need to train networks since too little training leads to under-fitting while too much training leads to overfitting training. It is recommended that training need to be stopped when generalization error starts to increase. This will reduce overfitting and improve generalization of deep neural networks [22]. Another approach is to train the network once for large number of training set or configure network to be under constrained, that it, have more capacity than required for the problem.

4. Results and Discussions

The analysis for both descriptive and inferential on the risk factors for wheezing among children is carried and model performance assessment conducted that enables the selection of the best model. The variables are presented, discussed and analysis carried on Demographic Health Survey.

4.1. Description of Study Variables

The variables considered and the analysis coding scheme of both the response and predictor variables are summarized in Table 1.

Table 1. Coding scheme of the study variables

Variable	Description	Coding and description
Wheezing	Response variable	0 – No-wheeze, 1 - wheeze
Age	Age of children in months	Continuous variable
Gender	Gender of children	0 – Female, 1 - Male
Breastfeeding	Exclusive breastfeeding	0 - Not exclusive, 1 - Exclusive
Smoking	Smoking tobacco exposure	0 – N0, 1 - Yes
Living	Difficult living conditions	1 - Low, 2 – Medium, 3 – High
Residence	Residence of children	0 – Rural, 1 – Urban
Atopy	Genetic vulnerability to develop allergy	0 – Absent, 1 - Present
Maternalage	Age of mother grouped in years	0 – 13 – 16, 1 – 17 - 20, 2 – 21 - 44
Pretermbirths	Babies born alive before 37 weeks	0 – N0, 1 - Yes

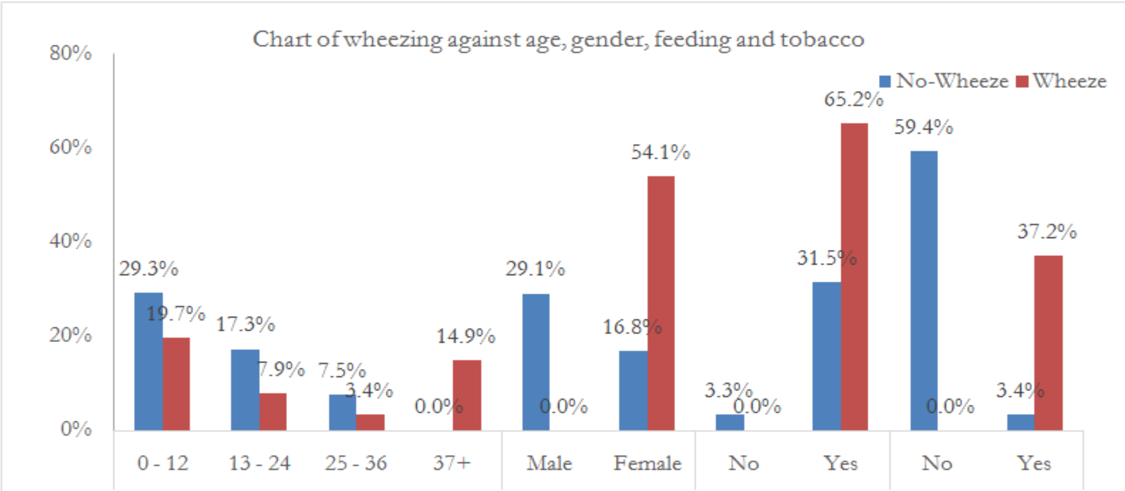

Figure 4. Chart of wheezing against age, gender, feeding and smoking

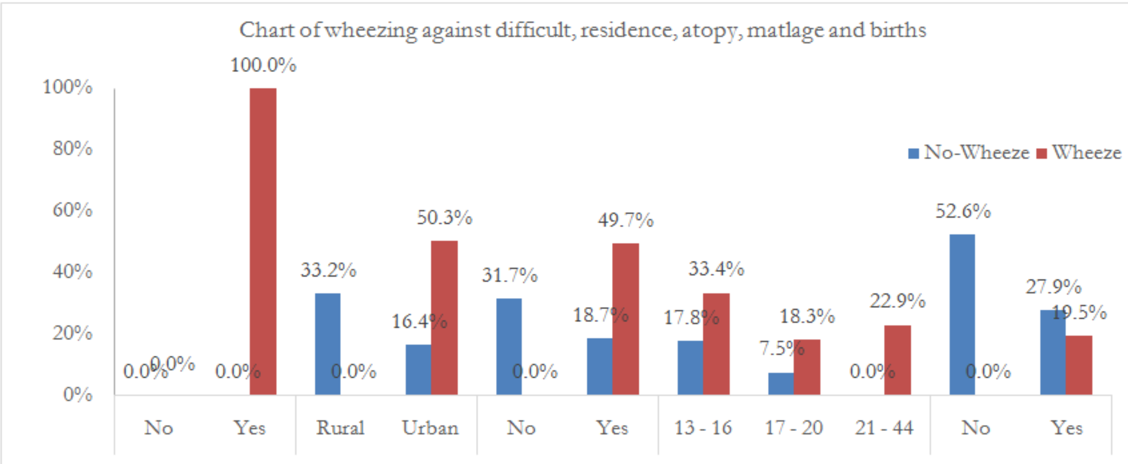

Figure 5. Chart of wheezing against living, residence, atopy, maternal age and preterm

4.2. Descriptive Analysis

The cross-sectional data collected indicates that there are higher cases of wheeze among children in the age of 0 – 12 months with higher risk of developing wheezing with 19.7% of cases, the female gender recorded higher risk of wheezing with 54.1% of the cases than male gender, among those with exclusive feeding the risk of developing wheezing is higher recording 65.2% of the cases while those exposed to tobacco smoking have higher risk of developing wheezing recording 37.2% of the cases as depicted in Figure 4.

Considering the children living under difficult conditions the risk is higher with 100% of the wheeze cases, children living in urban have a higher risk of developing wheeze recording 50.3% of the cases, the children with atopy have higher risk recording 49.7% of wheeze cases, children born in which the mother is younger, that is 13 – 16 years, the risk of developing wheeze is higher with 33.4% of the cases while preterm births have higher risk of developing wheeze recording 19.5% of the cases as depicted in Figure 5.

4.3. Selecting Risk Factors

In determining the set of variables that are potential risk factors in childhood wheezing modelling, the chi-square test was conducted for each of the variables with wheeze response variable using Pearson Chi-square test at 5% significance level. The results of the chi-square test are recorded as in Table 2. The results indicates that the variables feeding, tobacco smoking exposure and difficult living conditions are potential risk factors that may be considered in wheezing modelling. A further generalized linear multiple logistic regression is conducted and results are recorded in Table 3.

Table 2. Chi-squared test statistic and p values

Risk factors	X-Squared	df	P-value
Age	59.165	56	0.3608
Gender	0.623	1	0.4298
Feeding	421.111	1	0.0000
Smoking	374.211	1	0.0000
Living	79.502	2	0.0000
Residence	0.520	1	0.4712
Atopy	0.799	1	0.3713
Maternal	2.845	2	0.2411
Preterm	0.022	1	0.8820

Table 3. Generalized multiple linear regression estimated parameters

Coefficients	Estimate	Std. Error	Z Value	P(>|z|)
Intercept	1.5117	0.77487	1.951	0.0511
Age	-0.0213	0.00976	-2.182	0.0291
Feeding	-3.9437	0.50488	-7.811	0.0000
Smoking	1.8969	0.48469	3.914	0.0001
Living	-0.5422	0.25172	-2.154	0.0312
Maternal	0.5518	0.26210	2.105	0.0353

In the generalized linear multiple logistic regression the variables age, feeding, tobacco smoking exposure, difficult living conditions and maternal age are found to be significant and therefore can be considered in model. The age of children and maternal age factors were initially found not to be significant in the chi-square analysis. In this study the analysis however considers simple random sample without replacement as a recommendation in order to have a reduced cost of survey, greater speed of getting results, greater accuracy of results, greater scope and adaptability [23]. The significant variables and estimated parameters for the generalized multiple linear regression model are recorded as in Table 4.

Table 4. Generalized multiple linear regression estimated parameters

Coefficients	Estimated	Std. Error	Z Value	P value
Intercept	3.2198	1.1261	2.859	0.0043
Age	-0.0387	0.0159	-2.430	0.0151
Feeding	-4.9901	0.8016	-6.225	0.0000
Smoking	1.7851	0.6626	2.694	0.0071
Living	-0.8209	0.3722	-2.206	0.0274
Residence	1.0891	0.5329	2.044	0.0410

The Akaike Information Criterion (AIC) and Bayesian Information Criterion are commonly used to evaluate model performance and model complexity. The model performance may be evaluated using probabilistic framework, such as log-likelihood under the framework of maximum likelihood estimation while model complexity may be evaluated as the number of degrees of freedom or parameters in the model.

The AIC and BIC are methods for scoring and selecting a models and the model with the lowest AIC or BIC is selected. The AIC and BIC are computed using the following equations;

$$AIC = -2(LL)/N + 2(k/N) \text{ or } n(LL) + 2k \tag{11}$$

$$BIC = -2(LL) + k\log(N) \text{ or } n(LL) + k\log(n) \tag{12}$$

Where LL is the log-likelihood of the model, N and n is the number of examples in the training dataset and k is the number of parameters in the model (Brownlee, 2020).

The quantity calculated in BIC is different from AIC, although can be shown to be proportional to the AIC. Unlike the AIC, the BIC penalizes the model more for its complexity such that more complex models will have a worse or larger score and will, in turn, be less likely to be selected. The null deviance, residual deviance and Akaike Information Criterion (AIC) of the best generalized multiple logistic regression model were recorded as 502.44, 179.57 and 191.57 respectively.

4.4. Two-Stage Artificial Neural Network Model

In the proposed developed Two-Stage Artificial Neural Network (ANN) model in equations (9) and (10), the significant variables identified in the previous section that includes age, feeding, smoking, living and residence were considered in fitting the Two-Stage Artificial neural Network model. The search for optimal Artificial Neural Networks with one hidden layer with one neuron, two neurons, three neurons and four neurons were considered for five iterations and the results recorded in Table 5.

Table 5. Search for Artificial Neural Network Model in One-Hidden Layer

Iteration	Model	Error	Threshold	Steps
1	ANN(1)	2.455641	0.006769	2649
	ANN(4)	1.891908	0.008788	2197
2	ANN(2)	2.033283	0.005946	908
5	ANN(3)	1.968134	0.008997	2210

There are three frequently used performance measures commonly used in selecting the better model that includes absolute error, Root Mean Square Error and percent good classification measures. The percent good classification is the preferred measure of performance for classification in Artificial Neural Networks since it is consistent with the task. The Artificial Neural Network ANN (1), ANN (2) and ANN (3) recorded precision of 0.0% and accuracy of 65.1% while ANN (4) recorded precision of 97.1% and accuracy of 39.4%. In the proposed Two-Stage Artificial Neural Network model, ANN (1, 4) recorded precision of 97.1% and accuracy of 99.0% as recorded in Table 6. The architecture for the ANN (1) and ANN (4) are depicted as in Figure 6 and Figure 7 respectively while the estimated parameters for ANN (1) and ANN (4) are recorded in Table 7 and Table 8 respectively.

Table 6. Confusion Matrix for ANN Model with One-Hidden Layer

		Observed			
Network Model	Predicted	0	1	Precision	Accuracy
ANN(1)	0	380	204	0.0%	65.1%
	1	0	0		
ANN(2)	0	380	204	0.0%	65.1%
	1	0	0		
ANN(3)	0	380	204	0.0%	65.1%
	1	0	0		
ANN(4)	0	32	6	97.1%	39.4%
	1	348	198		
ANN(1, 4)	0	380	6	97.1%	99.0%
	1	0	198		

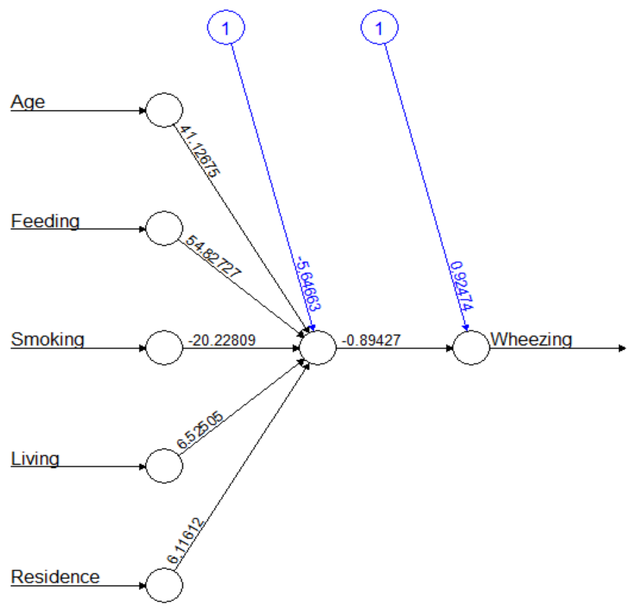

Figure 6. Architecture Artificial Neural Diagram, ANN (1)

Table 7. Weights for the Artificial Neural Network Model ANN (1)

Description	Weight	Description	Weight
Intercept.to.1layhid1	-5.64664	Living.to.1layhid1	6.52505
Age.to.1layhid1	41.12675	Residence.to.1layhid1	6.11612
Feeding.to.1layhid1	54.82727	Intercept.to.Wheezing	0.92474
Smoking.to.1layhid1	-20.22809	1layhid1.to.Wheezing	-0.89427

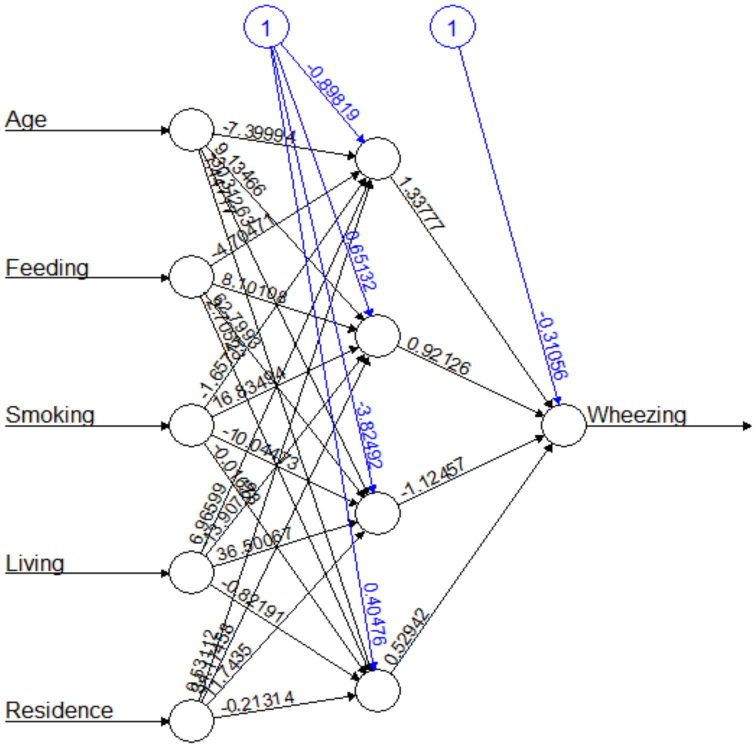

Figure 7. Architecture Artificial Neural Diagram, ANN (4)

Table 8. Weights for the Artificial Neural Network Model ANN (4)

Description	Weight	Description	Weight
Intercept.to.1layhid1	-0.89819	Smoking.to.1layhid3	-10.04473
Age.to.1layhid1	-7.39994	Living.to.1layhid3	36.50067
Feeding.to.1layhid1	-4.70471	Residence.to.1layhid3	11.74350
Smoking.to.1layhid1	-1.65717	Intercept.to.1layhid4	0.40476
Living.to.1layhid1	6.96599	Age.to.1layhid4	1.44717
Residence.to.1layhid1	0.53112	Feeding.to.1layhid4	2.70523
Intercept.to.1layhid2	0.65132	Smoking.to.1layhid4	-0.01678
Age.to.1layhid2	9.13466	Living.to.1layhid4	-0.82191
Feeding.to.1layhid2	8.10108	Residence.to.1layhid4	-0.21314
Smoking.to.1layhid2	16.83494	Intercept.to.Wheezing	-0.31056
Living.to.1layhid2	-13.90766	1layhid1.to.Wheezing	1.33777
Residence.to.1layhid2	34.17458	1layhid2.to.Wheezing	0.92126
Intercept.to.1layhid3	-3.82492	1layhid3.to.Wheezing	-1.12457
Age.to.1layhid3	-30.31263	1layhid4.to.Wheezing	0.52942
Feeding.to.1layhid3	62.79930		

The Artificial Neural Network models with two hidden layers were also considered in the study. The Artificial Neural Network ANN(1, 1) recorded precision of 88.2% and accuracy of 93.3%, ANN (2, 1) recorded precision of 74.4% and accuracy of 89.0%, ANN (3, 1) recorded precision of 19.6% and accuracy of 71.9%, ANN (4, 1) recorded precision of 12.7% and accuracy of 69.5%, ANN (2, 2) recorded precision of 87.7% and accuracy of 93.3% while ANN (3, 2) recorded precision of 17.6% and accuracy of 71.1%. The selection of the model and performance are recorded in Table 9 and Table 10 respectively.

Table 9. Search for Artificial Neural Network Model in Two-Hidden Layer

Iteration	ANN Model	Error	Threshold	Steps
2	ANN(2, 1)	7.154658	0.0090632	1996
	ANN(3, 1)	6.801973	0.0096053	3761
	ANN(4, 1)	6.244876	0.0091757	1371
5	ANN(1, 1)	8.115141	0.0095292	1277
	ANN(2, 2)	7.088962	0.0076923	1812
	ANN(3, 2)	6.578770	0.0081007	1577

Table 10. Confusion Matrix for ANN Model with Two-Hidden Layers

		Observed			
Network Model	Predicted	0	1	Precision	Accuracy
ANN(1, 1)	0	365	24	88.2%	93.3%
	1	15	180		
ANN(2, 1)	0	368	52	74.5%	89.0%
	1	12	152		
ANN(3, 1)	0	380	164	19.6%	71.9%
	1	0	40		
ANN(4, 1)	0	380	178	12.7%	69.5%
	1	0	26		
ANN(2, 2)	0	366	25	87.7%	93.3%
	1	14	179		
ANN(3, 2)	0	379	168	17.6%	71.1%
	1	1	36		

In the two-hidden layer Artificial Neural Network models, the performance of ANN (1, 1) model recorded better performance compared to the rest of the models and therefore selected for consideration as a possible model for modelling wheezing risk factors among children. The selected better performing Artificial Neural Network model ANN (1, 1) in the two-hidden layer is depicted as in the architecture shown in Figure 8 and the estimates of parameters are recorded in Table 11.

The ANN (1, 1) in the two-hidden layer Artificial Neural Network recorded precision of 88.2% and accuracy of 93.3% compared to Two-Stage Artificial Neural Network ANN (1, 4) in the one-hidden layer Artificial Neural Network recording precision of 97.1% and accuracy of 99.0% is therefore not a better based on the measures of precision and accuracy. Then based on this analysis, the developed Two-Stage Artificial Neural Network model is a better performing model for modelling wheezing risk factors and therefore recommended for consideration in classifying the children as being at risk of developing wheeze based on the identified significant risk factors for wheeze among children in Kenya.

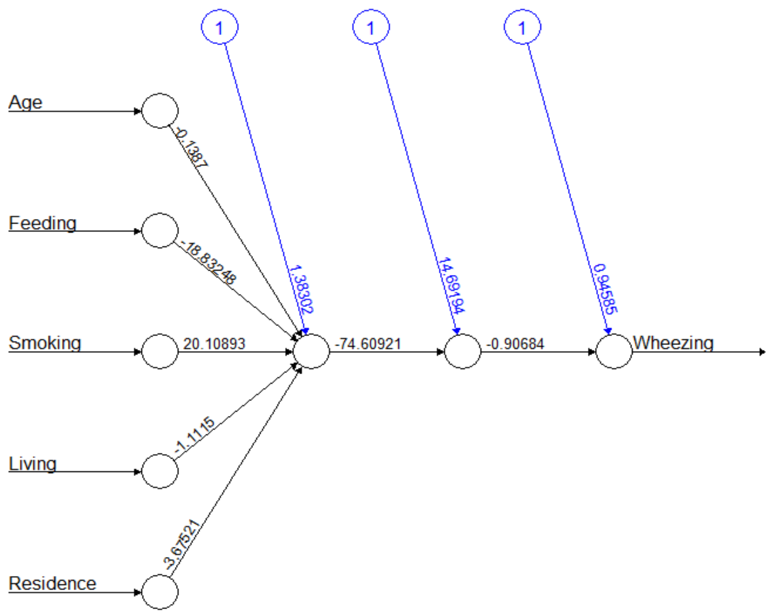

Figure 8. Architecture Artificial Neural Diagram, ANN (1, 1)

Table 11. Weights for the Artificial Neural Network Model ANN (1, 1)

Description	Weight	Description	Weight
Intercept.to.1layhid1	1.38302	Residence.to.1layhid1	-3.67521
Age.to.1layhid1	-0.13870	Intercept.to.2layhid1	14.69194
Feeding.to.1layhid1	-18.83248	1layhid1.to.2layhid1	-74.60921
Smoking.to.1layhid1	20.10893	Intercept.to.Wheezing	0.94585
Living.to.1layhid1	-1.11150	2layhid1.to.Wheezing	-0.90684

5. Conclusions, Recommendations and Further Study

5.1. Conclusion

The developed model in the study is to be used in classifying the wheeze cases among children using the identified significant risk factros of age, feeding, smoking, living and residence at Gatundu hospital, Kenya.

i) The logistic model recorded precision and accuracy of 89.2% and 93.3%

ii) The Artificial Neural Network of one-hidden layer with single node recorded precision and accuracy of 0.0% and 65.1%

iii) The Artificial Neural Network of one-hidden layer with four node recorded precision and accuracy of 97.1% and 39.4%

iv) The Artificial Neural Network of two-hidden layers with one neuron in first hidden layer and one neuron in the second hidden layer recorded precision and accuracy of 88.2% and 93.3%

v) The Two-Stage logistic Artificial Neural Network of one-hidden layer with either one neuron or four neurons recorded precision and accuracy of 97.1% and 99.0% respectively.

It is therefore clear that when using the classification criterion considered of both precision and accuracy measures the developed Two-Stage Artificial Neural Network model, ANN (1, 4) with one-hidden layer performs better as compared to the Generalized Multiple Logistic Regression model or the single stage Artificial Neural Networks having one-hidden layer or the single stage Artificial Neural Networks having two-hidden layers.

5.2. Recommendations

The recommendations based on the results of the study include the following;

i) Allocation of more resources for improvement of health facilities to collect enough data for analysis that will be useful in identifying other possible risk factors in order to improve the developed classifying model.

ii) The governments need to allocate more resources to improve the living conditions of the families in order to reduce the chances of children developing wheezing.

iii) The governments need to introduce programmes to educate persons on wheezing risk factors for instance dissemination on the limitation of exposure to smoking and importance of breastfeeding.

5.3. Further Study

In the extension of the study, the following may be considered in improving the results that have been achieved in extending the study;

i) A study need to be carried that that considers additional datasets that will be investigated for other possible risk factors that may include access to medical facilities and economic factors as possible wheezing risk factors.

ii) A study need to be carried that considers additional sampling techniques to ascertain whether there will be a significant improvement in the model performance.

iii) A study can be carried that considers other machine learning techniques of clustering, dimensionality reduction, transfer learning and bootstrap in modelling wheezing risk factors in order to determine the suitability of the models compared to the ones considered in this study.

References

[1] Division of Leprosy, T. a. (2011). *Guidelines for Asthma Management in Kenya.* Nairobi: Ministry of Health and Sanitation (DLTLD) and Kenya Association for the Prevention of Tuberculosis and Lung Diseases (KAPTLD).

[2] Taussing, L. M., Wright, A. L., Holberg, C. J., Halonen, M., Morgan, W. J., & Martinez, F. D. (2003). Tucson Children Respiratory Study: 1980 to Present. *The Journal of Allergy and Clinical Immunology*, 111(4): 661-675.

[3] Mugambi, E., Mech, J., & Kamau, E. (2018). Epidemiology and Risk Factors for Astma in Kenya. *Journal of Kenya Association of Physicians*, 1(2): 66-72.

[4] Martinez, F. D., Wright, A. L., Holberg, C. J., Morgan, W. J., & Taussig, L. M. (1992). Maternal age as a risk factor for wheezing lower respiratory illness in the first year of life. *American Journal of Epidermiology*, 136(10): 1258-1268.

[5] Moraes, L. S., Takano, O. A., Mallol, J., & Dirceu, S. E. (2013). Risk factors associated with wheezing in infants. *Journal of Pediatria*, 89(6): 559-566.

[6] Castro-Rodrguez, J. A., Holberg, C. J., Morgan, W. J., Wright, A. L., & Martinez, F. D. (2001). Increased incidence of asthmalike symptoms in girls who become overweight or obese during the school years. *American Journal of Respiratory and Critical Care Medicine*, 163(6): 1344-1349.

[7] Shephard, R. J. (2010). Lifestyle and Lung Health in Children and Adolescents: Children's Way of Life and Their respiratory Health. *American Journal of Lifestyle Medicine*, 4: 7-32.

[8] Mutius, E. V. (2016). The microbial environment and its influence on asthma prevention in early life. *The Journal of Allergy and Clinical Immunology*, 137(3): 680-689.

[9] Kotecha, S. J., Dunstan, F. D., & Kotecha, S. (2012). Long termrespiratory outcomes of late preterm born infants. *Seminars in fetal & neonatal medicine* (pp. 17(2): 77-81). Cardiff: Elsevier.

[10] Sherriff, A., Peters, T., Henderson, J., & Strachan, D. (2001). Rsik factor associations with wheezing patterns in children followed logitudinally from birth to 31/2 years. *International Journal of Epidemiology*, 30(6): 1473-1484.

[11] Azad, M. B., Vehling, L., Lu, Z., Dai, D., Subbaro, P., Becker, A., . . . Sears, M. (2017). Breastfeeding, maternal asthma and wheezing in the first year of life: a longitudinal birth cohort study. *The European Resipiratory Journal*, 49(5): 1602019.

[12] Guo, J., Zhu, W., & Wang, H. (2019). Risk Factors and prognosis of recurrent wheezing in Chinese young children; a prospective cohort study. *Allergy Asthma clin Immunol*, 15(38): 2019.

[13] Arif, A. A., Delclos, G. L., Lee, E. S., Tortolero, S. R., & Whitehead, L. W. (2003). Prevalence and risk factors of asthma and wheezing among adults; an analysis of the NHANES III data. *European Respiratory Journal*, 21: 827-833.

[14] Linneberg, A., Simonsen, J., Petersen, J., Stensballe, L., & Benn, C. (2005). Differential effcets of risk factors on infant wheeze and atopic dermatitis emphasize a diffrent etiology. *Journal of Allergy Clinical Immunology*, 117(1): 1-6.

[15] Mageto, T., Njuguna, E., & Osundwa, D. (2018). Modeling Diabetes Risk Factors (A Case Study of Focus Medical Centre in Kiambu, Kenya 2016). *Science Journal of Applied Mathematics and Statistics*, 6(1): 7-15.

[16] Yang, W. (2017). *Early Warning for Infectious Disease Outbreak: Theory and Practice.* Cambridge: Academic Press.

[17] Batarseh, F., & Yang, R. (2020). *Data Democracy: At the Nexus of Artificial Intelligence, Software Development and Knowledge Engineering.* Cambridge: Academic Press.

[18] Ghan, S. K., & Deepak, S. N. (2013). Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Mathematical Problems in Engineering*, 10, 55-66.

[19] Munoz-Bravo, S., Perez-Mendez, A., & Rivas-Echeverria, F. (2009). Artificial Neural Network Model for Estimating the Female and Male Population in Working Age in Venezuela. *WSEAS Transactions on Business and Economics*, 6, 177-188.

[20] Feng, J., & Shengnan, L. (2019). Perfomance Analysis of Various Activation Functions in Artificial Neural Networks. *Journal of Physics: Conference Series*, 1237: 1-6.

[21] Panchal, F., & Panchal, M. (2014). Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. *International Journal of Computer Science and Mobile Computing*, 3(11): 455-464.

[22] Brownlee, J. (2020). *Probability for Machine Learning: Discover How To Harness Uncertainty With Python.* San Francisco: Machine Learning Mastery.

[23] Singh, D., & Chaudhary, F. S. (2020). *Theory and Analysis of Sample Survey Designs.* New Delhi: New Age International Publishers.