# Undesired Child Nexus Poor Health Condition: An Application of Machine Learning

By

Urva Zanaib

Registration No: PIDE2019FMPHILETS11

Supervised By

Dr. Amena Urooj

**PIDE School of Economics**

**Pakistan Institute of Development Economics**

**Islamabad Pakistan**

**2023**

Pakistan Institute of Development Economics

# Pakistan Institute of Development Economics, Islamabad

***PIDE School of Economics***

## CERTIFICATE

This is to certify that this thesis entitled **"Undesired Child Nexus Poor Health Condition: An Application of Machine Learning"** submitted by **Ms. Urva Zanaib** is accepted in its present form by the School of Economics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in Master of Philosophy in Econometrics

Supervisor: Dr. Amena Urooj Signature: 

External Examiner: Dr. Tanweer Ul Islam Signature: 

Head,
PIDE School of Economics: Dr. Shujaat Farooq Signature: 

# Declaration

I Urva Zainab hereby state that my MPhil thesis titled Undesired Children Nexus Poor Health Condition: An Application of Machine Learning is my own work and has not been submitted previously by me for taking any degree from Pakistan Institute of Development Economics or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my Graduation the university has the right to withdraw my MPhil degree.

Date: 15 - 3 - 23

Urva Zainab

Signature of Student

Name of Student

Dedicated to my all-Family Member

## *ACKNOWLEDGEMENT*

*In the name of Allah, the most Merciful and Beneficent*

*First and foremost, praise is to ALLAH, the Almighty, the greatest of all, on whom ultimately, we depend for sustenance and guidance. I would like to thank Almighty Allah for giving me opportunity, determination, and strength to do my research. His continuous grace and mercy were with me throughout my life and ever more during the tenure of my research.*

*I would like to express my sincere gratitude to my advisor Dr. Amena Urooj for the continuous support of my research, for his patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my MPhil study.*

*I sincerely thank to my father and my sister for financial support during my whole career, special thanks to my brothers for their encouragement, moral support, personal attention, and care. I would like to acknowledge many people for helping me during my research.*

*Last but not the least, I would like to thank my friends (Sheza Muqadas and Faridoon Khan and all PHD fellows) for giving valuable suggestions during my study. I enjoyed spending time with them. Thanks for giving me such a joyful time.*

*Urva Zanaib*

# Table of Contents

# List of Figure

# List of Table

# ABSTRACT

One of the key indicators of human development in a child's poor health condition practically is childhood morbidity and mortality. Although child morality is not desired, Pakistan nevertheless has a high rate of child morbidity and mortality. One of the psychological factors that may be associated with child morality is an unplanned pregnancy, whether unwanted (the parent did not want any more or more children) or mistimed (the pregnancy occurred sooner than hoped). Unplanned pregnancies and births are two psychological issues that harm children's nutritional health (Shapiro-Mendoza et al., 2005); (Shaka et al., 2020).

The goal was to evaluate the impact of mothers' aspired status on the morbidity and mortality of children in Pakistan. We limited our analysis to children under 5 before the survey and used Pakistan demographic and health survey (2017-18), a national representative cross-sectional survey. By reducing the optimal number of children from total live births, we were able to estimate the undesired status (excess in boys, girls, both, and parity). We assessed morbidity (fever, diarrhea, cough, ARI, and SBR), nutritional status, and under 5 moralities. Finally, we perform machine learning techniques LDA, RF, SVM, and NN in the analysis of the data**.** The findings revealed that the overall percentage of the undesired child was 8%, 4%, 15%, and 27% for boys, girls, parity, and dual excess respectively. All the variable was associated with the undesired child. Child morbidity, fever, and cough were higher among the undesired children. We found little evidence that undesired children have acute respiration infection (ARI). We found very little evidence that an undesired child has a significant impact on childhood diseases. The ratio of child morality was less for boys but higher for girls.

**Keywords:** Machine learning, Support Vector Machine, Neural Network, Random Forest, Linear discrimination Analysis, Pakistan Demographic Health and Survey, Acute respiratory infections, antenatal care, postnatal checkup, and family planning**,** short rapid/or breath, severely stunted, severely underweight, Severely Wasted, wasted, Birth Size, treatment of diarrhea, treatment of fever/cough.

# CHAPTER 1

# INTRODUCTION

## 1. Introduction

Despite Pakistan's exceptionally high newborn morality rate, there isn't much evidence in the literature about the cause of baby survival there. Low middle income countries including Pakistan, India and Bangladesh share high burden of mortality (Ahmed et al., 2018). In Pakistan, there has been few research on newborn and child survival, but they have used a small number of factors and lacked a strong conceptual framework. Although machine learning is extensively used and performing pretty-well, ML-based algorithm heavily relies on manually created features. So, the goal of applying machine learning in the healthcare of undesired children and to enhance the effectiveness and quality of care of undesired children.

Undesired is defined as, when women deliver children more then she desired or preferred, this led to undesired child (Flatø, 2018). Strong preference for son and less for female child reluctance to undesired. There is strong desired for son and low for daughters, a situation is known as the "fertility squeeze," which linked to pre and postnatal discrimination against daughters. Because (Shaka et al., 2020) prove that undesired child suffers more from disease than wanted. Another thing the children face lack of attention of his/her mother. Thus, reduce the mother's ability to cope with children every day. So that is the reason to conduct this study as this area is not explored yet using ML techniques.

From statistics and machine learning there are many techniques that can be utilized both for inference and predication. Thus, the classical statistical models become less precis. In contrast, ML focus on the predication by using general algorithm to find the patterns in unwieldy data.

ML technique for teaching machines to handle the data more efficiently. When a machine learning model is used to make real-world predictions, it provides an impartial assessment of how well it will perform. Machine learning algorithms may automatically analyze and extract knowledge (Alam et al., 2021) ;(Kozareva & Montoyo, 2006) from big data. Even though these procedures have a wide range of possible applications in medicine. To handle large sample sizes, many techniques have evolved. In the health sector (Haq et al., 2020; Yu et al., 2010) supervised and non-supervised machine learning approaches are used to diagnose different kinds of diseases. This method enables the research to explore hidden patterns for predicting expected outcomes (Ahamed et al., 2021). Moreover, ML techniques are trained with datasets having different features and external variables.

The rational of using machine learning algorithms over conventional statistics is these techniques are largely flexible and have no prior assumptions, the traditional model are totally based on assumptions including error distributions, additivity of parameters within the explanatory variables and proportional hazards. These assumptions are not mostly met in clinical research. Machine learning models can easily address the interaction variables, while traditional statistical method address interaction between the determinant and confounders. Moreover, ML techniques integrate the data into predictions.

The supervised ML techniques that deal more with classification (Ahmed et al., 2021) include the following classifier: Nave Bayes Classier, Random Forest (RF), support vector machine, Decision Tree, Neural Network, K-Mean Clustering, Boosting, and more. To avoid overfitting, which occurs when a model performs well in a set of data used for training but not so well where data is unseen, the existence of outliers can affect outcomes and the impact of noise. To overcome this problem a different researcher uses a different method, cross-validation, and synthetic Gaussian noise data for error in the validation process of machine learning.

So, a method is needed which can accurately use machine learning techniques to predict the effect of the undesired child. Different ML algorithms such as SVM, NN, LDA, and RF models can be utilized for this purpose. The goal of this work is to find effective machine learning based models for health conditions of undesired children under five, utilizing the PDHS dataset. The further incentive is to find the best model with more accuracy predictions and the minimum errors. Another important factor included in this study are the prediction of the health status of the undesired child which is an important indicator of human development (Alam & Islam, 2022). It's critical assisting at-risk and ensure their retention, by reducing undesired childbearing. Nonetheless, it might be difficult to achieve but it is important to improve the child's health and survival.

A substantial number of studies have gone into predicting the health condition of an undesired child for different aims, like under-five child survival, morbidities, treatments, and nutritional status by adjusting key socioeconomic covariates. Unfortunately, the undesired child mortality and morbidity, especially malnourishment is worse (Alam & Islam, 2022). However, analyzing the undesiredness of children like, (Alam & Islam, 2022) studied in Bangladesh, has not been explored in Pakistan.

Undesiredness may be used to explain why female newborn morality is higher than a male morality among moms who have lower reproductive wants and why infant mortality is higher among the mothers which want one to three children. Infant of the excess gender are especially more vulnerable in situations when there is an unfavorable gender composition (Flatø, 2018), which would be more significant for infant mortality than unfavorable childbearing.

Undesiredness may account for a higher proportion of child mortality among women who have lower fertility desires. It can also be characterized by the accumulation of children in comparison to the ideal family as well as the gender composition as indicated by mothers.

The preference for the gender composition of children is more diverse in developing countries. Gender preferences are also important in determining Undesiredness. For instance, the conflict between a lower desire for a child and a higher desire for sons is generally direct to as fertility squeeze and is considered discrimination against girls. Unplanned births are the outcome of undesired pregnancies, which are classified as unwanted pregnancies (occurring when there is no desire for more children) and mistimed pregnancies (occurring at a time other than desired) are the crucial public health concerns that give rise to maternal mortality and morbidity across the world on the other hand (Organization, 2019); (Santelli, 2003); (D'Angelo, 2004); (Rackin, 2018).

More about how to accurately assess women's intentions to become pregnant is a matter of great discussion among social scientists. The common measurement is based on the mother's retrospective account of their intentions during pregnancy. Women who had already given birth at the time of this study, thus their responses may be influenced by the health of newborns, including their critical condition. Research has proven contrary to be true, certain cross sectional studies have revealed that children from unwanted pregnancies have greater survival rates. We believe that this is because mothers are less likely to report that deceased children from unwanted pregnancies.  A sub-Saharan African study confirms that mothers are less likely to indicate that their dead children were the consequence of unplanned pregnancies than mothers of living children.

Pakistan is facing maternal and child health issues for many years. Few studies focus on the Maternal (variables about the mother of the child) Factors as determinants of Child nutrition outcomes ((Harpham et al., 2005); (Costa, 2020); (Masibo, 2012); (Aboderin, 2011); (Teupser et al., 2010); (Black et al., 2008). Excessive son preferences in Asia countries have resulted in sex-selective abortions and excessive child mortality rates among girls (Anderson & Ray, 2010), specifically for girls with both many sisters and high parity (Guilmoto, 2009); (Gupta, 1987); (Jayachandran, 2015).

Female babies receive less care and have fewer family resources as compared to male infants resulting in a high death rate among females in Pakistan (Qadir et al., 2011). Gender role in Pakistan is usually well-delineated. From birth, son prefers, and daughter neglect reign supreme and gender preferences are more likely to implement when fertility goals are lower, which might have consequences for future sex ratio in childbirth and infant mortality disparities (Bongaarts, 1992). In Pakistan, unmet family planning needs have remained high over the last decades, with very minor improvements in contraceptive prevalence (Asif & Pervaiz, 2019). Every day, around 16,000 children under the age of five years die throughout the world, mostly because of avoidable diseases (United Nations Children Fund (UNICEF 2015). The health condition of undesirable children is in grave danger (Flatø, 2018). If a woman produces more children than she desires, the children may suffer from inadequate f proper care, as well as feelings of hatred for undesired offspring throughout pregnancy (Flatø, 2018). There is linked between childhood mortality and morbidity which includes malnutrition is likely to be significantly higher in unwanted children (Shaka et al., 2020). (Shaka et al., 2020) indicated that children from unplanned pregnancies were three times more likely to be stunted than children born from planned births.

This high infant mortality in Pakistan is linked to women's education, low income, and poor health service (Patel et al., 2021). The UN acknowledges the critical need to eliminate the avoidable newborn and child morality by 2030 (UNICEF 2015), the third aim of the Sustainable Development Goal is to ensure healthy lives and promote well-being. The death rate of children under five is already fallen below in 118 countries as per mentioned by SDGs and effects and pledges must be enhanced for the remaining nations (UNICEF 2016;

(Organization, 2019)). As result, lowering the number of undesired pregnancies and births will contribute to achieving the objective.

The basic objective of this study is to suggest concerned ministries and the population welfare department deduce ways of limiting the rising trends of undesired pregnancies, by educating the populace through certain print, electronic, and social media channels. This way, child morbidity and child mortality rate could also be lowered, keeping in mind the demographic and socio-economic factors.

Moreover, the United Nations recognizes that there is need to end preventable infant and child deaths between 2016 and 2030 (WHO, World health Organization 2019), and third goal of sustainable Development Goals (SDGs) focuses on the ensuring the healthy lives and promoting wellbeing (UN, World health Organization 2015). As of SDGs, all countries should reduce neonatal mortality and under five moralities to 12 or fewer and 25 or fewer per 1000 live births within 2030 (WHO, World health Organization 2016, UNICEF,2018). The under-five morality rate of 118 countries had already been below the SDGs target. Still, for the remaining countries, mainly central and southern Asia and sub-Saharan Africa, progress and promises will need to accelerate of achieve the target (WHO, World health Organization 2016, Hargreaves et al., 2019). Therefore, reducing unwanted pregnancy and childbearing will be helpful in reaching the goal.

## 1.1. Objectives of study

1. To build and compare the predication model for undesired children's health conditions using advanced machine learning techniques.
2. To determine the prevalence of undesired children vs desired children based on various demographic and socio-economic factors.

3. To evaluate the SDGs concerning the prevalence of undesired children in Pakistan.

## 1.2. Hypothesis

Ho: There isn’t enough evidence to conclude that undesired children experience better health circumstances.

H1: There is enough evidence to conclude that undesired children experience better health circumstances.

We will sub-divide this hypothesis according to the objectives of the study.

## 1.3. Research Question

1. How does the accuracy of the ML technique vary in predicting the health condition of the undesired PDHS dataset?

2. How can an effective ML algorithm be discovered to integrate that into an undesired child’s health?

## 1.4. Significance of the study

One of the most important predictors of child development in child health, particularly childhood mortality. Infant mortality rates are undesired in any country, but still, it is particularly high in underdeveloped countries. As a result, it is necessary to examine the influence of a mother’s child's desires status on childhood mortality and morbidity in Pakistan, as well as the probable variables linked with it.

## 1.5. Research gap

In the health sector, this area is not explored yet in Pakistan. So, this research makes little effect to investigate the impact of undesired children on childhood mortality and morbidity. We will examine the issue using nationally representative data and advanced approaches that may improve acceptance and applicability to other settings with comparable socioeconomic status. [(Alam & Islam, 2022); (Shaka et al., 2020); (Flatø, 2018)] utilize traditional methods to address the undesiredness of infants. To better assess the risk of morality, morbidity, and nutritional conditions among undesired children, the classifier method (RF, NN, LDA, SVM) (machine learning algorithm) will be used. These classifiers are expected to predict the results more accurately (Islam., et al 2020).

# CHAPTER 2

# LITERATURE REVIEW

Both machine learning and statistics employ various methods which can be used for prediction and inference. Machine learning techniques are used to describe the prediction goal however statistical techniques put emphasize inference (Bzdok et al., 2018), which can be obtained by appropriately fitting and projecting probability models. If the objective is prediction and optimization in complex data sets, then machine learning may be more advantageous than conventional statistical techniques, machine learning may be better than conventional statistical techniques since it has fewer statistical assumptions than conventional parametric approaches (linear relationships, absence of multicollinearity). The aim behind the prediction is to identify the variables that strongly influence prediction accuracy.

## 2. Introduction to Literature Review

Child mortality is higher for the firstborn child and last-born child (Montgomery, 1996), when investigating the relationship with the level of fertility, on the other hand, relationship with the level of morbidity and mortality among children and women using EDHS.

The relationship between undesired childbirth, health, and mother-childhood interaction was examined. Various studies have exhibited that women who had unintended pregnancies have worse bonding with their children from teenage to later life. Furthermore, (Barber et.al, 1999) studies explore not only unwanted children bear worse bonding and association with mothers. In contrast to this, other children in the family get affected too. The study concluded that unintended pregnancy decreases a mother's time and attention to their children, and this lays down the foundation for long-term dis-association.

Quality healthcare services provision is a provincial responsibility and the priority actions emanating from (National Health Vision Pakistan 2016-2025) would be in concert with the provincial needs, and expectations. The goal of this national health vision is to align with provincial objectives and aspirations. By overarching every aspect and by facilitating/advocating for the mobilization of financial and technical resources to ensure that essential health services are accessible to all citizens. On the other hand, the federal government is determined to support and facilitate the provinces in developing/ implementing their strategies.

Implications of unplanned pregnancy on the health of children and parents, given the prevalence of unplanned pregnancies and their potential impact on family health and well-being in both developing and developed nations. To control the confounding factors in this investigation, multivariate analyses were used. In this research (Gipson, 2008) including both mistimed and unwanted pregnancies, the impact of pregnancy desire on health outcomes was inconsistent.

To assess the association between child marriage, morbidity, and child mortality (Raj, 2010), bivariate analysis employed and exhibited strong correlations between post-natal child marriage and child, child diarrhea, malnutrition, low birth weight, and mortality. However, malnutrition is more likely in younger children born to mothers who are minors than those born to women who are of a specific age.

For such investigation and purposes, the latest nationwide data from India is used to investigate the negative consequences of intendedness for mothers' and children's health (Singh, 2012). This study includes the mother-fixed effects and found strong evidence that unwanted children face multiple risks right from their birth which include higher neo-natal, post-neo-natal, and early child death, there is also a high risk of stunting and being, not fully vaccinated. A study similar to this one focuses on how gender variations in percentage composition affect child health and survival (Mishia, 2014), (Das Gupta, 1997), (Chamarbagwala, 2011).

Logistic regression is used to determine the link between child marriage (before the age of 18) and mortality and morbidity in infants under five in Pakistan. (Nasrullah et al., 2014), analysis showed children born to younger moms, in lieu of child marriage, increases the chance of diarrhea. The nutritional condition and birth outcome of the pregnant teenage females in Tanzania found poor pregnancy outcomes. The study (Shirima et al., 2015) suggested a solution for such problems, that at the basic level of school, girls should be taught about reproductive health.

Furthermore, evaluation of nutritional condition among children under five who were born to adolescent and adult mothers in Ghana using logistic regression (Wemakor et al., 2018) compared the results. Children of adolescents were eight times more stunted, 3 times more wasted, and 13 times more underweight than adult mom’s children. Therefore, the risk of child malnutrition rises as the mother's age decreases.

Efforts in infant mortality or health situations are subject to preferences within the family, as the mortality rate is high in sub-Saharan Africa, where mother desire might be one of the factors. Using 79 DHS information, the study discovers that the child, unwanted or of no sex preference, by the mother relates to a higher mortality that is not caused by a consistent maternal factor but correlated with maternal preferences and varies consistently between siblings. As a result (Flatø, 2018) show excess in the morality of undesired children, accounts for 3.3 percent for male newborn and 4 percent for newborn female child.

There is a well-established link between unplanned pregnancy, mother, and child health.

However, the link between unplanned pregnancy and child malnutrition is not fully known and it might be relevant in counties like Bangladesh where both events are common. The results (Rahman et al., 2019) reveal that even after controlling other factors, the study shows that unplanned pregnancy is linked to an elevated risk of stunting, wasting, and being underweight among children under five and suggested that special attention needs to be paid to the issues (such as contraceptive failure) that promote undesired pregnancy directly or indirectly.

Multivariable mixed effects models are utilized to determine the stability of retrospective pregnancy, developing a hypothesis that among women who refused abortions and carried their pregnancies to term, would likely become more intended with time and have a lower rate of depression than those who had an abortion (Rocca et al., 2019). In response to the pregnancy result, women's views of their goals altered slightly.

For third-world countries like Bangladesh, malnutrition might be seen as a major problem. The purpose of the research is to categorize malnutrition using deep learning techniques to predictive modeling on important malnutrition traits to determine a child's nutritional status who is between the age of 0-59 months. BDHS (Bangladesh Demographic and Health Survey 2014) on children's nutrition is used with Artificial Neural Network technique. The most accurate technique for predicting wasting, underweight, and stunting is Artificial Neural Network. The concluding remarks by (Shahriar et al., 2019) utilizing deep learning to assess nutritional status is the best.

Another research study is aimed at how unplanned pregnancies, as well as, family and children's factors, affected the nutritional condition of children under five. (Shaka et al., 2020). Unplanned pregnancy was discovered to be one of the predictors of stunting. With the infant, from unplanned pregnancies, more likely to be stunted than children from planned pregnancies. Using multiple logistic regression, this study examines how the child's preference by mother's influenced childhood sickness and health. Different factors like the mother's education and income index were all found to be significant predictors of unwanted children (Alam & Islam, 2022). This study shows malnutrition and mortality were higher in undesired children. The higher infant mortality rate in Pakistan is linked to lack of women's education, poor economic situations, and low access of public healthcare facilities. Moreover, in Pakistan, health initiatives do not focus to address the most underserved population, such as women and children, particularly in rural areas (Patel et al., 2021).

One of the important measures of a woman's well-being is an intended pregnancy. In Bangladesh, married women experienced 28 percent of all pregnancies as unwanted (Hossain et al., 2022). This study sought to examine how well six different ML algorithms (RF, NB, SVM, KNN, ENR, and LR) performed when used to forecast unwanted pregnancies among married women in Bangladesh. The Elastic Net Regression (ENR) demonstrated the best outcomes and perhaps the most accurate classification for detecting unintended pregnancy taken into consideration. The finding may also help detect women who are at risk for unwanted pregnancy.

Another concern, all three forms of children's nutrition, including vitamin, over-eating, and micro-nutrient deficiencies are at a low level in Iran, suggested by a study. The most frequent nutritional issues cause child mortality. So, the study aims to develop a hybrid ML algorithm based on a food recommendation system to prevent malnutrition in children. In terms of target research, the findings of (Najaflou, 2021) are applicable and conclude that, in identifying malnutrition, the Decision Tree method is the best as its outcomes 98.5 percent accuracy.

### 2.1. Literature review for machine learning technique

By training and testing breast cancer database, (Lui et al., 2004) explored sample complexity in a linearly separable dataset. The analytics have emphasized two types of networks: single layer, and multilayer. The generalized results show both types of networks. The training sample size may be 1 with satisfactory outcomes. However, multilayer perform well in any training set, but a single layer needs the selection of a training sample. For multilayer threshold values for testing, the error is low. So, the results have suggested that it is possible to achieve good performance through training networks.

The analytical aspect of this study has emphasized the models that analyze species habitat interactions at the community level. Multivariate random forest (MRF) evaluates 541 estuaries with 24 predictors. Several predicted species habitat interactions were verified by the model findings. The results (Miller, 2014) identify, using advanced machine learning techniques like MRF are a more powerful method to model the species on spatial patterns. The focus of this work is to compare the performance of logistic and machine learning technique and the results proved that ML generally outperform linear regression in most cases, (Hurtado et al., 2016), the accuracy of the prediction is high for machine learning than linear regression.

Another research study focused on predicting the performance of machine learning and proving that ML is superior in the rare situation like identifying patients who are likely to suffer unplanned ICU (Intensive Care Unit) re-admissions are highly undesired, (Desautels et al., 2017) and therefore, increases the variation in care and increased mortality.

Machine learning regression techniques are increasingly being used to predict individualized behavioral outcomes. There are two main factors of this study. One is machine learning algorithm, and the other, it uses 700 samples (Cui & Gong, 2018). In the situation of the limited sample, the findings reveal that LASSO regression and OLS regression performed worse among other algorithms. Both regressions will perform when the sample size was increased. The findings give crucial insight into choosing ML techniques and individualized behavior. To provide information about basic concepts of machine learning techniques used in mental health situations and their real-world application (Cho, 2019). Among many algorithms, the Cho states that Support vector machine provides high accuracy. To identify malnutrition using a deep learning technique of predictive modeling on major malnutrition variables to predict malnutrition status in children aged 0-59 months. To achieve this, an artificial neural network technique is used, and it provides the best accuracy with stunting and wasting (Shahriar, 2019). To predict the performance of the students for different aims with small sample size, (Zohair,

2019) employed a machine learning technique to choose the most appropriate model, and the study demonstrated the efficacy of support vector machine and linear discriminant analysis algorithms in providing an acceptable small sample size producing in the situation of small sample sizes.

For the prediction of anemia status in children under 5, with common risk factors as characteristics of BDHS data is used. The study uses machine learning techniques such as CART, KNN, LR, LDA, RF, and SVM to predict anemia status. (Khan et al., 2019) discover that the RF method had the best classification accuracy and the K-NN method had the lowest accuracy of all the algorithms tested. Using the data from the BDHS 2014, this study assesses the efficacy of ML (RF, LR, CART, NN, SVM, and GBOOST) approaches in predicting stunting among children under the age of five.

Settling all this in mind, developing a peaceful community requires the healthy development of children. The purpose of this research was to predict infant nutritional condition by using the characteristics of the mother. This study uses a self-created dataset of underweight and normal weight infants along with 18 features to perform the ML techniques of LR, DT, K-NN, LDA, GNB, and SVM. The accuracy of the decision tree (DT) was 99 percent, the accuracy of K-NN was 85 percent, LR was 88 percent, LDA is 93 percent, SVM with 88 percent, and GNB is 86 percent. All the algorithms have performed incredibly well (Zakria Hussain et al., 2020).

Furthermore, the machine learning algorithm is used to predict the likelihood of a positive COVID-19 diagnosis. The study (de Moraes Batista, 2020) uses 235 adult patients in Brazil, and only investigated emergency cases. Several machine learning techniques were used, and results proved predicted performance of the support vector machines algorithm is the best in a limited sample size.

In developing nations, malnutrition is the leading cause of infant mortality. To determine the most significant determinants of undernutrition among children under 5 in Ethiopia, this article explored the effectiveness of ML methodologies. (Fenta et al.,2021) uses cross-sectional data, along with machine learning techniques: random forest (RF), elastic net, neural network (NN), least absolute shrinkage, and logistic regression. The optimal ML model was determined to be RF based on parent education, place of residence, and region in Ethiopia.

However, to analyze the most significant risk factors for stunting in Tanzanian children under five as well as the most effective classifier for predicting stunting (TDHS 2015) machine learning techniques are used. The model was built using the following algorithms: random forest (RF), decision tree, K-NN, SVM, and LR, and the accuracy of RF was higher among all the classifiers (Lucy Lawrence, 2021).

The 2014 Kenya Demographic and Health Survey and supervised ML algorithms were used to predict infant mortality. Various classification methods were employed. Random Forest, K-NN, and Logistic Regression were used. Random Forest had the highest accuracy; the results showed that the random forest model was the most effective one (Kioko, 2021).

With another study, machine learning is utilized to find the possible predictors associated with infant mortality in Bangladesh. To identify essential aspects of infant mortality, the Boruta method was applied. (Rahman et al., 2021) use decision trees, random forests, support vector machines, and logistic regression. In comparison to previous ML approaches, the random forest accurately predicted infant mortality, including the integration effects of variables.

Another similar literature uses machine learning (ML) algorithms to predict malnutrition risk characteristics (stunted, wasted, and underweight). Random forest shows more accurate results than other classifiers, whereas Logistic regression shows important factors to be statistically significant with stunting, wasting, and wasted children. (Rahman et al., 2021), research reveals that an LR-RF combination might be used to effectively classify and predict malnutrition status of children with more accuracy.

The EDHS is worth mentioning data sources for undernutrition, to predict the risk factor, multiple machine learning techniques (K-NN, RF, EXTRMEGBOOST, NN, and GLM) were used. The XGTRE method was a significantly superior ML algorithm for predicting childhood under-nutrition in Ethiopia as compared to other algorithms (Bitew et al., 2022).

Furthermore, machine learning algorithms were used to predict infant mortality in Rwanda using demographic and health surveys in another study. For predicting models, Random Forest, Decision Trees, Support Vector Machine, and Logistic Regression. This study preferred the RF model since it performed the best and the model was the best predictive model of infant mortality accuracy (Mfateneza et al., 2022).

In India, the malnutrition rate is extremely high. Predicting the birth weight of newborns allows parents and medical experts to get prepared to take prevention and take steps to promote the child's development. For this Gaussian Nave Bayes (GNB) and Random Forest was used. Self-generated data with 18 different mothers related- features was used (Jain et al.,2022). GNB's accuracy was 86% while RF was 100%, both approaches have shown significant results.

Using various ML techniques to predict the nutritional status of under 5 children in Bangladesh. The 2014 BDHS secondary dataset was used for analysis. Support vector machine, linear discrimination analysis, k-nearest neighbors, and random forest and logistic regression techniques have been explored to reliably predict the condition of children. Based on the results (Islam et al., 2022), the RF technique was marginally more effective than other machine learning (ML) in this study to predict the nutritional status of children under 5.

## 2.2 Literature Gap

With high rate of child morbidity and mortality in Pakistan, investing in infant health is tough within households, in which maternal preference play important role. However, (Flatø, 2018) proved the being born parity or of a gender undesired by mother relates to infant mortality and how this differential mortality varies between mother at different stages. (Alam & Islam, 2022); (shaka et al.,

2020) evident that the childhood morbidity, mortality, and malnutrition is higher in undesired child. Only (Flatø, 2018) ;(Alam & Islam, 2022) ;(Shaka et al., 2020) explored undesired children under five by using traditional statistical methods. Literature makes it quite evident that undesired children are a neglected aspect of our culture, in part because moms tend not to report them directly. Therefore, it is crucial to emphasize the unwanted children to lower the mortality and morbidity rates.

### 2.3 Summary of Literature Review

We conclude that different researchers estimated unintended pregnancies and their impact on the mother's health very few studies examined the association between undesired children and their health status, but no study is found that had used machine learning techniques on the undesired child. The structure and impact of this research study intended to fill the gap by explaining this area in the case of Pakistan. We intend to explore the association between undesired children and their health challenges in the case of Pakistan using machine learning techniques and identifying them.

# CHAPTER 3

# DATA AND METHODOLOGY

## Introduction

In machine learning, classification is one of the mostly applied task when dealing with high-of dimensional data which often leads to non-linearity and don't not meet the assumption traditional statistical procedures (Kampichler et al.,2010). The importance of feature selection increases, especially in data set with plenty of variables and features (Chen et al.,2020). This chapter discusses the machine learning techniques which will use in predicting the performance of the undesired child and their health. Section 3.1 describes each, the model used for predicting the undesired child health status in this study. In which an overview of all models (1) Linear Discrimination Analysis, (2) Random Forest, (3) Neutral Network, (4) Support vector machine is given. Section 3.2 describe the variables and data used for the study.

## 3.1 Materials and Methods

### 3.1.1 Source of Data and limitations

Pakistan Demographic and Health Survey of 2017-18 (PDHS) is used in this study. It was a probability sample survey of the Pakistan population that was conducted across the country. A two-stage cluster survey approach was utilized, with total clusters chosen in the first stage and a systematic sample of 12,815 (HHS) on average in the second stage. This survey gathered information on demographics, socioeconomic status, health, and nutrition. The 2017-18 PDHS gathered data from 12,364 ever-married women, with all ever-married women being asked to submit extensive information on their wishes and actual births in the previous four years. All such women are excluded from the survey who wanted more than nine children because having more than nine children is considered normal fertility, and the number of children is “up to God”. We utilized whether the last child was “wanted or not” for the validation of the measures of the undesired child.

### 3.1.2 Outcome variables

Child mortality includes early neonatal, late neonatal, post neonatal, infant, child, and underfive moralities, childhood morbidity includes fever, cough, diarrhea short and/or rapid breath, and acute respiration infection, treatments for morbidity (fever and cough, diarrhea), receiving postnatal checkup (PNC) within two months of the births, and malnutrition were undernutrition and over nutrition.

Early neonatal mortality (ENM) is the percentage of a child that die within the first six days of their birth. Late neonatal mortality (LNM) is the proportion of the child who died between the age of 0 and 28 days, while post neonatal mortality (PNM) is the percentage of the child between the ages of 29 and 364 days. The infant is the percentage of the child who died before reaching his/her first birth (0 and 365 days). The child percentage of the child died between after their first birth and before reaching their fifth birthday. The under-five mortality (U5M) is the percentage of children who died before they reach the age of five.

The morbidity data was obtained from those who had been identified with a condition (yes/no). The outcome variable fever was coded as 0 and 1, representing children who hadn’t fever coded as 0 and who had a fever within the last two weeks is coded as 1. The same technique is used for cough, diarrhea, and short rapid birth as they are for fever. The children under the age of five showing the symptoms of chest-related infection or fast breathing were used to compute the acute respiratory infection (ARI). With the most prevalent morbidities, we also include the postnatal child checkup, treatment of diarrhea, and fever/cough.

DHS calculates nutritional status using anthropometric measurements. The mother's estimation of the childbirth size now of the birth was used to determine the size, which was coded as large, larger than size, average, smaller, and smaller than size. Based on WHO child growth standard -2 SD or -3 SD below the mean for undernutrition and above +2 SD or +3 SD over nutrition, we also calculate stunting, severely stunted, wasting, severely wasting, underweight, and severely underweight.

**Table 1: Outcome Variables of the Study**

| Outcome Variables | |
|---|---|
| **Childhood Malnutrition** | **Child Morbidity** |
| Stunting | Fever |
| **Wasting** | Cough |
| Underweight | Diarrhea |
| Birth size | Acute Respiration Breath |
| **Child Mortality** | Short Rapid Breath |
| Early Neonatal | **Treatment of Childhood Morbidity** |
| Late Neonatal | Postnatal checkup baby within 2 months |
| Post-Neonatal | Treatment of Diarrhea |
| Infant | Treatment of Fever/Cough |
| Child Under 5 | Vitamin A in the last six months |

### 3.1.3 Predictor Undesiredness

The undesired child is a key predictor of the study, this predictor variable s is not directly reported, some research uses an undirected method to measure the undesired child. Excess and no-excess in children were the categories for the desired status of children. In DHS, women aged 15–49 were asked about the ideal number of children including gender composition, were questioned about their desired number of children, including gender composition, whether they wanted boys (Number of boys), girls (number of girls), or both boys and girls (number of boys and girls) [(PDHS 2017-18); (Alam & Islam, 2022); (Croft et al., 2018)].

The measure of an undesired child in this research is taken from a question in the DHS that asks all the women aged 15-49 to estimate how often children they would want if they could start again. Individuals are then questioned if they want male or female offspring, or children of both genders(parity). Several potential flaws in using the stated intended family size as a criterion of undesiredness will be considered. Nonetheless, the research contends that three categories of undesiredness may be used to analyze individual morality, morbidity, and nutritional status: excess gender composition, excess parity, and dual excess. This research also discusses how the occurrence of these three sorts of the undesired child varies depending on the reproductive wishes of the mother, and the wide range of gender preferences prevalent across the country.

The following are the indicators of undesired children, which are produced using a standard technique. Let Ci be the number of living children born to the same mother (including i who may or not maybe (Gi OR Bi)).

$$\boldsymbol{Ci} = \boldsymbol{Bi} + \boldsymbol{Gi} \qquad \textbf{(1)}$$

Additionally, let Cm indicates the mother’s desired family size, she also has gender desire for those children who include a certain number of boys (Bm), girls (Gm), and children of either gender (Nm) (mother might prefer either gender of the child). The following is the summary of the preferences:

$$Cm = Bm + Gm + Nm \tag{2}$$

Excess in parity can be defined as:

$$Epi = 1 \quad if\ Epi = Ci - Cm > 0 \tag{3}$$

Let the child’s gender i be s, which might be male or female. The number of children of the same sex as child Si (Bi or Gi), and the desire for children of that gender, Sm can then be calculated. Which result as excess in the composition is given as follows:

$$Esi = 1 \quad if\ Esi = Si - (Sm + Nm) > 0 \tag{4}$$

Children are in excess if the overall number of children or the number of same-sex children exceeds the desired number. Here's a short illustration. If a mother has six children and wishes to have four more. One girl, two boys, and one who might be of either gender make up her desired family. She gives birth to four sons initially. The first three boys are not in any way excessive. In terms of composition, the fourth boy is excessive, but not in terms of parity. After that, her son died. She then had a girl who is not in any way excessive. in fact, the final child she delivered is a state of dual excess.

A cross-sectional study is conducted using Pakistan Demographic and Health Survey 2017-18. This study is designed to predict the prevalence of undesired child health, to predict the health condition of the undesired child, and the factor associated with undesired child, we utilize Pakistan Demographic Health and Survey (PDHS) 2017-18. The data is split into two parts one is the training part and the other is testing. In this study, we are considering morbidities, mortalities, and nutritional status of undesired children as a dependent variable which are described below, and undesired child as an independent variable along with other co-factors. Undesiredness can be measured by an indirect approach in this study by using PDHS. Therefore, we have different models in this study. Whereas explanatory variables for determining the effect of the undesired child, we grouped the variables into demographic variables and socio-economic variables. These all variables are selected based on previous literature.

### 3.1.4 Explanatory Variables

The goal of this study is to determine the number of undesired children based on socioeconomic differences. We attempted to account for the possibility that undesired children are linked to parity, age of mothers, education, and wealth during measurement.

We select the co-factors based on the previous study (Alam & Islam, 2022); (Shaka et al., 2020); (Flatø, 2018) and considering demographics, socioeconomic, and developmental variables that influence children's health. However, numerous research has been carried out in Pakistan ((Khan et al., 2019) ;(Nasrullah et al., 2014); (Khan et al., 2016)) to determine the causes and variations in infant mortality, morbidity, and malnutritional associated with factors like mother education, region, birth order, gender, employment status, wealth index, and residence. Moreover, (Alam & Islam, 2022) identified the mother nutritional status, birth size, birth order, household size, mother education, mother care, wealth status, region (KPK, Punjab, Sindh, and Baluchistan) Kashmir and GB data is excluded from data, and residence (Urban and Rural area) as significant factor of undesired child health under five.

The factors for health and demographic are year of birth, the current age of the mother, sex of child is defined as girl or boy, and parity/ birth order is taken from 1 to 8. By sub-divide the current age of women which is given in 6 groups (18-24,24-29,30-34,35-39,40-44,45-49).

The mother education (primary, middle, high, and no education no), wealth index (poor, middle, and rich) we used the wealth index variable contrasted in the PDHS survey.

The person who usually chooses the respondent's health care, the person who usually decides on large household purchases, and the person who usually decides on visits to family or relatives were used to quantify women's empowerment. socioeconomic variables (Raj et al., 2010), (Alam & Islam, 2022). Furthermore, as programmatic variables impacting child health in Pakistan, we have antenatal care, postnatal checkup, visits by family planning professionals, and access to media including TV, Radio, and Newspapers categories as "Yes and No". Moreover, we had ANC, PNC, visits by (FP) workers in last six month reported as "Yes or No", and access to any media for FP information.

**Table 2: Explanatory variables of the Study**

| **Explanatory Variables** | |
|---|---|
| **Undesired Predictors** | **Socioeconomic and Demographic Predictors** |
| Excess in Girls | Birth Order |
| Excess in Boys | Sex of Child |
| Dual Excess | Mother Education |
| Excess in Parity | Working Status of Women |
| | Mother Age |
| | Women Empowerment |
| | Wealth Index |
| | Excess to Media |
| | Visit of Family Planning in Last Six Month |
| | Region |
| | Province |

### 3.1.5 Measuring undesired Child

The undesired child is a key predictor of the study, this predictor variable s is not directly reported, some research uses the undirected method to measure the undesired child. Excess and no-excess in children were the categories for the desired status of children. In DHS, women aged 15–49 were asked about the ideal number of children including gender composition, were questioned about their desired number of children, including gender composition, whether they wanted boys (number of boys), girls (number of girls), or both boys and girls (number of boys and girls) (PDHS 2017-18; (Alam & Islam, 2022); (Croft et al., 2018)). We identified four types of undesired children, as being excess in boy children, excess in the girl child, excess in both boys and girls, and excess in the total number of children. We measured undesired children using the conventional approach as found in existing literature ((Flatø, 2018); (Bongaarts, 1990); (Bongaarts, 1992); (Casterline et al., 2007)). Therefore, we used the built-in PDHS variable to check the measurement of undesired children “whether wanted last pregnancy” coded as “wanted then,” “wanted later,” and “wanted no more” which is more affected by rationalization (after birth, the undesired child may be desired) (Alam & Islam, 2022).

### 3.1.6 Methodology

Machine algorithms have been used in medical studies to predict a variety of childhood health statuses. (Shahriar et al., 2019), (Jaskari. et al 2020). Random forest (Rf), linear discrimination analysis, neural network, and support vector algorithm, has been utilized to assess the status of malnutrition, and illnesses (morbidity) and include their treatment and morality using common socio-economic factors.

However, research has focused on using machine learning approaches to create prediction models for childhood malnutrition and the machine learning algorithm predict the best model with higher more accuracy (Shahriar et al., 2019). But in health research, undesired child using ML algorithms is not explored before in case of Pakistan. Furthermore, in addition to the capabilities of standard clinical equipment, ML may be more capable of predicting, more accurately. Therefore, in this research well-known machine learning methods such as linear discrimination, random forest, neural network, and support vector machine is used to predict the health status of under-five children in Pakistan. Additionally, accuracy, sensitivity, and specificity are used to analyze the algorithms systemically. So, we have following are the models of this study:

## 3.2 Econometric Model

Multivariate analysis is involved multiple independent variables in the model and one outcome variable. Multivariate data deal with data reduction which highlights our second objective of the study.

Multivariate regression in supervised machine learning algorithm involving multiple data variables for analyses, as multivariate regression is extension of multiple regression with one dependent and multiple independent variables. Based on independent variables, we try to predict the output.

**General model:**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

Y= (y1, y2, ……, yn) is Outcome vector P×1

X= (X1, X2, …., Xp) which is design matrix

Bo = (B0, B1, …., Bp) is vector of coefficient

ei = (e1, e2, …., ep) error vector.

The final model of the study is given the detailed following the multivariate structure.

**Model for Child Morbidity**

$$Child\ Morbidity = \beta_0 + \beta_1 excess\ in\ girls + \beta_2 excess\ in\ boys + \beta_3 excess\ in\ parity + \beta_4 dual\ excess + \beta_5 mothereducation + \beta_6 working\ status + \beta_7 wealth\ index + \beta_8 Province + \beta_9 residing + \beta_{10} acess\ to\ media + \beta_{11} home\ visit\ by\ family\ planning + \beta_{12} childage + \beta_{13} birthorder + \beta_{14} womenempowerment + \beta_{15} Sex\ of children + \beta_{16}\ mother\ age$$

Outcome variables for child morbidity under five are Fever in last two weeks, Cough in last two weeks, Diarrhea in last two weeks, Acute Respiratory Infection, Short, Rapid Breath and predictor remain same for all the outcome variables.

**Model for Child Malnutrition**

$$Child\ Malnutrition = \beta_0 + \beta_1 excess\ in\ girls \beta_2 excess\ in\ boys + \beta_3 excess\ in\ parity + \beta_4 dual\ excess + \beta_5 mothereducation + \beta_6 working\ status + \beta_7 wealth\ index + \beta_8 Province + \beta_9 residing + \beta_{10} acess\ to\ media + \beta_{11} home\ visit\ by\ family\ planning + \beta_{12} childage + \beta_{13} birthorder + \beta_{14} womenempowerment + \beta_{15} Sex\ of children + \beta_{16}\ mother\ age$$

Outcome variables for child malnutrition under five are Severely Stunted, Stunted, Severely Wasted, Wasted, Severely Underweight, Underweight, and Birth Size. Covariates remains same for all the outcome variables.

**Model for Child Mortality**

$$Child\ Mortality = \beta_0 + \beta_1 excess\ in\ girls \beta_2 excess\ in\ boys + \beta_3 excess\ in\ parity + \beta_4 dual\ excess + \beta_5 mothereducation + \beta_6 working\ status + \beta_7 wealth\ index + \beta_8 Province + \beta_9 residing + \beta_{10} acess\ to\ media + \beta_{11} home\ visit\ by\ family\ planning + \beta_{12} childage + \beta_{13} birthorder + \beta_{14} womenempowerment + \beta_{15} Sex\ of children + \beta_{16}\ mother\ age$$

Outcome variables for child mortality under five are Early Neonatal Mortality, Late Neonatal Mortality, Post Neonatal Mortality, Infant Mortality, Child Mortality, Under 5 Mortality. Covariates remains same for all the outcome variables.

**Model for Treatment of Child Morbidity**

$$Treatment\ of\ Child\ Morbidity = \beta_0 + \beta_1 excess\ in\ girls \beta_2 excess\ in\ boys + \beta_3 excess\ in\ parity + \beta_4 dual\ excess + \beta_5 mothereducation + \beta_6 working\ status + \beta_7 wealth\ index + \beta_8 Province + \beta_9 residing + \beta_{10} acess\ to\ media + \beta_{11} home\ visit\ by\ family\ planning + \beta_{12} childage + \beta_{13} birthorder + \beta_{14} womenempowerment + \beta_{15} Sex\ of children + \beta_{16}\ mother\ age$$

Outcome variables for child mortality under five are Postnatal Checkup of Baby within 2 days, Treatment of Fever/Cough, Vitamin A in last Six Months, Treatment of Diarrhea. Covariates remains same for all the outcome variables.

Machine learning techniques are popular in detecting the disease. Supervised machine learning has potential to improve the efficiency. Supervised machine learning techniques are useful to extract the patterns. It takes direct feedback for prediction. Whereas supervised machine learning is categorized in classification and regression methos. SVM, RF, NN, KNN, DT (Chauhan.et al., 2021) and many more are popular supervised learning. The aim of the classification algorithm is to detect and predict the disease. On other hand unsupervised learning techniques does not feedback for predication. Unsupervised machine learning algorithms find the hidden pattern of data. Unsupervised learning technique like PCA are used for dimensionality reduction. Some supervised machine learning techniques are discussed below used in this study.

### 3.3 Linear discrimination Analysis

ML algorithms are model-free approaches for solving classification problems efficiently (Khan et al., 2019)**.** Linear Discriminant analysis is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. This method is used in statistics to recognize patterns and find the linear combination of features that characterizes or separates two or more classes (Rusdiana, 2016). Linear discrimination analysis (Tharwat et al., 2017) used to discover linear combination variables or features, which are helpful for dimensionality reduction. It is a dimension reduction technique, not a discriminant classifier, that considers a linear transformation that maximizes class reparability in a reduced dimensional space, with the criterion being to find dimension reduction that maximizes between-class scatter while minimizing withinclass scatter (Park, 2005, Luo & Chen, 2020). Some limitation of LDA is

- LDA is supervised machine learning, means it need labelled data to work.
- It assumes that all the classes are normally distributed and that the feature are uncorrelated.
- It also assumes that small sample size (smaller classes) needs to exceed to predict the variables and unequal sample size is acceptable.

Discriminant analysis is used where the cluster is known as priori. LDA is closely related to analysis of variance (ANOVA) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. However, ANOVA uses categorical independent variables and a continuous dependent variable, whereas discriminant analysis has continuous independent variables and a categorical dependent variable (i.e., the class label Logistic regression and probit regression are more like LDA than ANOVA is, as they also explain a categorical variable by the values of continuous independent variables. These other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the LDA method.

$$\boldsymbol{Y = W1X1 + W2X2 + \cdots + WnXn} \qquad \textbf{(5)}$$

Where Y is the discrimination score of morbidity, mortality, and malnutrition. W's the vector of the discriminant coefficient, and X's the vector of n independent variables of the study. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis.

## 3.4 Random Forest

Breiman in 1995 and 1996 invented the bagging approach, which aggregates many tree classifiers, each built on a bootstrap sample of the training data. It is used in regression as well as in classification problems.

Random forest has proven to be a very useful approach, the feature selection with higher number of variables can be handled easily (Chen et al., 2010). Random Forest is a classification algorithm that contain a random set of decision trees to amplify the given dataset with (Biau et al., 2008). RF is an extremely powerful prediction machine that has one of the best prediction results as well as a high level of variable importance. RF captures both linear and non-linear correlation with modification of the model parameters. RF analyzes multidimensional interactions among predictors since it utilizes TREE as building components suppose we have Yi (i=1...., N) and Xip (p=1...., p) are response and predictors variables, where Xip are the number of the pth node for the (undesired and other co-factors) ith gene, Yi is outcomes variable for ith condition.

- Binary (yes/no) questions, often known as splits, are questions that are expressed in terms of predictors and are used to partition the predictor space. Nodes are subsamples called leaves.
- Each tree is made of the bootstrap sample with replacements selected from the dataset.
- Within the framework of regression, a node impurity measure is related to the response variable.
- A spilled function ϕ (s, t) can be used to test each node split. The optimal split, which maximizes, has the most homogenous response distributions in the ensuring children's nodes between all other splits.

## 3.5 Neural Network

The NN is a common machine learning that resembles the structure of human neurons. McCelloch and Pitts, Widrow and Hoff, and Rosenblatt are all authors of NN (McCelloch, 1943,). The NN extracts linear combinations of inputs and passes them on the output to the next layer. The final output is compressed and utilized for classification using a logistic function. the basic component of neural networks delivers excellent prediction performance. For a network having many hidden layers, a NN may become more computationally intractable.

A neural network tries to make a prediction by connecting one or more output variables Y on one or more input variables X. In its most basic form, the input patterns are applied at the first layer, which also connects the network to its surroundings. The second layer is hidden layer, it uses the RBF. The network responses are applied to the activation patterns via the output layer, which linear:

$$\boldsymbol{Yi} = \sum\sum \boldsymbol{wmj}\mathbf{G}(||\mathbf{xi} - \mathbf{cm}||) + \mathbf{b} + \mathbf{eij} \tag{7}$$

Where Yi is the ith member of the YN, K output matrix, and no of RBNN responses. Wmj is the weight matrix, Xnm is input patterns, cm is the square centroid matrix and eij is the matrix of residuals.

**Assumptions**

- The distribution of the outcome variable must be of normal distribution.
- The variances of residual eij should be equal
- There should no correlation between residuals.

### 3.6. Support vector Machine

SVM is popular ML algorithm for classification. (Boser et al., 1992), (Vapnik, 1992) introduces SVM, and Cortes and Vapnik represented the current standard version (1995), (Cortes et al., 1995). SVM is a supervised ML approach used to solve classification and regression problems. The support vector machine is described as:

- The concept of SVM is to find optimal hyperplane.
- SVM separate the two considered classes by maximizing the distance between the classes and closest point.
- Another thing SVM do if any point lying on the boundary it calles as support vector. Moreover, middle of margin is called optimal hyperplane and the point lyes on wrong direction are weighted down.
- To find the optimal hyperplane, linear separation is used. Otherwise for non-linear kernel techniques are used to find optimal hyperplane.
- The non-linear used following function; radial, polynomial, and hyperbolic tangent function.

The goal of parametric classification is to characterize each class's usual feature space values or distribution of each class (Akay, 2009). SVM, on the other hand, concentrates solely on the training samples that are close in the feature space to the ideal class border (Cortes & Vapnik, 1995); (Vapnik, 2013); (Pal, 2012). SVM is frequently used and gives great prediction performance.

The problem may be avoided by repeatedly applying the classifier to every conceivable combination of classes; however, this means that processing time will grow exponentially as the number of classes grows (Croft et al., 2018); (Vapnik, 2013). In addition, this technique provides great discriminative power by the transformation of the input space into multi-dimensional space using specific non-linear functions known as kernels. We will use the kernel function, which converts the non-linear data space into linear data space, so the construction model of the SVM model is as follows:

$$\boldsymbol{yi} = \boldsymbol{W}.\,\boldsymbol{Xi} + \boldsymbol{\beta} \qquad \textbf{(8)}$$

Where Yi is outcome variables (malnutrition, morbidity, and treatment of morbidity) detail in table 1, X's are independent variables of the study, βis constant, and W are weight vector. In contrast to multi-logistic regression, the SVM method tends to classify things without assigning estimates of probabilities in the dataset. Yu et al., (2010) show, traditional statistical approaches may be outperformed by the SVM methodology.

### 3.7.1 Model evaluation

Four evaluation parameters were taken into consideration.

### 3.8 Confusion Matrix

A confusion matrix is a performance statistic that is typically used to evaluate the performance of the predicted data for which real values (or classifiers) are known (Khan et al., 2020).

Especially, the confusion matrix summarizes the results of classification issue predictions. Count numbers and percentages are used to describe the number of correct and incorrect predictions. So, it is key to the confusion matrix. When predictions are created, the confusion matrix shows how confused your predicted model is. It gives us insight not just into the classifier's error, but also into the types of errors that are made.

The confusion table is shown below:

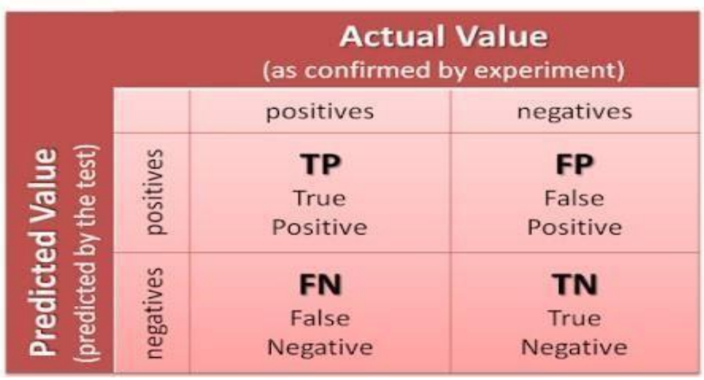


**Figure 1: confusion matrix**

### 3.9 Accuracy

Accuracy is the basis of estimating the performance of any predictive model. It estimates the ratio of right predictions to the total number of data points evaluated. This study was comprised of the best accuracies that were obtained by several ML algorithms after applying the feature selection as well as k-fold techniques. Where misclassification is equal to 1-accuracy Mathematically, accuracy can be calculated as the following:

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

### 3.10 Sensitivity

Sensitivity is defined as the proportion of real positive cases that got predicted as positive (or true positive). It is also termed recall. This implies that there will be another proportion of real positive cases, which would get predicted incorrectly as negative (termed as the false negative). This can also be presented in the form of a false-negative rate. Mathematically, sensitivity can be estimated as the following:

$$Sensitivity = \frac{\mathbf{Tp}}{+ \boldsymbol{FN}} \boldsymbol{TP}$$

### 3.11 Specificity

Specificity is the proportion of real negative cases that got predicted as negative (or true negative). This implies that there will be another proportion of real negative cases, which would get predicted as positive and could be termed as false positives. This can also be presented in the form of a false-positive rate. Mathematically, specificity can be estimated as the following

$$Specification = \frac{\mathbf{TN}}{\boldsymbol{TN + FP}}$$

### 3.12 Precision

One measure of machine learning is precision, the quality of a positive predication made by the model. Precision is calculated by dividing the number of true positive predication by the total number of true positive predication.

**TP**

$$\boldsymbol{Precision} = \frac{}{\boldsymbol{TP} + \boldsymbol{FP}}$$

# CHAPTER 4

# RESULTS AND DISCUSSION

High dimensional data, particularly in terms of many characteristics is now more prevalent in machine learning challenges (Wang et al., 2019). In machine learning Feature selection is most power tool of machine learning. There are four main reasons why feature selection is important:

(i) To simplify the model by reducing the parameters.

(ii) Decrease the training data set

(iii) Overfilling can reduce by enhancing generalization.

(iv) To avoid the curse of dimensionality

Besides, we evaluate the models and then compare each performance and accuracy of the four algorithms. The nationally representative data, Pakistan Demographic and Health Survey (PDHS) is used to estimate the model of four classification methods; linear discriminate analysis, Random Forest, Neural Network, and Support Vector Machine and evaluate these algorithms to check which model is the best in the empirical analysis of child health. For estimating the model, we divide the dataset into training and testing sets. 70 and 80 percent of data were included in training and 30 and 20 percent in testing.

To evaluate the prediction performance, we used the 5-fold and 10-fold cross-validation method which shows good performance in model selection. Then, we compare the different machine learning models based on their predicted accuracy. The confusion matrix summarizes the results of classification problem predictions and is used to illustrate the accuracy.

Following is the introduction of the chapter rest is divided into subsections. Section 4.2 will explain the results and discuss the estimated parameter discussed in Table 1 for the undesired child aged

0-5 years and evaluate the measurements of the estimated machine learning algorithm. Table 3 present descriptive statistics of the background characters associated with the undesired child. Almost all the background characteristics are significantly associated with the undesired child.

**Table 3: Descriptive Statistics of Predictors**

| Predictors | Mean | Variance | Standard Deviation | Skewness | Kurtosis |
|---|---|---|---|---|---|
| Excess in girls | 0.22 | .1711929 | .4137547 | 1.358796 | 2.846326 |
| Excess in parity | .1646859 | .1376813 | .3710544 | 1.808126 | 4.269319 |
| Excess in dual | .2275042 | .1758954 | .4193988 | 1.300011 | .690028 |
| Excess in boys | .1044143 | .0935914 | .3059271 | 2.587242 | 7.693823 |
| Mother education | .917657 | 1.197122 | 1.094131 | .7123503 | 2.025234 |
| Excess of media | .1918506 | .1551757 | .3939234 | 1.565179 | 3.449784 |
| Women empowerment | .1876061 | .336057 | .5797042 | 3.440711 | 14.7294 |
| Wealth index | 1.896435 | .7913042 | .8895528 | .2036838 | 1.295283 |
| Birth order | 1.485569 | .6320891 | .3995367 | .9427697 | 2.819061 |
| Residence | 1.528014 | .249427 | .4994266 | -.1122306 | 1.012596 |
| Visited by fieldworker | .549236 | .2477862 | .4977812 | -.1979058 | 1.039167 |
| province | 3.036503 | 4.354656 | 2.086781 | 1.131189 | 3.403336 |
| Mother age | 2.230051 | .3947801 | .6283153 | .1918056 | 3.080761 |
| Working status | .1561969 | .1319114 | .3631961 | 1.894013 | 4.587285 |
| Child age | 1.973684 | .8191879 | .90509 | .7739756 | 2.965189 |
| Child sex | .4626486 | .2488161 | .4988147 | .1498244 | 1.022447 |

In table 2, descriptive statistics of background characteristics of undesired child is given and we can see that no predictor share common mean.

## 4. Child morbidity

### 4.1 Random Forest

All morbidity data were obtained from those who had been diagnosed with a condition (yes/no) in the two weeks before the survey. Similarly, children under the age of five had a fever at any point in the two weeks before the survey was used to calculate fever. The same technique is used for assessing cough, and short and/or rapid breathing (SRB) as it is for fever. The symptoms of chest related short or rapid breathing of children under 5 were used to compute the acute respiratory infection (ARI).

**Prediction from Random Forest by Using Training Dataset**

It is an ensemble machine-learning technique in which the results of multiple decision trees are built and merged. This method provides better and improved predicted results. In a random forest, we have chosen 550 trees. If we increase the number of trees, the error rate remains the same. The parameter for child morbidity of the undesired and desired child is mentioned in Table 1.

The training set confusion matrix tables are shown in the appendix. The random forest (RF) algorithm 4.12 table shows that the percentage of undesired children who had a cough in the previous week before the survey is 5.32%, which is correctly predicted by the model, whereas 53.56% are children who had no cough in the last two weeks. The incorrect prediction of the model (359 undesired children had cough out of 6107 are incorrectly predicted). The class error of having no cough is 9.88%, and having cough in Table 4.12, is 86.8%. The percentage of children with no acute respiratory infection is 1.8% and who had an acute respiratory infection (ARI) is 7.4%. The error class for a group that had no acute respiration infection is 92.4% and the children had acute respiratory infection is 3.7%. The 00B error rate is 24.85 percent.

The undesired children who experienced fever in the last two weeks were found to be 6.28% and 52.02% are children who are not experiencing the fever. The out-of-bag sample of fever is 41.06%, with 77.1% correct prediction that belongs from had no short rapid breath (SRB) and had short rapid breath is 0.14% which are correctly predicted. We found that 77% of children had not experienced diarrhea, and 1% of children had experienced diarrhea accounted for by RF forest. The prevalence of symptoms of short rapid breath (SBR) is 0.14% with 22.72% out-of-bag error.

### 4.1.1. Prediction from Random Forest by Using Testing Dataset

In table 4.1 the model made a correct prediction of 57.9% for children who do not experience the fever and 38% prevalence of symptoms of fever excess in girls predicted by the model. About 3.17% excess in parity had experienced a cough and 58% had not had a cough. The prevalence of children who does not experience diarrhea is 76.09% and excess in boys has experienced the diarrhea in last two weeks show no prediction, 2% correct prediction that had no acute respiration infection, 73.3% the symptoms of acute respiration infection among the excess in girls. While excess in parity experiences 0.14% short rapid breath (SBR).

### 4.1. Confusion Matrix of Child Morbidity using Random Forest

| Confusion matrix of fever using Random Forest | | |
|---|---|---|
| | **No Fever** | **Fever** |
| **No Fever** | 57.80% | 3.30% |
| **Fever** | 0.32% | 38.40% |
| **Confusion matrix for Cough using Random Forest** | | |
| | **No Cough** | **Cough** |
| **No Cough** | 58.70% | 31.70% |
| **Cough** | 6.34% | 3.17% |
| **Confusion matrix for Diarrhea using Random Forest** | | |
| | **No Diarrhea** | **Diarrhea** |
| **No Diarrhea** | 81.40% | 0 |
| **Diarrhea** | 18.60% | 0 |
| **Confusion matrix for Short Rapid Breath using Random Forest** | | |
| | **No SBR** | **SBR** |
| **No SBR** | 79.10% | 20.02% |
| SBR | 0.71% | 0.10% |
| **Confusion matrix for Acute Respiration Infection using random forest** | | |
| | **No ARI** | **ARI** |
| **No ARI** | 2.03% | 2.90% |
| **ARI** | 21.70% | 73.30% |

### 4.1.2. Important Variables of Child Morbidity

However, we are more interested to know which variable is more important in the building of undesired child morbidity, mortality, and treatment of morbidity model. Two measures, mean decrease accuracy and mean decrease Gini are commonly used in the investigation of important variables under random forest. The mean decrease in the Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest.

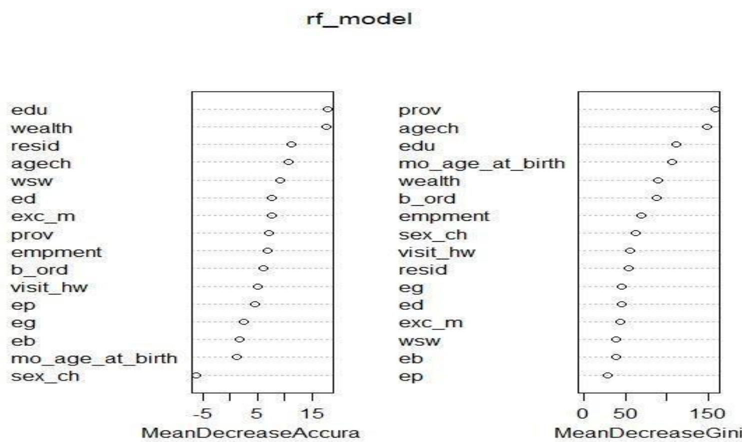


**Figure 2: Plot of Cough**

This is the fundamental outcome of the random forest, and it shows how important it is in classifying the data. The mean decreasing accuracy shows plot expresses how accuracy model losses by excluding the variable. The symptoms of cough in undesired children is higher belonging to middle class (17.7%) and non-educated (17.9%) women. The mean decreasing Gini plot expresses province is important.

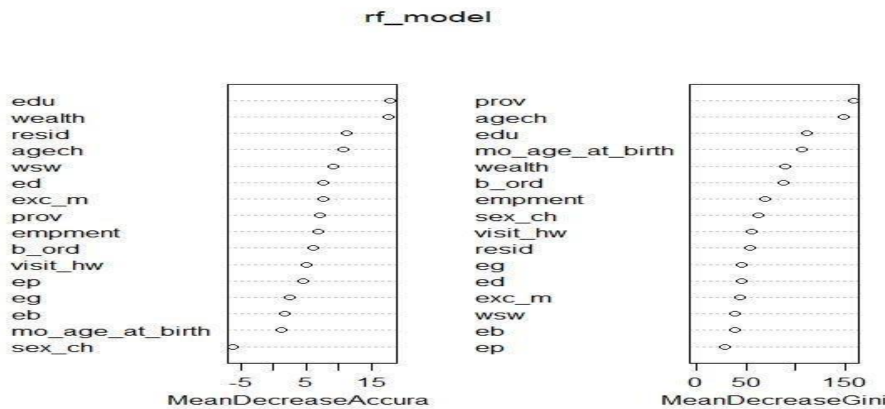


**Figure 3: Plot of Acute Respiration Infection**

This is the fundamental outcome of the random forest, and it shows how important it is in classifying the data. The mean decreasing accuracy shows plot expresses how accuracy model losses by excluding the variable. Figure 3 shows 36% female education, and 43% wealth index is through a significant determinant of undesired children suffering from ARI. The women from Baluchistan and KPK had, 35% excess in girls suffering from ARI.

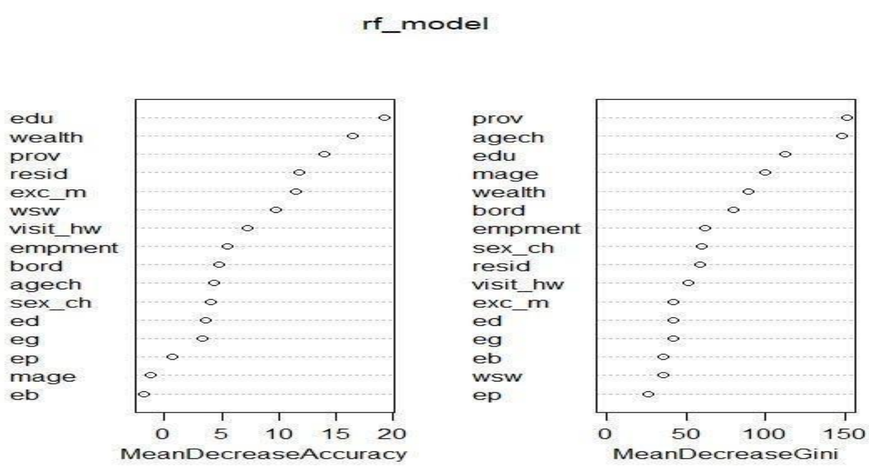


**Figure 4: Plot of Fever**

Figure 4 illustrates the important variable. According to the mean-decreasing accuracy plot mothers' education (30%) and 16.5% poor women had a higher rate of undesired children having fever, while the mean-decreasing Gini plot shows province 13% and child aged 4.2% play a significant role among dual excess children who experience fever.

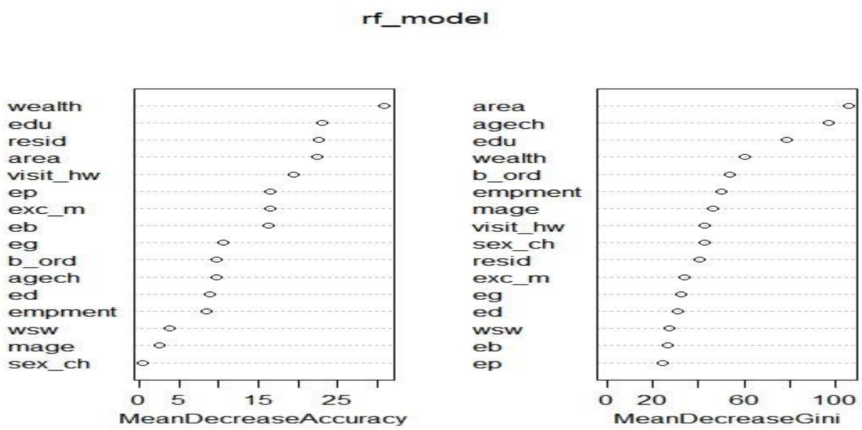


**Figure 5: Plot of Diarrhea**

The prevalence of diarrhea is higher among excess in undesired boys which is 16.3% are associated with 23% low education of mother, 30% poor women. Mean decrease Gini shows 29% women from urban area had higher undesired children (excess in boys) suffering from diarrhea.

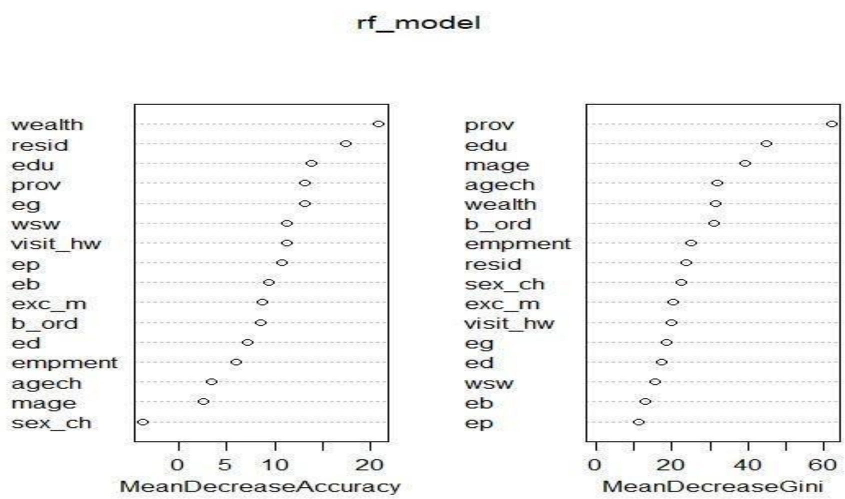


**Figure 6: Plot of Short Rapid breath**

However, according to mean decrease accuracy most important variable in modeling the experience of short rapid breath within the last two weeks among undesired child is the 21% wealth and 22% women belong to urban area. The mean decrease in the Gini coefficient measure of variable importance for estimating a target variable is how each variable contributes to the homogeneity of the nodes, 78% mother's having no education and 16% women belong to Baluchistan region had prevalence of short rapid breath among excess in girls (13%).

## 4.2 Linear Discrimination Analysis

Linear discrimination analysis for the undesired child under five for morbidity.

### 4.2.1. Prediction from LDA by Using Training Dataset

It is determined that 56.21% of correct predictions by the model correspond to no fever in two weeks while 29.41% belong to the undesired child having fever in the previous two weeks, 6.73% of children had a cough, in appendix confusion matrix 0.33% prediction that are no acute respiration infection for an undesired child, 75 % had ARI. The undesired children experienced 60.8% of diarrhea. All results of child morbidities using linear discrimination analysis are present in the appendix.

### 4.2.2. Prediction from LDA by Using Testing Dataset

In Table 4.2, it is determined that 58% of correct predictions by the model correspond to no fever in two weeks before the survey, while 3% of correct predictions are belonging to the group of undesired children having fever in the previous two weeks and 56.5% children who do not have cough and 1.6% children had cough is incorrectly predicted by LDA model.

Although 77.4% had no short rapid breath and 0.94% of the undesired child has short rapid breath problems, 48% had no acute respiration infection and 48% had acute respiration infection. 61.7% of undesired children experienced diarrhea mentioned in the table 4.2.

**4.2. Confusion Matrix of Child Morbidity using Linear Discrimination Analysis**

| **Confusion matrix of Fever using LDA** | | |
|---|---|---|
| | **No Fever** | **Fever** |
| **No Fever** | 58.02% | 3.90% |
| **Fever** | 35.03% | 3.03% |
| **Confusion matrix of cough using LDA** | | |
| | **No Cough** | **Cough** |
| **No Cough** | 56.50% | 3.53% |
| **Cough** | 38.20% | 1.60% |
| **Confusion Matrix for Diarrhea using LDA** | | |
| | **No Diarrhea** | **Diarrhea** |
| **No Diarrhea** | 2.94% | 30.80% |
| **Diarrhea** | 4.41% | 61.70% |
| **Confusion Matrix for SBR using LDA** | | |
| | **No SBR** | **SBR** |
| **No SBR** | 77.36% | 0.85% |
| **SBR** | 20.80% | 0.94% |
| **Confusion Matrix for ARI using LDA** | | |
| | **No ARI** | **ARI** |
| **No ARI** | 48.02% | 0.75% |
| **ARI** | 0.94% | 50.20% |

## 4.3. Neural Network

### 4.3.1. Prediction from Neural Network by Using Training Dataset

The undesired child who had a fever and not had a fever using neural network is represented in confusion matrix 4.1.11 in appendix child with symptoms of fever and correctly determined by the model is determined that 14%. The prevalence of symptoms of acute respiratory infection is 73% predicted by NN, and 3.87% had short rapid breath is which correctly predicted. About 1.05% of children had diarrhea.

### 4.3.2. Prediction from the Neural Network by Using Testing Dataset

We found 47% and 41.4% of undesired children did not have from fever or cough in the last two weeks and 11% and 12.2% had a fever within the last two weeks mentioned below in table 4.3. Most of the children had acute respiratory infection in children 66%. According to the NN model, there 1.7% of accurately predicted cases belong to short rapid breath and 76% of appropriate predictions come from cases where there is no short rapid breath.

### 4.3. Confusion Matrix of Child Morbidity using Neural Network

**Confusion matrix of fever using NN**

| | No Fever | Fever |
|---|---|---|
| **No Fever** | 46.90% | 28.40% |
| **Fever** | 13.30% | 11.30% |

**Confusion matrix of cough using NN**

| | No Cough | Cough |
|---|---|---|
| **No Cough** | 41.40% | 27.60% |
| **Cough** | 18.60% | 12.20% |

**Confusion matrix for ARI using NN**

| | No ARI | ARI |
|---|---|---|
| **No ARI** | 3.90% | 9.60% |
| **ARI** | 19.80% | 66.40% |

**Confusion Matrix for SBR using NN**

| | No SBR | SBR |
|---|---|---|
| **No SBR** | 75.90% | 19.10% |
| **SBR** | 3.20% | 1.70% |

## 4.4. Support Vector Machine

### 4.4.1. Prediction from support vector machine by Using Training Dataset

In the case of SVM Table 4.1.17 shows in the appendix with proper classification on diagonal determined that 57.22% correspond to an undesired child not suffering from fever in the last two weeks and 7.4% of an undesired child who experience fever in the last two weeks, 0.84% correct prediction that are belongs from had no acute respiration infection and had acute respiration infection is 76.1% which are correctly predicted. The cough, short rapid breath, and diarrhea experienced by the undesired child and predicted by the model are 6.6%, 0.4%, and 0.07%.

**Prediction from support vector machine by Using Testing Dataset**

Table 4.4 shows a confusion matrix with proper classification on the diagonal and miss classification on the other using the testing data set, 55.2% correct prediction by model, corresponding to the undesired child not suffering from fever in the last two weeks and correctly predicted undesired children are 4.2%, table 4.37 in which we can see that 58.3% correct prediction children who do not have a cough and 6.67% children had a cough. Table 4.4 shows no prediction for the prevalence of diarrhea in undesired children using a support vector machine (SVM). It is determined that 77.2% of correct predictions in table 4.4, correspond to correct predictions that belong to have no short rapid breath and 0.75% of correct predictions are belonging to the group of the undesired child has short rapid breath problem, 0% correct prediction that belongs from had no acute respiration infection and had acute respiration infection is 76% which are correctly predicted in table number 4.4.

**4.4. Confusion Matrix of Child Morbidity using Support Vector Machine**

| **Confusion matrix of Fever using SVM** | | |
|---|---|---|
| | **No Fever** | **Fever** |
| **No Fever** | 55.20% | 39.10% |
| **Fever** | 4.90% | 4.20% |

| **Confusion matrix of cough using SVM** | | |
|---|---|---|
| | **No Cough** | **Cough** |
| **No Cough** | 55.70% | 37.70% |
| **Cough** | 3.40% | 3.20% |

| Confusion Matrix for Diarrhea using SVM | | |
|---|---|---|
| | **No Diarrhea** | **Diarrhea** |
| **No Diarrhea** | 80.30% | 19.70% |
| **Diarrhea** | 0 | 0 |

| Confusion Matrix for ARI using SVM | | |
|---|---|---|
| | **No ARI** | **ARI** |
| **No ARI** | 0% | 23.70% |
| **ARI** | 0.28% | 76% |

| Confusion Matrix for SBR using SVM | | |
|---|---|---|
| | **No SBR** | **SBR** |
| **No SBR** | 77.20% | 19% |
| **SBR** | 3.03% | 0.75% |

**Discussion**

The previous study showed that the fever in undesired children (Alam & Islam, 2022) is 5.5% which is higher in our findings and ARI is higher in over findings.

The childhood morbidity and background characteristics of the selected covariate are mentioned in section 3.2 of this study; five different machine learning algorithms were applied to classify the children in the test dataset. All models were trained based on 5 and 10fold cross-validation. The core principle of this study is to predict the undesired child morbidity among children <5 years. These ML were applied using 70% of the individuals in each group and 30% testing group.

Table 4, the LDA classifier has an accuracy of 61% with 62% sensitivity and a specification of 43% in predicting child fever in the last two weeks. The SVM demonstrates an accuracy of 57%

with 91% of sensitivity and specification of 9%. While the cough within the last two weeks in undesired showed higher accuracy as compared to the (Alam & Islam, 2022), LDA performance the worse with an accuracy of 58% with 59% sensitivity and specification of 31% for children having cough, where NN has higher accuracy with 30.6% sensitivity and specification of 68%. In had cough in last two weeks, NN model performs better than another classifier. The prevalence of short rapid breath children under 5 is higher, as LDA has an accuracy of75%, SVM demonstrates an accuracy of 78%, RF accuracy is 79% and NN accuracy is 77% with 8%
sensitivity and specification of 95%.

However, a paper conducted by (Alam & Islam, 2022) showed that the acute respiratory infection (ARI) among excess in boys was higher, the study identifying the important feature showed ARI is higher among excess in girls. LDA has the best algorithms performance in predicting the ARI among the undesired child with an accuracy of 76% with 44% sensitivity and specification of 77%. Another classifier also performed better and almost the results were closed.

The prevalence of fever and diarrhea in the last two weeks is higher among undesired children. Whereas among the ML algorithms LDA model performs better than another classifier, and RF performs worse in case of diarrhea. LDA, the accuracy is 81% with 81.4% sensitivity and specification of 0%. RF accuracy is 76%. Overall performance of random forest and linear discrimination analysis accomplished the best results in predicting childhood morbidity.

**Table 4: PERFORMANCE OF VARIOUS CLASSIFICATION ALGORITHMS**

| model | Child morbidity | Fever in last two weeks | Had Cough in last two weeks | Diarrhea in last two weeks | Acute Respiration Infection | Short, Rapid Breath |
|---|---|---|---|---|---|---|
| RF | Accuracy% | 96.27% (0.9343, 0.9812) | 62% 95% CI: (0.5629, 0.6729) | 77% 95% CI: (0.7083, 0.8083) | 75% 95% CI: (0.7046, 0.7982) | 79.2% 95% CI: (0.7608, 0.822) |
| | Sensitivity % | 99.4% | 90.24% | 1.0000 | 8.537% | 99.11% |
| | Specificity | 92.13 % | 9.09% | 0.000 | 96.2% | 0.70% |
| LDA | Accuracy% | 61.06% 95%CI (57.83%, 64.22%) | 58.2% 95% CI: (52.5%, 63.74%) | 81.4% 95% CI: (76.62%, 85.62%) | 76.8% 95%CI (72.01%,81.08%) | 75% 95%CI: (67.93%, 81.21%) |
| | Sensitivity % | 62.354% | 59.6% | 81.43% | 44.4% | 75% |
| | Specificity% | 43.75 % | 31.25% | 0.0000 | 44.5% | 0.000 |
| NN | Accuracy% | 64.5% | 64.4% | 77.5% | 70.45% | 77.65% |
| | Sensitivity % | 28.42% | 30.6% | 1.99% | 0.000 | 8.2% |
| | Specificity | 77.9% | 67% | 98.6% | 13.63% | 95.93% |
| SVM | Accuracy% | 57.49% 95% CI: (55.19%, 59.77%) | 58.93% 95%CI (56.68%, 61.15%) | 80.34% 95%CI: (75.34%, 84.72%) | 76% 95%CI: (71.07%, 80.36%) | 78.05% 95%CI: (75.43%, 80.51%) |
| | Sensitivity % | 91.85% | 94.31% | 1.000 | 0.000000 | 99.76% |
| | Specificity | 9.648 % | 94.31% | 0.0000 | 99.61% | 0.000 |

## 4.5. Treatment of Child Morbidity

## 4.6 Random Forest

### 4.6.1. Prediction from Random Forest by Using Training Dataset

Confusion matrices illustrate that 55.1% of undesired children received any type of postnatal care within two months of birth. 10.27% of correct predictions correspond to such children who did not receive any treatment for diarrhea (TDIA), and 8.4% of correct predictions are belonging to the group of children who were treated for diarrhea. The out-of-bag (OOB) error of the model is 53.7%. the class error not having treatment of diarrhea (TDIA) is 0.77 and having treatment of diarrhea is 0.15. The confusion matrix is determined by the formula as mentioned in the equation, which shows that 5.5% are children who did not receive any treatment for fever/cough and 70% were treated for fever/cough. The OOB for the treatment of fever/cough is 27.46%. the class error was 80% for not having any treatment and 6.5% for having treatment of fever/cough. In terms of vaccination, around 42.7% had received vitamins within 6 months, 00B error rate is 36.62%.

### 4.6.2. Prediction from Random Forest by Using Testing Dataset

About 21% excess in girls received the postnatal checkup within two months after their birth which is also correctly specified by the model, 7.3% corresponds to such children who did not receive any treatment for diarrhea, about 69% excess in parity received treatment for diarrhea. Almost 52% are such children who are excess in girls and did not receive any treatment for fever/cough and 7% is the child who was treated for fever/cough. In terms of vaccination, around 20.45% had not received the vitamin within 6 months and 43% excess in the undesired dual child had received vitamin within 6 months.

### 4.5. Confusion Matrix of Treatment of Child Morbidity using Random Forest

**Confusion matrix of PBC using Random Forest**

| | No PBC | PBC |
|---|---|---|
| **No PBC** | 6.14% | 3.50% |
| **PBC** | 69.80% | 20.50% |

**Confusion matrix of TDIA using RF**

| | No TDIA | TDIA |
|---|---|---|
| **No TDIA** | 7.30% | 10.10% |
| **TDIA** | 34.10% | 69.10% |

**Confusion matrix for treatment of Fever/cough using RF**

| | No TFever/cough | TFever/cough |
|---|---|---|
| **No TFever/cough** | 52% | 7.16% |
| **TFever/cough** | 33.91% | 6.90% |

**Confusion Matrix for treatment of Fever/cough using RF**

| | No Vitamin | Vitamin |
|---|---|---|
| **No Vitamin** | 20.40% | 14.90% |
| **Vitamin** | 21.16% | 43.40% |

### 4.6.3 Important variable of Treatment of Childhood Morbidity

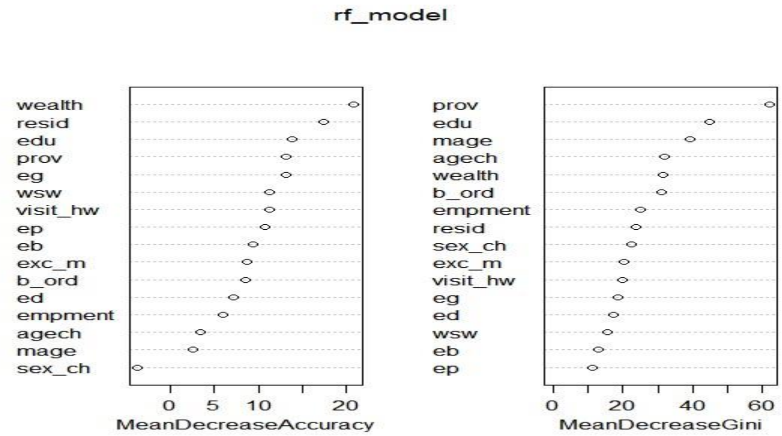


**Figure 7: Plot of PBC**

Random forest allows us to look at feature importance, which is how much the Gini index for feature, the more important it is. The undesired girls born to poor mother (22.06%) is lower in receiving any type of postnatal care. The women from urban area (17.4%), as well as from KPK (61%) are two significance factors associated with receiving lowest postnatal care among undesired children.

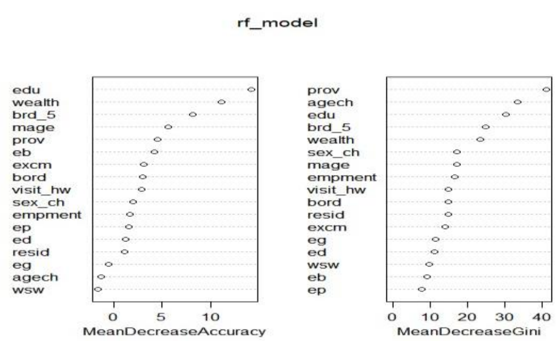


**Figure 8: Plot of Treatment of Diarrhea**

As plots shows, 5.1% excess in boys receive less treatment of diarrhea and the factor associated with undesired children receiving lowest treatment is 15.1% non-educated mother belonging to Baluchistan region 41%. On the other hand, the mean decrease Gini plot illustrates 33.3% child having age from 12month to 36 months don’t receive treatment of diarrhea.

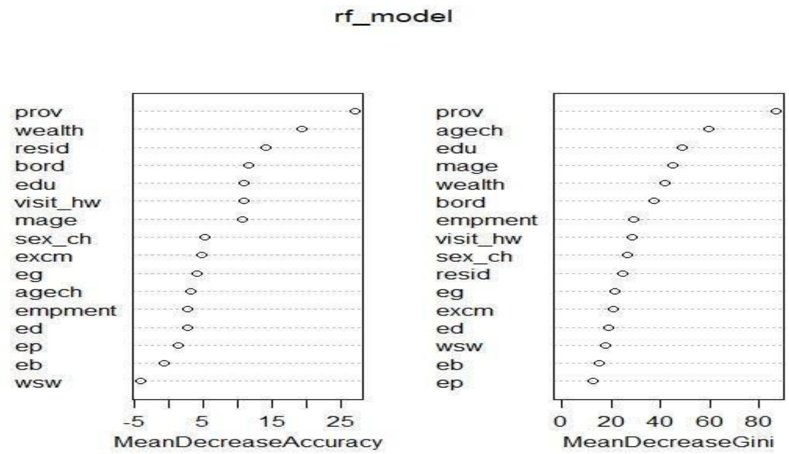


**Figure 9: Plot of treatment of fever/cough**

Random forest allows us to look at feature importance, which is how much the Gini index for a feature, the more important it is. The girls born to poor mother (22.06%) and from region Baluchistan (29%) and KPK (8.4%) significance factor in receiving lowest treatment of fever/cough.

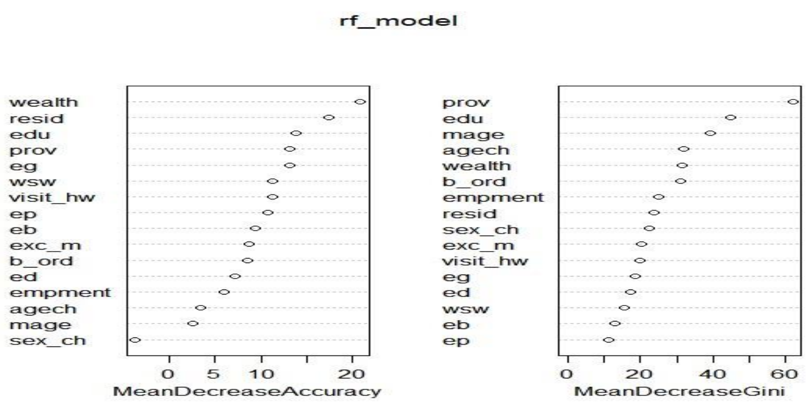


**Figure 10: Plot of vitamin A**

Figure 10 shows a plot of mean decrease accuracy and mean decrease Gini in context to vaccination, 16.1% excess in girls slightly received less vitamin A within six months preceding the survey who born to 20.1% poor women and 19% non-educated women who belonging to 16.2% Baluchistan's urban area.

## 4.7 Linear Discrimination Analysis

### 4.7.1 Prediction from LDA by Using Training Dataset

In appendix 4.1.27, the confusion matrix demonstrated 56.3% of correct predictions are belonging to the group of children who were checked during six months. Confusion matrix 4.50 indicates that 60.8% child who is treated for diarrhea, 1.57% correspond to such children who did not receive any treatment for fever/cough, and 72.88% of the child who was treated for fever/cough. Almost 19.3% had not received the vitamin within 6 months and 43.6% had received the vitamin within 6 months.

### 4.7.2 Prediction from LDA by Using Testing Dataset

About 10.8% of correct predictions correspond to such children who did not receive any checkups for two months and 58.93% of correct predictions are belonging to the group of children who were checked for two months. Using LDA, 2.9% correct prediction, correspond to such children who did not receive any treatment for diarrhea and the actual model also shows the same prediction, and 61.7% correct prediction are belonging to the group of children who were treated for diarrhea (TDIA). Misclassified rate of receiving treatment for diarrhea is 3%, and 26% correct prediction are belonging to the group of children who were treated for fever/cough (TFever/Cough). On diagonal represents correct classification, in the context of vaccination, around 21.5% had not received the vitamin within 6 months and 42% had received the vitamin within 6 months.

### 4.6. Confusion Matrix of Treatment of Child Morbidity using Linear Discrimination Analysis

| **Confusion matrix of PBC using LDA** | | |
|---|---|---|
| | **No PBC** | **PBC** |
| **No PBC** | 10.80% | 23.20% |
| **PBC** | 7.02% | 58.90% |
| **Confusion matrix of TDIA using LDA** | | |
| | **No TDIA** | **TDIA** |
| **No TDIA** | 2.94% | 4.40% |
| **TDIA** | 30.88% | 61.70% |
| **Confusion matrix for treatment of Fever/cough using LDA** | | |
| | **No TFever/cough** | **TFever/cough** |
| **No TFever/cough** | 1.60% | 72% |
| **TFever/cough** | 0.80% | 26% |
| **Confusion Matrix for treatment of Fever/cough using LDA** | | |
| | **No Vitamin** | **Vitamin** |
| **No Vitamin** | 21.60% | 18.18% |
| **Vitamin** | 18.18% | 42% |

### 4.8 Support Vector Machine

#### 4.8.1 Prediction from Support Vector Machine by Using Training Dataset

The postnatal checkup within two months received by undesired children shown by the confusion matrix in the appendix is 56.12% and 14.39% are not checked-in for two months. In the confusion matrix, under the SVM model, the rate of correct prediction is 0% which means the model indicates such children who did not receive any treatment for diarrhea, and 63.3% child are those who were treated for diarrhea. LDA model for training shows 96.5% correct prediction for undesired children who were treated for fever/cough, while 30.6% had not received the vitamin within 6 months and 31.45% had received the vitamin within 6 months.

### 4.8.2 Prediction from Support Vector Machine by Using Testing Dataset

The confusion matrix of PBC (postnatal checkup baby) is given in table 4.7, 14% of children have not received checkups for two months and correctly predicted by the support vector machine (SVM). The SVM model for the testing data set is given below, the table represents that no correct prediction, corresponds to such children who did not receive any treatment for diarrhea, and 65% correct prediction are belonging to the group of children who were treated for diarrhea (TDIA).

Throughout 4.4% of appropriate predictions come from treatment of fever/cough (TFever/Cough) and corresponding to such children who did not receive any treatment for fever/cough and 4% who were treated for fever/cough. In the table 4.7, around 28.1% had not received the vitamin within 6 months and 31.25% had received the vitamin within 6 months.

**4.7. Confusion Matrix of Treatment of Child Morbidity using Support Vector Machine**

| **Confusion matrix of PBC using SVM** | | |
|---|---|---|
| | **No PBC** | **PBC** |
| **No PBC** | 14.40% | 8.90% |
| **PBC** | 20.60% | 56.10% |
| **Confusion matrix of TDIA using SVM** | | |
| | **No TDIA** | **TDIA** |
| **No TDIA** | 0 | 0 |
| **TDIA** | 34.30% | 65.60% |
| **Confusion matrix for treatment of Fever/cough using SVM** | | |
| | **No TFever/cough** | **TFever/cough** |
| **No TFever/cough** | 4.30% | 5.30% |
| **TFever/cough** | 21.90% | 66.70% |

| Confusion Matrix for treatment of Fever/cough using SVM | | |
|---|---|---|
| | **No Vitamin** | **Vitamin** |
| **No Vitamin** | 28.10% | 27.10% |
| **Vitamin** | 13.40% | 31.25% |

### 4.9 Neural Network

#### 4.9.1 Prediction from Neural Network by Using Training Dataset

It is shown that 24.7% of correct predictions correspond to such children which are not checkups, 63.4% of correct predictions are belonging to the group of children who were checked, 54.9% of the undesired child received the treatment for diarrhea, children who did not received any treatment for their fever and cough fall into the 6.5% correct predicted category, while those who did receive treatment fall into 70.4% correct prediction category. Regarding vaccination, around 28% had not received the vitamin within 6 months and 42% had received vitamin within 6 months.

#### 4.9.2 Prediction from the neural network by Using Testing Dataset

Table 4.8 represent the result of correctly predicted element on diagonals, 15.7%children who are not checked for two months, and model actual model also shows children who a not checked, 54% of a child who was checked during six months. The overall predicted accuracy of the model is 69.9%. This confusion matrix indicates13% that of children did not receive any treatment for diarrhea and the actual model also says no children received the treatment and 38.8% child were treated for diarrhea. While 1.6% correct prediction, children who did not receive any treatment for fever/cough, and 66% received the treatment. The confusion matrix (table 4.8) is represented 24.4% had not received the vitamin within 6 months and 38.4% had received the vitamin within 6 months.

### 4.8. Confusion Matrix of Treatment of Child Morbidity using Neural Network

| Confusion matrix of PBC using NN | | |
|---|---|---|
| | **No PBC** | **PBC** |
| **No PBC** | 15.70% | 11.30% |
| **PBC** | 18.70% | 54.20% |
| **Confusion matrix of TDIA using NN** | | |
| | **No TDIA** | **TDIA** |
| **No TDIA** | 13.50% | 23.30% |
| **TDIA** | 24.20% | 38.80% |
| **Confusion matrix for treatment of Fever/cough using NN** | | |
| | **No TFever/cough** | **TFever/cough** |
| **No TFever/cough** | 1.60% | 5.60% |
| **TFever/cough** | 26.50% | 66.31% |
| **Confusion Matrix for treatment of Fever/cough using NN** | | |
| | **No Vitamin** | **Vitamin** |
| **No Vitamin** | 24.40% | 20.60% |
| **Vitamin** | 16.40% | 38.40% |

## Discussion

Efforts had been provided, to evaluate the impact of undesired children on child treatment of morbidity. Results in (table 5) the postnatal check using testing data from the LDA show an accuracy of 69.7% with 60% sensitivity and specification of 71%. The SVM demonstrates an accuracy of 70.5% with 41% of sensitivity and specification of 86%. RF accuracy is 96% with 99% sensitivity and specification of 92.1%, and NN accuracy is 69.9% with 82% sensitivity and specification of 45 %. In a post-neonatal check of the baby, the RF model performs better than another classifier and LDA performs worse. The accuracy of test data was seen in table 3 as 64% with 4% sensitivity and specification of 66%. The SVM has a 65% accuracy with 0% of sensitivity and a specification of 1%. The RF accuracy is 63% with 17%

sensitivity and specification of 87%. NN accuracy is 52% with 63% sensitivity and specification of 36%. The SVM model performs better in cases where diarrhea has been treated in the last two weeks. The SVM demonstrates an accuracy of 75.9% with 0% of sensitivity and a specification of 1%. RF accuracy is 72.4% with 16.7% sensitivity and specification of 92%. NN accuracy is 67.9% with 92% sensitivity and specification of 5.6%. In the treatment of cough/fever in the last two weeks, the SVM model performs better than another classifier. %. RF accuracy is 63.9% with 49% sensitivity and specification of 74%. NN accuracy is 62.8% with 65% sensitivity and specification of 57%. In receiving vitamin, A in 6 months, the RF model performs better than another classifier.

**Table 5: PERFORMANCE OF VARIOUS CLASSIFICATION ALGORITHMS**

| | model | Treatment of Diarrhea | PBC | VITAMIN A | Treatment of fever/cough |
|---|---|---|---|---|---|
| RF | Accuracy% | 72.25% 95%CI (65.07%,78.63%) | 63.36% 95%CI (65.07%,78.63%) | 72.47% 95%CI (65.07%,78.63%) | 63.92% 95%CI (65.07%,78.63%) |
| | Sensitivity% | 47.62% | 17.77% | 17% | 49.1% |
| | Specificity% | 85.47% | 87.2% | 92.5% | 74.45% |
| LDA | Accuracy% | 69.7% | 64.71% | 75.2% | 63.64% |
| | Sensitivity% | 66.86% | 60.61% | 40% | 20.83% |
| | Specificity% | 0.5469 | 0.823817 | 0.5 | 0.9691715 |
| NN | Accuracy% | 69.7% | 52.9% | 67.9% | 62.8% |
| | Sensitivity% | 82.69% | 63.4% | 92.2% | 65.06% |
| | Specificity% | 45.5% | 35.8% | 57% | 59.7% |
| SVM | Accuracy% | 70.56% | 65.6% | 75.97% | 59.38% |
| | Sensitivity% | 41.27% | 0% | 1% | 67.58% |
| | Specificity% | 86.32% | 1% | 0% | 53.53% |

**Child Malnutrition**

The findings are obtained from a collection of data that had been cleaned and integrated with 16 features taken from the literature review (Alam & Islam, 2022). The accuracy features were identified for WHZ, HAZ, and birth size. Testing was done using a 5-fold cross-validation method. Table1 shows the prevalence of malnutrition and background characteristics of the selected covariates. Four different machine learning methods are used to categorize the children in the test set as malnutrition or nutrition in the study. The predictive abilities of these algorithms were compared using performance measures such as specificity, accuracy, and sensitivity.

## 4.10 Random Forest

### 4.10.1 Prediction from Random Forest by Using Training Dataset

The prevalence of severely stunted under-five children is around 1.3% and the not severely stunted is 12.6% as presented in the appendix. The error class for a group had not severely stunted is 1.33% and that had severely stunted is 92.3%. The 00B error rate is 16.8%, stunted under-five children are around 17% and the not stunted is 46%. The error class for a group that had not stunted is 23.6% and that had stunted is 92.3%. The 00B error rate is 36.3%., in which we can see that 82% are not severely underweight children, will 2% correct prediction that belongs to the severely underweight child. The OOB error is 16.7%. no correct prediction that our child has a smaller birth size, 62% show correct prediction that belongs to the group of average birth size, and we can observe that no correct prediction comes from the larger birth size groups.

### 4.10.2 Prediction from Random Forest by Using Testing Dataset

The confusion matrix is represented (table 4.9) for the test dataset of stunted with 43.9% correct prediction, not stunted child, while 18.4% correct prediction are dual excess stunted child. The prevalence of severely stunted is around 1% and the not severely stunted is 79%. The prevalence of severely underweight and not severely underweight among undesired children is 5% and 81%. The prevalence of wasted is shown more in boys which is around 3.3% and severely wasted is 3.8% which is higher in undesired girls. The RF shows 3.5% excess among undesired boys is underweight. The birth size of the excess boy is average in appendix table 4.89 and larger than average, 72% and 0.07% respectively.

**4.9. Confusion Matrix of Child Malnutrition using Random Forest**

| Confusion matrix of severely stunted using Random Forest | | |
|---|---|---|
| | **No SS** | **SS** |
| **No SS** | 79.80% | 17.60% |
| **SS** | 1.50% | 1% |
| **Confusion matrix of stunted using Random Forest** | | |
| | **No Stunted** | **Stunted** |
| **No Stunted** | 43.90% | 20% |
| **Stunted** | 17.80% | 18.40% |
| **Confusion matrix for severely underweight using Random Forest** | | |
| | **No SUW** | **SUW** |
| **No SUW** | 81.30% | 12.70% |
| **SUW** | 2.83% | 5.03% |
| **Confusion Matrix for underweight using Random Forest** | | |
| | **No underweight** | **underweight** |
| **No underweight** | 73.20% | 20.30% |
| **Underweight** | 3% | 3.50% |
| **Confusion Matrix for wasted using Random Forest** | | |
| | **No Wasted** | **Wasted** |

| | | |
|---|---|---|
| **No Wasted** | 68.60% | 9.10% |
| **Wasted** | 19% | 3.31% |
| **Confusion Matrix for Severely wasted using Random Forest** | | |
| | **No Swasted** | **SWasted** |
| **No Swasted** | 94% | 6.01% |
| **SWasted** | 0% | 0% |

### 4.10.3 Important Variable plots for Child Malnutrition

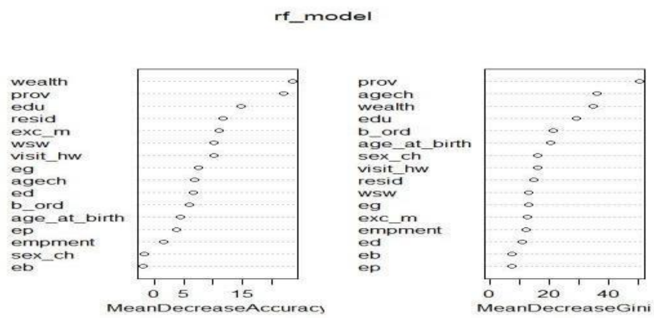


**Figure 11: Plot of Underweight child**

In building an underweight child model 23.5% wealth index (women from poor families) from 18.8% bulachistan and 9.4% Sindh are variable presented in the plot are statistically significant with 7.7% excess in girl with age 0-59 month (6.9%) are underweight.

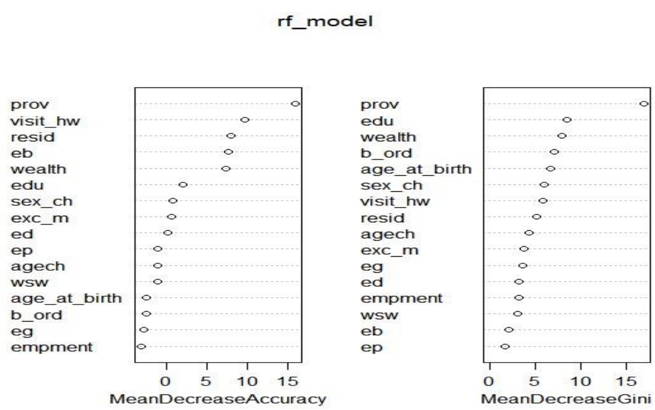


**Figure 12: Plot of wasted child**

In above plot important Variable associated with building the waste child model are region, mother education and visit family planning. The 16.7% women from Baluchistan region having non-educated 4% did not follow any family. As result 7.8% excess in boys.

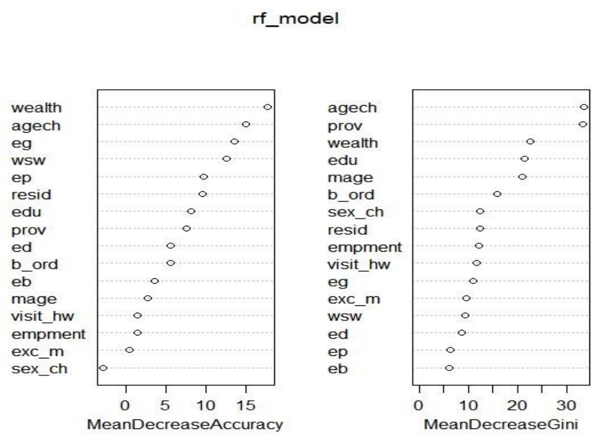


**Figure 13: Plot of severely stunted**

Figure 13, 8.2 % excess in girls with (12.2%) age 0-1 month had higher probability of severely stunted and Mean decreases accuracy show theses undesired children belong to 11.9% of poor mother from all the region (6%).

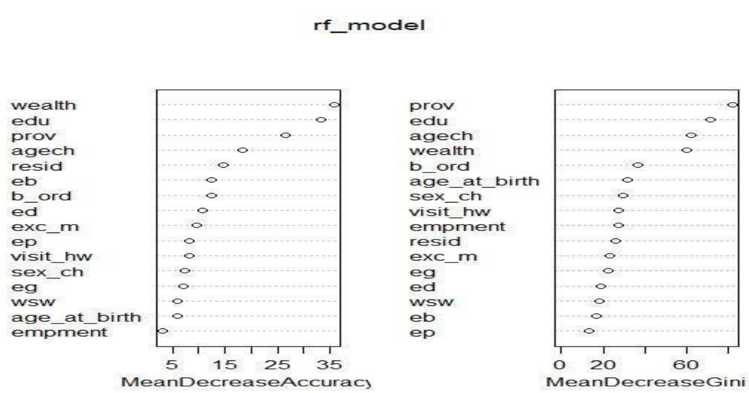


**Figure 14: Plot of stunted child**

Mean decrease accuracy is computed by a randomized tree exhibiting desirable properties for assessing the relevance of important variables. So, the plot of mean decrease accuracy computes wealth and the mother's education as an important variable for the stunted undesired child. While the mean decrease Gini plot also computes mother education as an important variable.

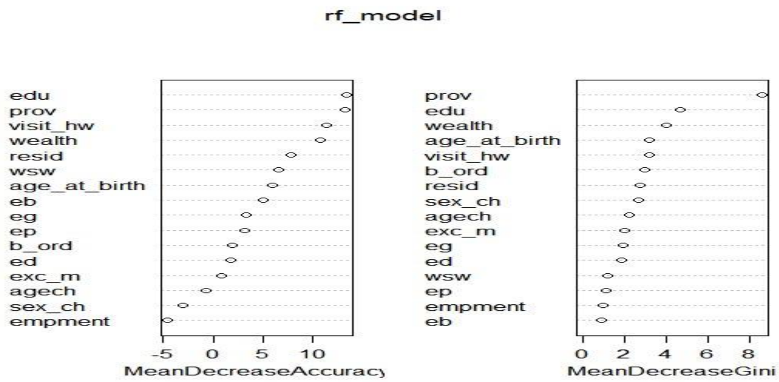


**Figure 15: Plot of severely wasted child**

The figure 15, mean decrease accuracy and mean decrease Gini shows 3.4% excess in girls and 4.96% excess in boys are severely wasted whom belonging to 13.3% illiterate mother's from KPK region (13.18%).

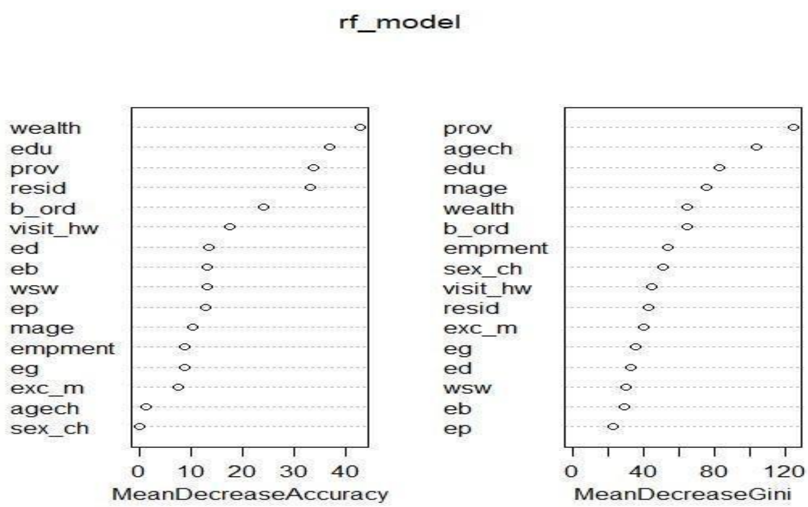


**Figure 16: Plot of birth size**

The random forest method identifies household wealth, mother education, child age, and the province as key variables in explaining the malnutrition of undesirable children under 5 in Pakistan in both plots in figure 16. The birth size of 8.8% excess in girls is average which is lower as compared to excess in boys and these girls born to 44.8% poor families whose mother are mostly uneducated (41.9%).

## 4.11 Linear Discrimination Analysis

### 4.11.1 Prediction from LDA by Using Training Dataset

In table 4.1.47, the confusion matrix is represented for the training dataset .in which table off-diagonal represents miss classification observation, and on diagonal represents correct classification, the prevalence of severely stunted under-five children is around 1.5% and the not severely stunted is 81%, 57% the not severely stunted child, will 15.3% are stunted child. 91% the not severely underweight child, will 4.5% belong to the severely underweight child, no correct prediction of a child having a smaller birth size, 74% show correct prediction that belongs to the group of average birth size, and we can have observed that 0.4% correct prediction come from the larger birth size groups. Whereas 42% of children are severely underweight.

### 4.11.2 Prediction from LDA by Using Testing Dataset

In the table, confusion matrix the prevalence of severely stunted under-five children is around 2%, and the not severely stunted is 81%, which is correctly predicted by the model. The LDA classifier shows 3.3% of undesired children has a prevalence of being severely underweight. while the model shows no prediction for wasted child and severely wasted, only 1% correct prediction shown by the model of the severely underweight child all table shown in the appendix. The LDA model shows no prediction for smaller and larger birth sizes, 75% shown in (appendix 4.98) correct prediction that belongs to the group of average birth size. The stunted children predicated by LDA is 13%.

### 4.10. Confusion Matrix of Child Malnutritional using Linear Discrimination Analysis

**Confusion matrix of severely stunted using LDA**

| | No SS | SS |
|---|---|---|
| **No SS** | 81% | 0.16% |
| **SS** | 17.3% | 2% |

**Confusion matrix of stunted using LDA**

| | No Stunted | Stunted |
|---|---|---|
| **No Stunted** | 42.7% | 10.5% |
| **Stunted** | 21.2% | 13% |

**Confusion matrix for severely underweight using LDA**

| | No SUW | SUW |
|---|---|---|
| **No SUW** | 68% | 26.4% |
| **SUW** | 2.17% | 3.3% |

## 4.12 Support vector machine

### 4.12.1 Prediction from support vector machine by Using Training Dataset

In table 4.99 in the appendix, the prevalence of severely stunted under-five children is around 0.8% and the not severely stunted is 82 % and 3.7% of severely underweight children. 53% correct predictions that belong from the not stunted child, 16.3% correct prediction in table 4.100 for the stunted child. The model predicted a 1.9% prevalence of severely wasted undesired and a 1% prevalence of wasted among children under 5.

### 4.12.2 Prediction from support vector machine by Using Testing Dataset

In the table 4.11, the confusion matrix is represented the prevalence of stunted under-five children is around 15% and the not stunted is 50%.  In addition to this 78% of correct prediction of severely underweight children 3.5% and not severely underweight child was 77% and 3.1% of undesired girls have a prevalence of severely wasted according to the model prediction.

The group of children having smaller birth sizes is not predicted by the model,74% show correct prediction that belongs to the group of average birth size, and we have observed that no correct prediction comes from the larger birth size groups (appendix 4.1.63).

**4.11. Confusion Matrix of Child Malnutritional using Support Vector Machine**

| Confusion matrix of stunted using SVM | | |
|---|---|---|
| | **No Stunted** | **Stunted** |
| **No Stunted** | 50% | 23.10% |
| **Stunted** | 11.70% | 15.10% |
| **Confusion matrix of severely underweight using SVM** | | |
| | **No SUW** | **SUW** |
| **No SUW** | 77% | 16.40% |
| **SUW** | 3.05% | 3.50% |
| **Confusion matrix for severely wasted using SVM** | | |
| | **No Wasted** | **S-Wasted** |
| **No Wasted** | 93.10% | 3.70% |
| **S-Wasted** | 0 | 3.10% |

### 4.13 Neural Network

### 4.13.1 Prediction from Random Forest by Using Training Dataset

In the table, confusion matrix (appendix) is represented for testing dataset in which table off diagonal represent miss classification observation and on diagonal represent correct classification, the prevalence of severely stunted under five child is approximately 0.6% while the not severely stunted is 86%. 1.5% of the children are underweight, 20.9% not underweight. The model predict 45% children are not stunted child, while 10% correct prediction that are belongs from the stunted child.  The prevalence of severely wasted is higher among the undesired children which is almost 9.7% and 4.120 tables in appendix, illustrate no prediction is made by model in for wasted child.

### 4.13.2 Prediction from Random Forest by Using Testing Dataset

The model says 3% children are severely stunted, 12% children are stunted. The prevalence of severely underweight children is 2.5%. The model shows correct prediction of prevalence of severely wasted among undesired is 9%. Neural network model shows 5% excess in girls are underweight.

**4.12. Confusion Matrix of Child Malnutritional using Neural Network**

| Confusion matrix of severely stunted using NN | | |
|---|---|---|
| | No SS | SS |
| No SS | 74% | 17.00% |
| SS | 5.90% | 3.00% |
| **Confusion matrix of stunted using NN** | | |
| | No stunted | stunted |
| No stunted | 53% | 26.30% |
| stunted | 9.21% | 12.00% |
| **Confusion matrix of severely underweight using neural network** | | |
| | No SUW | SUW |
| No SUW | 75.90% | 18.10% |

| | | |
|---|---|---|
| **SUW** | 3.40% | 2.51% |
| **Confusion matrix of underweight using neural network** | | |
| | **No underweight** | **underweight** |
| **No underweight** | 64.30% | 22.00% |
| **underweight** | 1% | 5.01% |
| **Confusion matrix of wasted using NN** | | |
| | **No SWasted** | **SWasted** |
| **No SWasted** | 84.20% | 5.30% |
| **SWasted** | 2% | 9.02% |

**Discussion**

The prevalence of malnutrition and five different machine learning algorithms were applied to classify the children in the test dataset as "malnourished" and "nourished." As in the table 6 accuracy, specification and sensitivity is displayed. Using LDA, the accuracy of severely stunted is s 82.2% with 82% sensitivity and specification of 5%. The SVM demonstrate accuracy of 81.1% with 99.5% of sensitivity and specification of 0.8%. RF accuracy is 80.8% with 98% sensitivity and specification of 5.53%. NN accuracy is 77.1% with 14.7% sensitivity and specification of 92%. In the case of severely stunted of undesired child LDA model perform better than another classifier and neural network. In case of stunted undesired child among all the classifiers, SVM accomplished best results with accuracy of 65 % and sensitivity and specification of 81% and 39%. The LDA for underweight follow the accuracy of 69% with sensitivity 4%. The predicted accuracy of SVM is 81% for prevalence of Underweight among undesired children and specification of 86%. The random forest algorithm in table 5 showed higher accuracy with 85% in predicting severely underweight.

Using LDA, the results show in children under 5 severely wasted is higher and its accuracy is 97%. In predicting the performance of nutritional status among undesired child SVM classifier has higher accuracy rate. If we compare our result with previous research (Alam & Islam, 2022);(Flatø, 2018) the undesired girls were significantly associated with severely underweight and severely stunted which is 50% and 41%.

**Table 6: PERFORMANCE OF VARIOUS CLASSIFICATION ALGORITHMS**

| | model | S | SS | UW | SUW | W | SW | BS |
|---|---|---|---|---|---|---|---|---|
| **RF** | Accuracy% | 62.34% | 80.86% | 76.74% | 85.15% | 63.90% | 94.65% | 94.65% |
| | Sensitivity% | 71.20% | 98.17% | 76.27% | 23.20% | 49.20% | 1 | 0.43% |
| | Specificity% | 48.07% | 5.31% | 14.67% | 97.40% | 74.45% | 0 | 99.70% |
| **LDA** | Accuracy% | 63.62% | 82.28% | 69.20% | 82.60% | 84.80% | 97.53% | 75.31% |
| | Sensitivity% | 66.86% | 60.61% | 40% | 20.83% | 54.29% | 0 | 76.31% |
| | Specificity% | 0.5469 | 0.823817 | 0.5 | 0.9691715 | 0.6981 | 0 | 0.8187 |
| **NN** | Accuracy% | 64.47% | 69.90% | 69.27% | 78.36% | 83.47% | 92.50% | |
| | Sensitivity% | 31.33% | 82.69% | 18.60% | 12.12% | 0% | 0% | |
| | Specificity% | 85.06% | 45.54% | 88% | 95.65% | 95.86% | 98.24% | |
| **SVM** | Accuracy% | 65% | 70.56% | 81.19% | 76.80% | 59.38% | 95.06% | 74.1 |
| | Sensitivity% | 81.07% | 41.27% | 99.59% | 17.64% | 67.58% | 95.06% | 1% |
| | Specificity% | 39.48% | 86.32% | 0.90% | 96.18% | 53.53% | 59.20% | 0% |

## 4.14 Child Morality

On the basic of timing of deaths in last four years before to the survey, childhood morality was determined. Early neonatal is defined as if a child died with a firth week of the birth. Late neonatal mean child died within one-month (0-28) days, while post neonatal died between 29 and 365 days. Together, late neonate and post neonatal make up infant mortality. The under-five is measured as if a child died before researching his/her fifth birthday.

## 4.15 Random Forest

### 4.15.1 Prediction from Random Forest by Using Testing Dataset

The testing sets confusion matrix output is represented in table of early neonatal child, which shows the prevalence of early neonatal of undesired children found to be 3.05% which is not correctly predicted using testing dataset, whereas 96.9% are children who are not found to be early neonatal and correctly specified by the model, but model show no prediction from early neonatal of undesired children using random forest. Similarly, for late neonatal, post neonatal, infant, child and under 5 random forests show no prediction by using test dataset.

## 4.16 Linear Discrimination Analysis

### 4.16.1 Prediction from LDA by Using Testing Dataset

By using linear discrimination analysis (LDA) Early neonatal (ENN) predict no results. Confusion matrix shows prediction of late neonatal (LNN).  The matrix shows no prediction for late neonatal of undesired child and 96.6% correct prediction shown for no late neonatal child.

However, post neonatal and infant mortality also shows no results among undesired child by using this model. The prevalence of child morality is 2% among the excess in undesired girls which is Predicted by LDA model, while 98.1% correct prediction, correspond to prevalence of no child mortality among undesired child, whereas, (Alam & Islam, 2022) results are much higher then or study (5.4%) among undesired children.

**4.13. Confusion Matrix of Child Mortality using Linear Discrimination Analysis**

| **Confusion matrix of child mortality using LDA** | | |
|---|---|---|
| | **No child** | **child** |
| **No child** | 98% | 0.00% |
| **child** | 0.00% | 1.90% |

### 4.17 Neural network

#### 4.17.1 Prediction from neural network by Using Testing Dataset

The prevalence of early neonatal among the undesired child is around 0.05% which is correctly predicted by the model and no early neonatal morality predicted by neural network (NN) is 96%. The prevalence of late neonatal morality is 0.05 % as shown in confusion matrix and 96% are those undesired children predicted by model, who didn’t die at age of 1 month. The infant mortality of undesired child was 1% which is correctly predict by model presented in confusion matrix and no undesired child die at the age of 1 year and the model predict 94% infant mortality. Whereas rest of other morality demonstrate no result when predicted by neural network model. Pakistan has very little evidence on child mortality when we are comparing (Alam & Islam, 2022). The prevalence of early neonatal is 2.4% and infant mortality is 6.5% which is significantly higher than over findings.

### 4.14. Confusion Matrix of Child Mortality using Neural Network

| Confusion matrix for Early Neonatal using NN | | |
|---|---|---|
| | **No ENN** | **ENN** |
| **No ENN** | 96% | 3.00% |
| **ENN** | 0.60% | 0.05% |
| **Confusion Matrix for Late Neonatal** | | |
| | **No LNN** | **LNN** |
| **No LNN** | 96% | 3.26% |
| **LNN** | 0.68% | 0.05% |
| **Confusion matrix for Infant using NN** | | |
| | **No Infant** | **Infant** |
| **No Infant** | 94.60% | 4.30% |
| **Infant** | 0.90% | 0.16% |

**Support Vector Machine**

The results provided in (all confusion matrix in appendix) of child morality show no prediction. To capture the relationship between morality and undesired children, support vector machine performs extremely poor. The predicted performance of our model (SVM) was evaluated in collection of observation is not used to test our model, that is out of sample. Among all other machine learning algorithms SVM was worse.

For validation of results mentioned in chapter, we have discussed the results with one of the NGO in Islamabad with the head of NGO her working experience was around 15 years. NGOs are one of essential management to improve the social and developmental activities. The present study's objectives and results were discussed. Our results reveled that this important and neglected issue of the society. In real life this issue is very common every third house in Pakistan is facing. The problem highlighted by the

NGO is that mother cant says this because so many factors are significantly associated with her life. Only the way to lessen is the awareness of undesired children in women and strong message for the prefer of son is needed. Almost same results were concluded from the qualitative study.

# CHAPTER 5

# QUALITATIVE RESEARCH TOWARD POLICY

The focus of this section of the study is to shed light on qualitative aspect which includes practical health care facilities and efforts to explain what, how, and why. The objective of this study is to build a bridge between two intersecting areas of research, social indicator, and health-related areas. For this, here we proved the importance of social and health areas, one by empirical study, and some of the focus is diverted to qualitative research. This was a targeted sample, focusing on the women who are facing such circumstances. We choice public and private hospitals using snowball sampling and target the women who are pregnant and as well delivered the baby. Out of that woman we can know which are welling to answer that the child is unwanted or wanted. For qualitative research, we have conducted interviews from different hospitals which include public and private. On average 30 plus interviews were conducted with pregnant women who delivered their baby.

The efforts are put to assess the effect of undesired children on childhood mortality, morbidity, and nutritional problems in Pakistan. The qualitative findings suggest that age, education, residence, religion, household wealth index, husband's education, and some other factors including access to media, and working status of women have an association with undesired children. Because there is a less ratio given to women's opinions in getting pregnant and having another baby. Therefore, a wide number of pregnancies are categorized as undesired, and children born out of this pregnancy will then be unequally treated and might face childhood morbidity and malnutrition. That's the cause of a high graph of childhood diseases.

During interviews, the common aspect that is revealed by most of the mothers having three or four babies, is that she doesn't want another baby. Their health condition of is deteriorating day by day due to poor health facilities. Though, upon request of not revealing their personal information, in fear of their relatives and in-laws, they disclosed that they are not even ready for third or fourth child, but they must bear the pressure of society, family, and their husband.

In a peered interview, the ratio of unplanned babies was slightly higher among the mothers who were expecting a fourth or fifth baby, while the ratio of planned babies was less among them. This ratio of comparison is stated in table 5.1 below. This implies that the preferred family size in Pakistan is 3 or 4.

**Table 8: Results of qualitative research**

| Unplanned babies | Planned babies |
|---|---|
| 87.5% | 12.5% |

Another aspect that the pregnant women highlighted and especially those who declared their babies as unwanted, their babies' birth size was diagnosed to be low and most of the babies were suffering from fever. That is another proof of being unwanted and a thing that they were doing by force and not with their willingness.

## 5.1. Policy Recommendation

Health is the foremost part of life either of the mother or child. This research revealed that child morbidity and malnutrition have a higher ratio in Pakistan. We found very little effect on child mortality, but qualitative research findings show child mortality is also higher in the case of Pakistan. The findings illustrate the idea of an ideal family concerning initiatives to alleviate child morbidity and malnutrition in Pakistan. Limiting undesirable pregnancies and having fewer children may help to lessen some child morbidity, mortality, and malnutrition of children, and with the right use of contemporary contraception, unwanted pregnancy can be avoided. The consequences of reducing child and under-five mortalities, this will help in achieving the SDGs. However, due to sex preferences (especially of male child), undesired children are born. Many women frequently become pregnant to have the desired sex of children, even if it means giving birth to unwanted sons or daughters. Therefore, there should be a strong and clear message against gender preference.

The tagline, “No unwanted pregnancies are desirable” is another option. Our empirical results show a significant attention is needed to put over mother’s education, that will be the way to shift focus for lowering child morbidity, mortality, and malnutrition in future policy action. The mothers with higher levels of education reduced the chance of undesired children as well as child morbidity and mortality. A policy response to decrease mortality among children born in families larger than the ideal would necessitate an understanding of the reasons for having children beyond stated ideals. We will provide a path for policymakers, planners, and health providers to resolve the issues of child health and address the flaws in the health system.

# CHAPTER 6

# CONCLUSION AND LIMITATIONS

Pakistan Demographic and Health Survey, commonly called PDHS covers every aspect of the demographic and social arena. Here, an effort is put to cover sociodemographic, socioeconomic, health, and nutritional aspects with a machine learning approach to detect the health status of the undesired child under five. This way, impact of an undesired child on childhood mortality, morbidity, and nutritional issues in Pakistan are studied. The empirical results suggest that child morbidity, fever, and cough were highly occurring factors among undesired children.

The results of sociodemographic factor suggest that certain provinces have their own social norms and traditions, where male chauvinism and male dominant society pose a high occurrence of undesired children. Where women are less likely be the part of family plannings. Socioeconomic factor has emphasized the economic status of households and in a society like Pakistan, where majority of households has a smaller number of bread winners and monthly income lies under 200 USD, the requirements of proper care and medication is not met, and child diseases are way more common than other Asian developing nations.

Gender preference (particularly for a male child) is another factor of more undesired children. Health status lies at stake, as unwanted pregnancies have a significant link to low birth weight.

The mother's apprehension about the right diet for an unwanted pregnancy could be the explanation, resulting in increased nutritional difficulties. As a response, most unwanted offspring may face inequitable treatment, resulting in nutritional deficiencies and health issues.

This way, child malnutrition will be the contributor to elevated risk of stunting and being underweight. With the application of machine learning techniques, the health condition of the undesired children given some risk factors, the algorithms highlight the health problems of the undesired child and findings provide a baseline for policy implication for family planning, raising health status, improving child mortality index, and contributing to healthy children with resolution of their health problems.

## 6.1 Limitations and Strengths

Being the first ever research in this aspect and inclusion of every significant factor of the society of Pakistan using PDHS data, this study possesses some strengths. The generalization to other comparable socioeconomic level is a must as the analysis as provided a base for such purpose. To predict the undesirable state of the child's health, machine learning techniques are used. Despite such strong base, there are some drawbacks as well. There is no policy implication for health improvement, and for mothers' ethical perspectives as unwanted pregnancy later turn into wanted after birth.

The causality and health problems of mother and undesired child isn't well explained with the results as a quantitative study would not be able to determine how a mother handled an unintended pregnancy and the absence of the child. So, qualitative studies and prospective cohort studies produce the best results.

## 6.2 Policy implication

The government needs to reemphasize the family planning program to reduce childhood mortality, morbidity, and malnutrition due to unwanted pregnancies lead to unwanted childbirth. Limiting unwanted pregnancies may reduce the causes of mortality of undesired children, and unwanted pregnancies may be overcome with the proper utilization of modern contraception. This will help achieve the SDG of reducing infant and under-five mortality. However, undesired children are most likely to give birth due to a desired sex composition. This desire of male child resulted in many women may frequently becoming pregnant and giving birth to undesired sons or daughters. Therefore, a strong message should be provided about gender preferences. Another must do factor is improving the health facilities for mothers and children to decrease the likely chances of child mortality, morbidity, unwanted pregnancies, low birth weight, abnormal growth of children, and most probably, the equitable treatment to every child. The action plan to policy maker is there must be a strong slogan that unwanted pregnancies are desirable and strong preference of son than daughter. As a Muslim we have a strong believe daughters are blessing of Allah.

# References

Aboderin, O. S., Fadare, O. G., Kumuyi, G., Lawal, O., & Babtunde, O. (2011). Pedagogical appraisal of Internet and computer usage among secondary school teachers and students in Southwestern part of Nigeria. *World Journal of Young Researchers, 1*(4), 54-59.

Ahmed, N., Ahammed, R., Islam, M. M., Uddin, M. A., Akhter, A., Talukder, M. A. A., & Paul, B. K. (2021). Machine learning based diabetes prediction and development of smart web application. *International Journal of Cognitive Computing in Engineering*, *2*, 229-241.

Ahmed, J., Alam, A., & Raynes-Greenow, C. (2018). Maternal empowerment and healthcare access determines stillbirths and early neonatal mortality in Pakistan: analysis of demographic and health survey 2012-13. *Journal of Global Health Reports, 2*.

Alam, M., & Islam, M. (2022). Is there any association between undesired children and health status of under-five children? Analysis of a nationally representative sample from Bangladesh. *BMC pediatrics, 22*(1), 1-14.

Alam, M. S. B., Patwary, M. J., & Hassan, M. (2021, February). Birth mode prediction using bagging ensemble classifier: A case study of bangladesh. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)* (pp. 95-99). IEEE.Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert systems with applications*, *36*(2), 3240-3247.

Anderson, S., & Ray, D. (2010). Missing women: age and disease. *The Review of Economic Studies, 77*(4), 1262-1300.

Asif, M. F., & Pervaiz, Z. (2019). Socio-demographic determinants of unmet need for family planning among married women in Pakistan. *BMC public health*, *19*(1), 1-8.

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., De Onis, M., Ezzati, M., . . . Group, C. U. S. (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *The lancet, 371*(9608), 243-260.

Barber, J. S., Axinn, W. G., & Thornton, A. (1999). Unwanted childbearing, health, and motherchild relationships. *Journal of health and social behavior*, 231-257.

Bongaarts, J. (1990). The measurement of wanted fertility. *Population and Development Review*, 487506.

Bongaarts, J. (1992). Do reproductive intentions matter? *International Family Planning Perspectives*, 102-108.

Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public health nutrition*, *25*(2), 269280.

Biau, G., Devroye, L., & Lugosi, G. (2008). Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research*, *9*(9).

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144152).

Bzdok, D., Krzywinski, M., & Altman, N. (2018). Machine learning: supervised methods. *Nature methods*, *15*(1), 5.

Casterline, J. B., & Laila, O. (2007). El-Zeini. 2007.“. *The estimation of unwanted fertility.*

*Demography, 44*(4), 729-745.

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S.-H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry investigation, 16*(4), 262.

Chauhan, T., Rawat, S., Malik, S., & Singh, P. (2021, March). Supervised and unsupervised machine learning based review on diabetes care. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 581-585). IEEE.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Costa, J. C., Maia, M. F. S., & Victora, C. G. (2020). Ciência & Saúde Coletiva-25 years: contributions to pregnancy, delivery, and childhood studies. *Ciência & Saúde Coletiva, 25*, 48134830.

Croft, T. N., Marshall, A. M., Allen, C. K., Arnold, F., Assaf, S., & Balian, S. (2018). Guide to DHS statistics. *Rockville: ICF, 645*.

Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*, *178*, 622637.

D'Angelo, D. V., Gilbert, B. C., Rochat, R. W., Santelli, J. S., & Herold, J. M. (2004). Differences between mistimed and unwanted pregnancies among women who have live births. *Perspectives on sexual and reproductive health, 36*(5), 192-197.

de Moraes Batista, A. F., Miraglia, J. L., Donato, T. H. R., & Chiavegatto Filho, A. D. P. (2020). COVID19 diagnosis prediction in emergency care patients: a machine learning approach. *medRxiv*.

Desautels, T., Das, R., Calvert, J., Trivedi, M., Summers, C., Wales, D. J., & Ercole, A. (2017). Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. *BMJ open*, *7*(9), e017199.

Flatø, M. (2018). The differential mortality of undesired infants in sub-Saharan Africa. *Demography, 55*(1), 271-294.

Fenta, H. M., Zewotir, T., & Muluneh, E. K. (2021). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. *BMC Medical Informatics and Decision Making*, *21*(1), 1-12.

Gipson, J. D., Koenig, M. A., & Hindin, M. J. (2008). The effects of unintended pregnancy on infant, child, and parental health: a review of the literature. *Studies in family planning*, *39*(1), 18-38.

Guilmoto, C. Z. (2009). The sex ratio transition in Asia. *Population and Development Review, 35*(3), 519549.

Hargreaves, S., Rustage, K., Nellums, L. B., McAlpine, A., Pocock, N., Devakumar, D., & Aldridge, R. W. (2019). Occupational health outcomes among international migrant workers: a systematic review and meta-analysis, Lancet Glob Health.

Harpham, T., Huttly, S., De Silva, M. J., & Abramsky, T. (2005). Maternal mental health and child nutritional status in four developing countries. *Journal of Epidemiology & Community Health, 59*(12), 1060-1064.

Hazra, A., & Gogtay, N. (2016). Biostatistics series module 1: Basics of biostatistics. *Indian Journal of Dermatology, 61*(1), 10.

Hossain, M., Habib, M., Saleheen, A. A. S., Kamruzzaman, M., Rahman, A., Roy, S., ... & Rukon, R. (2022). Performance Evaluation of Machine Learning Algorithm for Classification of Unintended Pregnancy among Married Women in Bangladesh. *Journal of Healthcare Engineering*, *2022*.

Islam, M., Haque, M., Iqbal, H., Hasan, M., Hasan, M., & Kabir, M. N. (2020). Breast cancer prediction:

a comparative study using machine learning techniques. *SN Computer Science*, *1*(5), 1-14.

Islam, M. M., Rahman, M. J., Islam, M. M., Roy, D. C., Ahmed, N. F., Hussain, S., ... & Maniruzzaman, M. (2022). Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh. *International Journal of Cognitive Computing in Engineering*, *3*, 46-57.

Jain, S., Khanam, T., Abedi, A. J., & Khan, A. A. (2022, February). Efficient Machine Learning for Malnutrition Prediction among under-five children in India. In *2022 IEEE Delhi Section Conference (DELCON)* (pp. 1-10). IEEE.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., & Arriaga-Weiss, S. (2010). Classification in conservation biology: a comparison of five machine-learning methods. *Ecological Informatics*, *5*(6), 441-450.

Khan, J. R., Chowdhury, S., Islam, H., & Raheem, E. (2019). Machine learning algorithms to predict the childhood anemia in Bangladesh. *Journal of Data Science*, *17*(1), 195-218.

Khan, A., Khan, F. U., Israr, A., & Zardari, S. J. F. H. (2020). Determinants of Customer Satisfaction in Online Booking Hotels: An Evidence from Asian Countries. (12).

Khan, G. N., Turab, A., Khan, M. I., Rizvi, A., Shaheen, F., Ullah, A., . . . Yaqoob, M. (2016). Prevalence and associated factors of malnutrition among children under-five years in Sindh, Pakistan: a crosssectional study. *BMC nutrition*, *2*(1), 1-7.

Kozareva, Z., & Montoyo, A. (2006). *Paraphrase identification on the basis of supervised machine learning techniques.* Paper presented at the international conference on natural language processing (in Finland).

Kioko, C. (2021). *Supervised machine learning approaches to predict infant mortality: a case study of the 2014 Kenya Demographic and Health Survey* (Doctoral dissertation, University of Nairobi).

Liu, B., Cui, Q., Jiang, T., & Ma, S. (2004). A combinational feature selection and ensemble neural network method for classification of gene expression data. *BMC bioinformatics*, *5*(1), 1-12.

Lucy Lawrence, S. (2021). *Predicting stunting status among children under five years: The case study of Tanzania* (Doctoral dissertation, University of Rwanda).

Luo, S., & Chen, Z. (2020). A procedure of linear discrimination analysis with detected sparsity structure for high-dimensional multi-class classification. *Journal of Multivariate Analysis*, *179*, 104641.

MacKay, D. M., & McCulloch, W. S. (1952). The limiting information capacity of a neuronal link. *The bulletin of mathematical biophysics*, *14*(2), 127-135.

Masibo, P. K., & Makoka, D. (2012). Trends and determinants of undernutrition among young Kenyan children: Kenya Demographic and Health Survey; 1993, 1998, 2003 and 2008– 2009. *Public health nutrition*, *15*(9), 1715-1727.

Mfateneza, E., Rutayisire, P. C., Biracyaza, E., Musafiri, S., & Mpabuka, W. G. (2022). Application of machine learning methods for predicting infant mortality in Rwanda: analysis of Rwanda demographic health survey 2014–15 dataset. *BMC Pregnancy and Childbirth*, *22*(1), 1-13.

Montgomery, M. R., & Lloyd, C. B. (1996). Fertility and maternal and child health. In *The impact of population growth on well-being in developing countries* (pp. 37-65). Springer, Berlin, Heidelberg.

Nasrullah, M., Zakar, R., Zakar, M. Z., & Krämer, A. (2014). Girl-child marriage and its association with morbidity and mortality of children under 5 years of age in a nationallyrepresentative sample of Pakistan. *The Journal of pediatrics, 164*(3), 639-646.

Organization, W. H. (2019). Children: reducing mortality; 2019. *Reference Source*.

Pal, M., & Foody, G. M. (2012). Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5*(5), 1344-1355.

Patel, K. K., Rai, R., & Rai, A. K. (2021). Determinants of infant mortality in Pakistan: evidence from Pakistan Demographic and Health Survey 2017–18. *Journal of Public Health, 29*(3), 693-701.

Park, C. H., & Park, H. (2005). A relationship between linear discriminant analysis and the generalized minimum squared error solution. *SIAM Journal on Matrix Analysis and Applications*, *27*(2), 474-492.

Qadir, F., Khan, M. M., Medhin, G., & Prince, M. (2011). Male gender preference, female gender disadvantage as risk factors for psychological morbidity in Pakistani women of childbearing age-a life course perspective. *BMC public health*, *11*(1), 1-13.

Rackin, H. M., & Morgan, S. P. (2018). Prospective versus retrospective measurement of unwanted fertility: Strengths, weaknesses, and inconsistencies assessed for a cohort of US women. *Demographic research*, *39*, 61.

Rahman, M., Nasrin, S. O., Rahman, M., Rahman, A., Mostofa, G., Jesmin, S. S., & Buchanan, F. (2019). Maternal pregnancy intention and its association with low birthweight and pregnancy complications in Bangladesh: findings from a hospital-based study. *International health*, *11*(6), 447-454.

Raj, A., Saggurti, N., Winter, M., Labonte, A., Decker, M. R., Balaiah, D., & Silverman, J. G. (2010). The effect of maternal child marriage on morbidity and mortality of children under 5 in India: cross sectional study of a nationally representative sample. *Bmj*, *340*.

Rocca, C. H., Wilson, M. R., Jeon, M., & Foster, D. G. (2019). Stability of retrospective pregnancy intention reporting among women with unwanted pregnancies in the United States. *Maternal and child health journal*, *23*(11), 1547-1555.

Rahman, S. J., Ahmed, N. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Maniruzzaman, M. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *Plos one*, *16*(6), e0253172.

Rahman, A., Hossain, Z., Kabir, E., & Rois, R. (2021, October). Machine Learning Algorithm for Analysing Infant Mortality in Bangladesh. In *International Conference on Health Information Science* (pp. 205-219). Springer, Cham.

Rusdiana, S. (2016, October). Linear discriminant analysis (LDA) to clarifying vacant land allocations of Baiturraman district in Banda Aceh city. In *AIP Conference Proceedings* (Vol. 1775, No. 1, p. 030071). AIP Publishing LLC.

Santelli, J., & Rochat, R. (2003). Hatfield- Timajchy. *K., Gilbert, B. C., Curtis, K., Cabral, R., Hirsch, JS and Schieve, L*, 94-101.

Shapiro-Mendoza, C., Selwyn, B. J., Smith, D. P., & Sanderson, M. (2005). Parental pregnancy intention and early childhood stunting: findings from Bolivia. *International journal of epidemiology, 34*(2), 387-396.

Shaka, M. F., Woldie, Y. B., Lola, H. M., Olkamo, K. Y., & Anbasse, A. T. (2020). Determinants of undernutrition among children under-five years old in southern Ethiopia: does pregnancy intention matter? A community-based unmatched case-control study.

*BMC pediatrics, 20*(1), 1-10.

Singh, A., Chalasani, S., Koenig, M. A., & Mahapatra, B. (2012). The consequences of unintended births for maternal and child health in India. *Population studies, 66*(3), 223-239.

Shirima, C. P., Kimanya, M. E., Routledge, M. N., Srey, C., Kinabo, J. L., Humpf, H.-U., . . . Gong, Y. Y. (2015). A prospective study of growth and biomarkers of exposure to aflatoxin and fumonisin during early childhood in Tanzania. *Environmental health perspectives, 123*(2), 173-178.

Shahriar, M. M., Iqubal, M. S., Mitra, S., & Das, A. K. (2019, July). A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh. In *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 145-149). IEEE.

Song, F., Mei, D., & Li, H. (2010, October). Feature selection based on linear discriminant analysis. In *2010 international conference on intelligent system design and engineering application* (Vol. 1, pp. 746-749). IEEE.

Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications, 30*(2), 169-190.

Teupser, D., Baber, R., Ceglarek, U., Scholz, M., Illig, T., Gieger, C., . . . Huster, D. (2010).

Genetic regulation of serum phytosterol levels and risk of coronary artery disease.

*Circulation: Cardiovascular Genetics, 3*(4), 331-339.

UNICEF, WHO. World Bank Group, United Nations. Levels and Trends in Child Mortality: Report 2018, Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation. 2018

United Nations. The global strategy for women's, children's, and adolescents' health (2016–2030). World Health Organization; 2015. doi:10.1017/CBO9781107415324.004.

Vapnik, V. (1999). *The nature of statistical learning theory*: Springer science & business media.

Wemakor, A., Garti, H., Azongo, T., Garti, H., & Atosona, A. (2018). Young maternal age is a risk factor for child undernutrition in Tamale Metropolis, Ghana. *BMC research notes, 11*(1), 1-5.

Wang, X. D., Chen, R. C., Yan, F., Zeng, Z. Q., & Hong, C. Q. (2019). Fast adaptive k-means subspace clustering for high-dimensional data. *IEEE Access, 7*, 42639-42651.

World Health Organisation. Child mortality and causes of death. WHO. 2016. https://www.who.int/gho/child_health/mortality/en/. Accessed 28 Jul 2019

World Health Organization (WHO). Children: reducing mortality. 2019. https://www.who.int/newsroom/fact-sheets/detail/children-reducing-mortality. Accessed 22 Aug 2020.

Zohair, L. M. A. (2019). Prediction of Student'sperformance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-18.

Zakria, M., Afzal, M., Hussain, M., & Gillani, S. (2020). Assess knowledge, attitude, and practices of mothers regarding infant and child breastfeeding in the community. *Globus An International Journal of Medical Science, Engineering and Technology*, *9*(2), 15-23.

https://extranet.who.int/countryplanningcycles/sites/default/files/planning_cycle_repository/pakistan/national_health_vision_2016-25_30-08-2016.pdf

# Appendix

## Child morbidity 1. Random forest

**Table 4.11: Confusion matrix of fever using Random Forest**

Type of random forest: classification
Number of trees: 550
No. of variables split 3
OOB estimate of error rate: 41.09%

Confusion matrix:

| | No fever | fever | Class Error |
|---|---|---|---|
| No fever | 3050 | 420 | 0.1210375 |
| fever | 1987 | 401 | 0.8320771 |

**Table 4.12: Confusion matrix of cough using Random Forest**

Type of random forest: classification
Number of trees: 550
No. of variables split: 3
OOB estimate of error rate: 41.12%

Confusion matrix:

| | No Cough | Cough | Class Error |
|---|---|---|---|
| No Cough | 3271 | 359 | 0.0988981 |
| Cough | 2152 | 325 | 0.8687929 |

**Table 4.13: Confusion matrix of Diarrhea using Random Forest**

Type of random forest: classification
Number of trees: 550
No. of variables split: 3
OOB estimate of error rate: 22.1%

Confusion matrix:

| | No diarrhea | diarrhea | Class Error |
|---|---|---|---|
| No diarrhea | 4551 | 11 | 0.00241122 |
| diarrhea | 1283 | 11 | 0.99149923 |

**Table 4.14: Confusion matrix of Acute Respiration Infection using random forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split 3<br>OOB estimate of error rate: 24.85% | | | |
| Confusion matrix: | | | |
| | No ARI | ARI | Class Error |
| No ARI | 15 | 183 | 0.92424242 |
| ARI | 24 | 611 | 0.03779528 |

**Table 4.15 Confusion Matrix of Short Rapid Breath using Random Forest**

| | | | |
|---|---|---|---|
| Type of random forest:<br>classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate:<br>22.72% | | | |
| Confusion matrix: | | | |
| | No SBR | SBR | Class Error |
| No SRB | 2176 | 22 | 0.0100091 |
| SRB | 619 | 4 | 0.9935795 |

**Linear Discrimination Analysis**

**Table 4.16: Confusing matrix of Fever using LDA**

| | **No Fever** | **Fever** |
|---|---|---|
| **No Fever** | 3459 | 179 |
| **Fever** | 2334 | 181 |

**Table 4.17: Confusing matrix of cough using LDA Confusion matrix:**

| | **No Cough** | **Cough** |
|---|---|---|
| **No Cough** | 3648 | 193 |
| **Cough** | 2308 | 207 |

**Table 4.1.8 Confusion matrix of Diarrhea using LDA**

| | No Diarrhea | Diarrhea |
|---|---|---|
| **No Diarrhea** | 59 | 435 |
| **Diarrhea** | 41 | 830 |

**Table 4.1.9: Confusion matrix of short rapid breath using LDA**

| | No SBR | SBR |
|---|---|---|
| **No SBR** | 2760 | 0 |
| **SBR** | 764 | 11 |

**Table 4.1.10: Confusion matrix of ARI using LDA**

| | No ARI | ARI |
|---|---|---|
| **No ARI** | 4 | 276 |
| **ARI** | 5 | 893 |

**Neural Network**

**Table 4.1.11: Confusion matrix of fever using NN**

| | No fever | fever |
|---|---|---|
| **No fever** | 2401 | 1337 |
| **fever** | 496 | 689 |

**Table 4.1.12: Confusion matrix of cough using NN**

| | No Cough | Cough |
|---|---|---|
| **No Cough** | 2815 | 1490 |
| **Cough** | 833 | 973 |

**Table 4.1.13: Confusion matrix of Diarrhea of Neural Network**

| | No Diarrhea | Diarrhea |
|---|---|---|
| **No Diarrhea** | 4045 | 1111 |
| **Diarrhea** | 13 | 53 |

**Table 4.1.14: Confusion matrix of diarrhea using NN for testing**

| | No Diarrhea | Diarrhea |
|---|---|---|
| **No Diarrhea** | 711 | 194 |
| **Diarrhea** | 10 | 4 |

**Table 4.1.15: Confusion matrix of ARI using Neural network**

| | No ARI | ARI |
|---|---|---|
| **No ARI** | 93 | 31 |
| **ARI** | 145 | 733 |

**Table 4.1.16: Confusion matrix of SBR using Neural network**

| | No SBR | SBR |
|---|---|---|
| **No SBR** | 2293 | 539 |
| **SBR** | 49 | 116 |

**Support vector Machine**

**Table 4.1.17: Confusion matrix for fever using SVM**

| | No fever | fever |
|---|---|---|
| **No fever** | 2471 | 1425 |
| **fever** | 99 | 323 |

**Table 4.1.18: Confusion matrix of Cough using SVM**

| | No Cough | Cough |
|---|---|---|
| **No Cough** | 2637 | 1509 |
| **Cough** | 72 | 300 |

**Table 4.1.19Confusion matrix of Diarrhea Support Vector Machine**

| | No Diarrhea | Diarrhea |
|---|---|---|
| **No Diarrhea** | 4551 | 1307 |
| **Diarrhea** | 0 | 4 |

**Table 4.1.20: Confusion matrix of ARI using support vector machine**

| | No ARI | ARI |
|---|---|---|
| **No ARI** | 7 | 191 |
| **ARI** | 1 | 634 |

**Table 4.1.21: Confusion matrix of SBR using SVM**

| | No SBR | SBR |
|---|---|---|
| **No SBR** | 1933 | 525 |
| **SBR** | 0 | 10 |

**Treatment of Child Morbidity**

**Table 4.1.22: Confusion matrix of postnatal checkup of baby using Random Forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: | | | |
| Confusion matrix: | | | |
| | No PBC | PBC | Class Error |
| No PBC | 576 | 664 | 0.1210375 |
| PBC | 341 | 1941 | 0.8320771 |

**Table 4.1.23: Confusion Matrix of treatment of Diarrhea using Random Forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 37.81% | | | |
| Confusion matrix: | | | |
| | No TDia | TDia | Class Error |
| Not TDia | 93 | 311 | 0.769802 |
| TDia | 106 | 593 | 0.1516452 |

**Table 4.1.24: Confusion matrix of treatment of fever/cough using random forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 27.46% | | | |
| **Confusion matrix:** | | | |
| | No tfever/cough | tFever/cough | Class Error |
| **No tfever/cough** | 112 | 460 | 0.8041958 |
| **tFever/cough** | 95 | 1354 | 0.0655625 |

**Table 4.1.25: Confusion matrix of treatment of fever/cough using random forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 36.62% | | | |
| **Confusion matrix:** | | | |
| | No vitamin | vitamin | Class Error |
| **No vitamin** | 581 | 598 | 0.5072095 |
| **Vitamin** | 435 | 1207 | 0.2649208 |

## Linear Discrimination Analysis

**Table 4.1.27: Confusion matrix of postnatal checkup of baby using LDA**

| | **No PBC** | **PBC** |
|---|---|---|
| **No PBC** | 554 | 749 |
| **PBC** | 341 | 2058 |

**Table 4.1.28: Confusion Matrix of treatment of Diarrhea of using LDA**

| | No Tdia | Tdia |
|---|---|---|
| **No Tdia** | 59 | 435 |
| **Tdia** | 41 | 830 |

**Table 4.1.29: Confusion matrix using Linear Discrimination Analysis**

| | No TFever/cough | tFever/cough |
|---|---|---|
| **No TFever/cough** | **38** | **665** |
| **TFever/cough** | **52** | **1761** |

**Table 4.1.30: Confusion matrix using LDA Confusion matrix:**

| | No Vitamin | Vitamin |
|---|---|---|
| **No Vitamin** | **682** | **790** |
| **Vitamin** | **514** | **1539** |

**Support vector machine**

**Table 4.1.31: Confusion matrix of PBC using Support vector machine**

| | No PBC | PBC |
|---|---|---|
| **No PBC** | **507** | **302** |
| **PBC** | **733** | **1980** |

**Table 4.1.32: Confusion Matrix of treatment of diarrhea using SVM**

| | No Tdia | Tdia |
|---|---|---|
| **No Tdia** | **0** | **0** |
| **Tdia** | **404** | **699** |

**Table 4.1.33: Confusion Matrix using Support Vector Machine**

| | **No TFever/cough** | **TFever/cough** |
|---|---|---|
| **No TFever/cough** | 0 | **0** |
| **TFever/cough** | 62 | **1741** |

**Table 4.1.34: Confusion Matrix using Support Vector Machine**

| | **No Vitamin** | **Vitamin** |
|---|---|---|
| **No Vitamin** | **855** | **736** |
| **Vitamin** | **324** | **879** |

**Neural Network** **Table 4.1.35: Confusion matrix of PBC using NN**

| | **No PBC** | **PBC** |
|---|---|---|
| **No PBC** | 681 | 281 |
| **PBC** | **431** | 1745 |

**Table 4.1.36: Confusion matrix for treatment of Diarrhea using NN**

| | **No Tdia** | **Tdia** |
|---|---|---|
| **No Tdia** | 209 | 107 |
| **Tdia** | 207 | 638 |

**Table 4.1.37: Confusion matrix treatment of fever/cough using neural network**

| | **No TFever/cough** | **TFever/cough** |
|---|---|---|
| **No TFever/cough** | **140** | **34** |
| **TFever/cough** | **457** | **1507** |

**Table 4.1.38: Confusion matrix of using Neural Network**

| | No Vitamin | Vitamin |
|---|---|---|
| **No Vitamin** | **869** | **478** |
| **Vitamin** | **387** | **1263** |

## Child Malnutrition

**Table 4.1.39: Confusion matrix of severely stunted using random forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 16.87% | | | |
| **Confusion matrix:** | | | |
| | **No SS** | **SS** | **Class Error** |
| **No SS** | 1183 | 16 | 0.01334445 |
| **SS** | 228 | 19 | 0.92307692 |

**Table 4.1.40: Confusion matrix of stunted using random forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 36.3% | | | |
| **Confusion matrix:** | | | |
| | **No SS** | **SS** | **Class Error** |
| **No SS** | 1141 | 353 | 0.236278 |
| **Stunted** | 539 | 421 | 0.5614583 |

**Table 4.1.41: Confusion matrix of severely underweight using random forest**

| | | | |
|---|---|---|---|
| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 16.7% | | | |
| **Confusion matrix:** | | | |
| | **No suw** | **suw** | **Class Error** |

| | | | |
|---|---|---|---|
| **No suw** | 1064 | 9 | 0.008387698 |
| **suw** | 208 | 14 | 0.936936937 |

**Table 4.1.42: Confusion matrix of underweight using random forest**

| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 16.7% | | | |
|---|---|---|---|
| **Confusion matrix:** | | | |
| | **No underweight** | **underweight** | **Class Error** |
| **No underweight** | 1105 | 39 | 0.03409091 |
| **underweight** | 308 | 45 | 0.87252125 |

**Table 4.1.43: Confusion matrix of severely wasted using random forest**

| | **No sw** | **sw** |
|---|---|---|
| **No sw** | 450 | 0 |
| **sw** | 30 | 8 |

**Table 4.1.44: Confusion matrix of wasted using random forest**

| | **No wasted** | **wasted** |
|---|---|---|
| **No wasted** | 678 | 61 |
| **wasted** | 5 | 12 |

**Table 4.1.45: Confusion matrix of birth size using random forest**

| Type of random forest: classification<br>Number of trees: 550<br>No. of variables split: 3<br>OOB estimate of error rate: 25.56% | | | |
|---|---|---|---|
| **Confusion matrix:** | | | |
| | **small** | **average** | **large** |
| **small** | 0 | 364 | 4 |
| **average** | 0 | 3838 | 9 |
| **large** | 0 | 941 | 0 |

**Table 4.1.46: Confusion matrix of birth size using Random Forest testing data set**

| | small | average | large |
|---|---|---|---|
| **small** | 0 | 82 | 0 |
| **average** | 0 | 951 | 229 |
| **large** | 0 | 32 | 0 |

**Linear Discrimination Analysis**

**Table 4.1.47: Confusion matrix of severely stunted using LDA**

| | No SS | SS |
|---|---|---|
| **No SS** | 484 | 107 |
| **SS** | 32 | 6 |

**Table 4.1.48: Confusion matrix of severely stunted using LDA**

| | No Stunted | Stunted |
|---|---|---|
| **No Stunted** | 1523 | 346 |
| **Stunted** | 722 | 471 |

**Table 4.1.49: Confusion matrix of severely underweight using LDA**

| | No SUW | SUW |
|---|---|---|
| **No SUW** | 1218 | 179 |
| **SUW** | 33 | 61 |

**Table 4.1.50: Confusion matrix of underweight using LDA**

| | No underweight | underweight |
|---|---|---|
| **No underweight** | 1611 | 15 |
| **underweight** | 489 | 14 |

**Table 4.1.51: Confusion matrix of birth size using LDA**

| | small | average | large |
|---|---|---|---|
| **small** | 0 | 429 | 0 |
| **average** | 0 | 4800 | 229 |
| **large** | 0 | 1171 | 22 |

**Table 4.1.52: Confusion matrix of underweight using LDA**

| | No underweight | underweight |
|---|---|---|
| **No underweight** | 205 | 70 |
| **underweight** | 28 | 16 |

**Table 4.1.53: Confusion matrix of severely wasted using LDA**

| | No SWasted | SWasted |
|---|---|---|
| **No SWasted** | 158 | 4 |
| **SWasted** | 0 | 0 |

**Table 4.1.54: Confusion matrix of wasted using LDA**

| | No Wasted | Wasted |
|---|---|---|
| **No Wasted** | 206 | 37 |
| **Wasted** | 0 | 0 |

**Table 4.1.55: Confusion matrix of birth size using LDA testing**

| | small | average | large |
|---|---|---|---|
| **small** | 0 | 21 | 0 |
| **average** | 0 | 241 | 229 |
| **large** | 0 | 58 | 22 |

**Support vector machine**

**Table 4.1.56: Confusion matrix of severely stunted using SVM**

| | No SS | SS |
|---|---|---|
| **No SS** | 1199 | 236 |
| **SS** | 0 | 11 |

**Table 4.1.57: Confusion matrix of stunted using SVM**

| | No Stunted | Stunted |
|---|---|---|
| **No Stunted** | 1310 | 558 |
| **Stunted** | 184 | 402 |

**Table 4.1.58: Confusion matrix of severely underweight using SVM**

| | No SUW | SUW |
|---|---|---|
| **No SUW** | 1218 | 17 |
| **SUW** | 33 | 61 |

**Table 4.1.59: Confusion matrix of underweight using SVM**

| | No underweight | underweight |
|---|---|---|
| **No underweight** | 1141 | 358 |
| **underweight** | 4 | 5 |

**Table 4.1.60: Confusion matrix underweight using Support vector machine**

| | No underweight | underweight |
|---|---|---|
| **No underweight** | 428 | 150 |
| **underweight** | 0 | 0 |

**Table 4.1.61: Confusion matrix of severely wasted using SVM**

| | No SUW | SUW |
|---|---|---|
| **No SUW** | 598 | 6 |
| **SUW** | 24 | 12 |

**Table 4.1.62: Confusion matrix of severely underweight using SVM**

| | No Wasted | Wasted |
|---|---|---|
| **No Wasted** | 1107 | 56 |
| **Wasted** | 24 | 12 |

**Table 4.1.63: Confusion matrix birth size using Support vector machine for testing**

| | small | average | large |
|---|---|---|---|
| **small** | 0 | 0 | 0 |
| **average** | 48 | 455 | 111 |
| **large** | 0 | 0 | 0 |

**Neural Network** **Table 4.1.64: Confusion matrix severely stunted using NN**

| | No SS | SS |
|---|---|---|
| **No SS** | 1416 | 189 |
| **SS** | 30 | 10 |

**Table 4.1.65: Confusion matrix stunted using NN**

| | No Stunted | Stunted |
|---|---|---|
| **No Stunted** | 1373 | 629 |
| **Stunted** | 121 | 331 |

**Table 4.1.66: Confusion matrix severely underweight using NN**

| | No SUW | SUW |
|---|---|---|
| **No SUW** | 1479 | 216 |
| **SUW** | 38 | 78 |

**Table 4.1.67: Confusion matrix underweight using NN**

| | No underweight | underweight |
|---|---|---|
| **No underweight** | 1338 | 253 |
| **underweight** | 55 | 164 |

**Table 4.1.68: Confusion matrix of severely wasted using NN**

| | No SW | SW |
|---|---|---|
| **No SW** | 609 | 14 |
| **SW** | 1 | 26 |

**Table 4.1.69: Confusion matrix of wasted using NN for testing**

| | No Wasted | Wasted |
|---|---|---|
| **No Wasted** | 101 | 15 |
| **Wasted** | 5 | 0 |

**Table 4.1.70: Confusion matrix birth size using NN**

| | small | average | large |
|---|---|---|---|
| **small** | 0 | 329 | 0 |
| **average** | 48 | 4789 | 111 |
| **large** | 0 | 1271 | 45 |

**Child Morality**

**neural network**

**table 4.1.71: Confusion matrix for Early Neonatal using training**

| | No ENN | ENN |
|---|---|---|
| **No ENN** | 4388 | 91 |
| **ENN** | 1 | 21 |

**Table 4.1.72: Confusion matrix for Late Neonatal using training**

| | No LNN | LNN |
|---|---|---|
| **No LNN** | 1870 | 28 |
| **ELNN** | 1 | 0 |

**Table 4.1.73: Confusion matrix for Post Neonatal using testing**

| | No PNN | PNN |
|---|---|---|
| **No PNN** | 4374 | 58 |
| **PNN** | 68 | 2 |

**Table 4.1.74: Confusion matrix for Infant using testing**

| | No Infant | Infant |
|---|---|---|
| **No Infant** | 4321 | 144 |
| **Infant** | 9 | 27 |

**Table 4.1.75: Confusion matrix for child using testing**

| | No child | child |
|---|---|---|
| **No child** | 4430 | 58 |
| **child** | 1 | 13 |

**Table 4.1.76: Confusion matrix for early neonatal using testing**

| | No ENN | ENN |
|---|---|---|
| **No ENN** | 1841 | 51 |
| **ENN** | 0 | 0 |

**Table 4.1.77: Confusion matrix for late Neonatal using testing**

| | No LNN | LNN |
|---|---|---|
| **No LNN** | 1870 | 28 |
| **LNN** | 1 | 0 |

**Table 4.1.78: Confusion matrix for post neonatal using testing**

| | No PNN | PNN |
|---|---|---|
| **No PNN** | 1252 | 12 |
| **PNN** | 0 | 0 |

**Table 4.1.79: Confusion matrix for post neonatal using testing**

| | No LNN | LNN |
|---|---|---|
| **No LNN** | 1836 | 36 |
| **LNN** | 1 | 0 |

**Table 4.1.80: Confusion matrix for post neonatal using testing**

| | No Infant | Infant |
|---|---|---|
| **No Infant** | 1221 | 43 |
| **Infant** | 0 | 0 |

**Table 4.1.81: Confusion matrix for child using testing**

| | No child | child |
|---|---|---|
| **No child** | 1871 | 58 |
| **Child** | 0 | 0 |

**Table 4.1.82: Confusion matrix for early neonatal using SVM**

| | No ENN | ENN |
|---|---|---|
| **No ENN** | 1841 | 53 |
| **ENN** | 1 | 0 |

**Table 4.1.83: Confusion matrix for post neonatal using SVM**

| | No PNN | PNN |
|---|---|---|
| **No PNN** | 5059 | 73 |
| **PNN** | 0 | 0 |

**Table 4.1.84: Confusion matrix for late neonatal using SVM**

| | No LNN | LNN |
|---|---|---|
| **No LNN** | 1371 | 170 |
| **LNN** | 1 | 0 |

**Table 4.1.85: Confusion matrix for infant using SVM**

| | No Infant | Infant |
|---|---|---|
| **No Infant** | 4926 | 210 |
| **Infant** | 0 | 0 |

**Table 4.1.86: Confusion matrix for child using SVM**

| | **No Child** | **Child** |
|---|---|---|
| **No Child** | 4430 | 71 |
| **Child** | 0 | 0 |

**Table 4.1.87: Confusion matrix for early and late neonatal using LDA**

| | **No ENN** | **ENN** |
|---|---|---|
| **No ENN** | 1869 | 0 |
| **ENN** | 15 | 0 |
| | No LNN | LNN |
| **No LNN** | 308 | 12 |
| **LNN** | 0 | 0 |

**Table 4.1.88: Confusion matrix for post neonatal using LDA**

| | **No PNN** | **PNN** |
|---|---|---|
| **No PNN** | 315 | 0 |
| **PNN** | 15 | 0 |

**Table 4.1.89:Confusion matrix for infant using LDA**

| | **No Infant** | **infant** |
|---|---|---|
| **No infant** | 304 | 0 |
| **Infant** | 16 | 0 |

**Table 4.1.90: Confusion matrix for early neonatal using NN**

| | No ENN | ENN |
|---|---|---|
| **No ENN** | 1841 | 5 |
| **ENN** | 0 | 0 |

**Table 4.1.91: Confusion matrix for LATE neonatal using NN**

| | No LNN | LNN |
|---|---|---|
| **No LNN** | 1836 | 0 |
| **LNN** | 1 | 0 |

**Table 4.1.90: Confusion matrix for post neonatal using NN**

| | No PNN | PNN |
|---|---|---|
| **No PNN** | 1875 | 24 |
| **PNN** | 0 | 0 |

**Table 4.1.90: Confusion matrix for child and infant using NN**

| | No Child | Child |
|---|---|---|
| **No Child** | 1871 | 28 |
| **Child** | 0 | 0 |
| | No Infant | Infant |
| **No Infant** | 1866 | 28 |
| **Infant** | 5 | 0 |