Nutritional Status Prediction in Neonate Using Machine Learning Techniques: A Comparative Study

Zakir Hussain(✉) and Malaya Dutta Borah

Department of Computer Science and Engineering,
National Institute of Technology Silchar, Cachar, Silchar 788010, Assam, India
zak08hussain@gmail.com, malayaduttaborah@gmail.com

Abstract. For proper physical and mental development, nutrition has vital role to play. Proper development of child both physically and mentally is essential for building a peaceful society. This study has focused on prediction of nutritional status in neonate using the features of mother. The machine learning techniques namely Logistic Regression, Decision Tree, K-Nearest Neighbor, Linear Discriminant Analysis, Gaussian Naïve Bayes and Support Vector Machine have been implemented using a self created dataset having 445 instances and eighteen features. The dataset has a label with two classes- under weight and normal weight. We have got an accuracy of 88% for Logistic Regression, 99% for Decision Tree, 85% for K-Nearest Neighbor, 93% for Linear Discriminant Analysis, 86% for Gaussian Naïve Bayes and 88% for Support Vector Machine. All the techniques have shown magnificent performance.

Keywords: Nutritional status · Nutritional status in neonate · Nutritional status in new born · Features of mother · Malnutrition · Under-weight

1 Introduction

In low-and-middle-income countries like India, both extremes of malnutrition are faced by the children due to rapid change in food system [8]. Most of the times the term malnutrition is used to mean under-nutrition, but actually it mean both under-nutrition and over-nutrition [3]. There are many reasons to develop malnutrition. Some of those are intake of diet or amount of calorie consumption, imbalanced calorie intake while suffering from disease, complications occurred due to illness such as less absorption and excessive loss of nutrient and sometimes all these causes can also be combined [14]. Also malnutrition influences in losing physical and mental potential and decreases the potential to handle stress [4]. Food is the main source of nourishing the body by supplying the necessary chemicals. Refusal to the food increases the probability to suffer from malnutrition. Early identification of the susceptibility toward suffering from malnutrition in

A. Bhattacharjee et al. (Eds.): MIND 2020, CCIS 1241, pp. 69–83, 2020.
https://doi.org/10.1007/978-981-15-6318-8_7

latter ages is very necessary as almost 80% of brain development happens within two years from the birth. If this development is disturbed by any means, then the children is likely to be the victim of the consequences of less development of the brain. Other consequences like higher infection, higher complication, increased loss of muscle, reduced wound healing, increased morbidity and mortality are also associated. As a result the performance of the children in all aspect gets reduced. This will create a human resource with less efficiency and less compatibility in certain fields. This will hamper the economic development of the country.

As stated in the report of World Health Organization (WHO), India is one of the countries having large number of children and adults suffering from the consequences of malnutrition [15]. The degree of malnutrition is very high in India. A WHO and UNICEF review in 2018 suggested that the Sustainable Development Goals (SDG) of eliminating all types of malnutrition by 2030 was desirable but not achievable and, on the basis of trends so far, recommended targets for the malnutrition indicators up to 2030 [13]. To ensure a curative measure for this problem is- early detection of the possibility for a child to be affected by malnutrition and treat accordingly. For early detection, the study and experiment can be carried out on the mother carrying the child in her womb. Some unique features are visible on the mother who is going to deliver a child with certain probability of getting affected by malnutrition. Those features can be considered for machine learning techniques to classify the children in certain classes for similar kind of treatment.

1.1 Motivation

If we look back we find that a large number of unsocial, inhuman and menacing activities happen in the under developed and developing countries like India. The count of abduction, kidnapping, accident, addiction, theft, extremist activity and other crimes have got an enhancement from past few years. Many causes are there to foster these activities but nutritional state of the body can be thought of as one of the major causes. If the above cases are analysed it can be found that the assailant is a victim of some sort of physical or mental disorder. Among other reasons, nutritional state of the body is highly responsible for these disorders. It is always said that a healthy mind resides in a healthy body. For a healthy body there must have proper balance in the nutrients. A person with good health and good mental state posses the ability to analyse the situations and causes with positive note and can handle those with positive attitude. A healthy mind can help a person to get right direction to decide what to accept and what to reject. Malnutrition in either the form of under-nutrition or over-nutrition can disrupt the development of physical and mental health. It has been expected that if the malnutrition can be overcome in early ages, then the rate of these irksome activities can be reduced significantly and we can dream for a better society.

1.2 Problem Statement

As the above mentioned problem is a problem that must be addressed for better society, many theoretical researches have been carried out. Causes of the problem is also assumed and medical science is trying to identify and solve the problems with their traditional approaches. The traditional approach is very time consuming and costly, and people underlying in Below Poverty Line (BPL) can not afford. Use of the improved techniques in this area is very less. Application of machine learning techniques in addressing these problems can be expected to reduce the cost as well as time. Machine learning techniques have the potential to predict the nutritional status in neonate better and as a result doctors can take quality decision for right treatment. In this research, machine learning techniques namely Logistic Regression, Decision Tree, K-Nearest Neighbor, Linear Discriminant Analysis, Gaussian Naïve Bayes and Support Vector Machine have been used to get maximum possible prediction using some important features. Since this is considering the prediction for the neonate, so the study will mainly focus on the mother.

The main objective of this research is to alert the parent by predicting nutritional status of neonate. The aim of this research is to find the solution for the following research questions to achieve the above objectives:

- Is it possible to identify the nutritional status of neonate?
- Is it possible to predict nutritional status of neonate before birth by using features of mother?
- Can machine learning techniques correctly predict nutritional status of neonate?

The first and the second questions are totally dependent on medical science. To get the answers of these questions, two Medical and Health Officers have helped a lot. To answer the third question, data were collected from Geramari MPHC, Dhubri, Assam and Kazigaon SD, Kokrajhar, Assam and machine learning techniques were applied.

2 Literature Review

Literature review for this study has focused on the prediction of nutritional status of neonate. Some surveys have also been carried out for the babies under the age of five years and also for adults suffering from the consequences of malnutrition. This is to extract the analysis reports and different techniques used for prediction. The study has covered a vast area related to prediction and classification considering different aspects. In [5], the authors have studied about the prediction of infant's weight in maternal hypertensive and non-hypertensive condition with Naïve Bayes method. For performance measurement they have used WEKA [5]. In [12], the authors have worked on a model based on the logical decision tree algorithm and decision tree algorithm. They have claimed that logical decision tree algorithm had the highest predictive capabilities with respect

to recall [12]. Also they have mentioned that the model based on the decision tree algorithm with low pruning had the highest precision [12]. The study in [1] aimed to asses maternal risk factors linked with low birth weight neonates. Also they have intended to compare Random Forest with Logistic Regression. They had carried out the experiment on 600 volunteer pregnant women [1]. They have identified four top rank variables: age of pregnancy, body mass index during the third three months of pregnancy, mother's age and body mass index during the first three months of pregnancy [1]. In their experiment they have claimed that Random Forest outperformed Linear Regression. The study in [7] aimed at prediction and classification of low birth weight data in Indonesia. They have used Binary Logistic Regression and Random Forest approach in IDHS (Demographic and Public Health Survey in Indonesia) data of 2012 [7]. They have considered the features like place of residence, time zone, wealth index, education level of mother, education level of father, age of the mother, job of the mother, number of children [7]. They have claimed that Linear Regression showed good performance in prediction, but poor performance in classification but Random Forest had good performance for both prediction and classification [7]. The study in [11] aims to compare logistic regression and data mining techniques. They also aimed to identify promising predictor variables as well as to come up with a decision support system to help the physicians for making better decision in case of low weight child birth [11]. They have used Logistic Regression, Support Vector Machine, Neural Network, Naïve Bayes, Decision Tree, Random Forest, and Data mining techniques [11]. They have carried out the experiment on data from Baystate Medical centre, Springfield, Massachusetts of 1986. They had used 189 instances with 11 attributes like ID-identification number, Mother's age in years (AGE), the weight before pregnancy (LWT), number of doctor visits during the first trimester of pregnancy (FTV), race (RACE), lifestyle information such as smoking (SMOKE), a history of previous preterm delivery (PTL), the existence of uterine irritability (UI), and hypertension (HT) [11]. They have identified highly influenced variables to predict LBW as- Mother's last weight before becoming pregnant, Mother's age, Number of doctor visits during the first trimester, parity [11]. The study in [17] has concentrated on extracting useful information from health indicators of pregnant woman for early detection of potential low birth weight cases using machine learning techniques. They have used Bayes minimum error rate and Indian healthcare data was used to construct decision rules [17]. They have used 18 attributes and got an accuracy of 96.77%. In study [2], they have mentioned about the development of Artificial Neural Network for predicting birth weight. They used Multi-layer concept topology on some birth cases in hospitals [2]. They have used the features like age, smoke, race, weight (lbs) before pregnancy, uterine irritability, number of doctor visits in 1st trimester, hypertension and claimed 100% accuracy. The study in [6] attempts to analyse malnutrition based on food intake, wealth index, age group, educational level, occupation. They have applied Decision tree, Artificial Neural Network in a dataset of family health survey having 254 instances and 9 attributes [6]. They have carried out the experiment for children under age of

five and found 68.50% accuracy for ID3, 77.17% for Random Forest and 77.17% for Multilayer perceptron. In [14], the authors have studied about the prediction of the mortality rate in surgical patients suffering from malnutrition. they have used data mining models like J48, ADTree and KNN [14]. In [3], the author has used decision tree technique for rule generation to help the medical experts to reduce the malnutrition condition among children under the age of 5 in developing countries. In [16], the authors have talked about the use of rule based classification along with agent technology to detect malnutrition in children. In [6], the authors have used decision tree and artificial neural networks to classify dataset of family health survey. They have studied about the nutritional status of children aged under five [6]. In [9], the authors have carried out logistic regression to identify the probabilities of explaining malnutrition by the features extracted by machine learning techniques from Indian Demographic and Health Survey dataset. In [10], the authors have used Bayesian Gaussian regression model for analysing the effect of selected socio-economic, demographic, health and environmental covariates on malnutrition for child under five years of age.

3 Methodology

3.1 Proposed Architecture

The proposed architecture in Fig. 1 shows the work flow from the data collection to the performance evaluation of different machine learning techniques. The required dataset has been created from the collected data and then the same has been inspected, cleaned, transformed and modelled as per the need of the used methods or techniques. After pre-procession the required features have

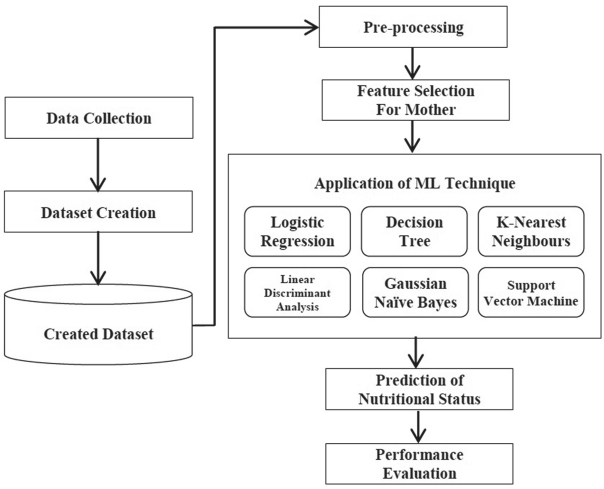

Fig. 1. Block diagram of proposed approach

been selected. Next the machine learning techniques have been applied on the dataset. The applied techniques have predicted and created the result. Based on the result, performance evaluation has been carried out for different methods using statistical tools.

3.2 Data Collection and Dataset Creation

The required data for this experiment has been collected from two government health centres. The authority of the health centres provided the hand written registers maintained for recording the information. The health centres keep track of a mother right from the start of pregnancy to the baby reaches 5 years of age. The work is for prediction and identification of nutritional status of neonate. So we have collected data of mothers. We have collected the data up to the birth of the baby as our objective is to predict the nutritional status of the neonate. The necessary data have been picked up and the dataset has been created on our own. The dataset is named as CBWDB.csv. The details of the dataset can be found in Table 1.

Table 1. Details of the created dataset

Name of the dataset:	CBWDB.csv
No. of instances:	445
No. of class labels:	02 L and N: L-under weight, N-normal weight
No. of attributes:	18
Details of attributes:	SEC Socio-Economic Condition
	Age(years) Age of mother at the time of pregnancy
	Height(cm) Height of mother at the time of pregnancy
	BGroup Blood Group of mother
	Parity No. of previous pregnancy crossing the period of viability
	ANC Antenatal Check
	Iwt(kg) Initial weight of mother
	Fwt(kg) Final weight of mother (Last ANC)
	IBP_sys Initial systolic Blood Pressure
	IBP_dias Initial diastolic Blood Pressure
	FBP_sys Final systolic Blood Pressure (Last ANC)
	FBP_dias Final diastolic Blood Pressure (last ANC)
	IHb(gm%) Initial Hemoglobin level
	FHb(gm%) Final Hemoglobin level (Last ANC)
	BS(RBS) Blood Sugar (Random)
	Term/Preterm Term: 37 to 40 weeks, Preterm: <37 weeks
	Sex Sex of new born baby
	BWt(kg) Baby birth weight
Missing value:	There were some missing values. Those have been replaced by the average of the nearest values.
Source:	Geramari MPHC, Dhubri, Assam, India and Kazigaon SD, Kokrajhar, Assam, India

3.3 Pre-processing and Feature Selection

While collecting the data from the health centres we noticed that there were so many columns recording different types of information about mother and child. All the information recorded in the register were not suitable for applying the machine learning techniques. Finally we have selected eighteen number of features and one label for classification. While creating the dataset, it was noticed that some data were not inserted in the register. As a result the data were missing in the created dataset. To get rid of the null value in the created dataset, the blank spaces were filled with the average of the values present in the nearby cells.

After thorough study and discussion with doctors, eighteen features were selected for the application of machine learning techniques. Those features are: SEC, Age(years), Height(cm), Bgroup, Parity, ANC, Iwt(kg), Fwt(kg), IBP_sys, IBP_dias, FBP_sys, FBP_dias, IHb(gm%), FHb(gm%), BS(RBS), Term/Preterm, Sex, BWt(kg). Along with these features one more column namely LNH has been considered for recording the nutritional status of the neonate. Table 1 can be referred for the details of these features.

3.4 Used Machine Learning Techniques

There are many machine learning algorithms available to implement classification and prediction problems. This study is about the prediction of the nutritional status of neonate. The following machine learning techniques have been implemented in a dataset created considering the features of mother. The dataset has a field denoting two different labels.

Logistic Regression. Logistic regression works when target is categorical. Here, output is strictly 1 or 0. For linear regression the hypothesis is $z = wx + b$, but for logistic regression it is the sigmoid of hypothesis of linear regression. That is, $h_\theta(x) = sigmoid(z)$. The hypothesis used here is the estimated probability. Mathematically this can be represented as-

$$h_\theta(x) = P(Y = 1|X; \theta) \tag{1}$$

This means, probability that $Y = 1$ given X which is parameterized by θ.
From the above equation we can write-

$$P(Y = 1|X; \theta) + P(Y = 0|X; \theta) = 1$$
$$\text{or } P(Y = 0|X; \theta) = 1 - P(Y = 1|X; \theta)$$

In this case, linear regression model is fitted with the data, then the data is acted upon by a logistic function. This logistic function predicts the targeted categorical dependent variable. Logistic regression comes in different categories like Binary, Multinomial and Ordinal Logistic Regression. Decision boundary is used to predict the class of a data. This means, a threshold can be fixed and based on that threshold, obtained estimated probability can be classified into classes.

Decision Tree. Decision tree are grown through an iterative splitting of data into discrete groups [5]. Decision tree which are used to predict categorical variables are called classification tree because it plays instances in categories [5]. Decision tree which is used to predict continuous variables are called regression tree [5]. Some of the decision tree are ID3, C4.5, C5.0, Quest, CART, and CHAID [5]. The tree is formed by selecting some attributes using Attribute Selection Measures. The attribute with best score is considered as splitting attribute. Information gain, Gain ratio and Gini index are the popular selection measures.

Information gain can be expressed as follows:
Information gain = Entropy before split of the dataset – Average entropy after split of the dataset
Mathematically information gain can be represented as follows:

$$Info(K) = -\Sigma_{i=1}^{m} p_i log_2 p_i \tag{2}$$

where, p_i is the probability that a random tuple in K belongs to class C_i.

$$Info_B(K) = \Sigma_{j=1}^{V} \frac{|K_j|}{|K|} \times Info(K_j) \tag{3}$$

$$Gain(B) = Info(K) - Info_B(K) \tag{4}$$

where, $Info(K)$ is the "average amount of information" that is needed to identify the class label of a tuple in K.

$\frac{|K_j|}{|K|}$ is the weight of j^{th} partition.

$Info_B(K)$ is the "expected information" that is required to classify a tuple in K based on partition by B.

The attribute B with the highest information gain i.e. $Gain(B)$ is chosen as the "splitting attribute" at node N.

This is generally used by ID3 (Iterative Dichotomiser) decision tree algorithm.

Gain Ratio normalizes the information gain to manage the issue of bias. The mathematical expression for this is as follows:

$$SplitInfo_B(K) = -\sum_{j=1}^{v} \frac{|K_j|}{|K|} \times \log_2 \left(\frac{|K_j|}{|K|} \right) \tag{5}$$

where, $\frac{|K_j|}{|K|}$ is the weight of the j^{th} partition, and v denotes number of discrete values in attribute B
The gain ratio can be defied as

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)}$$

This index is used by C4.5 algorithm. C4.5 algorithm is familiar as J48, and it is available in WEKA data mining tool.

Gini Index is to consider a binary split for each attribute. The mathematical expression for this is as follows:

$$Gini(K) = 1 - \Sigma_{i=1}^{m} P_i^2 \tag{6}$$

This index is basically used by CART (Classification and Regression Tree). This study has been implemented using Scikit-Learn and Scikit-Learn uses an optimized version of CART.

K-Nearest Neighbours. K-Nearest Neighbours (KNN) algorithm is supervised in nature. This algorithm can be used for both classification and regression problems. This algorithm works with an assumption that similar objects exist in close proximity. KNN uses the idea of similarity. Sometimes similarity is also called as distance, proximity, or closeness. Depending on the problem, there may have many ways of calculating the distance. However, Euclidean distance or straight line distance is familiar. So, it can be concluded that the working principle of KNN is to find distances between a query and all other examples in the data and to select K number of examples nearest to the query.

Linear Discriminant Analysis. Linear Discriminant Analysis (LDA) works by reducing the number of dimensions (i.e. variables or feature) in a dataset keeping as much information as possible. Basic steps along with mathematical representation followed by LDA can be as follows:

– Calculation of the "within class scatter matrix" using

$$M_w = \sum_{i=1}^{c} M_i \tag{7}$$

where c is the total number of distinct classes and

$$M_i = \sum_{s \in K_i}^{n} (s - m_i)(s - m_i)^T$$

and

$$m_i = \frac{1}{n_i} \sum_{s \in K_i}^{n} s_k$$

where s is a sample (i.e. row) and n is the total number of samples for a given class.
"Between class scatter matrix" is calculated using

$$M_B = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T \tag{8}$$

where,

$$m_i = \frac{1}{n_i} \sum_{x \in K_i} s_k$$

$$m = \frac{1}{n} \sum_{i}^{n} s_i$$

- Solving the generalized eigenvalue problem for $M_W^{-1}M_B$ to obtain the linear discriminant.
- Sorting of the eigenvalues from highest to lowest and selection of k eigenvectors. These are sorted as the eigenvectors having highest eigenvalues convey the most information.
- Creation of a matrix M_{new} with first two eigenvectors. Then

$$Y = X \cdot M_{new}$$

where X is a $s \times d$ matrix with s samples and d dimensions, and Y is a $s \times t$ matrix with s samples and $t(t < s)$ dimensions. That means Y is composed of the LDA components. Y can now be called as the new feature space.

Gaussian Naïve Bayes. Naïve Bayes classifier is a probability based machine learning model. Its crux is based on the Bayes theorem. i.e.

$$P(c|F) = \frac{P(F|c)P(c)}{P(F)} \tag{9}$$

Here, c can be considered as class variable which is supposed to be predicted and F can be considered as parameters or features. In this case, it is assumed that the predictors or features or parameters are independent, i.e. the presence of one particular feature does not affect the other. Hence it is called naive. There can have many instances of F. Let us consider that F has n number of instances denoted by $f_1, f_2, \ldots, f_n$. So, substituting F by its instances, we can write the above equation as

$$c = argmax_c P(c) \Pi_{i=1}^{n} P(f_i|c) \tag{10}$$

For Gaussian Naïve Bayes classifier, the above equation changes to a equation containing Gaussian function and looks like the following:

$$P(f_i|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} exp\left(\frac{(f_i - \mu_c)^2}{2\sigma_c^2}\right) \tag{11}$$

Support Vector Machine. Though Support Vector Machine (SVM) can be used for both regression and classification tasks, but it is widely used in classification. It has the ability to produce significant accuracy with less computation power. SVM basically finds a hyper-plane in an K-dimensional space (K is the number of variables or features) that distinctly classify the data points. To distinguish two classes of data points, there may have many hyper-plane. But the objective is to choose a plane that has maximum margin from the data points of both the classes. Maximum distance helps to classify the future unseen data points with more confidence. The dimension of the hyper-plane is dependent on the number of features. Data points that are closer to the hyper-plane, influence the orientation and position of the hyper-plane are called the support vectors.

In case of logistic regression, the output of linear function is squashed to the range of [0,1] using sigmoid function. In case of SVM, the output of linear function is considered directly. One class is maintained for the values greater than 1 and another is for the value -1. So, the range of SVM is $[-1, 1]$.

3.5 Experimentation

We have carried experiments on our proposed method using python. The popular tool called Scikit-Learn that comes with python has been used to implement different machine learning algorithms. The details of the experimental setup is shown in Table 2. The dataset has some fields that are containing alphanumeric values. The implemented algorithms don't work with alphanumeric values. That is why some of the attributes like SEC, BGroup, Term/Preterm and Sex has not been considered during the implementation. So, out of eighteen attributes we have selected fourteen attributes for the experimentation.

Table 2. Details of system and software packages used

Operating System	: Ubuntu 16.04 LTS
Language	: Python 3.5
Core Library	: Pandas
Library for visualisation	: Matplotlib and Seaborn
Library for ML	: Scikit-Learn
IDE	: jupyter notebook

4 Results and Discussions

The background study shows that limited works have been carried out in predicting the nutritional status of the neonate. Most of the studies have been carried out to predict and analyse the nutritional status of the child under the age of five years. Here, a novel study has been carried out to predict the nutritional status of the neonate with two labels namely under weight and normal weight using the features of mother. One more label called over-weight was also been included but due to less number of samples in the dataset for this category, the prediction accuracy for this label was very less. This level has been omitted for the time being and that has been left for the future work. This study had expected that certain features or attributes of mother during the pregnancy (from conceive to birth) can give a platform to predict the nutritional status of the neonate. In this study, six numbers of machine learning algorithms have been implemented in our created dataset and as expected we have got impressive results. The Table 3 shows the performance of six algorithms trained and tested on this dataset. Figure 2 can be referred for the pictorial view of the performance measures.

Now let us see whether the answers for the research questions are obtained or not. The first question was about the identification of the nutritional status of neonate. This is a theoretical question based on medical science. After thorough study and discussion with the doctors, it has been found that it is possible to identify the nutritional status of neonate from the birth weight. Birth weight of neonate has lot to say about the health status.

Table 3. Performance of different ML algorithms

ML Algorithm	Accuracy(%)		Precision		Recall		F1 Score	
	Train	Test	Label1	Label2	Label1	Label2	Label1	Label2
Logistic Regression	86	88	1.00	0.87	0.36	1.00	0.53	0.93
Decision Tree	98	99	1.00	0.99	0.95	1.00	0.98	0.99
K- Nearest Neighbour	87	85	0.69	0.87	0.41	0.96	0.51	0.91
Linear Discriminant Analysis	94	93	0.89	0.94	0.73	0.98	0.80	0.96
Gaussian Naïve Bayes	86	86	0.64	0.91	0.64	0.91	0.64	0.91
Support Vector Machine	90	88	0.83	0.88	0.72	0.88	0.76	0.86

The second question was about the possibility of predicting nutritional status of neonate by using features of mother. As we know, before birth the baby is totally dependant on mother for nutrition. The physical and mental status of mother will have significant influence on the baby. The physical status of mother can be understood by investigating some features of the mother. It is clear that by analysing the features of mother, the birth weight can be predicted and so the nutritional status.

The third question was about the correctness of prediction by machine learning techniques. To get the answer, this experiment has been carried out and for six number of popular classification and prediction algorithms of machine learning has been trained and tested. It has been found that among Logistic Regression, Decision Tree, K-Nearest Neighbours, Linear Discriminant Analysis, Gaussian Naïve Bayes and Support Vector Machine; Decision Tree has shown the best performance in terms of Accuracy, Precision, Recall and F1-Score. From Table 3, it is clear that machine learning techniques can predict the nutritional status with very good accuracy.

4.1 Comparison with Other Studies

It has been found that limited studies have been carried out for nutritional status prediction for neonate. Most of the studies are for the nutritional status of child below five years of age. Different researchers have implemented different algorithms on different datasets and varied number of attributes. One of the researchers has obtained 95% accuracy for Random Forest [1], another researcher has used Minimum Error Rate classifier and got an accuracy of 96.77% [17], another researcher got 80.372% of accuracy using Naïve Bayes classifier [5]. A researcher has claimed 100% accuracy using Artificial Neural Network [2]. In our study we have implemented six algorithms. We have got the accuracy of 88% for Logistic Regression, 99% for Decision Tree, 85% for K-Nearest Neighbors, 93% for Linear Discriminant Analysis, 86% for Gaussian Naïve Bayes and 88% for Support Vector Machine. Figure 3 can be referred for the plot of the features for Support Vector Machine using Principal Component Analysis.

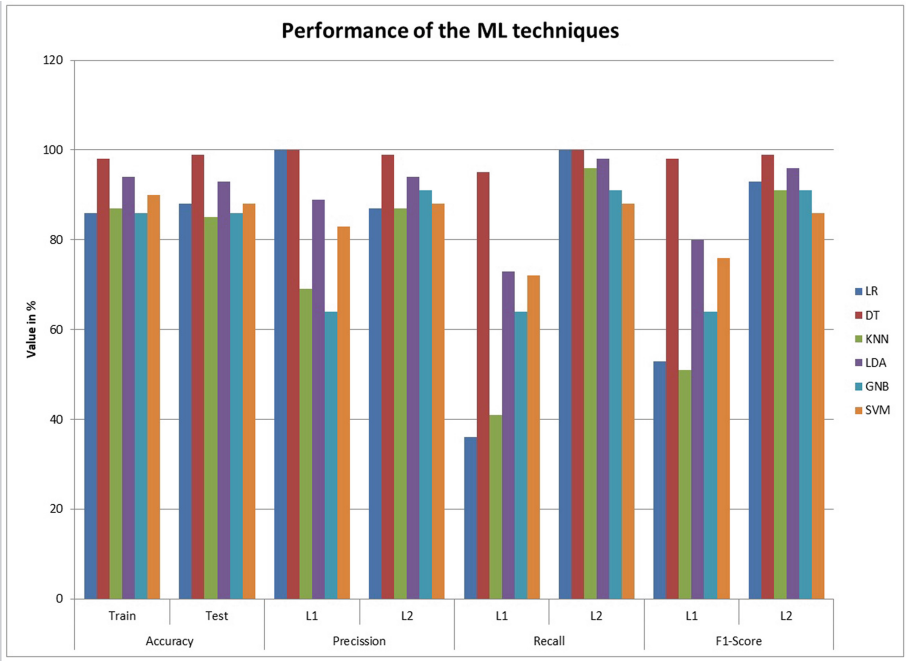

Fig. 2. Bar chart of performance measures

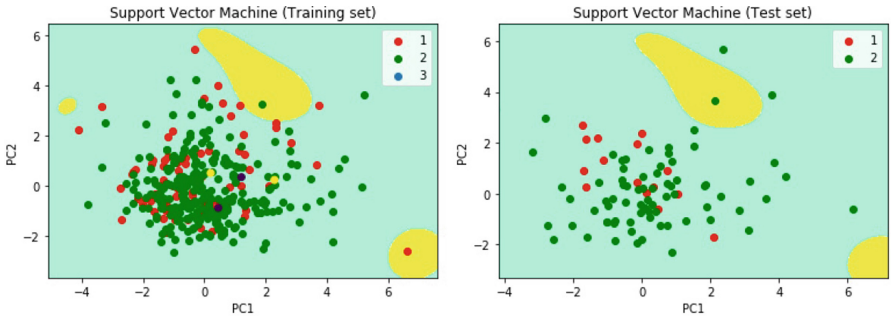

Fig. 3. Plot of the features for SVM using PCA

5 Conclusion and Future Work

Nutritional status plays vital role in signaling the development of physical and mental health. Proper development of child both physically and mentally is essential for building a peaceful society. Prediction regarding the nutritional status of neonate can alert the parent and doctors to take preventive and curative measures. In this study, six machine learning techniques have been implemented using a small dataset. All the techniques have shown impressive accuracy in prediction. As a future work, the dataset will be expanded and the third level

of nutritional status will be taken into consideration. Also other state-of-the-art machine learning techniques like neural network, deep learning can also be implemented.

Acknowledgment. The data required for this work has been provided by Geramari MPHC, Gauripur, Dhubri, Assam, India and Kazigaon SD, Kokrajhar, Assam, India. Authors would like to offer heartfelt thank to the authorities of both the organizations.

References

1. Ahmadi, P., et al.: Prediction of low birth weight using random forest: a comparison with logistic regression. J. Paramedical Sci. (JPS) **8**, 36–43 (2017)
2. Al-Shawwa, M., Abu-Naser, S.S.: Predicting birth weight using artificial neural network. Int. J. Acad. Health Med. Res. (IJAHMR) **3**(1), 9–14 (2019)
3. Ariyadasa, S.N., Munasinghe, L.K., Senanayake, S.H.D., Fernando, M.: Knowledge extraction to mitigate child malnutrition in developing countries (Sri Lankan context). In: 2013 4th International Conference on Intelligent Systems, Modelling and Simulation, pp. 321–326, January 2013. https://doi.org/10.1109/ISMS.2013.23
4. Dalal, A.: DC Dutta's textbook of gynaecology and textbook of obstetrics. J. Obstet. Gynecol. India **66**, 303–304 (2016). https://doi.org/10.1007/s13224-016-0848-4
5. Desiani, A., Primartha, R., Arhami, M., Orsalan, O.: Naive Bayes classifier for infant weight prediction of hypertension mother. J. Phys. Conf. Ser. **1282**, 012005 (2019). https://doi.org/10.1088/1742-6596/1282/1/012005
6. Duraisamy, T.D., Sudha, P.: Identification of malnutrition with use of supervised datamining techniques-decision trees and artificial neural networks. Int. J. Eng. Comput. Sci. **3**, 2319–7242 (2014)
7. Faruk, A., Cahyono, E., Eliyati, N., Arifieni, I.: Prediction and classification of low birth weight data using machine learning techniques. Indonesian J. Sci. Technol. **3**, 18–28 (2018). https://doi.org/10.17509/ijost.v3i1.10799
8. Haughton, P., Drysdale, C.: Twin presence of obesity and under-nutrition reflects shifts in food systems (2019). https://www.who.int/news-room/detail/16-12-2019-more-than-one-in-three-low-and-middle-income-countries-face-both-extremes-of-malnutrition. Accessed 20 Jan 2020
9. Khare, S., Kavyashree, S., Gupta, D., Jyotishi, A.: Investigation of nutritional status of children based on machine learning techniques using Indian demographic and health survey data. Procedia Comput. Sci. **115**, 338–349 (2017). https://doi.org/10.1016/j.procs.2017.09.087
10. Mohammed, S., Asfaw, Z.: Bayesian Gaussian regression analysis of malnutrition for children under five years of age in Ethiopia, emdhs 2014. Arch. Public Health **76**, 21 (2018). https://doi.org/10.1186/s13690-018-0264-6
11. Senthilkumar, D., Paulraj, S.: Prediction of low birth weight infants and its risk factors using data mining techniques. In: Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management Dubai, United Arab Emirates (UAE), March 2015
12. Sudha, P.: An efficient identification of malnutrition with unsupervised classification using logical decision tree algorithm. Int. J. Innov. Res. Comput. Commun. Eng. **4**, 1311–1315 (2016)

13. Swaminathan, S., et al.: The burden of child and maternal malnutrition and trends in its indicators in the states of India: the global burden of disease study 1990–2017. Lancet Child Adolesc. Health **3**, 855–870 (2019). https://doi.org/10.1016/S2352-4642(19)30273-1
14. Watcharapasorn, P., Kurubanjerdjit, N.: The surgical patient mortality rate prediction by machine learning algorithms. In: 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–5, July 2016. https://doi.org/10.1109/JCSSE.2016.7748844
15. WHO: 2018 global nutrition report: Shining a light to spur action on nutrition (2019). https://www.unicef.org/rosa/press-releases/2018-global-nutrition-report-reveals-malnutrition-unacceptably-high-and-affects. Accessed 26 Dec 2019
16. Xu, D., Ganegoda, U.: Rule based classification to detect malnutrition in children. Int. J. Comput. Sci. Eng. **3**, 423–429 (2011)
17. Yarlapati, A.R., Roy Dey, S., Saha, S.: Early prediction of LBW cases via minimum error rate classifier: a statistical machine learning approach. In: 2017 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 1–6 (May 2017). https://doi.org/10.1109/SMARTCOMP.2017.7947002