A Review in Prediction of Malnutrition Status Using Data Mining Techniques

S.Dhivya[1*], T. A. Sangeetha[2]

[1]Assistant professor in Computer Applications Kongu Arts and Science College (Autonomous),Erode, Tamilnadu, India.

[2]Head and Associate professor in Computer Applications Kongu Arts and Science College (Autonomous), Erode, Tamilnadu, India.

[*]**Corresponding author:** S.Dhivya, [1]Assistant professor in Computer Applications Kongu Arts and Science College (Autonomous), Erode, Tamilnadu, India , Tel: 9500506005; E-mail: s.dhivya87@gmail.com

Abstract

Child nutritional deficiency comes under various facts such as malnutrition, low birth weight, infant and young child feeding, iodine, and Vitamin deficiency. In recent years malnutrition is a widespread problem in global level. This research produce about the review of data mining techniques is used to predict malnutrition status of young children. The root cause of child malnutrition varies across the regions in every country because of various impacts like life style, food intake and environmental changes, maternal care and also motherhood care. The most the research, the results predict the malnutrition status using some anthropometric parameters of preschool age from 5 to under10 age of children and clinical sign parameters are considered to predict the best accuracy. Recently data mining uses the proposed method of clustering, classification techniques, Regression and machine learning algorithms to predict the malnutrition status with highest accuracy using anthropometric parameters (height, weight and age) for stunted (Low height-for-age) and wasted (Low weight-for-height),underweight (Low weight- for-age) and clinical sign attributes to predict the statuses of malnutrition with and without transformed attributes.

Keywords: Child Malnutrition; Data mining; Clustering; Classification; Anthropometric Parameters; Wasted; Stunting; Underweight.
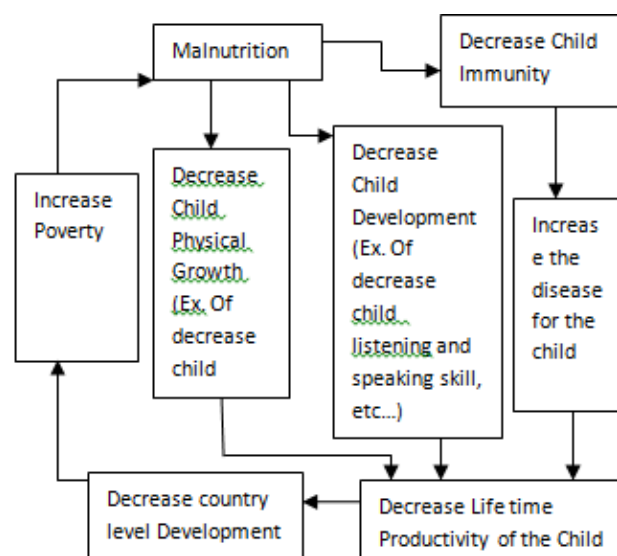
Introduction

Data Mining in Healthcare

Big data analytics is the major impact in recent research sector. In healthcare platform Big data act as a major role in processing the unstructured data. In medical discipline, immunization is the biggest challenges for all living humans. Immunization system varies based on human ages [child, adult, aged people].In [6] to analyze the child immunization system with the help of vaccine provided by the various government and also address the effect of vaccine whether it is useful to improve immunity level or not . From this results its help to choose the successful choice of vaccine to improve the achievement of health organization. In the field of medicine the data are stored in a database in unstructured format. In this case

the proposed method of data mining is useful to predict the results in structured format using feature mining method. In many situation drugs given to the pediatric patients based on results of the impact of medicine. In this method [7] using Pediatric DB to analyze the data to leads the right choice of medicine at right time.

The main advantages of data mining in medical field are to extract the valuable information and to predict the results to take right choice of decision in critical situation. Exist data mining techniques like classification, clustering and regression is used to handle large number of data sets that are available in medical field. The main aim of this research is to detect the malnutrition status and impact of the children to improve the awareness about malnutrition among the people and improve the child health status among the nation. Based on this prediction the child health is analyzed whenever needed and take the correct choice of decision at right time to improve child health

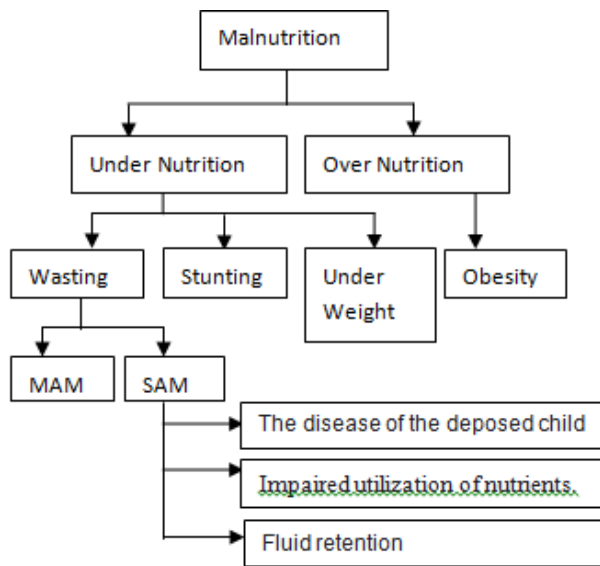Figure1: Impact of Malnutrition

Classification of Malnutrition

Malnutrition can broadly classify into two categories such as Under nutrition and Over nutrition. Under nutrition mainly classify into three types namely wasting (low weight for height),

Stunting (low height for age), Under weight (low weight for age) and obesity can causes because of over nutrition problem. Wasting can comes under two types MAM (moderate acute malnutrition) and SAM(severe acute malnutrition).SAM can causes in three forms such as The disease of the deposed child, Impaired utilization of nutrients and Fluid retention. The disease of the deposed child malnutrition mainly suffer 1 to 5 year age child due to insufficient intake of proteins and carbohydrate. Impaired utilization of nutrients malnutrition can occur under one year of age children due to deficiency of calories and Fluid retention malnutrition happen due to lack of adequate proteins to maintain water distribution between blood and tissues.

Figure2: Classification of Malnutrition

Factors of Malnutrition

UNICEF said, the main causes of child malnutrition can be categorized into three main factors such as household food insecurity, inadequate care and unhealthy household environment, and lack of health care services.

Malnutrition factors are not same in all countries. Malnutrition is the foremost causes of child mortality among all developing countries including India. The study of [13] child nutritional status was using the available features in the Indian Demographic and Health Survey (IDHS) dataset to design the prediction model for malnutrition using machine learning approach. From this study the newly identified features can be used to identify the main influence of malnutrition.

Household Food Insecurity

Much research analysis said Food insecurity is a complex problem for malnutrition because most of the people do not have sufficient resources to meet their basic needs which lead the food insecurity in family. Some families has low income , lack of reasonably priced housing and suffered by high medical cost these factors causes some issues like chronic and acute malnutrition problems.

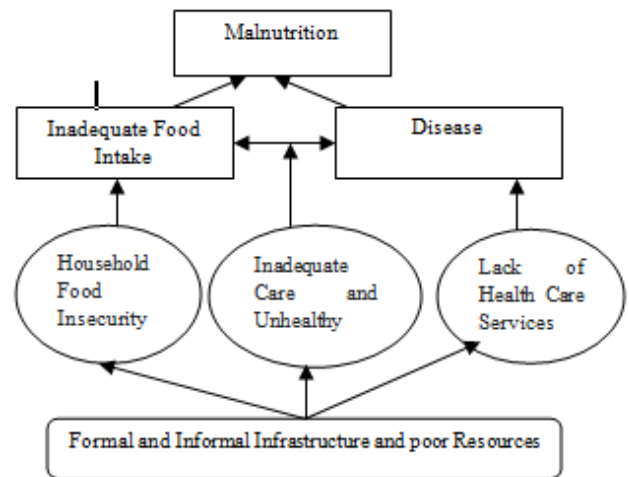Inadequate Care and Unhealthy Household Environment

Due to the low income of the family provide insufficient food supply for their child causes under nutrition problem and not proper breastfeeding create several health issues for the child. Unhealthy household environment such as lack of pure water consumption, poor sanitation facility, lack of housing facility and improper waste disposal or drainage causes several health issues such as fever, cough, Vitamin deficiency, etc.

Lack of Health Care Services

Every nation provides several health care services to improve the nutritional status of the child among the nation. Some time health care services not reached properly to the poor economic families due to improper communication and un-education. Government provide the health care services based on their people environmental status because the environmental status not same in all places across the country. Based on their needs to provide the necessary health care services to the needy peoples then improve the child welfare and their economical status.

To improve the child nutritional status follows the following three main initiatives such as 1. Improve the Women's nutrition before and after pregnancy and improve the quality of breast milk .2. To ensure children get vitamins and minerals for their needs. 3. To promote and support breastfeeding for the first 6 months of child's.

Figure3: The UNICEF Framework for Malnutrition

Related Works

Research has been carried out on different aspects of child health using various datasets among various countries in the world. Various data mining techniques are used in different child health related datasets to predict the expected results. In 2017 The Indian –DHS (Demographic Health Survey) Data set was used by Nair Akash Anilkumar, Deepa Gupta, Sangita Khare, Deepika Manippady Gopalkrishna, Amalendu Jyotishi [1] to predict the characteristics and causes of malnutrition across Indian States. The main concern was used four explanatory variables BMI (Body Mass Index), WAZ (Weight for Age), HAZ (Height for Age), and WHZ (Weight for Height) and divided the data into various clusters used k-means clustering for each of the variables. Based on cluster similarity values formed a cluster of

states to found the top ranking variables which are the most reasonable factor for malnutrition in each cluster. Here found the variables which are having similar ranking and which are having different ranking among the cluster. Those variables are finally considered in cluster policies to take a correct choice of decision to improve the child health.

The research by md Mehrab Shahriar, Mirza Shaheen Iqubal, samerat Mitra,Amit Kumar Das[2] to predict the malnutrition status of the children used deep learning approach. It has been used Demographic and Health Survey of children data. In this study used three anthropometric parameters namely wasting, underweight and stunting. For this study used Artificial Neural Network (ANN) to predict the best accuracy with anthropometric parameters such as HAZ, WAZ, WHZ. In feature extraction 16 most significant features were extracted and used python – numpy library for data preprocessing.This study used 'Tensorflow' and 'Keras' to build the framework model. In this approach used ANN and four machine learning algorithms namely SVM (Support Vector Machine) Classifier, Decision tree classifier, Navie bayes classifier and Random Forest classifier to predict the best accuracy and finally shows the result that the deep learning mechanism is a powerful technique to predict the best accuracy with the method of multiclass- multi label classification.

Data mining Based prediction of malnutrition [3] to predict the malnutrition status of the children less than five years used various data mining algorithms such as random Forest, PART rule induction, Navie Bayes classification and Logistic Regression statistical method. This study takes anthropometric parameters (height, Weight, Age, BMI) and clinical sign attributes to predict the best accuracy. Most of the research done by using statistical models but it does not give the best prediction. In this study used predictive models using data mining techniques to predict the malnutrition status. For this study clinical sign attribute grade (+) oedema considered as severe acute malnutrition to predict the malnutrition status using data mining classification algorithms. In data mining techniques knowledge discovery process of Domain understanding, Data understanding, Data preparation, Data mining algorithms and evaluation process these are consists to find out the best results. From this study calculated anthropometric indices Z – score value of weight-for-age- Z score(WAZ), height-for-age-Z score(HAZ), Weight-for-height- Z score(WHZ0, has calculated for each children using the equation 1.

Z – score = H-M /σ

[H- Indicate height, m – refers median of WHO's reference population, σ- standard deviation of WHO's reference population]. Here equation 1 is only applicable for normal distribution, the weight has non –normal distribution, in this case weight-for-age and weight-for-height Z score calculated by equation2

Z- score = (W-M)L – 1/ L*S

[W- indicates weight, S- coefficient of variation, L- skewness value]

In this study final dataset attributes such as (sex, age, weight, height, MUAC(mid-upper arm circumference),province, age group) and transformed attributes(HAZ, WAZ and WHZ) were used to predict the malnutrition status. For this study applied synthetic minority oversampling technique (SMOTE) to create balanced data classes. Final result from this, four separate models have been created for each status using RP,PART, NB and LR and finally found that the Random Forest and PART rule induction classifiers performed well with the result of highest accuracy.
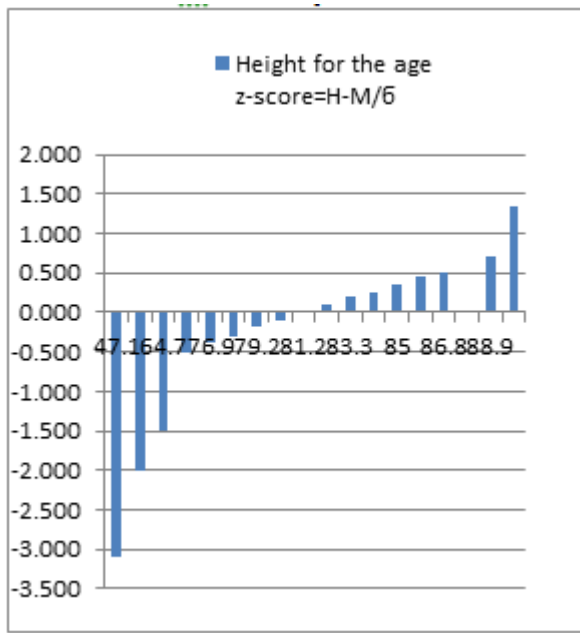
AGE	WEIGHT	HEIGHT	Height for the age	Weight for the age
			z-score=x-median/ϭ	(W-M)L – 1/ L*S
3 mth	2.6	47.1	-3.1	-0.77
6 mth	5.3	59.1	-2.009	-0.36
9 mth	6.7	64.7	-1.5	-0.2
12 mth	9.6	75.7	-0.5	0
13 mth	9.9	76.9	-0.391	0.01
14 mth	10.1	77.9	-0.3	0.02
15 mth	10.3	79.2	-0.182	0.02
16 mth	10.5	80.2	-0.091	0.02
17 mth	10.7	81.2	0	0.02
18 mth	10.9	82.2	0.091	0.03
19 mth	11.2	83.3	0.191	0.04
20 mth	11.3	84	0.255	0.04
21 mth	11.5	85	0.345	0.05
22 mth	11.7	86.1	0.445	0.06
23 mth	11.9	86.8	0.509	0.07
24 mth	10.1	81.6	0.036	0.02
36 mth	11.8	88.9	0.7	0.06
48 mth	13.5	96	1.345	0.18

Table 1: Z-Score (Height-for-Age and Weight-for-Age) for Sample Male Child data under below five age

Height-for-Age Z-score Classification:

-1 < HAZ < 0 : Normal(Well Nourished)

-2 < HAZ < -1 : Marginally Stunted

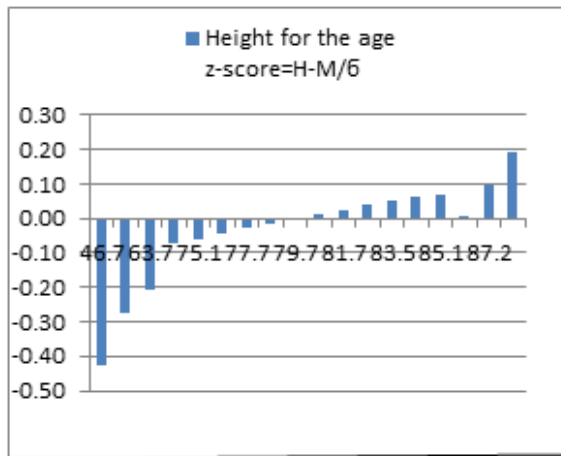-2 < HAZ < -3 : Moderately Stunted

HAZ< -3 : Severely Stunted

Figure 4: Graphical form of Height-for-Age Z-score for Male under five age

Weight-for-Age Z-score Classification:

-1 < HAZ < 0 : Normal(Well Nourished)

-2 < HAZ < -1 : Marginally Underweight

-2 < HAZ < -3 : Moderately Underweight

HAZ< -3 : Severely Underweight

Figure 5: Graphical form of Weight-for-Age Z-score for Male under five age

AGE	WEIGHT	HEIGHT	Height for the age	Weight for the age
			z-score=H-M/б	(W-M)L − 1/ L*S
3 mth	2.6	46.7	-0.43	-0.59
6 mth	5	58.4	-0.27	-0.31
9 mth	6.2	63.7	-0.21	-0.2
12 mth	9.2	74.1	-0.07	0
13 mth	9.5	75.1	-0.06	0.01
14 mth	9.7	76.4	-0.04	0.02

15 mth	9.9	77.7	-0.03	0.02
16 mth	10.2	78.4	-0.02	0.02
17 mth	10.4	79.7	0	0.03
18 mth	10.6	80.7	0.01	0.03
19 mth	10.8	81.7	0.03	0.03
20 mth	11	82.8	0.04	0.04
21 mth	11.3	83.5	0.05	0.05
22 mth	11.5	84.8	0.07	0.06
23 mth	11.7	85.1	0.07	0.07
24 mth	9.6	80.1	0.01	0.01
36 mth	11.2	87.2	0.1	0.05
48 mth	12.9	94.5	0.19	0.14

Table 2: Z-Score (Height-for-Age and Weight-for-Age) for Sample Female Child data under below five age.

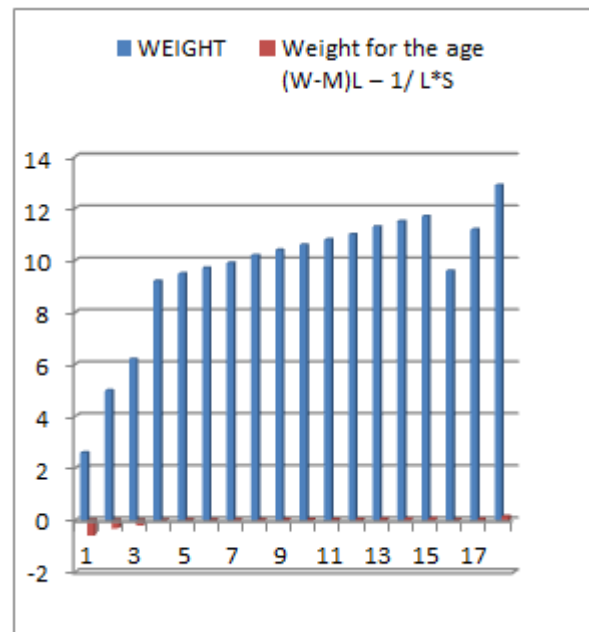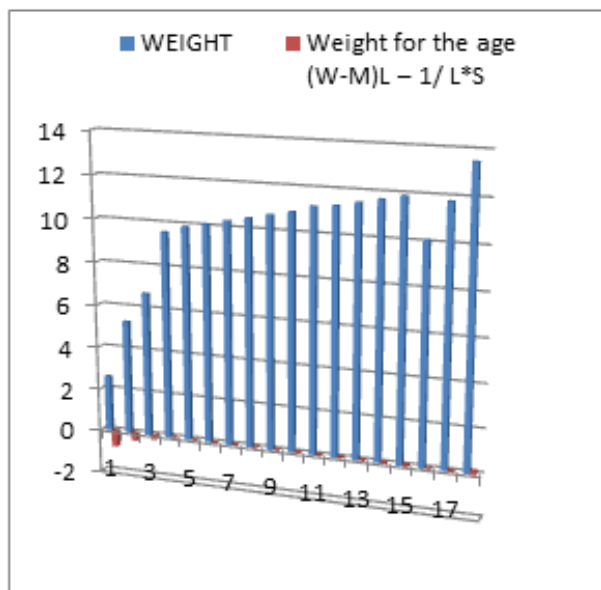**Figure 6:** Graphical form of Height-for-Age Z-score for Female under five age

Figure 7: Graphical form of Weight-for-Age Z-score for Female under five age

In [4], author analyzes malnutrition status of the toddler using k-means clustering algorithm. This research used 6 months to 72 months old infants data. Here using the parameters namely height, weight and age of the toddlers. Here k-means clustering algorithm split the nutrition data into five clusters namely good nutrition, moderate nutrition, malnutrition, more nutrition and obesity. It can be used to identify nutrition status of toddlers easily.

Nutrition is the most important factor for the child growth and normal functions of organs. The prevalence of malnutrition remained increased in all developing countries. The aim of this study [5] is to understand the spatial heterogeneity and meso-scale correlates of malnutrition across Indian districts. From this study first, explored the social and demographic correlates of malnutrition. Second, it's taken the survey under geographical location. Third, district level status of malnutrition has calculated. This study used National Family Health Survey round four data for analyzing the results among district level in India. Three anthropometric parameters of nutritional status of children namely stunted (Low height-for-age) and wasted (Low weight-for-height), underweight (Low weight- for-age) were used in this study. In this study used descriptive statistics models and Regression models were used to understand the status of malnutrition and its correlates. Final result of this study that the spatial clustering of malnutrition is found in this geographical data where poverty is high, BMI level among women is below normal, Women's education is low. From these findings, in order to increase sanitation, reduction of poverty and increase maternal nutrition can help to reduce child malnutrition in India.

Table3: Findings and Methodology of Related works

Ref.No.	Data mining Technique	Algorithm	Attributes / parameters	Findings
1	Data Mining Clustering	K- means Clustering	Anthropometric parameters such as height,Weig	Finding top ranking of variables which are the most
			ht,Age & BMI	reason for malnutrition
2	Data mining Classification and ANN Approach	Deep Learning, SVM, Decision Tree Classifier, Naive Bayes Classifier, Random Forest Classifier	Anthropometric parameters(WHZ, WAZ, HAZ)	Finding the result of Deep learning method produced best accuracy in prediction of malnutrition
3	Data mining classification	Random Forest, PART rule Induction, Naïve Bayes and Logistic Regression	Anthropometric(height, Weight,Age) and clinical sign rsparamete(oedema)	Random Forest and PART rule Induction performed well with the result of highest accuracy
4	Data mining clustering	K- means clustering	Anthropometric parameters such as height,Weight and Age	Nutrition data can be divided into 5 clusters namely good, moderate, more nutrition, malnutrition and obesity to find the nutrition status of the toddlers easily.
5	Statistic Model and Regression Model	Regression and Clustering	Anthropometric Parameters	Find Women low BMI level, High Poverty and Women's low education are the most correlates of child malnutrition in India

Vi. Data mining application in healthcare

Determining Viability of Treatments

Data mining Techniques in healthcare play a big role to analyze very huge dataset for predict the exact result. It is used to compare the symptoms, causes of disease and reasonable factors to give the best treatment option for a given patient conditions and their illness. In data mining techniques using various predictive or descriptive models to find the best result by comparing patients under different treatment protocols and analyze the result to provide the best treatment option which one is more effective to recover soon from illness. For this case data mining predict and provide standard treatment protocols for certain diseases.

Detecting and Eliminating Fraud

In healthcare sector, data mining focus on identifying insurance fraud and wrong medical claims. In this case data

mining establish normal patterns, then identifying unusual pattern in clinical medical claims, labs and others. The data mining techniques used in mobile healthcare services [11] provide decision support system to assess and predict the fraud in medical claims by approval officer.

Evaluate Treatment Effectiveness

In data mining various predictive methodologies can be used to find the best accuracy of result in various healthcare sectors. Using some Clustering, Classification and Regression methodologies and compare the results to find the best accuracy finally. In healthcare sector data gathered from same resemblances of symptoms and causes of diseases from more patients and compare the results using some prediction analyzes methods and identifies which are the most reasonable factors for their illness and based on this result, it's suggest to give the best treatment options effectively with lowest cost. In[10] data mining techniques collected large number of patients healthcare data in the process of diagnosis and treatment of intensive care unit and analyze which are the methods provides more benefits for patients safety and quality.

Save Lives of Patients using Predictive Medicine

Data mining techniques using statistical analyzes methods and predictive methods to find the best results. Compare to statistical analyze the data mining mostly using predictive methods in healthcare data analyzes to get the best accuracy of result. Prediction of disease in starting stage is very effective to save the patient life and minimize the treatment cost. In data mining using various classification algorithms [8] for disease prediction and widely used data mining technique of Decision Tree Algorithm [9] which is used to disease prediction with best accuracy of results.

Manage Customer Relationship

Now a day's customer relationship management is very essential in healthcare marketing.

In many applications healthcare providers play the role of promote disease education, wellness services and prevention measurement, in this case data mining techniques analyze the data and give the best prediction results to the healthcare providers, then the customer can easily get the best services from the healthcare providers. In [12], propose the machine learning framework to formulate patient satisfaction problem to identify the strong factors.

Recent days Pharmaceutical companies also get a benefit from data mining healthcare services. In global, large number of pharmaceutical companies provides the services to the customers. By manually the pharmaceutical companies are very difficult to track the information about which physicians prescribe which drugs and for what purpose and also difficult to find customer responses and physicians feedback about the drugs. In this case apply data mining techniques to huge number of genomic data and easily track which physicians prescribe which drugs for what purposes and pharmaceutical companies can easily decide whom to target for each particular medicine and also determine patients response to drug therapy and data

mining techniques shows to the healthcare providers which drugs give the more effective results.

Proposed Methodology

Most of the research for predicts the malnutrition status using various clustering, classification methodologies to find the result best accuracy. Most research using anthropometric parameters for analyzing result. This research in future will use the anthropometric parameters, clinical sign attributes and factors of malnutrition to find the malnutrition status and find major factors of malnutrition. In Data mining process first select the dataset and then choose the parameters to analyze the result. Next processes is Feature Extraction and transforms the data after implements the algorithm and then compare the result then finally choose the best accuracy of the result.
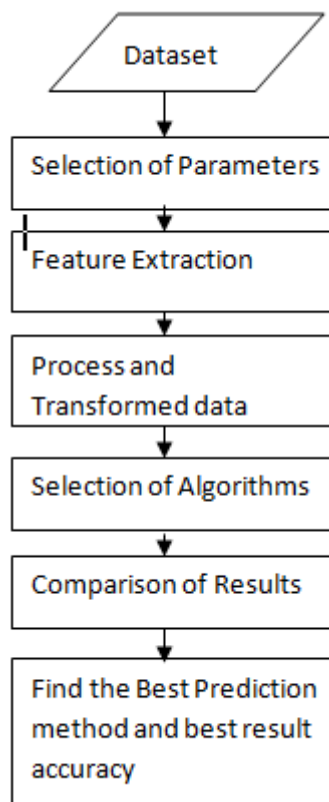

Figure8: Flow Graph of Proposed Method Execution

Conclusion

Child malnutrition is the biggest challenges across the country. Malnutrition causes due to various factors such as Poverty, Unhealthy Household Environment, Insufficient health care services, etc. Prediction of child malnutrition status and factors of malnutrition using various data mining algorithms such as clustering, classification and Regression with various parameters like anthropometric parameters (Height, Weight, and Age), Clinical sign attributes .Using these algorithms and compare these results and finally choose the best prediction model. The prediction of results can be used to the policy makers and healthcare providers to make the correct decision at right time.

References

1. Nair Akash Anilkumar, Deepa Gupta, Sangita Khare, Deepika Manippady Gopalkrishna, Amalendu Jyotishi, "Characteristics and Causes of Malnutrition across Indian States: A Cluster Analysis Based on Indian Demographic and Health Survey Data",IEEE, 2017.

2. .Md Mehrab Shahriar, Mirza Shaheen Iqubal, Samrat Mitra, Amit Kumar Das," A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh",IEEE, 2019.

3. Ziaullah Momand, Pornchai Mongkolnam, Pichai Kositpanthavong, Jonathan H. Chan," Data Mining Based Prediction of Malnutrition in Afghan Children", IEEE, 2020.

4. Sri Winiarti, Herman Yuliansyah, Aprial Andi Purnama, "Identification of Toddlers' Nutritional Status using Data Mining Approach", International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2018.

5. Junaid Khan ,Sanjay K Mohanty," Spatial heterogeneity and correlates of child malnutrition in districts of India", Khan and Mohanty BMC Public Health (2018) 18:1027- Open Access.

6. Nikhita Siringi, Shilpi Sharma," Child Immunization Using Data Analysis",IEEE, 2017.

7. Shantanu Deshmukh, Natalia Khuri," PediatricDB: Data Analytics Platform for Pediatric Healthcare",IEEE, 2018.

8. Emrana Kabir Hashi, Md. Shahid Uz Zaman and Md. Rokibul Hasan," An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques",IEEE, 2017.

9. M. Deepika, Dr. K. Kalaiselvi," A Empirical study on Disease Diagnosis using Data Mining Techniques",IEEE, 2018.

10. YONGCHANG GAO, CHENFEI SUN, RUICAN LI, QINGZHONG LI, LIZHEN CUI, BIN GONG," An Efficient Fraud Identification Method Combining Manifold Learning and Outliers Detection in Mobile Healthcare Services", IEEE, 2017.

11. Ning Liu, Soundar Kumara, Eric Reich, "Gaining Insights Into Patient Satisfaction Through Interpretable Machine Learning", IEEE Journal of Biomedical and Health Informatics, Volume: 25, Issue: 6, June 202

12. Sangita Khare , S Kavyashree , Deepa Gupta , Amalendu Jyotishi," Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data", Published by Elsevier B.V,2017