REVIEW ARTICLE e-ISSN: 2349-0659 p-ISSN; 2350-0964

A Review in Prediction of Malnutrition Status Using Data Mining Techniques

S. Dhivya*, T. A. Sangeetha

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 at a global level. This research produces a review of data mining techniques is used to predict the malnutrition status of young children. The root cause of child malnutrition varies across the regions in every country because of various impacts such as lifestyle, food intake, environmental changes, maternal care, and also motherhood care. Most of 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 the 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: Anthropometric parameters, Child malnutrition, Classification, Clustering, Data mining, Stunting, Underweight, Wasted

Asian Pac. J. Health Sci., (2022); DOI: 10.21276/apjhs.2022.9.1.41

Introduction

Data Mining in Healthcare

Big data analytics is a major impact in the recent research sector. In healthcare platforms, Big data act as a major role in processing unstructured data. In the medical discipline, immunization is the biggest challenge for all living humans. The Immunization system varies based on human ages (child, adult, aged people). In[1] to analyze the child immunization system with the help of vaccines provided by the various government and also address the effect of vaccines whether it is useful to improve immunity level or not. Results from this, it is helped to choose the successful choice of vaccines to improve the achievement of health organizations. In the field of medicine, the data are stored in a database in an unstructured format. In this case, the proposed method of data mining is useful to predict the results in a structured format using the feature mining method. In many situations, drugs are given to pediatric patients based on the results of the impact of medicine. This method[2] using Pediatric DB to analyze the data to leads to the right choice of medicine at right time.

The main advantages of data mining in the medical field are to extract valuable information and to predict the results to take the make the right choice of decision in a critical situation. Exist data mining techniques such as classification, clustering, and regression are used to a handle large number of data sets that are available in the 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 the child health

Impact of Malnutrition among Nation

Malnutrition factors are not the same in all countries. Malnutrition is the foremost cause of child mortality among all developing countries including India. The study of[3] child nutritional status was using the available features in the Indian Demographic and Health Survey dataset to design the prediction model for malnutrition using a machine learning approach. From this study, the newly identified features can be used to identify the main influence of malnutrition. Figure 1 represents cycles of malnutrition impacts that the child have malnutrition means that affect the entire functionality of their life.

Classification of Malnutrition

Figure 2 states that the classification of malnutrition status. Malnutrition can broadly classify into two categories undernutrition and overnutrition. Under nutrition mainly classifies into three types namely wasting (low weight-for-height [WHZ]), Stunting (low height-for-age [HAZ]), Underweight (low weight-for-age [WAZ]), and obesity can be caused because of over-nutrition problems. Wasting can come under two types moderate acute malnutrition and severe acute malnutrition (SAM). SAM can cause in three forms as the disease of the deposed child, impaired utilization of nutrients, and Fluid retention. The disease of the deposed child malnutrition mainly suffers 1–5 year age child due to insufficient intake of proteins and carbohydrates. Impaired utilization of nutrients malnutrition can occur under 1 year of age children due

Department of Computer Applications, Kongu Arts and Science College (Autonomous), Erode, Tamil Nadu, India.

Corresponding Author: S. Dhivya, Department of Computer Applications, Kongu Arts and Science College (Autonomous), Erode, Tamil Nadu, India. E-mail: s.dhivya87@gmail.com

How to cite this article: Dhivya S, Sangeetha TA. A Review in Prediction of Malnutrition Status Using Data Mining Techniques. Asian Pac. J. Health Sci., 2022;9(1):164-170.

Source of support: Nil

Conflicts of interest: None.

Received: 01/11/2021 **Revised:** 20/11/2021 **Accepted:** 05/12/2021

to deficiency of calories and Fluid retention malnutrition happen due to lack of adequate proteins to maintain water distribution between blood and tissues.

Factors of Malnutrition

In Figure 3 UNICEF represent the framework for malnutrition status. 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 the same in all countries. Malnutrition is the foremost cause of child mortality among all developing countries including India. The study of[3] child nutritional status was using the available features in the Indian Demographic and Health Survey dataset to design the prediction model for malnutrition using a 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 people do not have sufficient resources to meet their basic need which leads the food insecurity in the family. Some families have low income, lack reasonably priced housing, and suffered from high medical costs these factors causes some issues such as chronic and acute malnutrition problems.

Inadequate Care and Unhealthy Household Environment

Due to the low income of the family provide an insufficient food supply for their child causes undernutrition problems and not proper breastfeeding to 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, and Vitamin deficiency.

Lack of Health Care Services

Every nation provides several health care services to improve the nutritional status of the child among the nation. Sometime health care services are not reached properly to the poor economic families due to improper communication and un-education. The Government provides health care services based on their people's environmental status because the environmental status is not the 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's 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 the child's.

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 Demographic Health Survey (DHS) Data set was used by Anilkumar *et al*.[4] to predict the characteristics and causes of malnutrition across the Indian States. The main concern was to use four explanatory variables Body Mass Index (BMI), WAZ, HAZ, and WHZ, 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 rankings and which are having different rankings among the cluster. Those variables are finally considered in cluster policies to take a correct choice of decision to improve the child's health.

The research by Shahriar *et al*.[5] to predict the malnutrition status of the children used a deep learning approach. It has been used Demographic and Health Survey of children data. 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, and 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. This approach used ANN and four machine learning algorithms namely Support Vector Machine Classifier, Decision tree classifier, Naïve 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[6] to predict the malnutrition status of children <5 years used various data mining algorithms such as Random Forest, PART rule induction, Naive 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 is done using statistical models, but it does not give the best prediction. This study used predictive models using data mining techniques to predict the malnutrition status. For this study, clinical sign attribute grade (+) edema is 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 WAZ score, HAZ score, WHZ0 score (has calculated for each child using Equation 1.

$$Z\text{-score} = H\text{-}M/\sigma \quad (1)$$

(H: Indicate height, m: Refers median of the 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, WAZ and WHZ Z score calculated by equation 2.

$$Z\text{-Score} = (W\text{-}M)^{L}\text{–}1/L*S \quad (2)$$

(W: Indicates weight, S: Coefficient of variation, L: Skewness value)

Table 1 represents Findings and Methodology of related works. In this study, final dataset attributes such as (sex, age, weight, height, mid-upper arm circumference, province, age

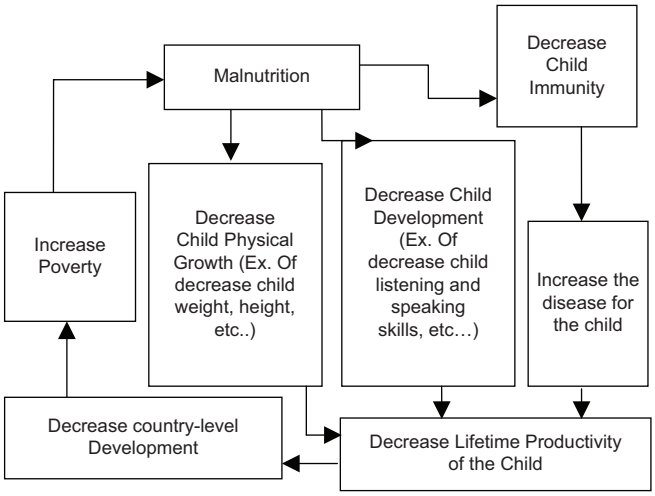

Figure 1: Impact of malnutrition

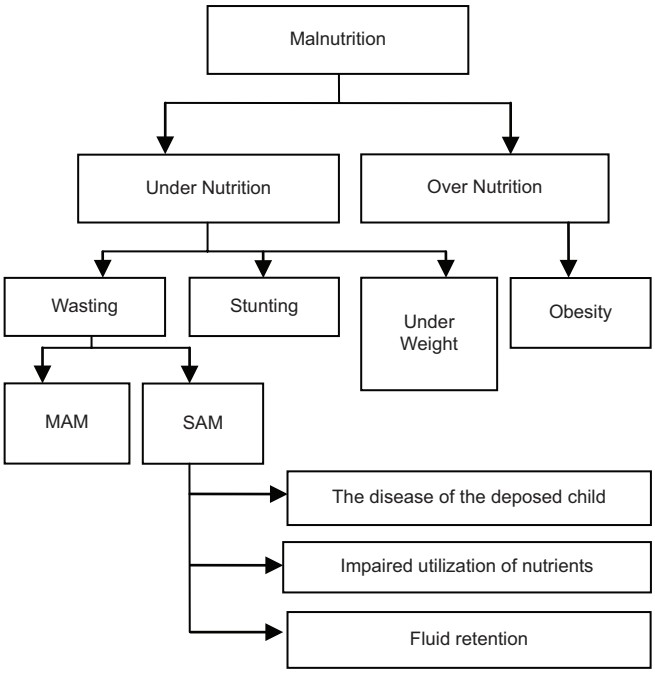

Figure 2: Classification of malnutrition

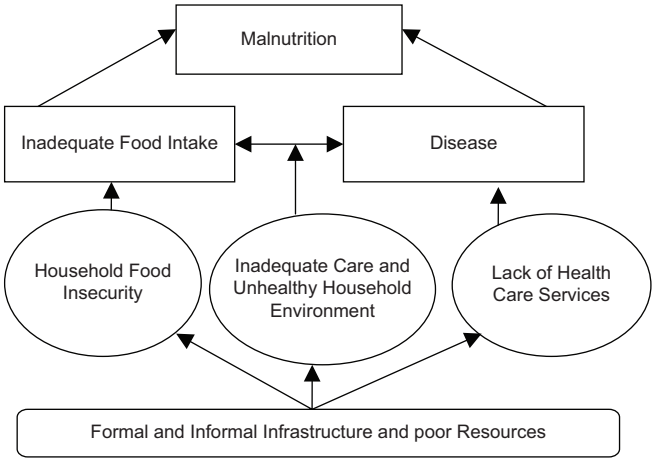

Figure 3: The UNICEF Framework for malnutrition

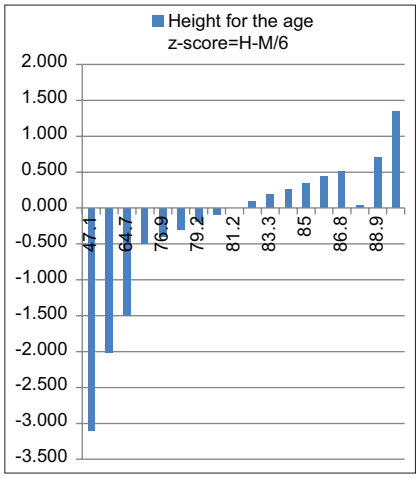

Figure 4: Graphical form of Height-for-Age Z-score for Male under-five age

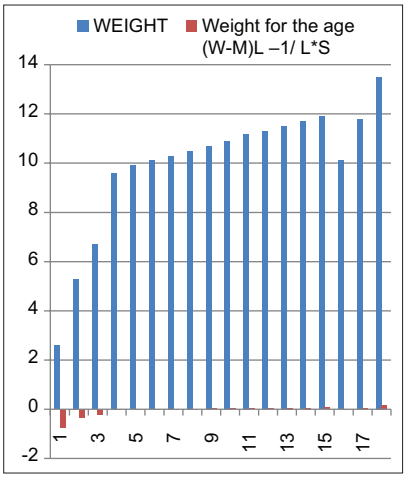

Figure 5: Graphical form of Weight-for-Age Z-score for Male under-five age

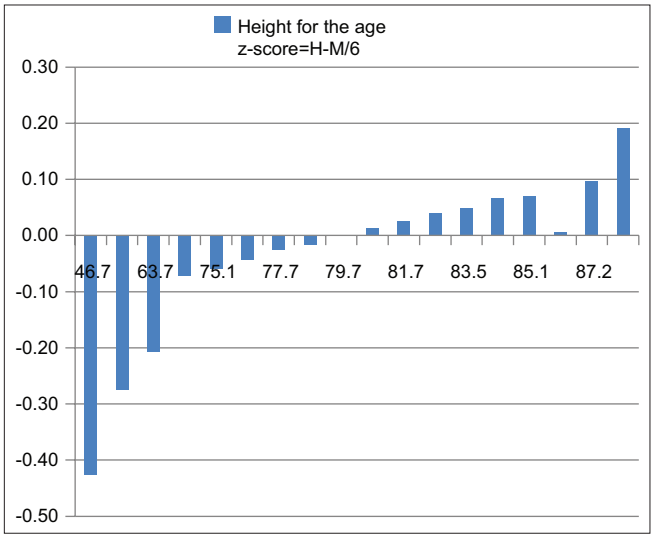

Figure 6: Graphical form of Height-for-Age Z-score for Females under-five age

Table 1: 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, Weight, Age and BMI	Finding top ranking of variables which are the most 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 the Deep learning method produced the best accuracy in the prediction of malnutrition
3	Data mining classification	Random Forest, PART rule Induction, Naïve Bayes and Logistic Regression	Anthropometric (height, Weight, age) and clinical sign parameters (edema)	Random Forest and PART rule Induction performed well with the result of the 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

SVM: Support vector machine, BMI: Body mass index, ANN: Artificial neural network, WHZ: Weight-for-height, HAZ: Height-for-age, WAZ: Weight-for-age

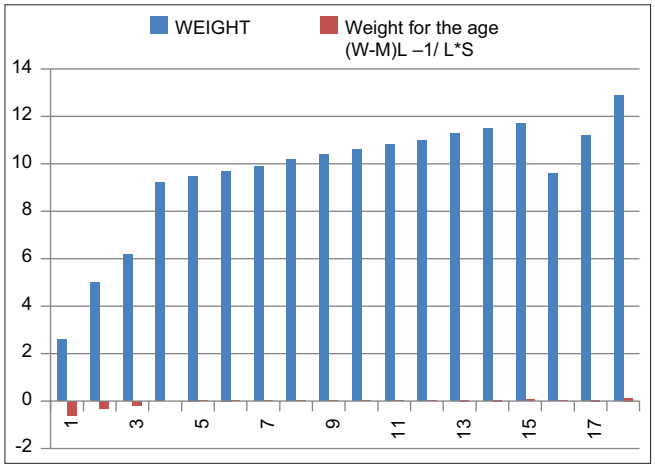

Figure 7: Graphical form of Weight-for-Age Z-score for Females under five age

group), and transformed attributes (HAZ, WAZ and WHZ) were used to predict the malnutrition status. For this study applied the synthetic minority oversampling technique to create balanced data classes. The 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 the highest accuracy.

In this study, final dataset attributes such as (sex, age, weight, height, 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 the synthetic minority oversampling technique to create balanced data classes. The 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 the highest accuracy.

In this survey has taken the account of some child DHS data to calculate z-score for height-for- age and WAZ using the formula of (1) and (2)

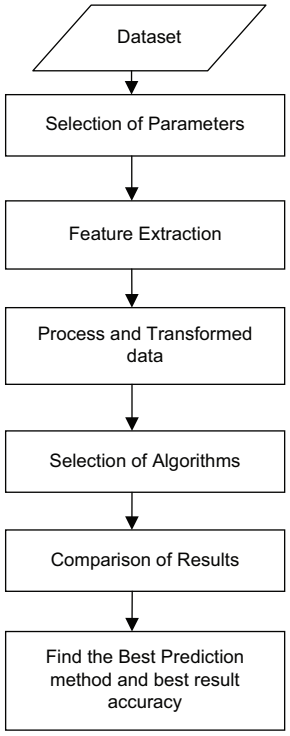

Figure 8: Flow graph of proposed method execution

Figure 4 represent Graphical form of Height-for-Age Z-score and Figure 5 represent Graphical form of Weight-for-Age Z-score for Male under- five age. It can be drawn based on the data of table (2).

HAZ Z-score Classification:

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

$-2 < HAZ < -1$: Marginally Stunted

$-2 < HAZ < -3$: Moderately Stunted

$HAZ < -3$: Severely Stunted

WAZ Z-score Classification:

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

$-2 < HAZ < -1$: Marginally Underweight

$-2 < HAZ < -3$: Moderately Underweight

Table 2: Z-Score (Height-for-Age and Weight-for-Age) for sample male child data under below five age

Age	*Weight*	*Height*	*Height for the age z-score=x−median/σ*	*Weight for the age ([W-M]^{L}−1/L*S)*
3 mth	2.6	47.1	−3.100	−0.77
6 mth	5.3	59.1	−2.009	−0.36
9 mth	6.7	64.7	−1.500	−0.20
12 mth	9.6	75.7	−0.500	0.00
13 mth	9.9	76.9	−0.391	0.01
14 mth	10.1	77.9	−0.300	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	0.000	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	0.700	0.06
48 mth	13.5	96	1.345	0.18

Table 3: Z-Score (Height-for-Age and Weight-for-Age) for sample female child data under below-five age

Age	*Weight*	*Height*	*Height for the age z-score=H-M/σ*	*Weight for the age (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	−0.20
12 mth	9.2	74.1	−0.07	0.00
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	0.00	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	0.10	0.05
48 mth	12.9	94.5	0.19	0.14

HAZ < −3: Severely Underweight

Figures 6 and 7 represent Graphical form of Height-for-Age, Weight-for-Age Z-score for female child under age five. It can be drawn based on the data of table (3).

In,[7] the author analyzes the malnutrition status of the toddler using a k-means clustering algorithm. This research used 6–72-months-old infant 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 the nutrition status of toddlers easily.

Nutrition is the most important factor for the child's growth and normal functions of organs. The prevalence of malnutrition remained increased in all developing countries. The aim of this study[8] is to understand the spatial heterogeneity and mesoscale correlates of malnutrition across Indian districts. This study first, explored the social and demographic correlates of malnutrition. Second, it's taken the survey under geographical location. Third, the district-level status of malnutrition has been 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 HAZ) and wasted (low WHZ), underweight (low WAZ) were used in this study. This study used descriptive statistics models and Regression models were used to understand the status of malnutrition and its correlates. The final result of this study is 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, increasing sanitation, reduction of poverty, and increase maternal nutrition can help to reduce child malnutrition in India. Tables 2 and 3 can be constructed based on the parameters of child height and weight under the age of five using the formula of (1) and (2).

Data Mining Applications in Healthcare

Determining Viability of Treatments

Data mining Techniques in healthcare play a big role to analyze very huge datasets to 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 condition and 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 predicts and provides standard treatment protocols for certain diseases.

Detecting and Eliminating Fraud

In the healthcare sector, data mining focus on identifying insurance fraud and wrong medical claims. In this case, data mining establishes normal patterns, then identifying an unusual pattern in clinical medical claims, labs, and others. The data mining techniques used in mobile healthcare services[9] provide a decision support system to assess and predict the fraud in medical claims by approval officers.

Evaluate Treatment Effectiveness

In data mining, various predictive methodologies can be used to find the best accuracy of results in various healthcare sectors. Using some clustering, classification, and regression methodologies and compare the results to find the best accuracy finally. In the healthcare sector data gathered from the 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 suggested to give the best treatment options effectively with the lowest cost. In[10] data mining techniques collected a large number of patients' healthcare data in the process of diagnosis and treatment of intensive care unit and analyze which are the methods that provide 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

analysis the data mining mostly using predictive methods in healthcare data analyses to get the best accuracy of the 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[11] for disease prediction and widely used data mining technique of Decision Tree Algorithm[12] which is used to disease prediction with the 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 promoting 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, and then the customer can easily get the best services from the healthcare providers. In,[13] propose the machine learning framework to formulate patient satisfaction problems to identify the strong factors.

In recent days, pharmaceutical companies also get a benefit from data mining healthcare services. In a global, a large number of pharmaceutical companies provide 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 physician's feedback about the drugs. In this case, apply data mining techniques to a 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 predicts the malnutrition status using various clustering, classification methodologies to find the result best accuracy. Most research using anthropometric parameters for analyzing the result. This research in the 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 the data mining process first, select the dataset and then choose the parameters to analyze the result. The next process is Feature Extraction and transforms the data after implements the algorithm and then compares the result then finally choose the best accuracy of the result. . Figure 8 represent that the flow graph of proposed method execution.

Conclusion

Child malnutrition is the biggest challenge across the country. Malnutrition causes due to various factors such as Poverty, Unhealthy Household Environment, Insufficient health care services, etc. Predicting of child malnutrition status and factors of malnutrition using various data mining algorithms such as clustering, classification, and Regression with various parameters such as 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 by the policymakers and healthcare providers to make the correct decision at right time.

References

1. Siringi N, Sharma S. Child Immunization Using Data Analysis. United States: IEEE; 2017.
2. Deshmukh S, Khuri N. PediatricDB: Data Analytics Platform for Pediatric Healthcare. United States: IEEE; 2018.
3. 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. Netherlands: Elsevier, B.V; 2017.
4. Anilkumar NA, Gupta D, Khare S, Gopalkrishna DM, Jyotishi A. Characteristics and Causes of Malnutrition Across the Indian States: A Cluster Analysis Based on Indian Demographic and Health Survey Data. United States: IEEE; 2017.
5. Shahriar M, Iqubal MS, Mitra S, Das AK. A Deep Learning Approach to Predict Malnutrition Status of 0-59 Month's Older Children in Bangladesh. United States: IEEE; 2019.
6. Momand Z, Mongkolnam P, Kositpanthavong P, Chan JH. Data Mining Based Prediction of Malnutrition in Afghan Children. United States: IEEE; 2020.
7. Winiarti S, Yuliansyah H, Purnama AA. Identification of toddlers' nutritional status using data mining approach. Int J Adv Comput Sci Appl 2018;9:164.
8. Khan J, Mohanty SK. Spatial heterogeneity and correlates of child malnutrition in districts of India. Khan Mohanty BMC Public Health 2018;18:1027.
9. Gao Y, Sun C, Li R, Li Q, Cui L, Gong B**.** An Efficient Fraud Identification Method Combining Manifold Learning and Outliers Detection in Mobile Healthcare Services. United States: IEEE; 2017.
10. Zhao S, Liu P, Tang GX, Guo Y, Li G. Development and Application of an Intensive Care Medical Data Set for Deep Learning. United States: IEEE; 2020.
11. Hashi EK, Zaman SU, Hasan R. An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques. United States: IEEE; 2017.
12. Deepika M, Kalaiselvi K. An Empirical Study on Disease Diagn using Data Mining Techniques. United States: IEEE; 2018.
13. Liu N, Kumara S, Reich E. Gaining insights into patient satisfaction through interpretable machine learning. IEEE J Biomed Health Inform 2020;25:2215-26.

BIOGRAPHIES

S. Dhivya[4] received M.C.A. Degree from Anna University, Coimbatore, and M.Phil Degree from Bharathiar University, Coimbatore, and currently Pursuing Ph.D (PT) in Kongu Arts and Science College(Autonomous), Erode, TN, India. She is currently working as Assistant Professor at Kongu Arts and Science College(Autonomous), Erode, TN, India. She has 10 years of teaching experience.

Dr.T.A. Sangeetha[5] received M.Sc (CS) Degree from Bharathiar University, Coimbatore, and M.Phil Degree from Bharathiar University, Coimbatore, and Ph.D from Mother Teresa Women's University, Kodaikanal, TN, India. She is currently working as Head and Associate Professor at Kongu Arts and Science College (Autonomous), Erode, TN, India. She has 20 years of teaching and 18 years of research experience. She has guided 20 M.Phil students in the area of Computer Science and now guiding 3 Ph.D (PT). She has presented Articles at various National and International Conference and published articles in various Journals.