Asian Journal of Research in Computer Science

14(4): 107-118, 2022; Article no.AJRCOS.93713
ISSN: 2581-8260

Predicting Malnutrition Status of Under-Five Children in Dhamar Governorate, Yemen Using Data Mining Techniques

Zakarya Abdullah Abdullah Al-Ansi [a*] and Basheer Mohamad Al-Maqaleh [a]

[a] Faculty of Computers and Informatics Thamar University, Republic of Yemen.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2022/v14i4296

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/93713

Original Research Article

Received 06 September 2022
Accepted 13 November 2022
Published 17 November 2022

ABSTRACT

Malnutrition is characterised by the insufficient intake of certain nutrients and the inability of the body to absorb or use these nutrients. This health problem keep going to be a real challenge among children under five years of age in developing countries, including Yemen, despite good aids provided. So, malnutrition is a health problem that significantly participates to child mortality rate in Yemen. The overall prevalence of malnutrition among children in Dhamar Governorate has significantly higher rates compared to other Yemeni governorates.
In this paper, an intelligent predictive system using data mining classification techniques such as J48 decision tree, Bagging and Multi-Layer Perceptron Neural Network (MLPNN) for predicting malnutrition status of under-five children in Dhamar Governorate is proposed.
The main objective of the present paper is to study these classification techniques to predict the 2018-2019 Dhamar Governorate, Yemen Demographic and Health Survey (DGYDHS) dataset and find an efficient technique for prediction. This dataset is imbalanced, so Synthetic Minority Over-sampling TEchnique (SMOTE) is utilised to balance the dataset.
The obtained results were evaluated by the famous performance metrics like Accuracy, TP (True Positive)-rate, FP (False Positive)-rate, Precision, F-Measure, Receiver Operating Characteristics (ROC) graph and execution time. The obtained results revealed that the three classifiers with all attributes have higher predictive accuracy and are generally comparable in predicting malnutrition cases.

**Corresponding author: E-mail: zikoalansi@gmail.com;*

Keywords: Malnutrition; data mining; classifications techniques confusion matrix.

1. INTRODUCTION

"Malnutrition is one of the main causes of death among children under five years of age and one of the most common factors threatening children's life and health" [1,2]. "Children malnutrition is a major public health problem in developing countries" [1]. "It contributes to child morbidity and mortality, poor intellectual and physical development of children, reduces resistance to diseases, and consequently affects negatively on development" [2].

"Malnutrition among children is a major health problem in Yemen" [3]. Currently, Yemen country has one of the highest rates of malnutrition in the world [4]. "In 2021 some 2.25 million Yemeni children under-five are malnourished and the national prevalence of under-five stunting is 46.4%, which is significantly greater than the developing countries average of 25%" [3,4]. "Yemeni children under-five wasting prevalence of 16.4% is also greater than the developing countries average of 8.9%" [4]. "Therefore, child malnutrition has long been recognized as one of the most serious problems in Yemen; however, national-level data on levels and determinants of malnutrition is scarce as Severe Acute Malnutrition (SAM) hits record levels in Yemen with a devastating toll on children under five" [3]. There is few official information about malnutrition in Dhamar Governorate except that it is considered as the worst affected governorate by severity based on the prevalence of acute malnutrition at or above 29.32%, so this issue needs more attention and evaluation [4].

In order to decrease the mortality from malnutrition on children under-five age there should be a quick and effective detection techniques, especially in developing countries like Yemen where there is a lack of specialists and wrongly diagnostic cases are on the rise.

"In recent years, Knowledge Discovery in Database (KDD), which includes data mining techniques has become a common research tool for medical researchers who seek to analyze data and extract useful information for effective decision making" [5,6]. "More precisely, predicting the outcome of a disease is one of the most interesting and challenging tasks to design and develop data mining applications" [6,7].

The propose of the current study is, therefore, to develop an intelligent predictive system using data mining classification techniques for predicting malnutrition status under-five children in Dhamar Governorate, Yemen using 2018-2019 DGYDHS dataset. In this work, three techniques such as J48 decision tree, Bagging and Multi-Layer Perceptron Neural Network (MLPNN) are used.

The rest of this paper is organized as follows. Section II presents the related works. Section III describes the proposed system architecture. The computational results and discussions are presented in Section IV. Section V concludes the paper with future works.

2. RELATED WORK

"Several studies to malnutrition diagnosis using data mining techniques are proposed in" [4,5,7,8,9,10]. These studies have applied different techniques to the given problems and gained different results for depending on the type of techniques used.

In [5], "an identification of malnutrition with the use of supervised data mining techniques was presented. The data mining techniques used were ID3 decision trees, Random Forest (RF) and Artificial Neural Networks (ANNs) and achieved accuracy reached, 68.50%, 77.17% and 77.17%, respectively".

In [7], "a predicting system under nutrition status of under-five children using data mining techniques: The case of 2011 Ethiopian demographic and health survey was proposed. This study achieved the objective of building predictive system using data mining classifiers such as J48 decision tree, Naïve Bayes and PART rule induction classifiers with accuracy such as, 92.24%, 89.68% and 92.62%, respectively".

Data mining based prediction of malnutrition in Afghan children was proposed in [8]. In their study, data mining approaches such as RF, PART rule induction, and Naïve Bayes classifiers were developed to predict the malnutrition status of children under five years of age in Afghanistan.

"Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh was developed" [11]. The data records from the 2014 Bangladesh Demographic and Health Survey (BDHS) were used and LR has been proved to accurately predict malnutrition status among children.

In [12], “a prevalence of malnutrition among infants in Sana’a city was determined. A descriptive cross-sectional survey was conducted among 400 Yemeni infants aged less than one year and the data was analysed using SPSS software. The study concluded that the prevalence of malnutrition among infants was very serious problems in Sana’a city, Yemen”.

In [13], “a mathematical model to estimate the incidence of child wasting in Yemen was presented. The ongoing civil war in Yemen has severely restricted imports of food and fuel, disrupted livelihoods and displaced millions, worsening already high pre-war levels of food insecurity”.

An assessment of the nutritional status of under-five years children at Aden province was proposed in [9]. The collected data was analysed using SPSS software to identify certain socio-demographic factors that influence the nutritional status of under-five children. The study showed that Moderate Acute Malnutrition (MAM) and SAM are areal public health problem in pre-schoolers of the province of Aden, Yemen.

“Prevalence of malnutrition among children in rural Dhamar Governorate: A pilot retrospective study was developed in” [4]. This study was aimed to estimate the prevalence of acute malnutrition among children aged 6 to 59 months in rural Dhamar Governorate, Yemen.

A nutrition survey in Eastern and Western Dhamar Governorate report was presented in [10]. In this survey, using the SMART surveys was used to establish the nutrition situation in Dhamar governorate. They determine some of the factors influencing malnutrition, and identify some of the public health services accessible to the Dhamar population.

The above related works revealed that the wide variety of problems in the health sector are making the use of data mining techniques. But there is no study that addressed the malnutrition problems using data mining techniques in any region in Yemen. The main advantage of applying data mining techniques over the statistical approaches used in previous research is that they enable to predict malnutrition level rather than only identifying risk factors for a specific outcome at group level. Therefore, the work presented in this paper is considered as the first attempt to predict the malnutrition status of under-five children in Dhamar Governorate from the 2018-2019 DGYDHS dataset using data mining techniques such as J48 decision tree, MLPNN and Bagging.

3. METHODOLOGY

3.1 The Poposed System

The proposed system architecture consists of six parts as shown in Fig. 1.

These six parts are as follows:

3.2 Understanding the Problem Domain Part

In this part, discussions with domain experts from Ministry of Public Health and Population (MPHP) and Dhamar Health Office (DHO), were conducted as the main source and reviewing different documents, books and journal articles that focus on data mining techniques in health care as well as malnutrition status of children was also used as supporting sources.

3.3 Understanding the Dataset Part

It is very important to acquire high quality data, which highly counts on the quality of the data collection process. The source data employed for this study purpose was 2018-2019 DGYDHS dataset. The 2018-2019 DGYDHS was conducted under the support of the MPHP and DHO.

The 2018-2019 DGYDHS collected information on the population and health situation which covers family planning, infant, child, adult and maternal mortality, maternal and child health, nutrition, and women’s empowerment.

This dataset comprised four data parts (i.e., children’s records, women records, household member records and death’s records) in excel file format. Each data included other attributes such as children’s data. In children’s data, there were household attributes, women attributes, birth attributes and measurements of anthropometric attributes with the total of 477 attributes. This dataset was a nationally Yemen survey of 1208 instances on children collected in order to classify malnutrition, oedema, and others in the study from women MUAC and under five year children on 477 attributes.

The original dataset from Emergency Nutrition Assessment (ENA) was exported to excel file because WEKA data mining tool does not accept ENA format and whose size amounts to 11.8 MB before any processing activity was done on it. These 477 attributes were not only on malnutrition but on oedema, vaccination, breast feeding, child preference, nutrition (under-five, adult and women), etc.

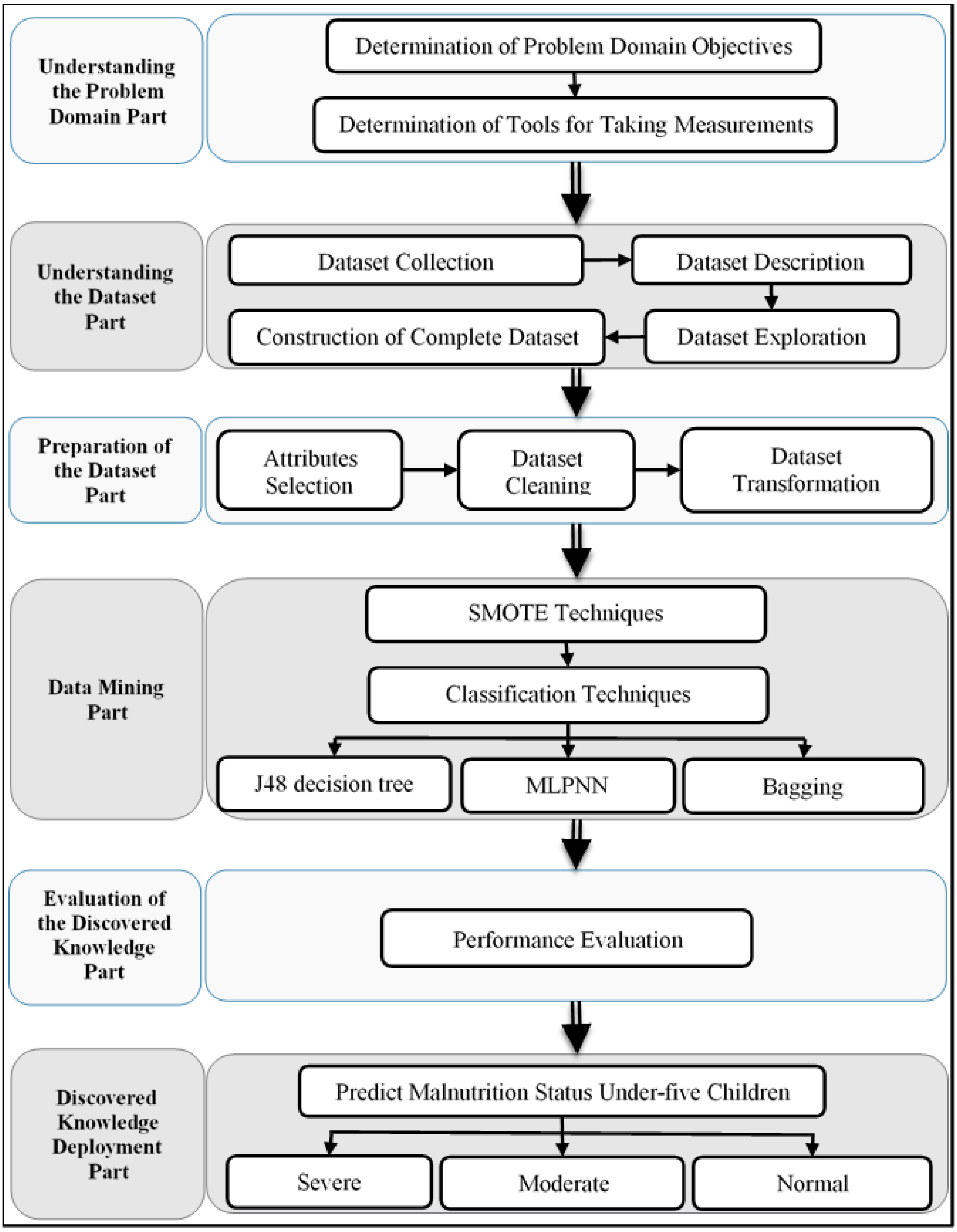

Fig. 1. The proposed system architecture

Table 1. The selected attributes for malnutrition classification

N.o	Field name	Data type	Descriptions
1	Survey-Year	Numeric	Survey Year
2	Cluster	Numeric	Clusters number
3	Team	Numeric	Team number
4	Child-no	Numeric	Child's of number
5	Woman-no	Numeric	Women of number
6	Sheet-no	Numeric	Sheet of number
7	Sex	Nominal	Sex of child
8	Months	Numeric	Child's age in months
9	Weight	Numeric	Child's weight in kilograms
10	Height	Numeric	Child's height in centimeters
11	MUAC	Numeric	Mid-Upper Arm Circumference measurement
12	Oedema	Nominal	The case of EDEMA in children
13	WAZ-WHO	Numeric	Weight / Age standard deviation (WHO)
14	HAZ-WHO	Numeric	Height / Age standard deviation (WHO)
15	WHZ-WHO	Numeric	Weight / Height standard deviation (WHO)
16	Malnutrition	Nominal	Predicting cases of malnutrition

3.4 Dataset Preparation Part

Data preparation step included attributes selection, data cleaning and transformation were used to make the data more suitable for mining part. From total of 477 attributes, 41 attributes related to malnutrition were selected. Also, from total 41 attributes, 8 least important attributes and 17 attributes that have more than 50% missing values were eliminated. The final 16 attributes were selected with the help of domain expend and extensive literature review. The malnutrition class values are normal, severe and moderate. Table 1 shows the selected attributes.

In data cleaning step, 18 records without class information were removed and 136 records were also removed due to deviations from the intended target group of the purpose of this study.

After the data was cleaned, it was transformed into WEKA format including Comma Separated Value (.csv) and Attribute Relation File Format (.arff).

Furthermore, a discretization process that converts continuous valued attributes to discrete values was used. In this case four anthropometric indices Height for Age Z-score (HAZ), Weight for Age Z-score (WAZ), Weight for Height Z-score (WHZ) and Mid-Upper Arm Circumference (MUAC) based on 2006 World Health Organization (WHO) growth multicenter standards were discretized [1].

3.5 Data Mining Part

In this part, the final 2018-2019 DGYDHS dataset was imbalanced as the distribution of its class attribute values are not approximately equally represented such as 889 (84%) instances with the normal class 126 (12%) instances with the moderate class and 39 (4%) instances with the severe class. In this case, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the dataset. As a result, the best performance was obtained with balanced dataset of the total of 2017 instances (889 normal, 504 moderate and 624 severe).

Furthermore, suitable classification techniques for developing a predictive system were selected. These techniques were J48 decision tree, Bagging and MLPNN.

Decision tree is a common classifier that is simple and easy to implement. The obtained results from decision tree were easier to read and interpret. Pruning is the method in machine learning that reduces the size of decision trees by deleting branches of the tree that obtain little power to classify instances [14]. J48 decision tree technique is the implementation of C4.5 decision tree in WEKA data mining tool. It deals with continuous and discrete features with missing values [15].

An ANNs is an information processing system that has special performance characters in common with biological neural networks. This processing system includes a large number of tiny processors to handle data processing. MLPNN is a finite acyclic graph. The nodes are neurons with logistic activation. It is one of the most commonly used ANN classifier algorithms [15,16].

Bootstrap Aggregating (Bagging) is an ensemble generation technique that uses variations of samples used to train base classifiers [5,17]. For each classifier to be generated, Bagging selects (with repetition) N samples from the training set with size N and train a base classifier. This is repeated until the desired size of the ensemble is achieved [17].

These techniques were implemented using WEKA 3.8.3 tool. due to its simplicity, proficiency in discovering, analysis and predicting patterns [18,19]. Therefore, WEKA has recognized as a benchmark tool in machine learning community as it has achieved a popular acceptance within many academic areas and it has become a widely utilized tool in the development of data mining applications [6,18,19]. These three techniques were chosen based on their distinguished performance during test and they have literature review support [20].

3.6 Evaluation of the Discovered Knowledge Part

After mining the required patterns, the interpretation and evaluation of the mined patterns was accomplished. The interpretation was concerned with whether the detected pattern was interesting or not. It was verified whether it has knowledge or not. The performance of the techniques was measured and evaluated using accuracy, TP Rate, FP Rate, Precision, F-Measure and ROC area.

3.7 Discovered Knowledge Deployment Part

Finally, discovered knowledge can be disseminated by using different techniques such as J48 decision tree, Bagging and MLPNN. Therefore, the overall research design developed classifiers that predict the malnutrition status of under-five children (probability of normal, severe or moderate). To enhance visualization and ease of interpretation, the obtained results were displayed in suitable formats such as tables, diagrams and charts.

4. RESULTS AND DISCUSSION

The dataset with 16 attributes and 2017 records/instances after applying SMOTE was used in the first way. In the second way, these three classifiers were fed with reduced dataset with 10 attributes. These 10 reduced attributes were child-no, child-age, weight, height, MUAC, oedema, WAZ-WHO, HAZ-WHO, WHZ-WHO and malnutrition that are highly relevant in predicting malnutrition statues of under-five children from 2018-2019 DGYDHS dataset.

All the experiments were done on a full training datasets and 10-fold cross validation was used for randomly sampling the training and test sets. In 10-fold cross-validation, the complete dataset was randomly split into mutually exclusive subsets of approximately equal size. The classification model is trained and tested 10 times.

The performances of the three classifiers were evaluated using the standard metrics of Accuracy, TP-rate, FP-rate, Precision, F-measure and ROC graph which were calculated using confusion matrix. The confusion matrix is a useful tool for analysing how good the classifier can accurately identify instances of several classes.

The experiments were conducted on a Legion laptop with Intel core TM i7 @ 2.6GHz processor and 16GB RAM. The performance of the proposed system is demonstrated below:

4.1 The Experimental Results

4.1.1 Experiment one

The first experiment was developed to evaluate the performance of a J48 pruned tree classifier. In this experiment, two ways were considered. On the first way, the proposed classifier was run on a full training set containing 2017 instances with 16 attributes. It took 0.03 second to build the model. The second way was executed on a full training set containing 2017 instances with 10 selected attributes. It took 0.02 second to build the classifier. This tree is simple and faster than the experiment conducted on full training set with all attributes. Table 2 shows the confusion matrix for experiment one. Also, the performance measures used are given in Table 3.

The results prove that the performance of J48 pruned classifier was better on the all attributes as it achieved classification accuracy equals to 96.23 %.

Table 2. Confusion matrix for J48 pruned classifier

Model	Confusion matrix			
	Normal	**Moderate**	**Severe**	**Actual**
J48 pruned with all (16) attributes	872	16	1	Normal
	24	460	20	Moderate
	2	13	609	Severe
J48 pruned with (10) selected attributes	877	12	0	Normal
	28	451	25	Moderate
	4	11	609	Severe

Table 3.Detailed performance measures for J48 pruned classifier

Model	Accuracy	TP-Rate	FP-Rate	Precision	F-Measure	ROC area
J48 pruned with all (16) attributes	96.23%	96.2%	2.0%	96.2%	96.2%	97.7%
J48 pruned with (10) selected attributes	96.03%	96.0%	2.2%	96.0%	96.0%	97.4%

Table 4. Confusion matrix for J48 unpruned classifier

Model	Confusion matrix			
	Normal	**Moderate**	**Severe**	**Actual**
J48 unpruned with all (16) attributes	869	19	1	Normal
	23	461	20	Moderate
	2	15	607	Severe
J48 unpruned with (10) selected attributes	874	15	0	Normal
	30	449	25	Moderate
	4	13	607	Severe

Table 5. Detailed performance measures for J48 unpruned classifier

Model	Accuracy	TP-Rate	FP-Rate	Precision	F-Measure	ROC area
J48 unpruned with all (16) attributes	96.03%	96.0%	2.0%	96.0%	96.0%	97.8%
J48 unpruned with (10) selected attributes	95.69%	95.7%	2.3%	95.7%	95.6%	97.2%

Table 6. Confusion matrix for Bagging classifier

Model	Confusion matrix			
	Normal	**Moderate**	**Severe**	**Actual**
Bagging with all (16) attributes	886	3	0	Normal
	32	451	21	Moderate
	7	12	605	Severe
Bagging with (10) selected attributes	882	6	1	Normal
	38	444	22	Moderate
	7	13	604	Severe

Table 7. Detailed performance measures for Bagging classifier

Model	Accuracy	TP-Rate	FP-Rate	Precision	F-Measure	ROC Area
Bagging with all (16) attributes	96.28%	96.3%	2.2%	96.3%	96.2%	99.5%
Bagging with (10) selected attributes	95.69%	95.7%	2.6%	95.7%	95.6%	99.4%

Table 8. Confusion matrix for MLPNN classifier

Model	**Confusion matrix**			
	Normal	Moderate	Severe	Actual
MLPNN with all (16) attributes	834	47	8	Normal
	52	433	19	Moderate
	15	37	572	Severe
MLPNN with (10) selected attributes	825	54	10	Normal
	51	396	57	Moderate
	13	59	552	Severe

Table 9. Detailed performance measures for MLPNN classifier

Model	**Accuracy**	**TP-Rate**	**FP-Rate**	**Precision**	**F-Measure**	**ROC Area**
MLPNN with all (16) attributes	91.18%	91.2%	4.6%	91.3%	91.2%	97.5%
MLPNN with (10) selected attributes	87.90%	87.9%	5.9%	87.9%	87.9%	96.5%

4.1.2 Experiment two

The performance of a J48 unpruned decision tree classifier in predicting malnutrition of under-five children was evaluated in this experiment. Table 4 and Table 5 summarize the results. Also, the results showed that the performance of J48 unpruned classifier with all attributes was slightly better than the performance of the same classifier with reduced attributes in all selected performance measures. It took 0.02 second to build the classifier.

4.1.3 Experiment three

This experiment was executed to evaluate the performance of Bagging classifier. As mentioned earlier, two ways were considered and the computational obtained results are given in Table 6 and Table 7.

The performance of Bagging classifier was better with the all attributes in all performance evaluation measures used as shown in Table 7, and it took 0.09 second to build the classifier.

4.1.4 Experiment four

In this experiment, the performance of MLPNN in predicting malnutrition was evaluated. The obtained results are given below in Table 8 and depicts performance evaluation measures used are shown in Table 9.

The obtained results showed that MLPNN classifier performance with all attributes has 91.18% accuracy, and took 4.53 seconds to build the classifier. The performance with selected attributes has 87.90% accuracy and the execution time was 2.7 seconds. MLPNN consumed longer execution time because it contains huge number of parameters (speed of learning, number of hidden layers, input neurons, number of iteration, updating weights, etc.) to set up.

4.2 Comparative Study

4.2.1 Performance evaluation and comparison of used classifiers

One of the objectives of the present study is the evaluation and comparison of performance of the used classifiers in order to choose the classifier, which performs the best.

Based on the experimental design, Table 10 summarises the experimental results that were obtained using three different classifiers with all and reduced attributes.

The first experiment shows that J48 pruned decision tree technique with all attributes is capable of predicting malnutrition as normal, moderate and severe with an accuracy of (96.23%). The second experiment by J48 unpruned decision tree classifier with all attributes also has significant effect on the classification and prediction of malnutrition (has accuracy of 96.03%). The third experiment shows that J48 pruned decision tree algorithm with 10 selected attributes with an accuracy of 96.03%. The fourth experiment by J48 unpruned decision tree classifier with 10 selected attributes also has significant effect on classification and prediction of malnutrition (has accuracy of 95.69%).

Table 10. Summarization of experimental results with three selected classifiers

NO	Model	Accuracy	TP-rate	FP-rate	Precision	F-measure	ROC area
1	J48 pruned with all (16) attributes	96.23%	96.2%	2.0%	96.2%	96.2%	97.7%
	J48 pruned with (10) selected attributes	96.03%	96.0%	2.2%	96.0%	96.0%	97.4%
2	J48 unpruned with all (16) attributes	96.03%	96.0%	2.0%	96.0%	96.0%	97.8%
	J48 unpruned with (10) selected attributes	95.69%	95.7%	2.3%	95.7%	95.6%	97.2%
3	Bagging with all (16) attributes	96.28%	96.3%	2.2%	96.3%	96.2%	99.5%
	Bagging with (10) selected attributes	95.69%	95.7%	2.6%	95.7%	95.6%	99.4%
4	MLPNN with all (16) attributes	91.18%	91.2%	4.6%	91.3%	91.2%	97.5%
	MLPNN with (10) selected attributes	87.90%	87.9%	5.9%	87.9%	87.9%	96.5%

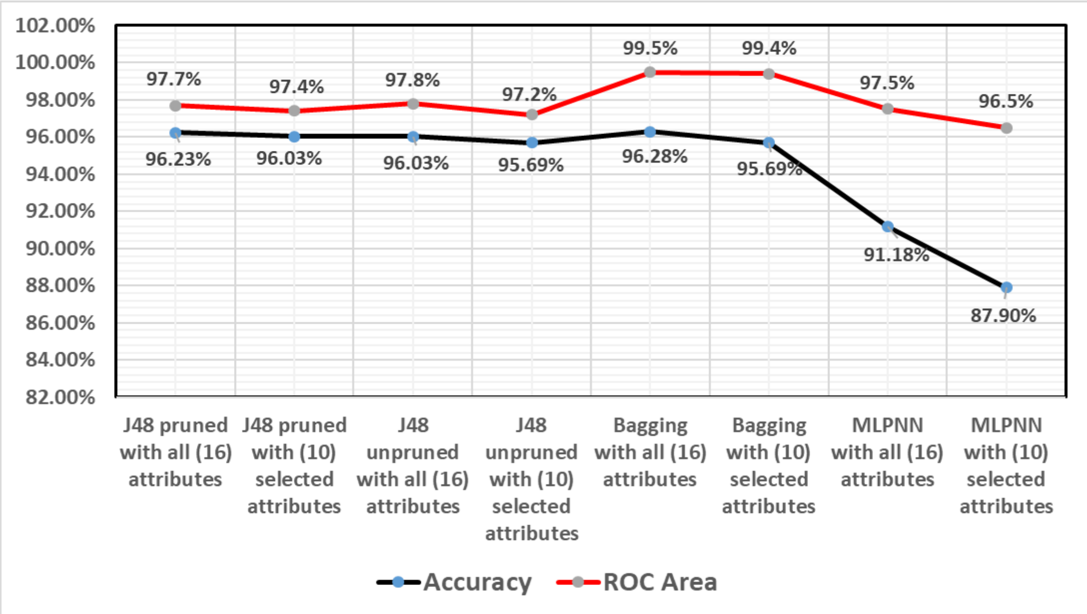

Fig. 2. Classifiers performance comparison using accuracy and ROC graph

The fifth and sixth experiments were developed to evaluate the performance of the Bagging classifier in predicting malnutrition with accuracy performance of (96.28% and 95.69%), respectively. Finally, the seventh and eighth experiments were designed to evaluate the performance of the MLPNN classifier in predicting malnutrition with accuracy performance of 91.18% and 87.90%, respectively. In terms of their accuracy performance, all classifiers have a good competitive performance except MLPNN with all attributes and selected attributes generate results with accuracy of 91.18% and 87.90%, respectively.

"The next task in testing the classifier to decide which one of the three classifiers constitutes a better model/classifier of the 2018-2019 DGYDHS dataset was evaluated by using ROC area analysis. ROC is the main indicator during classifier performance selection" [15].

In ROC analysis, when the curve is near the upper left that would indicate a better classifier. In this case, the ROC area performance of the

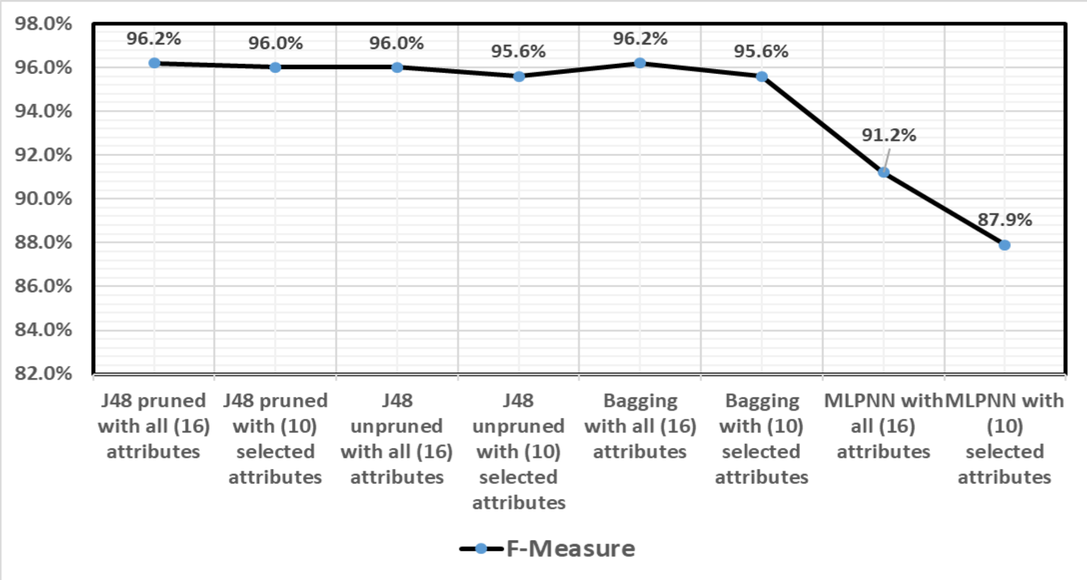

Fig. 3. Classifiers performance comparison using F-Measure

Table 11. The comparative study with other related systems

System	Classifier used	Data mining tool	(Accuracy)
A. Zahirzda et al. (2021) [21]	The three most popular techniques include Naïve Bayes, Random Forest, and Support Vector Machine (SVM)	Predicting Diarrhea	69.85%NB, 77.10%RF, and 72.61 %SVM
D. Thangamani P. Sudha et al. (2014) [5]	ID3 decision tree Random Forest MLP	WEKA software	68.50% ID3, 77.17% RF, and 77.17% MLP
S.M.J. Rahman et al. (2021) [22]	logistic regression (LR)	WEKA software	88.3% stunted 87.7%, wasted 85.7%, and underweight
S. Khare et al. (2017) [23]	Based cross-sectional Used enter and analyses data	WEKA software	70.34% underweight, 62.95% stunted, and 82.59% wasted
A. Alemu et al. (2016) [24]	J48 decision tree Naïve Bayes Bagging Artificial Neural Network	WEKA software	92.24% J48, 89.96% NB, 92.52% Bagging, and 91.67% ANN Respectively
Z. Markos (2014) [7]	J48 decision tree Naïve Bayes part rule	WEKA software	92.24% J48, 89.68% NB, and 92.62% PR
The prosed system	J48 decision tree Bagging MLPNN	WEKA software	96.23% J48, 96.28% Bagging, and 91.18% MLPNN

classifiers shows that Bagging, J48 unpruned decision tree and MLPNN with all attributes scored the highest area of 99.5%, 97.8% and 97.5%, respectively. The lowest ROC keeps account in MLPNN with selected attributes which is 96.5%. Fig. 2 shows the performance of the eight classifiers with respect to accuracy and ROC graph.

F-Measure balances both the concerns of Precision and Recall in one number. In this context, Bagging and J48 pruned with all attributes classifiers score 96.2% in terms of F-Measure compared to other classifiers as shown in Fig. 3.

Generally, the Bagging with all attributes is more understandable and less complex to human than other classifiers. Therefore, the performance of Bagging classifier with all attributes gives precious information in predicting nutritional status when compared to other classifiers.

4.2.2 Comparative study with other related systems

The performance of the proposed system with other related systems is evaluated and compared using predictive accuracy. In medical domain, it is crucial to search for accurate classifier that will decrease the error rate in diagnosis to make perfect decisions.

The performance of the related systems in this comparative study is data dependent for each system as shown in Table 11.

These results prove that the performance of the proposed system is competitive and outperforms the other related systems in terms of predictive accuracy.

5. CONCLUSION AND FUTURE WORK

Medical data mining in health care is regarded as an important task that needs to be extracted accurately and efficiently. In this paper, an intelligent predictive system using J48 decision tree, Bagging and MLPNN for malnutrition status under-five children in Dhamar Governorate from 2018-2019 DGYDHS dataset is proposed. The most effective classifier to predict malnutrition status of under-five children appears to be a Bagging classifier with all (16) attributes with a classification accuracy of 96.28% and ROC area of 99.5%. The performance of the proposed system outperformed other related systems in terms of predictive accuracy. This predictive system should be able to support medical decisions and make more information available to the specialists and provide new knowledge in this field.

One of the most future research directions would be developing an incremental learning framework to automated prediction of malnutrition status of under-five children. This framework can be designed to update the discovered patterns as new data arrive at different points of time in order to discover such interesting patterns. It can also be developed as a real time system, so the system may generate a real time prediction which definitely enables the specialists to minimize malnutrition among children. Furthermore, this study is considered as a demographic and health survey dataset. So that, future direction might need to discover knowledge in other domain areas such as clinical datasets by using deep learning techniques.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1 Isanaka S, et al. Routine amoxicillin for uncomplicated severe acute malnutrition in children. New England Journal of Medicine. 2016;374(5):444–453.

2 Tsegaye D, Tamiru D, Belachew T. Factors associated with dietary practice and nutritional status of pregnant women in rural communities of Illu aba Bor zone, Southwest Ethiopia. Nutrition and Dietary Supplements. 2020;12(103):10–2147.

3 IPC Acute Malnutrition Analysis Report, Jan 2020- March 2021, Integrated Food Security Phase Classification; 2021. Available:www.ipcinfo.org

4 Aljarmouzi HM, et al. Prevalence of malnutrition among children in rural dhamar governorate: a pilot retrospective study featured. Annals of Medicine and Health. 2020;2(2):32–34.

5 Thangamani D, Sudha P. Identification of malnutrition with use of supervised data mining techniques-decision trees and artificial neural networks. International Journal of Engineering and Computer Science. 2014;3(7):1270–1274.

6 Gupta GK. Introduction to data mining with case studies. 3rd Edition, PHI Learning Put. Ltd.; 2019.

7 Markos Z. Predicting under nutrition status of under-five children using data mining techniques: The case of 2011 Ethiopian demographic and health survey. Journal of Health & Medical Informatics. 2014; 5(2):152–154.

8 Momand Z, et al. Data mining based prediction of malnutrition in Afghan children. In 12th International Conference on Knowledge and Smart Technology (KST 2020). 2020;12–17.

9 Nabil QM. Al-Hajj et al. Assessment of the nutritional status of under-five years (6-59 months) children friendship teaching Hospital, Aden Province, Yemen. World Journal of Pharmaceutical Research. 2021;10(5):47-59.
10 Dhamar Governorate, Yemen: Eastern and Western Dhamar Nutrition Survey, MoPHP, United Nations Children's Fund (UNICEF); 2013.
11 Talukder A, Ahammed B. Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition. 2020;78:11-19.
12 Al-Rabeei NA, et al. Pervalance of malnutrition among infants in Sana'a City, Yemen. Ulutas Medical Journal (UMJ). 2015;1(1):1-6.
13 Hussein RA, et al. A mathematical model to estimate the incidence of child wasting in Yemen. Conflict and Health. 2021; 15(1):1–13.
14 Bramer M. Principles of data mining. Springer- Verlag London-Limited; 2007.
15 Manimekalaai K. A proficient heart disease prediction method using different data mining tools. International Journal of Engineering Science and Computing. 2016;6(3):2676-2680.
16 Kirmani MM. Heart disease prediction using multilayer perceptron algorithm. International Journal of Advanced Research in Computer Science. 2017; 5(8):1169-1172.
17 Aggarwal CC, Sathe S. Theoretical foundations and algorithms for outlier ensembles. ACM SIGKDD Explorations Newsletter. 2015;17(1):24–47.
18 Jain D, Singh V. Utilization of data mining classification approach for disease prediction: a survey. International Journal Education and Management Engineering. 2016;6:45-52.
19 WEKA. Available:http://www.cs.Waikato.ac.nz/mc/weka/index.html
20 Kouam CE, et al. Perspectives for integration into the local health system of community-based management of acute malnutrition in children under 5 years: A qualitative study in Bangladesh. Nutrition Journal. 2014;13(1):22-24.
21 Zahirzda A, Chanmas G, Chan JH. A data mining model for predicting diarrhea in Afghan children. TechRxiv. 2021;4(1):111–116.
22 Rahman SMJ, et al. Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. PLoS One. 2021;16:160-166.
23 Khare S, et al. Investigation of nutritional status of children based on machine learning techniques using indian demographic and health survey data. Procedia Computer Science. 2017;115:338–349.
24 Alemu E, et al. Assessment of breastfeeding practices in ethiopia using different data mining techniques. Indian Journal of Computer Science and Engineering. 2016;7(1):1–6.

© 2022 Al-Ansi and Al-Maqaleh; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/93713