Detection of effective factors on the adult Health Literacy level with a meta-heuristic algorithm

Elham Enayati
* Instructor, Computer Sciences Department, Basic Sciences Faculty, University of Bojnord, Bojnord, Iran. (corresponding author)
elham.enayati@gmail.com

Zeinab Hassani
Instructor, Computer Sciences Department, Basic Sciences Faculty, Kosar University of Bojnord, Iran.

Mitra Moodi
Associate Prof, Health Education and Health Promotion Department, Birjand University of Medical Sciences, Birjand, Iran

Received: 15 October 2019
Accepted: 14 January 2020
Doi: 10.22038/jhl.2020.44764.1089

ABSTRACT

Background and Objective: Health literacy is a global factor in determining the level of social literacy of persons in the individual and social health. To improve health in a community, it is essential which measuring and identifying effective factors on the health literacy. The purpose of this study was to identify and determine the factors affecting the level of health literacy using a meta-heuristic algorithm.

Materials and Methods: In this paper, a hybrid model has been introduced which benefit of bat optimization algorithm and machine learning algorithms to determine the major factors on health literacy level. In our model, the support vector machine algorithm and K-nearest neighbor are used to classify the data. The data set has been extracted from a descriptive-analytic study that was performed on 2133 samples to measure adults' health in South Khorasan Province in 2016.

Results: In this study, the combination of bat optimization algorithm and K-nearest neighbor with 93.76% accuracy and 95.47% precision has been able to identify effective factors on health literacy. Combined bat optimization algorithm and support vector machine have been obtained with 96.25 percent accuracy and 97.99 percent precision. Analytical results show that two areas: access to the resources and use of resources are impressive in the low level of health literacy in South Khorasan province are effective.

Conclusion: Considering the performance of the intelligent proposed model show this tool can be used to analyze health literacy data. This model does not require statistical analysis, and can provide high-quality results quickly.

Paper Type: Research Article
Keywords: .

▶ **Citation:** Enayati E, Hassani Z, Moodi M. Detection of effective factors on the adult Health Literacy level with a meta-heuristic algorithm. *Journal of Health Literacy*. Winter 2020; 4(4):46-55.

Introduction

Health literacy is a well-known term in the health area. This term was first raised in 1970. Health literacy means the ability of individuals to obtain, interpret, and understand necessary health care services and recommendations in a way that helps people to have a healthy life (1). According to the World Health Organization's report in 2018, health literacy is divided into four levels of inadequate, problematic, adequate and excellent. Among them, inadequate and problematic are defined as limited health literacy (2).

People in a society with limited health literacy are unable to properly understand the explanation on food packages and prescribed drugs consumption guidelines. These people are restricted in expressing disease symptoms to physicians and paramedics and perform inefficient in the use of public therapeutic and health services. Also, in a society with a limited health literacy level, there will be more problems in epidemic cases of a disease and dealing with its prevalence (3).

Improving health literacy is one of the fundamental and effective criteria in developing hygiene level in a society (4). A research (2009) in the United States indicated that the costs incurred due to limited health literacy in US are between 3 and 5 percent of the costs of the public health system annually (5). According to this research, the additional costs for a person with limited health literacy in a year compared to a person with adequate literacy level are between $ 143-7798. While, in the same year only 12 percent of people of American society had adequate health literacy. Nowadays, in many countries of the world such as Canada, USA, Japan, Korea, Australia, Switzerland and... different researches are conducted in health literacy area.

Researches on measuring the health literacy rate of society and identifying the parameters related to the causes of weakness or strength of health literacy level in a society has a particular importance. Various data are always collected in these researches, which extracting information from the raw data, requiring to have high-precision and up-to-date tools and techniques. Machine learning techniques are one of the effective tools for recognizing different dimensions of data. These techniques are capable to categorize data sets and discover hidden patterns within them. These methods are good tools for increasing the precision in data analysis process as well as decision making process. So far, many researchers in the health care area have used these techniques (6.7).

In this study, we categorize data obtained from a research plan in adult health literacy area through proposing a machine learning model that itself includes a combination of several advanced data mining methods, and then detect effective features in this categorization. The selected features of the proposed model are the same effective factors on adult health literacy level in the target society. We will show that the proposed model with high precision is capable to categorize individuals in terms of health literacy level. This method also enable to categorize similar cases that have not been included in the proposed model during the training stage and also help the specialist to predict the behavioral patterns of individuals in this regard through detecting effective factors on adult health literacy of the society and recognize the performance of the target community in similar cases.

Materials and methods

A proposed smart model for detecting effective factors on adult health literacy of the community is presented in this section. Before that, the data set has been explained, and in the following bat optimization algorithms, support vector machine and K-nearest neighbor has been presented. Also in the proposed model section, we will express the details of presented model and then the results obtained from model simulation are presented in the findings section, and we summarize the results obtained from the present study in the discussion section.

The data set used in this research has been extracted from a descriptive-analytical study. This cross-sectional study has been performed on adult population (persons over 18 years) in South Khorasan province in 1395 and 2133 persons has participated in that study to assess health literacy rate. Persons included all referrals to all health-care centers covering by Brigand, Khuu, Darian, Zircon, Saroyan, Sarbisheh, Bashrouieh, Tabas, Ferdows, Ghaen and Nehbandan health centers that the data were collected using multi-stage random cluster sampling method after identifying the proportion of each center from the number of samples related to each city and village. The proportion of each center from the sample number was determined in this way that 278 head-clusters randomly were selected from rural and urban regions of South Khorasan province in proportion to the population of urban and rural regions of each city and 8 head-cluster including 4 males and 4 female over 20 years old and from each age group of 20-30, 30-40, 40-50, 50-60 one male and one female were selected and a questionnaire was completed. Other inclination criterion to the study was literacy for reading and writing, and those with severe visual or hearing impairment or with mental or perceptual disorders were excluded from the study. The Iranian Health Literacy Questionnaire (National Research Plan of Iranian Health Literacy Assessment) was used for data collection. This research plan has been implemented by the Health Ministry Promotion and Education Office with collaboration of Research Center for Modeling in Health of Medical Sciences University of Kerman in seven selected provinces and in some provinces voluntarily. The reliability and validity of the questionnaire were examined by relevant experts. The data is obtained from the information collected from the 56-question questionnaire that completed by the participants. Scoring criteria is based on the Likert scale, which in this regard individuals fall into three categories of good, moderate, and poor in terms of health literacy.

The questionnaire contains 56 questions in 9 dimensions of health literacy, which include scopes of access to resources (Q1) and resource utilization (Q2) 11 questions that depend on the response to score of zero or one (5 and less poor, 6-8 moderate, 9 and higher good), ability to read content (Q3) and ability to understand content (Q4) 19 questions that depend on the response to score of 1-4 (37 and less poor, 38-56 moderate, 57 and higher good), ability to interpret, judge and assess (Q5) 8 questions that depend on the response to score of 1-3 (11 or less poor, 12-17 moderate, 18 and higher good), ability to make decision and communicate (Q6), having health (Q7), individual empowerment (Q8), social empowerment (Q9) totally 18 questions that depend on the response to a score of 1-3 (26 and less poor, 27-40 moderate, 41 and higher is good). The data after collecting were entered into SPSS software version 22 and were analyzed using appropriate indices and graphs and statistical tests (Table 1)

Table 1. Statistical analysis of data

Scope	Dimensions	Mean	Standard deviation
Q1	access to resources	10.7	0.1
Q2	resource utilization	9.3	0.2

Q3	ability to read content	12.4	0.1
Q4	ability to understand content	13.6	0.07
Q5	ability to interpret, judge and assess	13.6	0.08
Q6	ability to make decision and communicate	14.00	0.08
Q7	having health	11.4	0.1
Q8	individual empowerment	5.9	0.1
Q9	social empowerment	11.5	0.1
	Total score of health literacy	12.00	0.06

In this study, we apply the results of this research plan, which is computed using statistical techniques, to compare the outputs resulted from the proposed model. The outputs of the proposed model are presented totally independent of the results of Table 1 and if the model results are close to the mentioned statistical outputs, the accuracy of model performance is verified.

Methods

K-Nearest-Neighbor (KNN):

This algorithm is one of the sample-based data classification methods that place in the series of lazy algorithms. In this method, to classify each sample, the algorithm looks for the sample K that has the most similarity to the mentioned sample, and the unknown sample label is the category label with the most frequency among the similar sample K. Different criteria can be used to assess the similarity and dissimilarity of samples with the number of different features. Among them, Euclidean distance is one of the most common criteria in which the distance between two samples is

$$\mathrm{X}_i = (x_{1i} . x_{2i} . \dots . x_{ni}) . \mathrm{X}_j = (x_{1j} . x_{2j} . \dots . x_{nj})$$

with number N features in the state space is calculated as followed:

Equation (1): $dist(\mathrm{X}_i, X_j) = \sqrt{\sum_{z=1}^{n} (x_{zi}, x_{zj})^2}$

Here K is considered the least value $dist(\mathrm{X}_i, X_j)$ for the sample of its nearest neighbors (1).

Support Vector Machine (SVM)

Support vector machine is one of the machine learning techniques. This technique was first introduced by Vapnik in 1995 (8). The support vector machine can be used to classify objects or pattern recognition for specific classes (9). SVM is a supervised learning algorithm based on the theory of statistical learning and its purpose is to determine a separator super-plate that separates the two classes from each other using the training dataset (10).The support vector machine takes the data to a space with a much higher dimensions through a Kernel function (ϕ) in order to classify high-complexity data. The kernel function depicts the data from the input space to a space with higher dimensions in such a way that it is possible to separate the data linearly in that space. Then, the transferred training data are separated by an optimized super-plate in that space. In fact, without any change in the nature of the optimization problem by applying the transfer function ϕ, the data are mapped to a space with higher dimensions to determine the level of decision making linearly in this space (11).

In the support vector machine, the dataset contains N training data which represents each member in the form of $X=(x_i . y_i)$, $(1 \le i \le N$ و $x_i \in R^d)$, and in that d is the training data dimensions and is their category. In this way, separator super-plate is obtained from Equation (2).

Equation (2): $w \cdot X + b = 0$

W is the weight vector and b are the bias. The training examples that have the shortest distance (based on Equation (2)) from the separator plate are called the support vectors.

Equation (3): $y_i(w \cdot x_i + b) \ge 1$

Therefore, it is necessary to determine the coefficients w and b to find the super-plate so that SVM has the best classification. Separator super-plate distance from the support vector is equal to 2/||w||. As a result, to maximize the distance, only need to minimize, which is replaced by $||w||^2$ for ease of operation. So we have:

Equation (4): $\min_{b,w} \frac{1}{2}\|w\|^2 \quad , \quad y_i\,(w \cdot x_i + b) \geq 1$

The equation constraint (4) is solved using Lagrange. This is a quadratic programming problem with constraints in the form of linear inequalities

Equation (5):

$$\min_{b,w} \frac{1}{2}\|w\|^2 - c\sum_{i=1}^{N} l_{0,1}[y_i\,(w \cdot x_i + b) - 1]$$

The value c is the penalty coefficient that controls the error rate and the distance of the super- plate with the support vectors. This problem is solved by the help of Lagrange's indefinite coefficients. So:

Equation (6):

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_i\gamma_j\alpha_i\alpha_j(x_i.x_j)$$

$$\sum_{j=1}^{N} \gamma_i\alpha_i = 0, \quad 0 \leq \alpha_i \leq C, i = 1,2,\dots,N$$

In SVM nonlinear problems, the kernel function is used to map the data to higher dimensions. Then we have:

Equation (7):

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \gamma_i\gamma_j\alpha_i\alpha_j K(x_i.x_j)$$

The function is the kernel function. Therefore, selecting kernel function plays an important role in influencing the model capability and usually polynomial and RBF kernels functions are used (12).

In this study, the RBF kernel function has been used to classify the support vector machine.

Equation (8):

$$K(x_i.x_j) = e^{-\frac{\|x_i,x_j\|^2}{2\sigma^2}}$$

The reason for this matter is the ability of this function to analyze data with high-dimensions compared to other kernels.

Bat Optimization Algorithm:

Today, meta-heuristic algorithms are used in many clustering problems. The bat algorithm is one of these algorithms that was first introduced in 2010 in Yang's article (13). This algorithm is inspired from bat performance method in hunting. The bat sends a sound at speed (), wavelength (), frequency (), and loudness () to find bait, and it can find its hunt using it. This algorithm, like other meta-heuristic algorithms, has a specific position for each bat (i) at any moment, which is then updated in the next period based on the best answer () until it approaches enough to the bait or the stopping condition of the algorithm is realized (13).

In this algorithm, each new solution in step (t) is calculated as follows:

Equation (9)

$$f_i = f_{min} + (f_{max} - f_{min})\beta,$$

$$v_i^t = v_i^{t-1} + (x_i^t - x_*)f_i\,,$$

$$x_i^t = x_i^{t-1} + v_i^t\,, \qquad \beta \in [0\ 1]\,, i = 1,\dots,n$$

After that a new solution was selected among the generated values, updating the position for each bat is performed using Equation (10):

Equation (10):

$$x_{new} = x_{old} + \varepsilon A_t\,,\ \varepsilon \in [-1\ 1]\,,\ A_t = A_t^i$$

Where is the mean loudness of all bats. Also, the equation of the change of loudness and pulse propagation rate is as follow:

Equation (11):

$$A_t^{i+1} = \alpha A_t^i\,,\ r_t^{i+1} = r_t^0[1 - \exp(\gamma t)]$$

In this equation α, γ are fixed numbers.

Purposed model

The main problem in this study is to identify the effective factors on adult health literacy level in the target community. This problem in machine learning methods is known as feature selection problem. The problem of feature selection is to find a subset with minimum number N features from a set with M features, so that N <M. This feature reduction is such that the selected features to achieve the mentioned purpose contain sufficient and necessary information. There are various methods to solve the feature selection problem. One of these methods is the use of meta-heuristic algorithms that is able to select features using an evaluator function. Classifier models have been used as evaluator functions in many studies (14-16).

So far, various optimization algorithms including genetic algorithm, particle swarm, harmony, etc. have been used for feature selection (17, 18). Every one of optimization algorithms have a specific approach to select feature. The optimization algorithms are directed towards the target with regard to the used evaluator function. In the present study, the bat binary algorithm was used to select the feature, although each of the mentioned meta-heuristic algorithms also provide similar efficiency. In (13, 19, 21), the efficiency of the bat binary algorithm is compared with other meta-heuristic algorithms such as particle swarm, gravitational search, etc. and shows the relative superiority results of this method over other used meta-heuristic algorithms.

A proposed model for detecting the effective factors on adult health literacy level is presented in this section. This model is a combination of bat binary meta-heuristic algorithm and SVM and KNN classifier algorithms. The dataset is pre-processed before entering the model. Then, we evaluate the pre-processed data with the proposed model. The steps of performing proposed method are as follows

1. Pre-processing data

1.1 Data Normalization: Producing efficient results depends on the appropriate data. Data features values are in different range or domain. At this stage, to integrate the data using the relation (12), the data are normalized in the interval [0.1].

Equation (12):

$$X = (x - x_{min})/(x_{max} - x_{min})$$

x_{max}, x_{min}, are the maximum and minimum feature values, respectively.

2. Hybrid algorithm

At this stage, the bat binary algorithm is used to detect the effective factors on adult health literacy. Also, in two different scenarios, using support vector machine classification algorithms (BatSVM) and k-nearest neighbor (BatKNN), the available data are categorized into three categories: good, moderate and weak.

In the case of any participant in the mentioned dataset in the Material Section, the number of n feature is obtained with regard to respond that has been given to the questions in the questionnaire. Each person in our proposed model is equivalent to a bat with n dimensions which its values are 0 and 1 according to the binary version of the bat algorithm. Value 1 in each dimension of the bat means existence the feature and value 0 means existence no feature in the used classification algorithms. With the features selected in m iteration of bat binary algorithm, the classifier algorithms select the best classification of adult health literacy data, which the best classifier is selected based on the accuracy criterion. In the following, the pseudo-code of the two proposed models (BatKNN / BatSVM) is expressed (Figure 1).

1.2. Initialization of Algorithm Parameters: At this stage, the initial values of the algorithm parameters including the number of iteration and the initial population size of the bats are initialized. The initial values of the proposed method parameters are shown in Table 2.

Table 2 Initial Values of Proposed Method

Parameters

Parameter	value
Loudness(	0.25
Pulse propagation rate(	0.1
Maximum replications number	25,50,100
Early bats population	20,25

2.2. Recalling for the Bat Binary Combined Algorithm and the Support Vector Machine Classification Algorithms and the K-Nearest Neighbor.

```
Objective function f(x), x = (x_1, ..., x_d)^T

Initialize Bat population: X_i (i = 1, 2, ..., n) rand(0 or
1), Maximum iterations, and Vi= 0
Define pulse frequency F_i
Initialize pulse rates r_i and the loudness A_i

Evaluate f(x_i) by SVM / KNN

 while t < Maximum iterations do
   Generate new solutions by adjusting frequency,
   and updating velocities and locations/solutions
by equations (9) to (11)

   Evaluate f(x_i) by SVM / KNN

   if (rand >r_i ) then
       Select the global best solution (Gbest) among
the available best solutions
       Generate a local solution around the selected
best solution
   end
  Generate new solution randomly
  if ((rand <A_i )and (F(X_i )<F(Gbest)) then
    Accept the new solutions

    Increase r_i and reduce A_i
  end
  Find the current Gbest and Rank the Bat
end

return result and best position as selection features
```

Figure 1. pseudo-code of the proposed model

The stopping criterion of the BatKNN / BatSVM hybrid model is equal to maximum number of iterations. The 10-fold validation method was used to evaluate the best features selected by SVM and KNN classifiers. The fitness rate of each bat is calculated by evaluating the accuracy value of the support vector machine classification algorithms and the k-nearest neighbor.

The performance of the proposed model can be evaluated by different criteria. In this article, Accuracy, Precision and Recall indices are used (Equations 13-15), which TP, FP, TN and FN are the number of true positive, false positive, true negative, and false negative in data classification respectively (8).

Equation (13): $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation (14): $$\text{Precision} = \frac{TP}{TP + FP}$$

Equation (15): $$\text{Recall} = \frac{TP}{TP + FN}$$

Simulations are performed in MATLAB 2017 software and systematic with the following specifications.

- Intel Core i7-7700 @ 2.81GHz processor,
- 64 GB memory
- Windows10 OS configuration

Results

In this section, the results obtained from implementation of the proposed model have been presented on dataset and using two SVM and KNN classifiers in Table 3.

As can be seen in Table 3, the proposed model in the first scenario of the bat binary algorithm using SVM classifier in the best response has been obtained 96.25 accuracy value for classifying individuals at different health literacy levels. Also the best response of the second scenario using the KNN classifier is 93/76 accuracy value for classification of bats (individuals). The obtained results show the superiority of the first scenario (BatSVM) over the second scenario (BatKNN) in terms of classification accuracy. In the case of the selecting effective factors on health literacy, the results obtained from the model implementation on input data, indicate questions set of Q1 and Q2 scopes as two effective factors on the health

Table 3. Simulation results of proposed model

Proposed algorithm	The Number of iterations	The Number of bats	accuracy	precision	recall
BatSVM	25	20	96.25	97.92	96.43
	50		96.25	98.20	96.17
	100		95.69	97.99	95.53
	25	25	95.73	97.13	96.38
	50		96.16	97.48	96.66
	100		95.55	97.48	95.79
BatKNN	25	20	93.34	93.24	96.45
	50		93.53	95.47	64.67
	100		93.11	94.18	95.23
	25	25	93.58	95.18	95.00
	50		93.76	94.10	69.26
	100		93.11	94.82	94.66

literacy level of South Khorasan province.

Discussion

Generally, the statistical results obtained from the descriptive-analytical study used in this article showed poor level of adult health literacy in South Khorasan province. According to the results, more attention should be paid to community health literacy in “health promotion and hygiene education programs” and appropriate intervention programs should be codified and performed with collaboration of other sections such as mass media, educational environments, etc. to increase health literacy of community individuals. Also, with regard to the statistical results of the study that high percentage of participants are poor in accessing information resources (physicians and health staff, the Internet, educational brochures, etc.), it is necessary to think some measures in health promotion programs of individuals to increase individuals’ access to the information resources that is understandable, and following that increase community health literacy.

In this study, a smart model was also used to evaluate data related to health literacy level. The model is derived from the two hybrid algorithms BatKNN and BatSVM. In this method, we were seeking to find the best model to classify the data in the way that classify health literacy data into three categories: good, moderate, and poor. In the usual statistical methods, the classification of information related to the level of health literacy is achieved through answering to the questions in the questionnaire by the participants and earned scores by them. In contrast to the statistical method, our proposed model extracts a pattern that is capable of detecting a person’s health literacy level without awareness the acquired scores. In addition, the pattern extracted by the model also enable to detect the effective factors on the level of health literacy in the province

The pattern discovered by the two proposed hybrid models have identified two scopes of Q1 and Q2 that Q1 scope considers accessing to resources and Q2 scope considers resources use effect on health literacy level of South Khorasan Province among 9 scopes more important and more effective. As the statistical results presented in Table 1 show, the level of health literacy in South Khorasan province is in poor limit and with regard to the mean scores of different scopes, two scopes of Q1 and Q2 have lower mean values among the other scopes. Therefore, the statistical results confirm the results obtained from the proposed model.

In most studies performed in health literacy domain in global level, data have been analyzed using statistical techniques (22-24). The level of health literacy have been investigated in many internal researches, such as the study of Mohammadi et al. (25) and Zareban et al. (26). The results in these researches also show that analyses have been merely based on statistical methods. Effective factors on the health literacy of diabetic patients in Iran have been examined in (25) and the level of women's health literacy in Sistan&Baluchestan province has been evaluated in (26). A review study which has been conducted in 2018 (27) on a series of researches related to assessing health literacy level in Iran and more than 27 articles have been examined in this field. This study points to the lack of an effective strategy on improving community health literacy level. In most of mentioned studies, there is a lack of methods that enable to find effective factors on health literacy level. Only in limited cases, using the machine learning techniques have been observed in studies related to health literacy. In a research, Kandula et al. (28) in 2011 have used a classification algorithm to classify data related to health literacy level. In the present study, our proposed method also detect effective factors on adult health literacy level , in addition classifying data using machine learning algorithms,. This model is smart and does not require statistical analysis and in this respect is different from other performed works. There are no specific restrictions in this method over other similar datasets.

According to studies of the authors of this research, the use of smart methods in the field of health education in our country is very rare, and the results of the researches are often as numerical and statistical analyses, which have their particular limitations in discovering hidden patterns. In contrast, the proposed model is a smart tool for analyzing data related to health literacy that can analyze data in the least possible time without the need to user knowledge, and help to improve community health literacy with accelerating in decision-making of related professionals. In the future, it will also take advantage from using deep learning to identify hidden patterns to identify levels of community health literacy

Acknowledgments: The authors of this article thank and appreciate all those who contributed them in preparing this study.

Competing interests: The authors declare that they have no competing interests.

Funding: No financial support was received for this study

References

1. Lael-Monfared E, Tehrani H, Moghaddam ZE, Ferns GA, Tatari M, Jafari A. Health literacy, knowledge and self-care behaviors to take care of diabetic foot in low-income individuals: Application of extended parallel process model. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2019;13(2):1535-41. https://doi.org/10.1016/j.dsx.2019.03.008 PMid:31336518
2. Kickbusch I, Pelikan JM, Apfel F, Tsouros AD, World Health Organization, editors. Health literacy: the solid facts. Copenhagen: World Health Organization Regional Office for Europe; 2013. 73 p. (The solid facts).
3. Ahmadzadeh sani t, Vahedian-Shahroodi M, Tehrani H, Esmaily H. Relationship between health literacy and nutrition among middle-aged women. Journal of Health Literacy. 2019;3(4):9-15. https://doi.org/10.22038/jhl.2019.36772.1019
4. Khaleghi M, Amin Shokravi F, Peyman N. The Relationship Between Health Literacy and Health-Related Quality of Life in Students. Iranian Journal of Health Education and Health Promotion. 2019;7(1):66-73. https://doi.org/10.30699/ijhehp.7.1.66
5. Kaestle CF, Campbell A, Finn JD, Johnson ST, Mikulecky LJ. Adult Literacy and Education in America. Natl Cent Educ Stat Adult Lit Educ Am. 1999;288.
6. Domadiya N, Rao UP. Privacy Preserving Distributed Association Rule Mining Approach on Vertically Partitioned Healthcare Data. Procedia Comput Sci. 2019;148:303-12. https://doi.org/10.1016/j.procs.2019.01.023
7. Patel S, Patel H. Survey of Data Mining Techniques used in

Healthcare Domain. Int J Inf Sci Tech. 2016 Mar 31;6(1/2):53-60. https://doi.org/10.5121/ijist.2016.6206

8. vapnik A, Chervonenkis A. The necessary and sufficient conditions for consistency in the empirical risk minimization method. Pattern Recognit Image Anal. 1991;1(3):283-305.
9. Zhang J, Tittel FrankK, Gong L, Lewicki R, Griffin RJ, Jiang W, et al. Support Vector Machine Modeling Using Particle Swarm Optimization Approach for the Retrieval of Atmospheric Ammonia Concentrations. Environ Model Assess. 2015;21(4). https://doi.org/10.1007/s10666-015-9495-x
10. Mandal I. SVM-PSO based Feature Selection for Improving Medical Diagnosis Reliability using Machine Learning Ensembles. In: Computer Science & Information Technology (CS & IT) [Internet]. Academy & Industry Research Collaboration Center (AIRCC); 2012 [cited 2019 Jun 14]. p. 267-76. https://doi.org/10.5121/csit.2012.2326 PMid:22683053
11. Huang T-M, Kecman V, Kopriva I. Kernel Based Algorithms for Mining Huge Data Sets. Springer Berlin Heidelberg; 2006. 267 p.
12. Du J, Liu Y, Yu Y, Yan W. A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms. Algorithms. 2017 May 17;10(2):57. https://doi.org/10.3390/a10020057
13. Yang X-S. A New Metaheuristic Bat-Inspired Algorithm. In: González JR, Pelta DA, Cruz C, Terrazas G, Krasnogor N, editors. Nature Inspired Cooperative Strategies for Optimization (NICSO 2010) [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010 [cited 2019 Jun 14]. p. 65-74. Available from: http://link.springer.com/10.1007/978-3-642-12538-6_6
14. Diao R, Chao F, Peng T, Snooke N, Shen Q. Feature Selection Inspired Classifier Ensemble Reduction. IEEE Trans Cybern. 2014 Aug;44(8):1259-68. https://doi.org/10.1109/TCYB.2013.2281820 PMid:24108493
15. Krishnapuram B, Harternink AJ, Carin L, Figueiredo MAT. A Bayesian approach to joint feature selection and classifier design. IEEE Trans Pattern Anal Mach Intell. 2004 Sep;26(9):1105-11. https://doi.org/10.1109/TPAMI.2004.55 PMid:15742887
16. Guodong Guo, Dyer CR. Simultaneous feature selection and classifier training via linear programming: a case study for face expression recognition. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003 Proceedings [Internet]. Madison, WI, USA: IEEE Comput. Soc; 2003 [cited 2020 Jan 13]. p. I-346-I-352. Available from: http://ieeexplore.ieee.org/document/1211374/
17. Koumi F, Aldasht M, Tamimi H. Efficient Feature Selection using Particle Swarm Optimization: A hybrid filters-wrapper Approach. In: 2019 10th International Conference on Information and Communication Systems (ICICS) [Internet]. Irbid, Jordan: IEEE https://doi.org/10.1109/IACS.2019.8809133
18. Sindhu R, Ngadiran R, Yacob YM, Hanin Zahri NA, Hariharan M, Polat K. A Hybrid SCA Inspired BBO for Feature Selection Problems. Math Probl Eng. 2019 Apr 2;2019:1-18. https://doi.org/10.1155/2019/9517568
19. Alam WU. Improved Binary Bat Algorithm for Feature Selection. Abo Akademi University; 2019.
20. Taha AM, Mustapha A, Chen S-D. Naive Bayes-Guided Bat Algorithm for Feature Selection. Sci World J. 2013; 2013:1-9. https://doi.org/10.1155/2013/325973 PMid:24396295 PMCid:PMC3874950
21. Nakamura RYM, Pereira LAM, Costa KA, Rodrigues D, Papa JP, Yang X-S. BBA: A Binary Bat Algorithm for Feature Selection. In: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images [Internet]. Ouro Preto, Brazil: IEEE https://doi.org/10.1109/SIBGRAPI.2012.47 PMid:21383680
22. Marimwe C, Dowse R. Health literacy test for limited literacy populations (HELT-LL): Validation in South Africa. Mo P, editor. Cogent Med [Internet]. 2019 Aug 8 [cited 2020 Jan 9];6(1). https://doi.org/10.1080/2331205X.2019.1650417
23. Rademakers J, Heijmans M. Beyond Reading and Understanding: Health Literacy as the Capacity to Act. Int J Environ Res Public Health. 2018 Aug 7;15(8):1676. https://doi.org/10.3390/ijerph15081676 PMid:30087254 PMCid:PMC6121645
24. Schrauben SJ, Wiebe DJ. Health literacy assessment in developing countries: a case study in Zambia. Health Promot Int. 2015 Dec 10; dav108. https://doi.org/10.1093/heapro/dav108 PMid:26660155
25. Mohammadi Z, Banihashemi AT, Asgharifard H, Bahramian M, Baradaran HR, Khamseh ME. Health literacy and its influencing factors in Iranian diabetic patients. Med J Islam Repub Iran. 2015;29:6.
26. Zareban I, Izadirad H, Jadga K. Evolution of health status and health services utilization in women in Baluchistan region of Iran. J Health Lit. 2016;1(2):71-82. https://doi.org/10.20286/jech-02036
27. Dadipoor S, Ramezankhani A, Aghamolaei T, Rakhshani F, Safari-Moradabadi A. Evaluation of health literacy in the Iranian population. Health Scope. 2018;7(3):e62212. https://doi.org/10.5812/jhealthscope.62212
28. Kandula S, Ancker JS, Kaufman DR, Currie LM, Zeng-Treitler Q. A new adaptive testing algorithm for shortening health literacy assessments. BMC Med Inform Decis Mak. 2011 Dec;11(1):52. https://doi.org/10.1186/1472-6947-11-52 PMid:21819614 PMCid:PMC3178473