Arch Iran Med. October 2018;21(10):460-465

http www.aimjournal.ir

ARCHIVES OF IRANIAN MEDICINE

Open Access

Original Article

The Prediction of Obstructive Sleep Apnea Using Data Mining Approaches

Zohreh Manoochehri, MSc[1]; Mansour Rezaei, PhD[2*]; Nader Salari, PhD[3]; Habibolah Khazaie, PhD[4]; Behnam Khaledi Paveh, MSc[5]; Sara Manoochehri, MSc[1]

[1]Student Research Committee, Kermanshah University of Medical Sciences, Kermanshah, Iran
[2]Department of Biostatistics, Social Development and Health Promotion Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran
[3]Department of Biostatistics, School of Nursing and midwifery, Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran
[4]Sleep Disorders Research Center, Kermanshah University of Medical Sciences Kermanshah, Iran
[5]Sleep Disorders Research Center, Kermanshah University of Medical Sciences Kermanshah, Iran

Abstract

Background: Obstructive sleep apnea (OSA) which is the most common sleep disorder breathing (SDB), imposes heavy costs on health and economy. The aim of this study was to provide models based on data mining approaches (C5.0 decision tree and logistic regression model [LRM]) and choose a top model for predicting OSA without polysomnography (PSG) devices that is a standard method for diagnosis of this disease, to identify patients with this syndrome payment.

Methods: In this cross sectional study, data was extracted from the medical records of 333 patients with sleep disorders who were referred to sleep disorders research center of Kermanshah University of Medical Sciences during the years 2012–2016. All patients underwent one night PSG. A stepwise LRM was fitted and its performance was compared with C5.0 decision tree with use of the criteria of accuracy, sensitivity and specificity.

Results: For C5.0 decision tree, accuracy was obtained 0.757 with 95% confidence interval (0.661, 0.838), sensitivity was 0.66 and specificity was 0.809. For LRM, these items were obtained 0.737 (0.639, 0.820), 0.693 and 0.78 respectively.

Conclusion: C5.0 decision tree showed better performance than LRM in diagnosis of OSA. So this model can be considered as an alternative approach for PSG.

Keywords: C5.0 Decision tree, Logistic regression, Obstructive Sleep apnea, Polysomnography, Sleep disorders

Cite this article as: Manoochehri Z, Rezaei M, Salari N, Khazaie H, Khaledi Paveh B, Manoochehri S. The prediction of obstructive sleep apnea using data mining approaches. Arch Iran Med. 2018;21(10):460–465.

Received: December 9, 2017, Accepted: May 26, 2018, ePublished: October 1, 2018

Introduction

Every living creature, in order to survive, needs to sleep for some hours a day. On average, every human being spends one third of his/her life sleeping. Meanwhile, lack of sleep can cause many problems including physical, emotional and psychological disorders. So far, a variety of sleep-related disorders are known, among which are insomnia, sleep disorder breathing (SDB) and parasomnia.[1] Obstructive sleep apnea (OSA) is the most common SDB that occurs during sleep and is caused by repeated partial or complete collapse of upper airway.[2,3] One study found that OSA affects 2%–26% of general population.[4] OSA may cause obesity, daytime sleepiness, high blood pressure, heart disease, and type II diabetes.[5-7] In addition to the adverse effects on health, OSA also imposes heavy burdens on economy, so that its overall impact on economy is much higher than its direct medical costs.[8] Furthermore, undiagnosed OSA causes a large proportion of deaths due to traffic and harmful industrial accidents.[3] More than 80% of patients with moderate to severe OSA have not been diagnosed while they are in need of medical treatment.[9] Therefore, OSA can be regarded as a major concern in the field of public health.[4] A standard way to diagnose this disease is to use polysomnography (PSG) device which records information related to electroencephalography (EEG), electromyography (EMG), electrocardiography (ECG), electroochlography (EOG), oxygen saturation, sound of snoring, airstream, and respiratory effort (chest and abdominal movements). To do so, the patients are required to spend one night in PSG laboratory.[10] But most patients with sleep disorders do not have the feasibility of using this device, because on one hand, there is a limited number of centers equipped with PSG devices and these centers accept a limited number of patients.[11] On the other hand, patients have to deal with a costly, time-consuming, and complex problem.[12] Therefore, due to the growing demand of people who need OSA diagnosis, use of simple alternative diagnostic methods seems necessary. One of these methods is classification which makes diagnosis by dividing the subjects into two groups: healthy and unhealthy. Classification is one of the main tasks in the field of data mining.[13] Full realization of the data potential requires algorithms that automatically derive the existing data pattern and lead to discovery of hidden knowledge within them. Accordingly, data mining as a tool

***Corresponding Author:** Mansour Rezaei, PhD; Department of Biostatistics, School of Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. Tel: 0833-8263048, Email: rezaei39@yahoo.com

which combines different methods including statistical analysis, machine learning methods, artificial intelligence, and data management systems, is responsible for this task.[14,15] So far, different classification algorithms have been added to the literature of this scope such as artificial neural networks, Support Vector Machine, decision tree, Bayesian methods, K-nearest neighbor, and logistic regression.[15] However, to date, no research has been carried out to predict OSA based on the clinical characteristics of patients using LRM and C5.0 decision tree algorithm, and compare their results. Therefore, in this study, we chose these two models fit to data obtained from clinical characteristics, and compared their performance using the criteria such as accuracy, sensitivity and specificity, in order to introduce the model with more robust results as an alternative method in predicting OSA.

Materials and Methods

Study Population

This cross sectional study used data of 333 patients who were referred to Sleep Disorders Research Center at Kermanshah University of Medical Sciences during 2012–2016. As the inclusion criteria, all patients should complain about sleep problems and spend one night in a PSG lab. This study used no exclusion criteria. Information required for each patient was taken from clinical records archives; questionnaires including information about age, sex, body mass index, neck circumference, waist circumference, tea consumption, cigarette consumption, underlying disease(s) (including hypertension, chronic headaches, heart disease, respiratory disease, neurological disease, and diabetes); Berlin Questionnaire score (low risk, high risk); Epworth Sleepiness scale score; Global Sleep Assessment Questionnaire (GSAQ) (snoring status, breathing stop in sleep, feeling of sadness or anxiety); and finally information obtained from PSG device, that determine the status of OSA (: with or without OSA) as response variable.[16]

Berlin questionnaire consists of 3 parts. The first part evaluates snoring, the second part measures the amount of fatigue and daily sleepiness, and information such as blood pressure and body mass index are presented in the third part. If the score is positive in at least two parts, then the individual is considered in the high-risk group for OSA.[17] In the current study, we used the same version that was applied by Amra et al.[18]

The Epworth Sleepiness Scale is a questionnaire containing of 8 items with a 4-point likert scale that is used to measure daytime sleepiness. The total score of this questionnaire varies from 0 to 24. Scores more than 11 indicate excessive daytime sleepiness (high-risk for OSA).[19] In the current study, we used the Iranian version of the ESS.[20]

Global Sleep Assessment Questionnaire (GSAQ) is a standard questionnaire including 11 items. GSAQ is used to screen sleep disorders such as insomnia, OSA and restless leg syndrome. Each item in this questionnaire examines the abundance of symptoms of sleep disorders over the past four weeks with response options: 'never', 'sometimes', 'usually', and 'always'. The reliability for this questionnaire has been reported 0.51 to 0.92.[21]

Statistical Analysis

Logistic Regression Model

Logistic regression model (LRM) is one of the main approaches in the field of classification which is based on theories of probability and statistics.[22] In this study, LRM is used as a model to identify patients with OSA and subjects without this syndrome. For this purpose, Y_i represents the target class and takes two values of 0 and 1. If, $Y_i = 1$ then the subject belongs to the group of patients with OSA; otherwise, he/she can be assigned to the group of subjects without OSA syndrome. Also, $\hat{\pi}_i = P(Y_i = 1)$ represents the probability of success or belonging to OSA class. In this case, logit probability of success is in linear relationship with explanatory variables. Here, parameters are estimated using maximum likelihood method. If the calculated probability is greater than a predetermined amount (usually 0.5), the subject belongs to the group of patients with OSA.[23]

Decision Tree Algorithm

Decision tree algorithm is one of the most common methods of data mining which is used as a precise classification tool.[24] Decision tree algorithm has a flowchart-like structure constructed of a root node (topmost node in the tree), a leaf node (which labels the target class), and an internal node (labeled with test condition).[25] Results of the rules made by decision tree are explained with logical "if" and "then" conditions. To derive the necessary rules, a path should be followed from the root node to the leaf node.[26] Decision tree algorithm has some advantages over other data mining methods: 1) easy to understand for user; 2) easier applicability; 3) usability in a variety of numerical and grouped data and 4) requiring less prior information. Decision tree has different types. The most common types are Quintal's regression trees (ID3, C4.5 and C5.0) and Breiman's Classification and Regression Tree (CART).[27] C5.0 decision tree, proposed by Quinaln in 1987, is a modified version of ID3 and C4.5 algorithms.[28,29] C5.0 decision tree has two main advantages over its two previous algorithms: first, it generates more precise rules, and second, it requires much less time to generate these rules.[30] In order to avoid overfitting learning problems, decision tree algorithm uses pruning techniques. To do so, C5.0 algorithm employs post-pruning method in which branches are pruned after full growth of the tree.[31]

After collecting data and excluding incomplete information, almost 70% of the total sample (234 patients) was considered for training purpose and the remaining 30% (99 patients) was used for testing the models. Data were analyzed in R3.2.2 software environment. In order to build a C5.0 decision tree, C50 package was installed and evaluation criteria's of the fitted models were measured by installing CARET package. LRM was fitted to data obtained from the training group using stepwise backward elimination method. Then, its performance was compared with C5.0 decision tree using criteria's accuracy, sensitivity and specificity (Eqs. 1–3) on test data.

$$\text{Acc}=\frac{TP+TN}{N}=\frac{TP}{N}+\frac{TN}{N} \quad (1)$$

$$\text{Sensitivity}=\frac{TP}{TP+FN} \quad (2)$$

$$\text{Specificity}=\frac{TN}{FP+TN} \quad (3)$$

Results

From a total of 333 subjects, 208 subjects were definitely diagnosed with OSA disease by PSG device and 125 subjects were diagnosed to be healthy. The mean age of the patients with OSA (46.64 ± 13.11) was more than the patients without OSA (38.66 ± 14.83). The average body mass index of patients with OSA (28.73 ± 4.57) was significantly higher than patients without OSA (25.59 ± 4.53) (*P* value < 0.001). 70.7% of patients with OSA were male, while only 29.3% of them were female (Tables 1 and 2).

In order to fit LRM, the stepwise backward elimination method was used to select variables in the final model. With this method, a group of six variables (age, neck circumference, hypertension, snoring status, feelings of anxiety or sadness, and Epworth Sleepiness Scale score) were selected to be entered into the model (Table 3).

In general, a total of 11 rules were generated with C5.0 decision tree (Table 4).

Among the studied variables, waist circumference, snoring status, sex, sleep apnea, Epworth Sleepiness Scale score, and neck circumference, respectively, had the greatest

Table 1. Comparison of Means of Quantitative Variables Between Healthy and Sick Groups

Variables	Groups		Values of Test Statistics	*P* Value
	Without OSA	With OSA		
Age (y)	14.83 ± 38.66	13.11 ± 46.64	T = -5.113	<0.001*
BMI (kg/m²)	4.53 ± 25.59	4.57 ± 28.73	T = -6.090	<0.001*
Neck circumference (cm)	3.53 ± 35.58	3.68 ± 39.17	T = -8.729	<0.001*
Waist circumference (cm)	16.54 ± 90.82	11.99 ± 100.56	T = -6.205	<0.001*
Cigarette consumption (No.)	4.97± 1.39	5.20 ± 1.73	U = 12579.5	0.376
Tea consumption (No.)	3.17 ± 3.52	3.41 ± 3.95	U = 11630.5	0.103
Epworth Sleepiness Scale score	4.79 ± 4.96	5.36 ± 7.94	U = 8630.5	<0.001*

BMI, body mass index; OSA, obstructive sleep apnea.
T is an abbreviation for *t* test statistic and U is an abbreviation for Mann-Whitney U test statistic.
Sign *Indicates significance level of 0.05.

Table 2. Relationship Between Qualitative Variables With Response Variable (OSA)

Variables		Group						*P* Value
		Without OSA		With OSA		Total		
		No.	%	No.	%	No.	%	
Sex	Female	68	54.4	61	29.3	129	38.7	<0.001*
	Male	57	45.6	147	70.7	204	61.3	
Hypertension	Yes	14	11.2	49	23.6	63	18.9	0.005*
	No	111	88.8	159	76.4	270	81.1	
Chronic headaches	Yes	14	11.2	25	12	39	11.7	0.822
	No	111	88.8	183	88	294	88.3	
Heart disease	Yes	11	8.8	23	11.1	34	10.2	0.510
	No	114	91.2	185	88.9	299	89.8	
Respiratory disease	Yes	6	4.8	15	7.2	21	6.3	0.381
	No	119	95.2	193	92.8	312	93.7	
Neurological disease	Yes	0	0	5	2.4	5	1.5	0.081
	No	125	100	203	97.3	328	98.5	
Diabetes	Yes	0	0	9	4.3	9	2.7	0.018*
	No	125	100	199	95.7	324	97.3	
Berlin Questionnaire score	Yes	90	72	81	38.9	171	51.4	<0.001*
	No	35	28	127	61.1	162	48.6	
Snoring status	Yes	25	20	126	60.6	151	45.3	<0.001*
	No	100	80	82	39.4	182	54.7	
Breathing stop in sleep (GSAQ)	Yes	14	11.2	80	38.5	94	28.2	<0.001*
	No	111	88.8	128	61.5	239	71.8	
Feelings of sadness or anxiety	Yes	80	64	99	47.6	179	53.8	0.004*
	No	45	36	109	52.4	154	46.2	

OSA, obstructive sleep apnea.
Sign *Indicates significance level of 0.05.

Table 3. Information for Fitting LRM With Stepwise Method

No.	Variable	B	SE(B)	Wald Statistics	OR	*P* Value
1	Constant	-8.645	2.027	-4.265	0.0001	<0.001*
2	Age	0.044	0.014	3.045	1.045	0.002*
3	Neck circumference	0.191	0.055	3.488	1.210	<0.001*
4	Hypertension	-1.235	0.572	-2.159	0.291	0.031*
5	Tea consumption	-0.111	0.065	-1.695	0.895	0.09
6	Chronic headaches	1.023	0.562	1.819	2.781	0.069
7	Snoring status	1.916	0.427	4.485	6.796	<0.001*
8	Feelings of anxiety or sadness	-1.202	0.374	-3.215	0.3	0.001*
9	Epworth Sleepiness Scale score	0.128	00.039	3.309	1.137	<0.001*

OR, odds ratio; SE, Standard error.
Sign * indicates significance level of 0.05.

Table 4. Rules Generated by C5.0 Decision Tree

No.	Generated Rules
1	If the subject has waist size ≤ 74, then the possibility of belonging to the group without OSA is 92%.
2	If the subject is male who snores during sleep with BMI ≤ 27.7 and Epworth Sleepiness Scale score ≤ 4, then the possibility of belonging to the group without OSA is 87%
3	If the subject has neck circumference >31, waist size ≤84, and Epworth Sleepiness Scale score ≤ 7, then the possibility of belonging to the group without OSA is 86%.
4	If the subject has high blood pressure, Epworth Sleepiness Scale score ≤ 4, and snores during sleep, then the possibility of belonging to the group without OSA is 86%.
5	If the subject's age ≤ 27, waist size>84, and Epworth Sleepiness Scale score ≤ 4, then the possibility of belonging to the group without OSA is 83%.
6	If the subject's age ≤ 17, waist size>84, and does not snore during sleep, then the possibility of belonging to the group without OSA is 80%.
7	If the subject does not snore during sleep, has feelings of anxiety or sadness, Sleepiness Scale score ≤ 4 and has low risk (according to the Berlin Questionnaire score), then the possibility of belonging to the group without OSA is 86%.
8	If the subject has 41< neck circumference ≤ 42 and does not snore during sleep, then the possibility of belonging to the group without OSA is 80%.
9	If the subject is male who does not snore during sleep, and without breathing stop in sleep (according to the GSAQ), then the possibility of belonging to the group without OSA is 74%.
10	If the subject is male with age > 44, neck circumference > 34.5, without blood pressure and breathing stop in sleep (according to the GSAQ), then the possibility of belonging to the group with OSA is 92%.
11	If the subject has waist size > 73, then the possibility of belonging to the group with OSA is 65%.

BMI, Body mass index; OSA, obstructive sleep apnea; GSAQ, Global Sleep Assessment Questionnaire.

effects on rules generated by C5.0 decision tree. Also, according to McNemar test results, there was no statistically significant difference between groups derived from C5.0 model and actual groups (subjects with/without OSA) (*P* value=1). This happened for LRM, as well (*P* value = 0.556). After fitting LRM and C5.0 decision tree on training data (n=234 patients), then for determination of performance of these models, we used the test data (n=99 patients). Then the performance of these models, for classify healthy and unhealthy individuals, were determined (Table 5). For the results to be more valid, we must first fit the models on training data, and then compare performance on test data. Using the results of test data, we determined the criteria for evaluating the models' performance (Table 6).

It should be noted that, according to McNemar test results, In terms of predicted class for healthy and unhealthy individuals there was a significant difference between performance of C5.0 and LRM (*P* value <0.001).

Discussion

In recent years, several studies have been conducted in order to achieve an applicable method as an alternative for PSG device which is a standard way to diagnose OSA. In some of these studies, classification methods, which are among non-parametric methods in the field of data mining and machine learning, have been used to diagnose this disease. The main advantage of this approach is that, when using machine learning methods, unlike parametrical statistical methods, it is not necessary to establish a set of assumptions (such as normality of distribution or equality of variance and etc).[32,33] In most of the studies which used classification models for predicting OSA disease, information obtained by different electrocardiogram signals is used for prediction of OSA.[34] Since extracting this information is both costly and time-consuming, in this study clinical features that are easily accessible using questionnaires and patients' medical records are employed as independent input variables for LRM and

Table 5. Matrix Combined Results for LRM and C5.0 on Test Data (n = 99)

Actual Class	Predicted Class by LRM			Predicted Class by C5.0		
	Positive	Negative	Total	Positive	Negative	Total
Positive	TP = 34	FN = 15	49	TP = 24	FN = 12	36
Negative	FP = 11	TN = 39	50	FP=12	TN = 51	63
Total	45	54	99	36	63	99

TP, true positive; TN, True negative; FP, False positive; FN, False negative; LRM, logistic regression model; C5.0= C5.0 decision tree.

Table 6. The Results of the Assessment Classifier Models LRM and C5.0

Model	Accuracy	95% CI for Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
LRM	0.737	(0.639,0.820)	0.693	0.78	0.756	**0.722**
C5.0	0.757	(0.661,0.838)	0.666	0.809	0.667	**0.809**

LRM, logistic regression model; C5.0, C5.0 decision tree.

C5.0 algorithm. Based on the obtained results, with regard to accuracy and specificity criteria, comparing the 2 reviewed models, C5.0 has better diagnosis performance than LRM in predicting OSA (dependent variable). Superiority of this model has been mentioned in some of the previous studies. In a research carried out by Hui Meng, performance of three data mining methods (LRM, artificial neural network, and C5.0) in predicting diabetes or pre-diabetes risk factors were compared. Among these methods, C5.0 had the best performance with accuracy of 77.87%, sensitivity of 80.68% and specificity of 75.13%.[35] Also, results of a study conducted by Delen et al for predicting breast cancer survival using three data mining methods including C5.0, LRM, and artificial neural network, indicated that C5.0 had superiority over other 2 models. In the current research, accuracy of decision tree and LRM was 93.6% and 89.2%, respectively. Acceptable accuracy obtained from this study can be attributed to employing either large number of cases (about 200 000 cases) or well-defined data[27]. Although accuracy and specificity of test data in C5.0 algorithm were greater than LRM, sensitivity for LRM (0.693) was obtained more than C5.0 algorithm (0.666). This means that, in this study, LRM had more power in correct diagnosis of patients with OSA. The result of this study is consistent with the result of a study conducted by Fernandez et al in which risk factors for infectious inflammation of the breast were predicted. In the aforementioned study, area under ROC curve was similar for LRM and decision tree model; however, LRM had better sensitivity than decision tree algorithm.[36] Although in most of the previous studies, as well as the current study, C5.0 model shows better diagnostic capability than LRM, there are cases where not only LRM does not have lower accuracy, but also it is preferred to decision tree model. For example, in a research which was carried out to predict bankruptcies of listed companies in Tehran Stock Exchange, performance of these two models were compared and the results showed that area under ROC curve for LRM was more than decision tree model.[37] Based on the results of this study, it can be concluded that C5.0 decision tree has exhibited better performance than LRM in diagnosis of OSA. Since one of the limitations of the current study was its small sample size, it is suggested that after implementing C5.0 decision tree model on bigger samples to achieve greater accuracy, this model be used as an alternative for PSG device to identify patients with OSA.

In conclusion, comparing two studied models, C5.0 algorithm with accuracy and specificity of 0.757 and 0.809, performed better than LRM with accuracy and specificity of 0.737 and 0.78, respectively. Therefore, C5.0 decision tree can be used as an alternative approach for PSG in diagnosis of patients with OSA.

Authors' Contribution

ZM, MR, NS, and HK contributed to designing the study, ZM, SM, and BK collected the data, and analyzed by MR, NS, and ZM. The final report and manuscript were written by ZM and MR. All the authors read and approved the version for submission.

Conflict of Interest Disclosures

The authors have no conflicts of interest.

Ethical Statement

The study was approved by research ethics committee of Kermanshah University of Medical Sciences. The written informed consent was obtained from all the participants. The authors declare there are no competing interests. The study was funded by Kermanshah University of Medical Sciences (grant No. 94507).

Acknowledgment

The author wishes to thank the personnel of the Research Center for Sleep Disorders for their valuable collaboration in data collection, and gratefully acknowledge the Research Council of Kermanshah University of Medical Sciences for the financial support (Grant Number: 94507). This study was performed in partial fulfillment of the requirements for postgraduate of Zohreh Manoochehri, in Department of Biostatistics, Faculty of Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran

References

1. Sadock BJ, Sadock VA, Ruiz P. Kaplan and Sadock's Comprehensive Textbook of Psychiatry. 9th ed. New York: Lippincott Williams & Wilkins; 2009.
2. van Houwelingen KG, van Uffelen R, van Vliet AC. The sleep apnoea syndromes. Eur Heart J. 1999;20(12):858-66. doi: 10.1053/euhj.1998.1484.
3. Qureshi A, Ballard RD, Nelson HS. Obstructive sleep apnea. J Allergy Clin Immunol. 2003;112(4):643-51. doi: 10.1016/j.jaci.2003.08.031.
4. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun

Y. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. Br J Anaesth. 2012;108(5):768-75. doi: 10.1093/bja/aes022.
5. Somers VK, White DP, Amin R, Abraham WT, Costa F, Culebras A, et al. Sleep apnea and cardiovascular disease: an American Heart Association/American College of Cardiology Foundation Scientific Statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council on Cardiovascular Nursing. J Am Coll Cardiol. 2008;52(8):686-717. doi: 10.1016/j.jacc.2008.05.002.
6. Kasai T, Floras JS, Bradley TD. Sleep apnea and cardiovascular disease: a bidirectional relationship. Circulation. 2012;126(12):1495-510. doi: 10.1161/circulationaha.111.070813.
7. Shaw JE, Punjabi NM, Wilding JP, Alberti KG, Zimmet PZ. Sleep-disordered breathing and type 2 diabetes: a report from the International Diabetes Federation Taskforce on Epidemiology and Prevention. Diabetes Res Clin Pract. 2008;81(1):2-12. doi: 10.1016/j.diabres.2008.04.025.
8. Hillman DR, Murphy AS, Pezzullo L. The economic cost of sleep disorders. Sleep. 2006;29(3):299-305. doi: 10.1093/sleep/29.3.299.
9. Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep. 1997;20(9):705-6.
10. Kryger M, Roth T, Dement W. Principles and practice of sleep medicine. 5th ed. Philadelphia; WB Saunders Company; 2011:291-302.
11. Gagnadoux F, Pelletier-Fleury N, Philippe C, Rakotonanahary D, Fleury B. Home unattended vs hospital telemonitored polysomnography in suspected obstructive sleep apnea syndrome: a randomized crossover trial. Chest. 2002;121(3):753-8. doi: 10.1378/chest.121.3.753.
12. Pang KP, Terris DJ. Screening for obstructive sleep apnea: an evidence-based analysis. Am J Otolaryngol. 2006;27(2):112-8. doi: 10.1016/j.amjoto.2005.09.002.
13. Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Syst Appl. 2014;41(4):1476-82. doi: 10.1016/j.eswa.2013.08.044.
14. Venkatadri M, Reddy LC. A Review on Data mining from Past to the Future. Int J Comput Appl. 2011;15(7):19-22.doi: 10.5120/1961-2623.
15. Tomar D, Agarwal S. A survey on Data Mining approaches for Healthcare. Int J Biosci Biotechnol. 2013;5(5):241-66. doi: 10.14257/ijbsbt.2013.5.5.25.
16. Manoochehri Z, Salari N, Rezaei M, Khazaie H, Manoochehri S, Khaledi Pavah B. Comparison of support vector machine based on genetic algorithm with logistic regression to diagnose obstructive sleep apnea. J Res Med Sci. 2018;23:65. doi: 10.4103/jrms.JRMS_357_17.
17. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med. 1999;131(7):485-91. doi: 10.7326/0003-4819-131-7-199910050-00002.
18. Amra B, Nouranian E, Golshan M, Fietze I, Penzel T. Validation of the persian version of berlin sleep questionnaire for diagnosing obstructive sleep apnea. Int J Prev Med. 2013;4(3):334-9.
19. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991;14(6):540-5. doi: 10.1093/sleep/14.6.540.
20. Sadeghniiat Haghighi K, Montazeri A, Khajeh Mehrizi A, Aminian O, Rahimi Golkhandan A, Saraei M, et al. The Epworth Sleepiness Scale: translation and validation study of the Iranian version. Sleep Breath. 2013;17(1):419-26. doi: 10.1007/s11325-012-0646-x.
21. Roth T, Zammit G, Kushida C, Doghramji K, Mathias SD, Wong JM, et al. A new questionnaire to detect sleep disorders. Sleep Med. 2002;3(2):99-108. doi: 10.1016/s1389-9457(01)00131-9.
22. Peng H, Liang D, Choi C. Evaluating parallel logistic regression models. Silicon Valley, CA, USA: Proc IEEE Int Conf Big Data; 2013:119-26. doi: 10.1109/BigData.2013.6691743.
23. Chang CD, Wang CC, Jiang BC. Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. Expert Syst Appl. 2011;38(5):5507-13. doi: 10.1016/j.eswa.2010.10.086.
24. Farid DM, Zhang L, Rahman CM, Hossain MA, Strachan R. Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. Expert Syst Appl. 2014;41(4):1937-46. doi: 10.1016/j.eswa.2013.08.089.
25. Bhargava N, Sharma G, Bhargava R, Mathuria M. Decision tree analysis on j48 algorithm for data mining. Proc Int J Adv Res Comput Sci softw Eng. 2013;3(6):1114-9.
26. Ameri H, Alizadeh S, Barzegari A. Knowledge Extraction of Diabetics' Data by Decision Tree Method. J Health Adm. 2013;16(53):58-72.
27. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med. 2005;34(2):113-27. doi: 10.1016/j.artmed.2004.07.002.
28. Sharma S. Cervical Cancer stage prediction using Decision Tree approach of Machine Learning. Int J Adv Res Comput Communi Eng.2016;5(4):345-8.doi:10.17148/IJARCCE.2016.5488.
29. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1(1):81-106. doi: 10.1007/bf00116251.
30. Is See5/C5.0 Better Than C4.5? 2009. http://rulequest.com/see5-comparison.html.
31. Yeh JY, Wu TH, Tsao CW. Using data mining techniques to predict hospitalization of hemodialysis patients. Decis Support Syst. 2011;50(2):439-48. doi: 10.1016/j.dss.2010.11.001.
32. Stylianou N, Akbarov A, Kontopantelis E, Buchan I, Dunn KW. Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. Burns. 2015;41(5):925-34. doi: 10.1016/j.burns.2015.03.016.
33. Biglarian A, Bakhshi E, Rahgozar M, Karimloo M. Comparison of artificial neural network and logistic regression in predicting of binary response for medical data The stage of disease in gastric cancer. J North Khorasan Univ Med Sci. 2011;3(5-S5):15-23. doi: 10.29252/jnkums.3.5.S5.15.
34. Almazaydeh L, Elleithy K, Faezipour M. Obstructive sleep apnea detection using SVM-based classification of ECG signal features. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:4938-41. doi: 10.1109/embc.2012.6347100.
35. Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. Kaohsiung J Med Sci. 2013;29(2):93-9. doi: 10.1016/j.kjms.2012.08.016.
36. Fernandez L, Mediano P, Garcia R, Rodriguez JM, Marin M. Risk Factors Predicting Infectious Lactational Mastitis: Decision Tree Approach versus Logistic Regression Analysis. Matern Child Health J. 2016;20(9):1895-903. doi: 10.1007/s10995-016-2000-6.
37. Hosseini SM, Rashidi Z. Bankruptcy Prediction of Companies listed Corporations in Tehran Stock Exchange by Using Decision Tree and Logistic Regression. Journal of Financial Accounting Research. 2013;5(17):105-30.