ORIGINAL PAPER

The Usage of Association Rule Mining to Identify Influencing Factors on Deafness After Birth

Azimeh Danesh Shahraki, Reza Safdari, Hamid Habibi Gahfarokhi, Shahram Tahmasebian

Tehran University of Medical Sciences

Corresponding author: Azimeh Danesh. Tehran University of Medical Sciences

ABSTRACT

Background: Providing complete and high quality health care services has very important role to enable people to understand the factors related to personal and social health and to make decision regarding choice of suitable healthy behaviors in order to achieve healthy life. For this reason, demographic and clinical data of person are collecting, this huge volume of data can be known as a valuable resource for analyzing, exploring and discovering valuable information and communication. This study using forum rules techniques in the data mining has tried to identify the affecting factors on hearing loss after birth in Iran. Materials and Methods: The survey is kind of data oriented study. The population of the study is contained questionnaires in several provinces of the country. First, all data of questionnaire was implemented in the form of information table in Software SQL Server and followed by Data Entry using written software of C # .Net, then algorithm Association in SQL Server Data Tools software and Clementine software was implemented to determine the rules and hidden patterns in the gathered data. Findings: Two factors of number of deaf brothers and the degree of consanguinity of the parents have a significant impact on severity of deafness of individuals. Also, when the severity of hearing loss is greater than or equal to moderately severe hearing loss, people use hearing aids and Men are also less interested in the use of hearing aids. Conclusion: In fact, it can be said that in families with consanguineous marriage of parents that are from first degree (girl/boy cousins) and 2nd degree relatives (girl/boy cousins) and especially from first degree, the number of people with severe hearing loss or deafness are more and in the use of hearing aids, gender of the patient is more important than the severity of the hearing loss.

Key words: Data mining, Text mining, Clinical Data mining, Data mining Deafness, Deafness, Hearing Loss, Deafness recognition pattern, SQL Server Data Tools, Clementine, Binary Classifier.

1. INTRODUCTION

Since hearing loss is the most common sensory defects in human and one in a thousand babies with hearing loss before they speak (1), can be concluded that the number of cases for this disease has been great around the world and in our country too and are facing with large volumes the data related to the disease hearing loss. With the development of technology, information technology with own achievements science has come to the aid of medical and data mining technology can be used to obtain new knowledge between huge amount of raw data of deafness (2, 3, 4).

In the early 1990s, knowledge discovery in databases and word processing was invented for the first time and since 1995, data mining was entered seriously in the statistic discussions (5). On the other hand a lot of data generated and stored every day. Also, the increasing speed of computers has been caused to create algorithms that have high power analyze without limitation in the capacity and speed of computers (6) that both of questions has been created the possibility of knowledge discovery from large databases (7). Information age is activated many organizations to collect large volumes of data. However, the data is useful when "meaningful information" or "knowledge" can be extracted from it (8). The increase of data has been caused to the rise of new opportunities in business. The field of data mining and knowledge discovery from the database has emerged as a new discipline in engineering and computer science (9). Different definitions has been provided by

different people of data mining but definition that mentioned and shared in most of the references, is to: extract information and knowledge, and discover hidden patterns from large and complex databases (7). Due to the fact that human medical data are the most valuable and sensitive data for exploration and analysis, analysis and knowledge of them should be done at high degree of accuracy and sensitivity (8) and is therefore one of the most important fields of medicine in which data mining is used. The widespread use of medical information systems and the explosive growth of medical databases for become more efficient have had the needs to traditional data analysis using computer-assisted analysis (9). The aim of this study was also searching for hidden relationships in a complete database of information of deafness thorough the country. Mainly the two main purpose of data mining technology are prediction and description. Predictive data mining using data sets product the models to explain the system that can be used to predict the performance of different variables. Therefore, the purpose of predictive data mining is generated the model that using of performance code, done functions as predict, ranking, estimate (10). Descriptive data mining generated new and non-trivial information based on the available data sets that described behavior patterns of variables. The purpose of descriptive data mining is obtained a complete understanding of the system under study using hidden patterns of it and relationships within data sets (11). So far, in the field of medical has done significant activities of Knowledge discovery and biological data of peoples and patients have been analyzed by different data mining methods and discovered and extracted hidden patterns and relationships in large volume of data (12). The most medical services in the use of data mining include: a) the effects of the drug on the disease and its side effects (13), b) diagnosis and prediction of various diseases such as diagnosis and prediction of various cancers (13), c) to determine the method of treatment (14), d) predict the success rate of medical interventions such as surgery (15), e) analysis of the existing data in health information systems (HIS) (16), f) analysis of medical images (17). This study by using data mining of deafness is processed to identify the effective factors on severity of deafness.

2. RESEARCH METHOD

With regard to nature of research which lies on use of data mining to determine and analyze the relationship between data of the deaf, data oriented method has been used as the research method. The main origin of the present research has embedded on discovery of knowledge from the databases under study. Therefore, standard of Cross Industry Standard Process for Data Mining (CRISP-DM) has been used to conduct the research (18). The statistical population consists of all the questionnaires gathered from different provinces, recorded with 1736 information records in software. Further, 346 questionnaires consisting of audiogram forms have been considered. Firstly, all the data in the questionnaires were implemented in form of information table in software SqlServer regarding type of the associated field and then all the operations of data entry were fulfilled via the software written in language C#.Net. These data entail series of records which include personal information of each of the deaf people as well as the medical information pertaining to deafness. It

should be noted that such information include a series of quantitative variables including age, percent of deafness and so on as well as a series of qualitative variables including type of relationship, consanguinity and son on (Table 1). Further, types of variables include numeric, string and bit variables. After preparation of data to the format requested in the software and algorithms used in the research, the data were entered into the software and then the considered techniques were applied on the data. The algorithm applied on the data includes "Association Rule". Therefore, the data were prepared to enter them into the algorithm and then the data were entered into the algorithm. To apply the algorithm on the database, data entry was considered for 72 information fields and then some fields were removed after the stage of preprocessing and some fields were changed, whereby the most effective information in this research were found as that of in Table 1 based on holding some sessions purposed for speech with experts and evaluation of the fields existing in the questionnaire together with Audiogram information.

How to calculate and detect degree of deafness among individuals

Row	Details of content of field
1	The unique code at each questionnaire
2	Gender of patient: male & female
3	Percent of deafness
4	Percent of Hearing Loss in right ear In right
5	Percent of Hearing Loss in left ear
6	Use of assistive devices such as hearing aids
7	*consanguinity* in parents
8	Type of *consanguinity*
9	Number of deaf brothers
10	Number of deaf sisters
11	high frequency sounds(250) at left ear
12	high frequency sounds(500) at left ear
13	high frequency sounds(1000) at left ear
14	high frequency sounds(2000) at left ear
15	high frequency sounds(4000) at left ear
16	high frequency sounds(8000) at left ear
17	high frequency sounds(250) at right ear
18	high frequency sounds(500) at right ear
19	high frequency sounds(1000) at right ear
20	high frequency sounds(2000) at right ear
21	high frequency sounds(4000) at right ear
22	high frequency sounds(8000) at right ear

Table 1. List of the fields after preprocessing

As the deafness frequency of individuals is of greater importance at the range of 500 to 2000 frequency, grouping was fulfilled using Figure 1. Regarding Figure 1 which represents the range of detection of hearing to deafness at aforementioned frequencies. To detect type of deafness among individuals, average loudness at frequencies 500, 1000 and 2000 from audiometry was reported and the obtained value was examined in Figure 1(loudness axis), whereby type of deafness was determined. In Figure 2, audiogram of the patient's
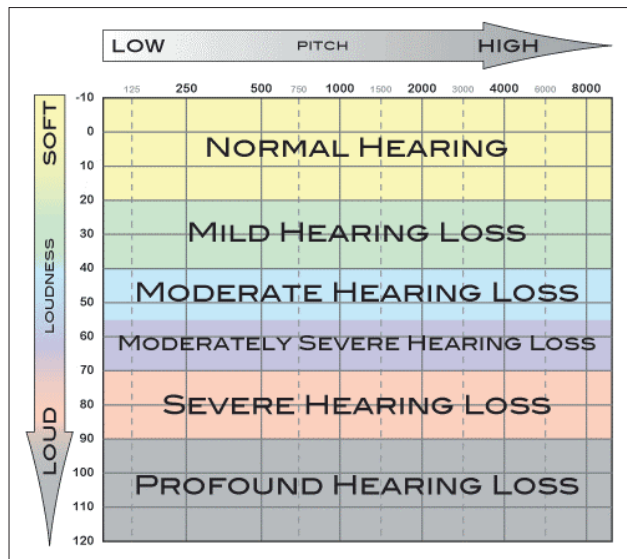
Figure 1. Range of frequency and sound loudness in detection of hearing to deafness

right ear with code 8 corresponding to Table 2 represents deafness of the patient. The mean of the specified range has been equal to 58 with moderately severe hearing loss based on Figure 1. Type of relationship has been mentioned with the title of cousin, yet numeric data has been stored in database.

According to the calculations above, the number of effec-

Right ear(code 8)						
Pitch	250	500	1000	2000	4000	8000
Loudness	40	50	60	65	75	80

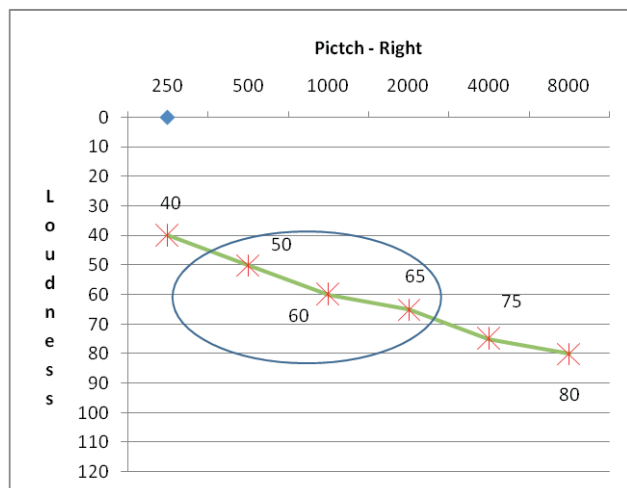Table 2. Deafness of patient via code 8

Figure 2. Audiogram of right ear of patient via code 8

tive fields is obtained based on Table 2.

Descriptive findings of research

cousin	Kinship=1
cousin	Kinship=2
cousin	Kinship=3
cousin	Kinship=4
cousin	Kinship=5
cousin	Kinship=6
Etc.	Kinship=7

Table 3. Comparison in number of deaf people and gender

With regard to the information in the questionnaires, it can

obtain a series of relative comparisons and frequency percent

Row	Details of content of field
1	Gender of patient: male & female
2	Number of deaf brothers
3	Number of deaf sisters
4	Average Hearing Loss
5	Hearing Loss in right ear
6	Hearing Loss in left ear
7	Use of assistive devices such as hearing aids
8	*consanguinity* in parents
9	Type of *consanguinity*

Table 2. List of fields after pre-processing

about the existing values, that it can compare them in form of the table below:

In Table 3, a comparison in number of the deaf people mentioned in the questionnaires through separation of gender among the statistical population has been displayed, so that it

Row	gender	No	%
1	Female	955	55.01
2	male	781	44.99
Sum		1736	100

Table 3. Relative frequency distribution for the gender of individuals

can observe that deafness more likely among the deaf women than deaf men for 10%. To define status of hearing loss and deafness regarding Table 1, it can say individuals with good hearing status to the ones with the ability for hearing in the frequencies ranging from 500 to 2000 and the loudness of

Row	Status of patient	No	%
1	Hearing loss	62	3.57
2	Deaf	1674	96.43
Sum		1736	100.00

Table 4. Relative frequency distribution for the hearing status of individuals

sounds ranging from 0 to 20. The individuals with hearing loss are called to those ones who have the ability for hearing in the frequencies ranging from 500 to 2000 and the loudness of sounds ranging from 20 to 70. The individuals with deafness are called to those ones with the ability for hearing in the frequencies ranging from 500 to 2000 and the loudness of sounds ranging from 70 to 120. In Table 4, a comparison in number of the deaf people mentioned in the questionnaires through separation of status of deafness among the statistical

Row	type of relationship	type of relationship	No	Percent of frequency
2	5	Cousin	49	2.82
3	6	Cousin	110	6.34
4	4	Cousin	139	8.01
5	2	Cousin	178	10.25
6	3	Cousin	210	12.10
7	1	Cousin	313	18.03
8	7	Etc.	733	42.22
Sum			1736	100.00

Table 5. Relative frequency distribution for the type of relationship of parents among the patients under study

population has been displayed, so that it can observe that the highest percent has been associated to the individuals with total deafness; 4% of individuals has been affected by hearing loss; Ultimately, a status of normal hearing has not been recorded in the questionnaires. In Table 5, it has been displayed that it can specify type of relationship based on type of relationship between parents. Further, as the table has been arranged based on the columns, it can deduce that the number of relative marriages has been witnessed among 60% of the individuals and marriage with strangers has been witnessed among 40% of the individuals. Since the field of relative marriage has been recorded in the questionnaire, it can examine the number of patients whose parents have had the hearing loss to the degree of 3 or 4 in Table 6. Findings from applying the algorithms of association rules with the purpose of determination of factors affecting severity of hearing. Algorithm "association" has been mentioned as a suitable method to discover the relationship between dependent variables, yet practically the obtained rules can be called to those rules

Row	Parents with relative marriage	No	Percent of frequency
1	Yes	1027	59.16
2	No	709	40.84
Sum		1736	100.00

Table 6. Relative frequency distribution for the relative marriage among the parents of the individuals under study

among which the rules with higher confidence and support have been selected, ultimately these rules should have been examined and confirmed by the experts. Support and confidence have been mentioned as two criteria for charismatic rule that represent utility and certainty of rule. In general, an association rule is considered in case it meets minimum support threshold and minimum confidence threshold, that these thresholds are determined by a specialized person. A more accurate definition for the aforementioned terms is as follow:

- Rule: if A and B enjoy support(S), S% of the transactions will include union of A and B.
- Rule: if A and B enjoy confidence(C), C% of the transactions will include union of A and B.

If the rules with confidence above 60% and support above 7% are considered, five rules below will be obtained.

3. DISCUSSION

The present research has examined the factors affecting deafness among the individuals via data mining, mentioned that to date no study has been conducted at this area. As the fields in the questionnaire were numerous in the records, a limited number of these fields were become beneficial in this study. Hence, new results have not been extracted, confirming the previous information. If more effective fields likewise the information from genetic experiments exist, it can acquire interesting results. For instance, if the genetic information of the records is filled in future, it can discover the relationship between patient's audiogram diagram and the genes which have a potential role in deafness and hearing loss. Expansion of knowledge and research at the area of individuals' hearing and factors affecting genetic deafness in the country resulted in reduction of surplus costs in genetic experiments. This applied research can be a step to know about

factors affecting genetic deafness in the country, under which it is hoped to acquire new and effective results by obtaining information at this area. With regard to the obtained rules as well as expert's view, it can deduce that two factors of number of deaf brothers as well as degree of relative marriage in parents have a severe effect on severity of deafness among individuals. Indeed, it can deduce that in the families with relative marriage with blood relationship, the number of individuals with severe hearing loss or deafness will be more.

4. CONCLUSION

With regard to area of research, the suggestions below have been proposed for the readers at this area:

- Overview of data mining approach by the expert systems to detect the factors affecting deafness among the individuals,
- Representation of the rules via data mining techniques including association rules to detect the relationship between factors affecting deafness,
- Representation of data mining techniques including neural network model to predict status of deaf people in future,
- Representation of prediction models of deafness using prediction based on data mining.

CONFLICT OF INTEREST: NONE DECLARED.

REFERENCES

1. Masic I, Ridjanovic Z, Pandza H, Masic Z. Medical Informatics. Avicena. Sarajevo, 2010: 255-278.
2. Lin WT, Wub YC, Zheng JS, Chen MY. Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital. Expert Systems with Applications. 2011 Sep; 38(9): 11078-11084.
3. Seifert JW. Data Mining: An Overview, Congressional Research Service. CRS Report for Congress, December 2004.
4. Tsai PSM, Chen CM. Discovering knowledge from large databases using pre-stored information. Information Systems. 2001 Mar; 26(1): 1-14.
5. Lindell Y, Pinkas B. Privacy Preserving Data Mining.. Lecture Notes in Computer Science. 2000; 1880: 36-54..
6. Gertisio C, Dussauchoy A. Knowledge Discovery from Industrial Data base. Journal of Intelligent Manufacturing. 2004; 15: 29-37.
7. Gilman.M Data Mining Overview, CEO, Data Mining Technologies Inc., 2000.
8. Shun Ngan P, Leung Wong M, Lam W, Sak Leung K, Cheng YC. Medical data mining using evolutionary computation. Artificial Intelligence in Medicine. 1999; 16:73-96.
9. Lavrac N. Selected techniques for datamining in medicine, Artificial Intelligence in Medicine. 1999; 16(1): 3-23.
10. Zupan BR. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform. 2008.
11. Selman A, Ebru Sezer B. Usage of data mining techniques in discovering the food consumption patterns of students and employees of university, World Conference on Information Technology, WCIT, 2010.
12. Hauben M. A decade of the data mining and still countingAdis dataInformation BV, 2010.
13. Bellaachia A, Guven E. Predicting Breast Cancer Survivability Using Data Mining Techniques, Department of Computer Science The George Washington University Washington DC, 2005.
14. McCabe RM, Adomavicius G, Johnson PE, Ramsey G, Rund E, Rush WA, O'Connor PJ, Sperl-Hillen JA. Using Data Mining to Predict Errors in Chronic Disease Care Agency for Healthcare Research and Quality (US), 2008.
15. Persson M, Lavesson N. Identification of Surgery Indicators by Mining Hospital Data: A Preliminary Study 20th International Workshop on Database and Expert Systems Applications. 2009: 323-327.
16. Chye Koh H, Tan G. Data Mining Applications in Healthcare. Journal of Healthcare Information Management. 2011; 19(2).
17. Antonie M, Zaïane OR, Coman A. Application of Data Mining Techniques for Medical Image Classification. ;In Proceedings of MDM/KDD. 2001: 94-101.
18. Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R. CRISP-DM 1.0 step by step dataminingguid, 2000..