An investigation of data mining techniques of the performance of a decision tree algorithm for predicting causes of traumatic brain injuries in Khatamolanbya Hospital in Zahdan city, 2012 to 2013

Ahmad Raeesi[1], Saeed Ebrahimi[2*], Leila Irfan Nia[1], Goli Arji[3], Moslem Askani[1]

ABSTRACT

Introduction: The aim of this study was to determine the performance of data mining techniques for predicting the causes of traumatic brain injuries in Khatamolanbya hospital, Zahdan city.

Method: In this cross–sectional, the study population included all patients who died of brain injury. Data were collected by the use of a researcher- made check list, provided under the direct observation of authorities in this area and analyzed by the data mining software of Clementine 12.0.

Results: According to the results of this algorithm, C5.0 decision tree algorithm has an accuracy of 81.4 percent, the highest precision; then, the algorithm is C & R(The Classification and Regression) with 77.8 percent.

Conclusion: Overall, it can be concluded from the decision tree algorithm that age is one of the leading causes of traumatic brain injuries . The results showed that all the cases involving traumatic lesions of the brain lead to the patient's death.. Although in some algorithms, some of the variables are important, they cannot be used alone as the main variable to be taken into account for the death of the patient.

Keywords: Data mining, Prediction of the factors of traumatic brain injuries

Introduction

Trauma is the main cause of death people, many of which are preventable. 16,000 people are daily losing their lives due to trauma (1). One of the growing and most common causes of death and disability worldwide, regardless of the level of development of countries, is traumatic lesions (2). Trauma is the leading cause of death before the age of 44 , and the fourth cause of death after cardiovascular diseases, neoplasms and respiratory diseases in all ages (3). Traffic incidence, attacks and fall are three most important factors of the head injuries in various societies.

Methods

The present study is an applied research using data analysis methods - descriptive analysis. The total number of patients who died in Khatamolanbya hospital due to brain lesions in the years 2012-2013 was 73 patients, 70 of whom had been already examined. Also, 70 cases who did not die due to brain lesions were studied. Census method has been used in this study. The instrument for data collection in this study was a researcher-made check list of researchers after reading various articles designed by experts. This is a list of 13 items which contains demographic data such as age and gender as well. It also contains items such as the admission season, time of admission and finally, based on the brain injury resulting in death or not, data were also classified. Methods of data collection included visiting the archives of Khatamolanbya Hospital (A) and case studies of patients who had suffered from a traumatic brain injury, respectively. Variables used in the study of data mining can be divided into two

[1]*Health Information Technology, Zahedan University of Medical Sciences, Zahedan, Iran.*

[2]*Health Information Technology, Shiraz university of Medical Sciences, Shiraz, Iran*

[3]*Instructor of Health Information Technology, Zabol University of Medical Sciences, Zabol, Iran*

*****Corresponding Author:***S Ebrahimi,** *Health Information Technology, Shiraz university of Medical Sciences, Shiraz, Iran. Email: Saeidebrahimi9@gmail.com.*

categories of target variable and predictor variable. The variable of death or non-death were used as target variable (dependent) and other variables as predictor variables (independent). After data collection from hospital patient records in Khatamolanbya (p), the data were put into SPSS software , version16 and clementine12.0, using Chi-square test and spearman test. The significance level was considered as 5%. 4 clementine12.0 software was used for modeling the decision tree algorithms (CHAID, QUEST, C5.0 and C & R tree).

Results

In this study, a total of 140 subjects (72.1% men and 27.9% women) participated. The mean age at injury , was 28.39 for men and 23.00 for women. As to the frequency of accidents, in spring the maximum number of traumatic lissions occurred (40.7%) and in winter the least(12.9%)

Decision tree algorithms

The decision tree algorithm C & R

Rules obtained from tree and the application of decision tree algorithm C & R

1. If both types of lesions are associated with multiple lesions, subdural hematoma, intracerebral hematoma, brain injuries, poisoning, and if your age is less than or equal to 54 and the injured person is driving or walking , as well as the place of occurrence is Zahdan city by the type of motorcycle or car, then the likelihood of death is 46 percent .

2. If multiple lesionssuch as subdural hematoma, intracerebral hematoma, brain injuries, poisoning are sustained if the injured person is younger or equal to 54 years of age, the injured person is driving or walking, the setting of the accident is outside Zahedan, motorcycle or car is used, then the possibility of death is 85 percent.

3. If multiple lesions are sustained including subdural hematoma , intra-cerebral hematoma , poisoning , and other brain injuries are unknown, age is less than or equal to 54 years, the activity of the person during the accident is unknown , and the place of occurrence or passengers is within the city Zahedan, the possibility of death is 30 percent.

4 . If multiple lesions including subdural hematoma , intracerebral hematoma , poisoning are sustained and other brain injuries are unknown, the age is less than or equal to 54, the activity done during the accident is unknown, and the place of occurrence or passengers out of the city of Zahedan, the death possibility is 54 percent,

5 . If Cantezhen injury, cerebral edema, external object in the brain, head injury or lesion within the skull, as well as the individual's age is less than 65, the death possibility is 4 percent.

6 . If Catazhen injury, cerebral edema, external object in the brain, head trauma or injury to the skull and also the person's age is more than 6, the death possibility is 100 percent.

QUEST decision tree algorithm

The algorithm is based on QUES, age is the highest cause of death from traumatic brain injuries and the accident location within the city or outside the city is the second one. The least important cause of death is the type of injuries.

Rules obtained from the application of QUEST decision tree algorithm

1. If a person's age is less than or equal to 36.61 and place of incident is Zahedan, then the probability is 28 percent.

2. If a person's age is less than or equal to 36.61, or the incident location is outside the city or unknown, then the likelihood of death is 62.22 percent.

CHAID Decision Tree Algorithm

Based on the CHAID algorithm, age is the highest cause of death from traumatic brain injuries, next being the accident location within the city or outside the city, and then the activity done at the time of accident is the least important cause of death.

Rules

1. If the age is less than or equal to 58 and the accident location is within the city of Zahedan and the person injured is a driver or passenger who is unknown, then the possibility of death is 21 percent.

2 . If the age is less than or equal to 58 and he accident location is with the city of Zahedan and the injured person is walking, then possibility of death is 70 percent.

3. If the age is less than or equal to 58 and the accident location is out of the town or city, then the possibility of death is 61 percent.

C5.0 Decision Tree Algorithm

According to C5.0 algorithm, the age is the highest cause of death in traumatic brain lesions and the type and what the person had been doing are the next causes, respectively.

Rules obtained from tree algorithm C5.0

1. If a person's age is less than or equal to 57, the type of lesion is multiple injuries, the driver is the injured person, the accident location is Zahedan, and type of vehicle is motorcycle, then the possibility of death is 80 percent.

2. If a person's age is less than or equal to57, the type of lesions is a multiple one, the injured person is the driver, and the accident location is outside the city of Zahedan, then the possibility of death is 82.35percent.

3. If a person 's age is less than or equal to 57, the type of lesions is a multiple one, the injured person is passenger, and the place of occurrence is outside Zahedan, then the possibility of the death is 54.54 percent.

4. If a person 's age is less than or equal to 57, the type of lesions is a multiple one, the person is injured passenger, and the accident location the city of Zahedan, then 32% is the possibility.

According to the results of the four algorithms as mentioned, C5.0 algorithm accuracy is 81.4 percent, the highest accuracy; the results are better than other algorithms and have higher reliability and accuracy. And then the algorithm of C &R is the most accurate one with 77.8 percent. Lesions were the variable that raised first in C & R and then age, and vice versa at C5.0. Overall decision tree algorithms results show that in traumatic brain injuries age is one of the leading causes of death in patients, and in all cases the first or second is decision tree algorithms . We conclude that if the age of the person is high or low, the tolerance to injuries during the incident is reduced. The mean age in the present study was 28, which is our youth generations. The fact demonstrates the need for planning the country wide programs for less brain traumas. In a

study which was performed in 2012 at Tehran University under the title of "Presentation of a Conceptual Method for Prediction of Heart Attacks through Data Mining", it was concluded that the general accuracy of decision tree models, C5.0, 83.28 in the survey , was 83.28 percent of the overall accuracy of the model C5 with the highest accuracy among the models implemented (4). In this study, the C5.0 model has the highest accuracy. Moreover, C5.0 decision tree model shows that if the vehicle is a motorcycle and the incident happened in the city of Zahedan, the possibility of death is very high. It might be due to the fact that in the city of Zahedan motorcycle is used more out of town than inside, as QUEST decision tree model revealed.

The model is a decision tree C & R if the type of waste, some associated injury and hematoma, subdural hematoma and intracerebral poisoning and cancer and other brain injuries are known, and these lesions are sustained within the city, the possibly of death in the city is 46% and that of outside the city is 85%. If Kantvzhn injury, brain edema and foreign bodies within the brain and the skull of a head injury is unknown and age is more than 65, traumatic lesions are most likely to lead to patient's death. The CHAID decision tree model shows that the possibility of death in pedestrian cases in Zahedan is 70% which is higher than that outside the city. This model also shows that the possibility of the driver's death in the city of Zahedan is 21%, which is lower than that out of town. The model is a decision tree C & R if the type of waste, some associated injury, hematoma, subdural hematoma, intracerebral poisoning, cancer and other brain injuries are known, and these lesions are sustained within the city, the possibly of death in the city is 46% and out of the city it is 85%, showing that the use of decision tree algorithms leads to better results for the prediction of traumatic brain injuries.

Conclusion

The results of this study showed that all possible cases of traumatic brain injuries are the result of the patients' death and all variables are dependent. Although some algorithms and some of the variables are important, they cannot be used alone as the main variable to take into account.

References

1. Theodoraki E-M. Innovative data mining approaches for outcome prediction of trauma patients. Journal of Biomedical Science and Engineering. 2010;03(08):791-8.
2. Meyer AA. Death and disability from injury: a global challenge. J Trauma. 1998;44(1):1-12.
3. Wu X, Hu J, Zhuo L, Fu C, Hui G, Wang Y, et al. Epidemiology of Traumatic Brain Injury in Eastern China, 2004: A Prospective Large Case Study. The Journal of Trauma: Injury, Infection, and Critical Care. 2008;64(5):1313-9.
4. Arji G. A conceptualmodelfordata-miningapproachtopredictmyocardialinfarctionin 1390. Tehran: SchoolofAlliedMedicalSciencesand Health Services; 1390.
5. Adel A, Ahmadi P, Sebt M. Desining Model for Choosing human resources with data mining approach].[In Persian]. Journalof Iranian Technology. 2010;2(4):5.
6. Andelic N, Sigurdardottir S, Brunborg C, Roe C. Incidence of Hospital-Treated Traumatic Brain Injury in the Oslo Population. Neuroepidemiology. 2008;30(2):120-8.
7. Bali RK. Clinical knowledge management: opportunities and challenges. IGI Global; 2005.
8. Chen H, Fuller SS, Friedman C, Hersh W. 2005;8.
9. Ho T. Data Mining and Data Warehousing. Prentice Hall; 2005.
10. 10. LaTour K, Eichenwald S. Health Information Management; Concepts, Principles, and Practice. AHIMA: Chicago. 2002:478-80.
11. Tavakoli N, Jahanbakhsh M. Opportunities and Challenges of HER Implementation in Isfahan [Project] [3. [In Persian]]. Isfahan: The university of Isfahan: School of Informatics and Management; 2010.
12. Wager K, Lee F. Glaser JP.Managing Health Care Information Systems: A Practical Approach for HealthCare Executives. New Jersey: John Wiley & Sons; 2005.
13. Wickramasinghe N, Gupta J, Sharma S. Creating Knowledge-Based Healthcare Organizations. Hershey: IdeaGroupInc (IGI); 2005.
14. Wyatt JC. Keeping up: learning in the workplace. Bmj. 2005;331(7525):1129-32.