Machine Learning Techniques in Predicting Delayed Pneumothorax and Hemothorax Following Blunt Thoracic Trauma

Ali Reza Khoshdel [1]; Hamidreza Bayati [2,*]; Babak Shekarchi [3]; Seyyed Ehsan Toossi [2]; Behnam Sanei [2]

[1]Department of Epidemiology, AJA University of Medical Sciences, Tehran, IR Iran
[2]Department of Surgery, Isfahan University of Medical Sciences, Isfahan, IR Iran
[3]Department of Radiology, AJA University of Medical Sciences, Tehran, IR Iran
*Corresponding author: Hamidreza Bayati, Department of Surgery, Isfahan University of Medical Sciences, Isfahan, IR Iran. Tel: +98-9102008954, E-mail: hamid.bayati@gmail.com

Background: Delayed pneumothorax and hemothorax are among the possible fatal complications of blunt thoracic trauma.
Objectives: Finding reliable criteria for timely diagnosis of high-risk patients has been an area of interest for researchers.
Material sand Methods: We gathered a database including 616 patients among which, 17 patients experienced the delayed complications. Employing four classification techniques, namely, linear regression, logistics regression, artificial neural network, and naïve Bayesian classifier, we tried to find a predictive pattern to recognize patients with positive results based on recorded clinical and radiological variables at the time of admission.
Results: First, without using machine learning techniques, we tried to predict the complications based only on a single variable. We recognized chest wall tenderness as the best single criterion that enables to classify all high-risk patients with 100% sensitivity (95% CI, 82-100). This criterion potentially excludes 57% (95% CI, 53-61) of low-risk patients from further observation. Then we used the machine learning techniques to assess the effect of all admission time variables together. According to our results, amongst the utilized techniques, logistics regression model enables not only to exclude 81% (95% CI, 77-84) of patients without complications from unnecessary observation, but also to recognize all patients with true positive results for pneumothorax and hemothorax (95% CI, 82-100).
Conclusions: Instead of serial chest X-ray, patients with blunt chest trauma could be initially evaluaed by a risk assessment model in order to avoid unnecessary work-up.

Keywords: Thorax; Pneumothorax; Hemothorax; Linear Model; Logistic Models

1. Bckground

Delayed pneumothorax (DPTX) and hemothorax (DHTX) are among possible fatal complications of blunt thoracic trauma (BTT). Although the incidence rates of DPTX and DHTX following BTT are somehow low and have been reported as frequent as 7.4% for DPTX and 2% to 6% in DHTX, serious consideration is required due to the high risk of mortality (1, 2). Current medical guidelines recommend the follow-up of seemingly high-risk patients with six-hour intervals serial chest X-Ray (CXR) (3). However, besides exposing the patients to excessive radiation and obtaining serial CXRs is not optimal and economical. In this respect, finding reliable criteria to classify high-risk patients for careful observation would be of great importance.

Rib fractures are recognized as an underlying factor for the delayed complications in different studies (4-7). Simon et al. found high prevalence of multiple or dis-placed rib fractures in patients with DHTX (8). Liman et al. discovered a correlation between number of fractured rib and DHTX occurrence (4). Sharma et al. emphasized on careful observation of these patients for well-timed diagnosis of DHTX (6, 7). However, to classify high-risk subjects accurately, considering the prevalence of rib fracture in patients with no delayed complication is also essential.

To exclude low-risk patients based on CXR findings, Rodriguez et al. investigated the diagnostic significance of the different clinical variables (9). They exploited features like mechanism of injury, intoxication, chest tenderness on palpation, crepitus, etc to classify high-risk complications. Using screening tests and based on the CXR findings, they reported the combination of tenderness on palpation and hypoxia as the best measure excluding 46% of patients. Shekarchi et al. recorded clinical and CXR -

---

Implication for health policy/practice/research/medical education:
The implication of this work is to predict delayed complications, namely, hemothorax and pneumothorax, following blunt thoracic trauma using machine learning classification techniques. In comparison to the usually performed six-hour interval serial chest X-rays, we tried to predict the delayed complications by using only admission time recorded variables. In this respect, high-risk patients can be recognized timely by keeping a lookout for reducing the risks of mortal delayed complications.

Table 1 . List of Classification Input Variables and Their Frequencies in Positive and Negative Classes [a,b]

Input Variable	True Positive (TP)	Sensitivity (TP/17) (95% CI)	False Positive (FP)	Specificity (1-FP/599) (95% CI)
Chest wall tenderness	17	100 (82-100)	257	57 (53-61)
Chest pain	16	94 (73-99)	434	27 (24-31)
Chest wall crepitation	4	24 (10-42)	5	99 (98-100)
Rib fracture	3	18 (6-41)	5	99 (98-100)
Subcutaneous emphysema	3	18 (6-41)	8	99 (97-99)
Abdominopelvic trauma	3	18 (6-41)	23	96 (94-97)
Chest wall Ecchymosis	2	12 (3-34)	26	96 (94-97)

[a] Abbreviation: CI, confidence interval.
[b] Sensitivity and specificity of single variable classification are also calculated with 95% confidence interval.

-variables of 680 patients (under publication) to predict the delayed complications. To assess the combination of variables, they applied logistic regression classifier with 64.7% sensitivity and 93% specificity. In this study, we generalized their conclusions by emphasizing on classification methods like artificial neural networks (ANNs) in order to determine better screening methods. In addition, we considered the possibility of developing a single variable-based recognition method for the delayed complications. Furthermore, we introduced a new formula enabling better screening accuracy.

ANNs provide a risk assessment tool with the capability of application in diagnosis, prognosis, focusing on recalling the incidence of rarely occurring disease profiles, and analysis of different treatment choices (5). Although there were only 20 published works concerning ANNs in medical practice until 1988, the method is regularly used in the medical field currently (10). However, the training process of ANNs necessitates different aspects, which may not be always available, leading to inconvenience (11). To overcome this problem, statistical tests are employed to evaluate the mapping confidence by dividing the data set to the training and validating subsets. In this study, we applied four classification techniques to find a predictive pattern for recognition of the high-risk patients using admission time recorded variables. The variables included radiological and clinical criteria mentioned in Table 1.

2. Objectives

We employed the dataset recorded by Shekarchi et al. from July 2009 to December 2010 in three hospitals. Only the patients who accepting to participate in the study along with meeting the inclusion criteria like no need for surgical interventions were included. Our analysis included 616 patients with BCT consisting of 422 (68%) males and 200 (32%) females who had 18 to 96 years of age (mean ± SD, 44.3 ± 20.0 years). The machine learning algorithms (explained in the Methods section) determined 17 subjects positive for delayed complications including nine cases with DHTX, seven with DPTX; moreover, it determined one case with delayed hemopneumothorax

from 599 patients with negative results. Table 1 displays the algorithm input variables as well as their frequencies in the high-risk and low-risk classes. Besides, sensitivity and specificity of single variable recognition and the corresponding 95% confidence intervals are also displayed.

3. Materials and Methods

Classification methods provided a mapping from the input space (See Table 1) to the categorical output space, i.e. positive and negative classes. Up to this point, we had employed four classification methods, namely, linear regression (LinReg), logistic regression (LogReg), ANNs, and naive Bayesian classifier (NBC) (12). The classification algorithms tried to learn the characteristics of the classes using the training data subset in a training phase. Then, the classification performance was tested on validation data subset to examine how the mapping could be generalized to new patterns. We trained an ANN with three and five neurons in the first and the hidden layers by minimizing classification error through the back-propagation algorithm. To train LinReg, LogReg, and NBC, we applied matrix pseudoinverse, iteratively reweighted least squares, and single variable histogram calculating algorithms, respectively.

To analyze the performance, we used the four well-known diagnostic test indices, namely, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In addition, ratio indices, namely, positive likelihood ratio (PLR) and negative likelihood ratio (NLR) as screening criteria were reported. Confidence intervals of diagnostic test and ratio indices were calculated using Wilson score method (13) and the method introduced by Simel et al. respectively (14).

4. Results

For each classification technique, we repeated the training phase 100 times with randomly chosen two-thirds of the data as training subset. Then, the best classifier based on having the highest receiver operating characteristics (ROC) curve area was selected. Table 2 reports the diagnostics results on all the data consisting of training and

Table 2. Diagnostics Accuracies and Corresponding Confidence Intervals Obtained by Four Classification Techniques

	LinReg	LogReg	ANN	NBC
Sensitivity (95% CI)	65 (41-83)	100 (82-100)	71 (47-87)	65 (41-83)
Specificity (95% CI)	97 (95-98)	81 (77-84)	97 (95-98)	97 (95-98)
PPV (95% CI)	65 (41-83)	49 (33-64)	38 (23-55)	38 (21-53)
NPV (95% CI)	99 (97-99)	100 (99-100)	99 (98-100)	99 (98-100)
PLR (95% CI)	21 (12-38)	5 (4-6)	21 (12-36)	19 (11-34)
NLR (95% CI)	0.36 (0.19-0.69)	0	0.3 (0.15-0.64)	0.37 (0.19-0.7)
ROC area	94.9	95.6	96.1	95

[a] Abbreviations: ANN, artificial neural network; CI, confidence interval; LinReg, linear regression; LogReg, logistic regression; NBC, naive Bayesian classifier; NLR, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value, ROC, receiver operating characteristics.

validation subsets; it provides an overall evaluation the delayed complications prediction in our subjects.

LogReg had the sensitivity of 100% (95% CI, 82-100) with the specificity of 81% (95% CI, 77-84) while the three other methods had high specificity of 97% (95% CI, 95-98) with much less sensitivity. LogReg led to the best NLR and NPV with high screening accuracy while LinReg, ANN, and NBC had comparable PLRs. Considering the risk of missing a high-risk patient, we were interested in recognizing all subjects with delayed complication; therefore, high sensitivity with reasonable specificity was important for our screening test. In fact, it would provide a tool to classify high-risk patients while removing many low-risk ones.

As stated before, the best screening accuracy was achieved by LogReg with sensitivity of 100% (95% CI, 82-100), specificity of 81% (95% CI, 77-84), PPV of 49% (95% CI, 33-64), and NPV of 100% (95% CI, 99-100). The model follows the below formula:

$$y = \frac{1}{1 + \exp(-z)}$$

Where z is defined as follows:

z = 25.01 Ch.Tend + 3.89 Ch.Pain + 2.01 Ch.Crep + 2.68 Rib. Frac + 2.19 Sub.Emph + 4.07 Abp.Tra + 1.32 Ch.Ecch -27.94 (The formula acronyms are defined in Table 1). For classification of the output, y should be compared with 0.5 as a threshold. It should be noted that although ANN provided a more complex method to model the data, LogReg outperformed it in terms of screening accuracy. We interpreted it by the fact that more complex models need more data for correct estimation of the model coefficients. In addition, model complexity increases the chance of trapping in local minima in training phase. Thus, with this number of patterns, LogReg that could be considered as a single neuron, outperformed the multilayer ANNs.

5. Discussion

In this study, we investigated the possibility of BTT de-

layed complications prediction based on admission-time recorded clinical and radiological variables. We used a dataset consisting of 17 patients with delayed complications and 599 patients without them whom were recorded in three hospitals from July 2009 to December 2010. Four classification algorithms were employed to find a predictive pattern for recognizing high-risk patients. To evaluate the results, diagnostics test indices namely sensitivity, specificity, PPV, NPV, PLR, and NLR with corresponding 95% confidence intervals were calculated.

In agreement with Rodriguez et al. (9), we recognized chest wall tenderness as the best single criterion enabling to classify all high-risk patients with sensitivity of 100% (95% CI, 82-100). This criterion potentially excluded 57% (95% CI, 53-61) of low-risk patients from further observation. In contrast with previous studies emphasizing on high sensitivity of the rib fracture (4, 6-8), this factor could only recognize 18% (95% CI, 6-41) of subjects with delayed complications in the our dataset.

We concluded that using the aforementioned LogReg formula identified all high-risk subjects and potentially excluded 81% (95% CI 77-84) of low-risk patients from serial CXR in the studied dataset. However, it should be noted that this was the primary and initiative result that should be validated and evaluated in larger and more comprehensive datasets before being put in practice.

Acknowledgements

Authors' Contribution

Alireza Khoshdel: supervisor, designing, reviweing Seyyed Ehssan Toossi: researcher; Hamidreza Bayati: data analyses and drafting, Babak Shekarchi: Specialist consultant, supervisor in radiology; Behnam Sanei: consultant in the emergency and surgical ward.

Financial Disclosure

Funding/Support

References

1. Lu MS, Huang YK, Liu YH, Liu HP, Kao CL. Delayed pneumothorax complicating minor rib fracture after chest trauma. *Am J Emerg Med.* 2008;**26**(5):551–4.
2. Misthos P, Kakaris S, Sepsas E, Athanassiadi K, Skottis I. A prospective analysis of occult pneumothorax, delayed pneumothorax and delayed hemothorax after minor blunt thoracic trauma. *Eur J Cardiothorac Surg.* 2004;**25**(5):859–64.
3. Legome E. *General approach to blunt thoracic trauma in adults.* 2011. Available from: http://laemeufba.files.wordpress.com/2011/06/general-approach-to-blunt-thoracic-trauma-in-adults.pdf.
4. Liman ST, Kuzucu A, Tastepe AI, Ulasan GN, Topcu S. Chest injury due to blunt trauma. *Eur J Cardio-Thorac Surg.* 2003;**23**(3):374–8.
5. Lisboa PJ. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw.* 2002;**15**(1):11–39.
6. Sharma OP, Oswanski MF, Jolly S, Lauer SK, Dressel R, Stombaugh HA. Perils of rib fractures. *Am Surg.* 2008;**74**(4):310–4.
7. Sharma OP, Hagler S, Oswanski MF. Prevalence of delayed hemothorax in blunt thoracic trauma. *Am Surg.* 2005;**71**(6):481–6.
8. Simon BJ, Chu Q, Emhoff TA, Fiallo VM, Lee KF. Delayed hemothorax after blunt thoracic trauma: an uncommon entity with significant morbidity. *J Trauma.* 1998;**45**(4):673–6.
9. Rodriguez RM, Hendey GW, Marek G, Dery RA, Bjoring A. A pilot study to derive clinical variables for selective chest radiography in blunt trauma patients. *Ann Emerg Med.* 2006;**47**(5):415–8.
10. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet.* 1995;**346**(8983):1135–8.
11. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med.* 2000;**19**(4):541–61.
12. Webb AR, Copsey KD. *Statistical Pattern Recognition.* Third ed: John Wiley & Sons, Ltd; 2011.
13. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med.* 1998;**17**(8):873–90.
14. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;**44**(8):763–70.