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 Sanet 2

1Department of Epidemiology, AJA University of Medical Sciences, Tehran, IR Iran
2Department of Surgery, Isfahan University of Medical Sciences, Isfahan, IR Iran
3Department 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

Received: February 16, 2014; Revised: April 6, 2014; Accepted: April 6, 2014

1. Background

Delayed pneumothorax (DPTX) and hemothorax (DHTX) are among the 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 displaced 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-
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 naïve 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
Khoshdel AR et al.

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

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

*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.

In this study, we investigated the possibility of BTT delayed 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

The authors would like to appreciate kind collaboration by the specialists and staff in the registered hospitals in Isfahan as well as the administrative officers in the deputy of research, AJA University of Medical Sciences for their support.

Authors’ Contribution

Alireza Khoshdel: supervisor, designing, reviewng
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

The data was adopted for advanced analyses from our ongoing work with permission from the research council of AJA University of Medical Sciences, Tehran, Iran.

Funding/Support

This work was supported by AJA University of medical sciences, Tehran, IR Iran.

References