Use of artificial neural networks to predict surgical satisfaction in patients with lumbar spinal canal stenosis

Clinical article

PARISA AZIMI, M.D.,1 EDWARD C. BENZEL, M.D.,2 SOHRAB SHAHZADI, M.D.,1 SHIRZAD AZHARI, M.D.,1 AND HASAN REZA MOHAMMADI, M.D.1

1Department of Neurosurgery, Shahid Beheshti University of Medical Sciences, Tehran, Iran; and 2Department of Neurosurgery, Cleveland Clinic Foundation, Cleveland, Ohio

Object. The purpose of this study was to develop an artificial neural network (ANN) model for predicting 2-year surgical satisfaction, and to compare the new model with traditional predictive tools in patients with lumbar spinal canal stenosis.

Methods. The 2 prediction models included an ANN and a logistic regression (LR) model. The patient age, sex, duration of symptoms, walking distance, visual analog scale scores of leg pain or numbness, the Japanese Orthopaedic Association score, the Neurogenic Claudication Outcome Score, and the stenosis ratio values were determined as the input variables for the ANN and LR models that were developed. Patient surgical satisfaction was recorded using a standardized measure. The ANNs were fed patient data to predict 2-year surgical satisfaction based on several input variables. Sensitivity analysis was applied to the ANN model to identify the important variables. The receiver operating characteristic–area under curve (ROC-AUC), Hosmer-Lemeshow statistics, and accuracy rate were calculated for evaluating the 2 models.

Results. A total of 168 patients (59 male, 109 female; mean age 59.8 ± 11.6 years) were divided into training (n = 84), testing (n = 42), and validation (n = 42) data sets. Postsurgical satisfaction was 88.7% at 2-year follow-up. The stenosis ratio was the important variable selected by the ANN. The ANN model displayed a better accuracy rate in 96.9% of patients, a better Hosmer-Lemeshow statistic in 42.4% of patients, and a better ROC-AUC in 80% of patients, compared with the LR model.

Conclusions. The findings show that an ANN can predict 2-year surgical satisfaction for use in clinical application and is more accurate compared with an LR model.

Key Words • lumbar spinal canal stenosis • prediction • surgical satisfaction • artificial neural network • logistic regression • technique

N early everyone experiences low-back pain (LBP) at some point in his or her life. A common cause of LBP is lumbar spinal canal stenosis (LSCS).4 The term LSCS implies that a portion of the spinal canal has narrowed.8 The symptoms of LSCS include pain, numbness, or weakness in the legs, groin, hips, buttocks, and lower back. Symptoms usually worsen when walking or standing (claudication).8 The ability to predict the surgical success rate is important in choosing the appropriate management of LSCS in these patients.

Medical informatics has been applied to develop prediction models to assess, diagnose, and treat these patients, such as logistic regression (LR) and artificial neural networks (ANNs) or neural networks (in short). One of the models, LR, is a traditional predictive tool. The other,

See the corresponding editorial in this issue, pp 298–299.

This article contains some figures that are displayed in color online but in black-and-white in the print edition.
Predicting surgical satisfaction with ANN in patients with LSCS

ANN, is a computational model based on the functioning of biological neural networks that can be used as nonlinear statistical data modeling tools with which the complex relationships between inputs and outputs (observed data) are modeled or patterns are found. The ANNs try to simulate the learning process of human beings; in other words they learn as humans learn, through observing samples. Actually they are made of a group of interconnected nodes (artificial neurons) that interact with each other based on predefined computational rules. Based on these rules, passing sample data (pairs of observed input/output data) through ANNs makes them modify their structure in such a way that they will be able to estimate the input/output relationship pattern of the systems under study. The resultant network at the end of this learning process can then be used to estimate or predict outputs for new inputs. This capability makes ANNs powerful tools for applications such as pattern recognition, medical diagnosis, financial applications, data mining, email spam filtering, and so on.

Different neuron arrangements and learning methods lead to different ANN types. The most common type is called a multilayer perceptron (MLP), which consists of 3 layers: 1) input layer; 2) hidden layer; and 3) output layer (Fig. 1). With learning processes, sample data are fed into the input layer and the network output is compared with the true expected output. The difference between the network’s actual output and expected output equates to an error that is used to modify the interconnection of neurons that are weighted based on a specific mathematical method. This method is called back-propagation. It enables the network to optimally emulate the model of the system under study. The resultant trained network can be used to predict or estimate output for new inputs. For medical applications, due to the limited number of treatment options available, these prediction models can potentially improve diagnostic accuracy, treatment decisions, and efficiency.

Relationships between prognostic factors and surgical satisfaction have not been previously investigated using ANNs. In this study we sought to develop an ANN model based on the age, visual analog scale (VAS) score for leg pain or numbness, walking distance, stenosis ratio (SR), the Japanese Orthopaedic Association (JOA) score, and the Neurogenic Claudication Outcome Score (NCOS). We also sought to determine whether ANNs perform better at predicting 2-year surgical satisfaction compared with LR in patients with LSCS.

Methods

Patients and Data Collection

This retrospective study population included 168 patients who had undergone surgery for LSCS between May 2007 and January 2012. The diagnosis of LSCS was established using clinical symptoms, neurological examinations, and imaging studies including plain radiography, MRI, and CT studies of the lumbar spine. All patients had the typical symptoms of LSCS, such as neurogenic intermittent claudication and leg pain and/or numbness. Diagnosis was confirmed by more than 1 spine surgeon in all patients. The level(s) of stenosis were explored on the MRI or CT studies. There were no restrictions on patient choice with regard to level(s) of LSCS, age, or other characteristics. The exclusion criteria were prior lumbar spine surgery and spinal anomalies.

Demographic data, including age, sex, and body weight were recorded. The walking distance (in meters) and duration of symptoms (in months) were recorded. A 100-mm VAS was used to measure both leg pain and numbness. The JOA score for assessing LBP was also calculated (total 29 points); a high JOA score indicates a better clinical outcome. The NCOS was determined (total 100 points); a high NCOS shows a better clinical outcome. Patient surgical satisfaction was recorded using standardized questionnaires as shown in Table 1.

Additional Measure

The cross-sectional area at the most stenotic level and at the pedicle level uninvolved by stenosis was measured using the Hamanishi technique (Fig. 2). The calculations were done by 2 independent surgeons, and these investigators were blinded to each other’s findings.

The SR was also calculated. In 1999, Lurencin and colleagues described the SR. It was used to determine the severity of stenosis. This ratio is the cross-sectional area of the canal at the axial MRI slice showing the greatest neurological compression at disc level over the cross-sectional area at the pedicle level above.

The ANN Model

The ANN was used and was based on the standard method and created using the Statistical Package for the Social Sciences (SPSS) software program. Our chosen model was an MLP ANN that is the most common type. An MLP ANN consists of a series of nodes arranged with
3 layers: 1) an input layer; 2) a hidden layer; and 3) an output layer. The MLP ANN used input (age, VAS score for leg pain or numbness, SR, walking distance, JOA score, and NCOS) and output data (surgical satisfaction) (samples for training set) to define (learn) the interrelationships among the data. Patients were partitioned by a 2:1:1 ratio to generate training, testing, and validation. Once the MLP ANN had been trained, it could then predict results from new sets of input data.

**Logistic Regression**

Traditional statistical analysis of variable significance was performed with standard LR on the same data set from which the ANN was assessed.

**Statistical Analysis**

All statistical analyses were performed using PASW Statistics 18 (version 18; SPSS, Inc., 2009). For each individual parameter and for comparisons of the ANN and LR models, the receiver operating characteristic–area under curve (ROC-AUC) was generated and used to calculate specificities, positive predictive value, and negative predictive value for the 2 models and parameters at 95% sensitivity. Discrimination capability was assessed by calculating the AUC from the ROC analysis. For each pair of ANN and LR models (trained and tested on the same data sets), Hosmer-Lemeshow statistics, ROC-AUC, and accuracy rate were calculated and compared using t-tests for the validation group (42 patients).

**Ethics Statement**

The research was approved by the ethics committee of Shahid-Beheshti University of Medical Sciences, Tehran, Iran.

**Results**

Demographic data for the patients with LSCS and their scores on the JOA test, the NCOS, and the SR are shown in Table 2. A total of 168 (59 male, 109 female) patients were included (mean age 59.8 ± 11.6 years, range 17–84 years) and were divided into training (n = 84), test-
Predicting surgical satisfaction with ANN in patients with LSCS

**TABLE 2: Demographic data and preoperative status in 168 patients with lumbar spinal stenosis*  

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>age (yrs)</td>
<td>59.8 ± 11.6</td>
<td>17–84</td>
</tr>
<tr>
<td>sex (% male)</td>
<td>35%</td>
<td></td>
</tr>
<tr>
<td>body weight (kg)</td>
<td>76.8 ± 11.1</td>
<td>54–101</td>
</tr>
<tr>
<td>duration of Sxs (mos)</td>
<td>46.0 ± 23.0</td>
<td>2–62</td>
</tr>
<tr>
<td>walking distance (m)</td>
<td>321.4 ± 218</td>
<td>0–600</td>
</tr>
<tr>
<td>VAS score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>leg pain (mm)</td>
<td>37.9 ± 27.6</td>
<td>10–100</td>
</tr>
<tr>
<td>leg numbness (mm)</td>
<td>57.6 ± 22.8</td>
<td>30–100</td>
</tr>
<tr>
<td>SR</td>
<td>0.46 ± 0.19</td>
<td>0.18–0.86</td>
</tr>
<tr>
<td>JOA score</td>
<td>12.9 ± 6.7</td>
<td>2–21</td>
</tr>
<tr>
<td>NCOS</td>
<td>27.2 ± 12.1</td>
<td>0–38</td>
</tr>
</tbody>
</table>

* Values are expressed as the mean ± SD unless otherwise noted. Sxs = symptoms.

**TABLE 3: Comparison of the ROC-AUC and predictive values of ANN and LR models, and individual parameters to predict surgical satisfaction in 168 patients*  

<table>
<thead>
<tr>
<th>Parameter/Model</th>
<th>ROC-AUC (%)</th>
<th>p Value†</th>
<th>Specificity (%)‡</th>
<th>PPV (%)‡</th>
<th>NPV (%)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.81</td>
<td>0.002</td>
<td>41</td>
<td>69</td>
<td>89</td>
</tr>
<tr>
<td>LR</td>
<td>0.77</td>
<td>0.002</td>
<td>34</td>
<td>63</td>
<td>82</td>
</tr>
<tr>
<td>SR</td>
<td>0.75</td>
<td>0.02</td>
<td>26</td>
<td>61</td>
<td>75</td>
</tr>
<tr>
<td>NCOS</td>
<td>0.63</td>
<td>0.25</td>
<td>21</td>
<td>56</td>
<td>59</td>
</tr>
<tr>
<td>JOA</td>
<td>0.61</td>
<td>0.34</td>
<td>19</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>VAS.P</td>
<td>0.54</td>
<td>0.46</td>
<td>16</td>
<td>55</td>
<td>54</td>
</tr>
<tr>
<td>VAS.N</td>
<td>0.53</td>
<td>0.78</td>
<td>14</td>
<td>53</td>
<td>52</td>
</tr>
<tr>
<td>walking distance</td>
<td>0.51</td>
<td>0.81</td>
<td>5</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>age</td>
<td>0.48</td>
<td>0.86</td>
<td>3</td>
<td>52</td>
<td>18</td>
</tr>
</tbody>
</table>

* NPV = negative predictive value; PPV = positive predictive value; VAS.N = VAS score for leg numbness; VAS.P = VAS score for leg pain.  
† Asymptotic significance on ROC-AUC analysis.  
‡ Specificity and predictive values at 95% sensitivity.
ing (n = 42), and validation (n = 42) data sets. Interrelationships between predictor variables (input nodes), hidden variables (3 of them in 1 hidden layer), and surgical satisfaction/dissatisfaction (output nodes) are illustrated in Fig. 3. Postsurgical satisfaction was 149 of 168 patients (88.7%) at 2-year follow-up.

The results of comparisons between the ANN and LR models and individual parameters are shown in Tables 3 and 4. The sensitivity analysis identified that SR is the important variable selected by the ANN model. Compared with the LR models, the ANN models had a better accuracy rate in 96.9% of patients, a better Hosmer-Lemeshow statistic in 42.4% of patients, and a better ROC-AUC in 80% of patients as well as better sensitivity.

Discussion

Predicting surgical satisfaction is most important during follow-up in patients with LSCS. Individual parameters such as VAS score for leg pain or numbness, SR, walking distance, JOA score, and NCOS score can provide a measure of prognostic utility; however, this study shows that the combination of these parameters in this ANN model could be used to make more precise predictions. The full structural details of the resulting ANN, which does not require further training, can be saved and used as a software estimator or predictor.

To date, there have been no studies that have analyzed 2-year surgical satisfaction in patients with LSCS based on the ANN model. In this study, in comparison with the LR model, the ANN model was more accurate in predicting surgical satisfaction (p < 0.001). With regard to accuracy rate (96.9%), it seems that the ANN model may be helpful to clinicians for surgical decision making and for self-assessment in patients with LSCS.

Maximizing the accuracy of the ANN for optimal clinical efficacy is an achievable end point that will be tracked in a follow-up study in which more patients and additional parameters will be used for the input layer of the ANN. Nevertheless, the ANN model introduced in this study is an acceptable test for predicting 2-year surgical satisfaction in patients with LSCS.

In this study, the important variable selected by the ANN was SR. Therefore, the SR parameter can be considered to be an effective variable to predict surgical satisfaction. However, further studies are needed to determine cutoff points of the SR for decision making in the management of this disease.

In this study, validation data sets (n = 42) were performed based on ANNs by using the standardized questionnaire introduced by Weiner et al. There are various instruments for measuring performance status or functionality in patients with LBP. The Oswestry Disability Index, the Roland Morris Disability Questionnaire, the NCOS, the JOA score, the Swiss Spinal Stenosis, the Core Outcome Measures Index, the STAart Back Screening Tool, the JOA Back Pain Evaluation Questionnaire, and the Short Form-36 scores are among well-known instruments for measuring functionality in patients with LSCS, and have been used to measure functionality in these patients. However, no existing standard cutoff points exist for the data acquired by these instruments for the assessment of surgical satisfaction in these patients. If standard cutoff points are subsequently made available, such instruments may be used.

A limitation of this study is that we were unable to identify all variables affecting surgical satisfaction in patients with LSCS to enter in the input layer in the ANN model. However, further studies of this model may consider the effect of a more detailed database that contains complications and clinical analysis findings, as well as more detailed results data, and may provide a more complete model. Second, this study is limited by its retrospective nature; we could not evaluate outcome measures. The third limitation is the comparatively small number of patients, especially in validation group. Finally, studies are needed to investigate the differences between long- and short-term follow-up of surgical satisfaction in an ANN model of these patients.

Conclusions

The study findings show that 2-year surgical satisfaction can be readily predictable by using the ANNs developed for use in clinical environments and that this model is more accurate compared with an LR model.

Acknowledgment

We thank the staff of the Neurosurgery Unit at Imam-Hossain Hospital, Tehran, Iran.

Disclosure

The authors declare that they have no competing interests.

Author contributions to the study and manuscript preparation include the following. Conception and design: Azimi. Acquisition of data: Azimi. Analysis and interpretation of data: Azimi. Drafting the article: Azimi. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Azimi. Study supervision: Benzel.

References

Predicting surgical satisfaction with ANN in patients with LSCS


Manuscript submitted July 19, 2013.
Accepted December 2, 2013.

Please include this information when citing this paper: published online January 17, 2014; DOI: 10.3171/2013.12.SPINE13674.

Address correspondence to: Parisa Azimi, M.D., Functional Neurosurgery Research Center of Shohada Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Sharadari St., Tajrish Square, Tehran 1989934148, Iran. email: parisa.azimi@gmail.com.