A Fuzzy Neural Approach to Classifying Low Back Disorders Risks

Jozef Zurada
Department of Computer
Information Systems
College of Business
University of Louisville
Louisville, KY 40292
jozef.zurada@louisville.edu

Donghui Shi
Department of Computer
Engineering
School of Electronics and
Information Engineering
Anhui University of
Architecture
Hefei, China 230601
shidonghui@aiai.edu.cn

Jian Guan
Department of Computer
Information Systems
College of Business
University of Louisville
Louisville, KY 40292
jeff.guan@louisville.edu

Abstract

Classification of low back disorders (LBDs) risks for common industrial lifting jobs is important to control and prevention of this common disability. Of particular interest to the researchers is the use of data mining methods in risk classification. This paper presents an adaptive neuro-fuzzy inference system (ANFIS) model for classifying LBDs risks. Though neuro-fuzzy inference modeling is extensively used in classification in other areas, its effect is little studied and understood in this area. In addition to presenting a new risk classification model for LBDs this paper also adopts a rigorous approach to data sampling, calibration and testing that are absent in many existing studies. The results indicate that the neuro-fuzzy approach is a viable method for LBD risk classification.

1. Introduction

Low back disorders (LBDs) have been recognized as the most common and most costly musculoskeletal disorder found in the workplace [1]. Understanding risk factors that contribute to LBDs is critical to better control and prevention of disability resulting from such disorders [1], [2]. A common approach is to identify the risk level of common LBD related job characteristics [2]. Data mining methods have been shown to be promising in risk level identification [3]. This paper presents a neuro-fuzzy model for classifying risk levels of common industrial jobs for their potential in causing LBDs. The best overall, low-risk, and high risk correct classification accuracy rates, which are within the 76-77% range, are consistent with the results reported in the Zurada [3] study and shows the feasibility of the ANFIS system in classification the risk of LBDs due to manual lifting jobs.

2. Classification of low back disorders risks

Contrary to common belief increased automation and use of robotics in the workplace has not reduced incidences of LBDs as disability caused by low back pain has continued to rise steadily [4]. As LBDs represent the most common and costly musculoskeletal disorder found in the workplace, extensive research has been conducted to identify the factors that contribute to LBDs [1]. Occupational factors such as lifting/forceful movement, bending and twisting, and whole-body vibration have been found to significantly increase chances of LBDs among workers [5], [6], [7]. A landmark study by Marras et al identifies five factors: lift frequency (LIFTFR) – the count of lifts/hour, twisting velocity (TV), external load moment (ELMOM), sagittal torso bending angle (STBA), and lateral velocity (LV) that are related to LBD risks [2]. These variables represent trunk motion and workplace factors. The resulting experimental data set is based on an analysis of over 400 industrial lifting jobs from about 50 different companies from which 235 jobs are selected. The data set contains 235 cases, which are categorized as high LBD risk and low LBD risk according to injury and medical records. 111 cases, 47.2%, are high risk and 124, 52.8%, are low risk. Therefore the output variable, RISK, has a binary value of high risk or low risk. Table 1 contains a sample of the data set. The availability of this data set has enabled researchers to focus on classification models to identify risk levels [3], [8], [9], [10], [11], [12]. Ergonomics interventions to control and prevent LBD risks depend to a large degree on correctly classifying the risk levels of these factors.
Table 1. Sample data set of LBD risk factors

<table>
<thead>
<tr>
<th>Job#</th>
<th>LIFTFR</th>
<th>TV</th>
<th>ELMOM</th>
<th>STBA</th>
<th>LV</th>
<th>RISK</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>200</td>
<td>18.8</td>
<td>169.9</td>
<td>31.8</td>
<td>88.48</td>
<td>High</td>
</tr>
<tr>
<td>1</td>
<td>762.5</td>
<td>6.1</td>
<td>19.3</td>
<td>45</td>
<td>42.8</td>
<td>Low</td>
</tr>
<tr>
<td>147</td>
<td>167.3</td>
<td>3.5</td>
<td>1.2</td>
<td>10.6</td>
<td>54.5</td>
<td>High</td>
</tr>
<tr>
<td>148</td>
<td>175</td>
<td>12.2</td>
<td>57.1</td>
<td>3.9</td>
<td>54.5</td>
<td>High</td>
</tr>
<tr>
<td>149</td>
<td>90.9</td>
<td>7.7</td>
<td>12.6</td>
<td>35</td>
<td>32.4</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>31.3</td>
<td>6.9</td>
<td>63.4</td>
<td>8.2</td>
<td>38.2</td>
<td>Low</td>
</tr>
</tbody>
</table>

As a result, a stream of research has emerged that focuses on building classification models based on the Marras data set to correctly identify a job as high risk or low risk\(^1\) [3, 8, 9, 10, 11, 12]. These models include neural networks, decision trees, \(k\)-nearest neighbor, ensemble models, random forest, and ant colony optimization. The existing studies report results with significantly varying overall classification accuracies [3]. For example, the overall rates were between 74.2% and 81.6%. The lowest and the highest rates for low risk jobs were 72% and 83%, respectively; whereas the best and the worst rates for high risk tasks amounted to 64% and 94.6%, respectively [3]. A main drawback in many of the existing studies is the use of a single partition of the data set into a training set and a test set. Furthermore, in some studies with more than one partition, the papers reported the results from the best partition. Single partition tends to make the results sample specific, and because of the small data set the classification rates may substantially differ even for somewhat different splits. In addition many of these studies do not provide details on design, calibration, and testing, thus making it more difficult to compare their results with those of other studies [8]. Though hybrid methods such as neuro-fuzzy models have proven to be effective classifiers in many different areas, they have been rarely used in LBD risk classification [8]. Akay et al. uses Neuro-Fuzzy Classification (NEFCLASS) [8] to build an LBD risk classifier. Though the results represent a slight improvement over those of previous studies, the lack of details on design, calibration, and testing makes it difficult to make meaningful comparisons.

This paper presents results of an LBD risk classifier using Adaptive Neuro-Fuzzy Inference System (ANFIS). The ANFIS model has not been used in LBD risk classification. In addition a rigorous data sampling approach is used to make the results more generalizable. We randomly split the original data set into training, validation, and test sets. The training set is used for initial model building, the validation set is used for model tuning, whereas the test set is used to obtain unbiased estimates of model’s performance. This partitioning procedure was repeated 30 times as the average cumulative accuracy rates measured over the number of partitions tended to stabilize after this number of runs. Therefore, the final results for the test sets reported in this paper are averaged over 30 runs. We used a standard 0.5 threshold to determine whether a lifting task is classified as low risk or high risk. In addition, we present ROC (receiver operating characteristics) charts for the best models, which show the general performance of the models at any cut-off in the [0,1] range. Our computer simulation included several scenarios. We built and tested the ANFIS models using backpropagation and hybrid techniques for parameters’ estimation. We also used different partitions for training and validation as well as variable reduction techniques to examine the performance of the models for fewer attributes. The classification accuracy rates show that the ANFIS approach can be useful in distinguishing between high and low risk jobs that cause LBDs.

3. A neuro-fuzzy model for classifying low back disorders risks

Neural fuzzy inference systems have emerged from the fusion of artificial neural networks and fuzzy inference systems. They have formed a popular framework for modeling real world problems including classification. ANFIS is one of the better known neuro-fuzzy inference systems [13]. Figure 1 shows the architecture of the ANFIS LBD risk classification model in this paper. The model has 5 (\(m\)) inputs representing the 5 job characteristics described in the last section. Each of the inputs has 2 (\(n\)) membership functions. The model has 5 layers with a Takagi, Sugeno, and Kang (TSK) type fuzzy inference system.

\(^1\) Please refer to Zurada [3] for an excellent review of the research on classification models.
Layer 1: The number of nodes (N) in the first layer is the product of the input size (m=5) and the number (n=2) of the membership functions for each input variable, or \( N = m \times n \). The output of each node is defined as

\[
O_j = \mu_{ij}(X_i), \text{ for } i = 1, m, j = 1, n
\]

where \( \mu_{ij} \) is the \( j^{th} \) membership function for the input \( X_i \), and is given as follows:

\[
\mu(X) = \exp \left\{ -\left[ \frac{x-c}{a} \right]^b \right\}
\]

where \( a, b, \) and \( c \) are the premise parameters.

Layer 2: the output in this layer represents the firing strength of each rule. The output is the product of all of its inputs as follows:

\[
O_k = W_k = \mu_{j1}(X_1)\mu_{j2}(X_2)\cdots\mu_{jn}(X_n)
\]

for \( k = 1, R \) and \( R \) is the number of rules.

Layer 3: the output in this layer normalizes the weighing factor of each of the input nodes \( k \) as follows:

\[
O_k = \bar{W}_k = \frac{W_k}{W_1 + W_2 + \cdots + W_R}
\]

Layer 4: the output of this layer represents a weighted value of the first order fuzzy if-then rule as follows:

\[
O_k = \bar{W}_k f_k
\]

where \( f_k \) is the output of the \( k^{th} \) fuzzy rule as follows:

\[
\text{If } (X_1 \text{ is } A_{11}) \text{ and } (X_2 \text{ is } A_{22}) \text{ and } \cdots \text{ and } (X_n \text{ is } A_{mn}) \text{ then } f_k = \sum_{i=1}^{m} p_{ij}X_i + r_k
\]

where \( p_{ij} \) and \( r_k \) are called the consequent parameters and \( j = 1, n \) and \( k = 1, R \).

Layer 5: Finally this single node layer computes the overall output \( (F) \) of the ANFIS model as the sum of all the weighted outputs of the previous layer as:

\[
O = F = \sum_{k=1}^{N} \bar{W}_k f_k
\]

where \( f_k \) represents the output of the \( k^{th} \) TSK-type rules as defined in layer 4.

The parameters, both the premise parameters and consequent parameters, are learned/optimized in the training process. Two parameter optimization methods are used in training. The first method is backpropagation and the second method is called hybrid method that uses a mixture of backpropagation and least squares. Once the parameters are learned, a simple cutoff rate of 0.5 is used in classification. If a testing case is classified as high risk and our model generates an output \( \geq 0.5 \), then the case is classified as high risk; otherwise it is classified as low risk. The next section describes the results of the simulation.
4. Discussion of results

Computer simulation was conducted using MatLab Fuzzy Logic Toolbox. The models were built, calibrated, and tested on the training, validation, and test sets, respectively. All three sets were independent. Thus, out of 235 samples in the original data set, 211 samples (90%) were randomly allocated to the training and validation sets and always about 24 random samples (10%) were set aside for the test set. The models were created for several different scenarios. First, out of 211 samples we used four different allocations for the training and validation sets. These allocations were (50% and 50%), (60% and 40%), (90% and 10%), and (95% and 5%) for the training set and validation set respectively. Second, we ran the models for a full set of five input variables. Several variable reduction techniques were consistent in showing that two variables, LIFTFR and STBA, in this order, always had the least predictive power. Consequently, we also ran models for a reduced set of four variables (LIFTFR eliminated) and three variables (LIFTFR and STBA eliminated). Finally, we applied two parameter optimization methods, backpropagation and hybrid. The hybrid method is a combination of backpropagation and least squares. To make the results as sample independent as possible we created 30 and 50 random partitions of training, validation, and test sets, and averaged the classification accuracy rates over the number of partitions.

We report the overall average correct classification accuracy rates as well as the average rates for high and low risk tasks at a standard 0.5 cutoff point. If the task is associated with a high risk task and a model generates the probability ≥ 0.5, the task is classified as high risk; otherwise it is low risk. If costs of misclassifications of high risk task as low risk and low risk task as high risk are equal, the standard 0.5 cut-off works fine. However, if costs are uneven, the performance of the models can be examined with ROC charts, which show the global performance of the models and a continuum of cutoff points in the range [0,1]. For example, if the cost of misclassifying a high risk task as low risk is 2.3 times greater than the cost of misclassifying a low risk task as high risk, a 0.3 threshold should be used. Thus, if a model generates a probability ≥ 0.3, the task will be categorized as high risk.

![Figure 2. The average cumulative rates measured over the number of partitions for the best model using backpropagation, (95%,5%) partition, and 3 attributes](image)

Dividing the small data set into three subsets (training, validation, and test) causes instability in the classification accuracy rates. To counter this effect and to obtain unbiased and realistic rates we created 30 and 50 different random generations of the three sets and then averaged the results over 30 and 50 runs,
respectively. Figure 2 depicts the cumulative average overall high risk and low risk correct classification accuracy rates measured over the number of runs/partitions for one of the models. The results show that the average rates tend to stabilize after about 30 runs, as can be seen in Figure 2.

Tables 2 and 3 depict the classification results from our computer simulation of the ANFIS system. It is clear from Table 2 that the backpropagation method outperforms the hybrid method. In addition the ANFIS system with 3 variables tend to classify better than the system with the full set of 5 variables. Also, the way the data is partitioned seems to matter because partitions (90%, 10%) and (95%, 5%) yielded the best results. The highest overall average classification accuracy rate, 76.4%, is obtained for backpropagation, 3 attributes, and the (95%,5%) partition. Table 3 presents the overall classification accuracy as well as the rates for high risk and low risk tasks for the best models only. Again, the backpropagation method appears to classify better than the other one and the classifier built on the (90%, 10%) and (95%,5%) partitions for 5 and 3 variables yielded the best results. The highest overall, low risk, and high risk rates are 76.4%, 75.8%, and 76.6%, respectively [backpropagation, (95%,5%) partition, 3 variables]. The second best rates are 75.1%, 76.1%, and 75.1% [backpropagation, (90%,10%) partition, 5 variables].

We present the ROC charts for the best models only (Figure 3). As mentioned before a typical probability cut-off is 0.5. The exact locations of the 0.5 cut-offs on the curves depends on the models and are located between the coordinates (0.23, 0.72) and (0.25, 0.79). Higher and lower cut-offs are located in the left bottom corner below these coordinates and upper right corner above the coordinates, respectively, of the curves. One can see that there are slight differences between the performances of the models for different cut-offs.

<table>
<thead>
<tr>
<th># of Variables</th>
<th>Optimization (Training, Validation)</th>
<th>Overall</th>
<th>Low Risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Variables</td>
<td>Hybrid (95%,5%)</td>
<td>72.8</td>
<td>74.8</td>
<td>70.4</td>
</tr>
<tr>
<td></td>
<td>Backpropagation (90%,10%)</td>
<td>75.1</td>
<td>76.1</td>
<td>75.1</td>
</tr>
<tr>
<td>4 Variables</td>
<td>Hybrid (90%,10%)</td>
<td>73.7</td>
<td>73.8</td>
<td>74.0</td>
</tr>
<tr>
<td></td>
<td>Backpropagation (95%,5%)</td>
<td>74.5</td>
<td>75.3</td>
<td>73.8</td>
</tr>
<tr>
<td>3 Variables</td>
<td>Hybrid (60%,40%)</td>
<td>74.9</td>
<td>75.1</td>
<td>74.4</td>
</tr>
<tr>
<td></td>
<td>Backpropagation (95%,5%)</td>
<td>76.4</td>
<td>75.8</td>
<td>76.6</td>
</tr>
</tbody>
</table>

5. Conclusions

This paper presents a first ANFIS model for LBD risk classification using a data set from a landmark study [2]. Computer simulation shows that the ANFIS model for detecting risks of LBDs due to manual tasks is a viable option. The model uses both the backpropagation method and hybrid method for parameters’ optimization. A rigorous data sampling, calibration, and testing procedure is followed in the training and testing of the ANFIS model. The computer simulation indicates that 3 variables (reduced from the original 5 variables) and a (95%, 5%) allocation for training and validation sets generate the best classification results. The classification rates were 76.4% for overall tasks, 75.8% for high risk tasks, and 76.6% for low risk tasks. These results are similar to the results produced by the neural network model with 3 variables in the comprehensive study by Zurada [3].
Future studies may explore the areas under ROC curves and the rules generated by ANFIS. The rules are important because they permit interpretation of the data and contribute to better understanding of types of lifting jobs which cause LBDs. The ability to interpret the rules could help efforts to set guidelines to prevent injuries.

![ROC Charts](image)

**Figure 3.** The ROC charts for the best models with 5, 4, and 3 attributes.

6. References


[4] R. A. Deyo, "Low-back pain is at epidemic levels. Although its causes are still poorly understood, treatment choices have improved, with the body's own healing power often the most reliable remedy," *Scientific American*, p. 49, 1998.


[8] D. Akay, M. A. Akcayol, and M. Kurt, "NEFCLASS based extraction of fuzzy rules and


