Classification of patients with pain based on neuropathic pain symptoms: Comparison of an artificial neural network against an established scoring system

Michael Behrman [a], Roland Linder [b], Amir H. Assadi [a], Brett R. Stacey [c], Misha-Miroslav Backonja [d],*

[a] *Department of Mathematics, University of Wisconsin – Madison, Madison, WI, USA*
[b] *Institute for Medical Informatics, University of Luebeck Ratzeburger Allee 160, D-23538 Luebeck, Germany*
[c] *Oregon Health and Science University Portland, OR, USA*
[d] *Department of Neurology, University of Wisconsin – Madison, 600 Highland Avenue, Madison, WI 53792, USA*

Abstract

Wider use of pain assessment tools that are specifically designed for certain types of pain – such as neuropathic pain – contribute an increasing amount of information which in turn offers the opportunity to employ advanced methods of data analysis. In this manuscript, we present the results of a study where we employed artificial neural networks (ANNs) in an analysis of pain descriptors with the goal of determining how an approach that uses a specific symptoms-based tool would perform with data from the real world of clinical practice. We also used traditional statistics approaches in the form of established scoring systems as well as logistic regression analysis for the purpose of comparison. Our results confirm the clinical experience that groups of pain descriptors rather than single items differentiate between patients with neuropathic and non-neuropathic pain. The accuracy obtained by ANN analysis was only slightly higher than that of the traditional approaches, indicating the absence of nonlinear relationships in this dataset. Data analysis with ANNs provides a framework that extends what current approaches offer, especially for dynamic data, such as the rating of pain descriptors over time.
© 2006 European Federation of Chapters of the International Association for the Study of Pain. Published by Elsevier Ltd. All rights reserved.

Keywords: Chronic pain; Classification; Artificial neural networks; Logistic regression; Pain intensity rating

1. Introduction

Pain assessment and characterization is the main focus of both the clinical and laboratory behavioral study of pain. It is based on patients' ability to provide descriptive statements about qualities of pain and other associated abnormal sensations through clinical history, assessment instruments, and pain intensity ratings. The validity of a variety of pain intensity ratings has been investigated in numerous studies over the past few decades. Although pain intensity ratings scales have been validated, an emerging consensus warns that simple pain ratings carry many hidden individualized meanings for each particular patient rather than generalizable information about each specific sensation (Williams et al., 2000). Conceptually, pain descriptors and symptoms that they represent would seem to offer more specific information. This is well recognized in the

---

* Corresponding author. Tel.: +1 608 263 5445; fax: +1 608 263 0412.
  E-mail address: backonja@neurology.wisc.edu (M.-M. Backonja).

clinical arena, but it is yet to be proven that this information contributes to an improvement in pain assessment, especially when attempts are made to utilize specific pain descriptors for elucidation of underlying mechanisms. The development of specific questionnaires such as the Neuropathic Pain Questionnaire (NPQ) (Krause and Backonja, 2003) for the assessment of neuropathic pain, has advanced pain research. At the same time detailed questionnaires require investigators to deal with significant increases in the amount of self-report data. To facilitate this task, summary scores help differentiate neuropathic pain (NP) from non-neuropathic pain (nonNP).

Calculating a score seems to be rather archaic in the light of more advanced computer-based approaches such as artificial neural networks (ANN). ANN are assumed to offer an advantage in dealing with nonlinear relationships that may exist among pain descriptors, i.e., relationships that do not obey a relation like "the higher descriptor A the higher descriptor B". For instance, the majority of biological phenomena are nonlinear, e.g., most relations between hormones and their effects are considered nonlinear, as are all events that are subject to threshold, such as neuronal spiking (Fallahati et al., 2001, 2002). Thus, the objective of the present study is to evaluate if ANN can improve upon current scoring systems. Since the ANN training set and the established scoring systems are based upon different patient data, logistic regression (LR) is used on the same data as the ANN. By employing LR, a conventional statistical method, we are able to establish to what extent nonlinearities are important in the data.

Artificial Neural Networks have emerged as a powerful tool for solving statistical regression and pattern classification problems in a multitude of domains: for comprehensive state-of-the-art expository surveys (Arbib, 2002), for comprehensive treatment (Príncipe et al., 2000), and for biological pain networks (Devor, 2002). The main concept of the ANN is step-wise improvement of the estimates for model parameters, so that the pattern classification errors are minimized. The heuristics behind most variants of ANN are loosely based on networks of biological neurons in which Hebb's plasticity mechanism is regarded as the basis for learning. The model parameters are reminiscent of the strength of synaptic messages passing among neurons. Step-wise adjusting of the parameters is called *training the network*, and often is implemented via a popular architecture called a multi-layer perceptron (MLP). To train an MLP, an efficient algorithm called Back Propagation (Rumelhart et al., 1986; Príncipe et al., 2000) has been adapted and refined in dozens of ways. Numerous properties of neural network models, such as adaptation and robustness (Haykin, 1999; Príncipe et al., 2000; Devor, 2002), complement traditional statistical methods, and in some cases such as non-parametric regression, ANN is the preferred methods of data analysis.

2. Methods

2.1. Patients and the diagnostic procedure

Patients from two academic pain clinics were recruited for this study: University of Wisconsin Hospital and Clinics in Madison, Wisconsin and Oregon Health and Science University (OHSU) in Portland, Oregon. The respective institutional review boards at each institution approved the study, which was conducted from 2000 to 2002.

All new patients presenting to the respective pain clinics with chronic pain of six months or longer duration were approached to participate in the study and those who signed consent forms were given a set of questionnaires. The questionnaires included an early version of the Neuropathic Pain Questionnaire (NPQ) and the Neuropathic Pain Scale (NPS). The early version of NPQ was utilized because it was the only version available at the time. The later version of NPQ was the result of additional analysis and revisions that were performed at a later date (Krause and Backonja, 2003). Patients, who had more than one component of pain, such as radicular pain in the leg as well as mechanical low back pain, were asked to concentrate on and rate the most severe component of their pain. This procedure is in accord with the procedures suggested by the literature (Backonja and Stacey, 2004; Bennett, 2001; Rasmussen et al., 2004).

Patients were evaluated by pain fellowship-trained pain clinicians with pain medicine certification. Medical and pain diagnoses as well as classification of patients' primary pain into neuropathic or non-neuropathic were made on the basis of comprehensive medical and pain history, physical examination and laboratory tests. The clinical pain diagnosis for each patient served as the accepted correct diagnosis for this study. The results of the neuropathic pain-specific questionnaires (NPS and NPQ) were not known to the examiners at the time of visit and were not used as the basis of diagnosis.

The primary criterion for the diagnosis of neuropathic pain was the presence of pain accompanied by sensory and motor abnormalities or only sensory abnormalities, in patients with an appropriate history consistent with a neurological injury or disorder. In the case of diagnostic uncertainty, laboratory and imaging studies were utilized to confirm the diagnosis. For patients who experienced pain that was thought to be due to co-existing neuropathic and non-neuropathic pain mechanisms, such as radicular low back pain, classification was made on the basis of the most predominant component of pain.

Patients were excluded if they had a condition that could not be diagnosed or receive a mechanistic classification (neuropathic vs. non-neuropathic), if they could not specifically rate their pain, if they had diffuse total-body pain, or if incomplete medical records compromised our ability to diagnose the patient's condition.

2.2. Scoring systems

The Neuropathic Pain Questionnaire (NPQ), including early version used in this study, was developed previously using factor analysis and discriminant analysis on 382 patients with purely neuropathic and non-neuropathic pain (Krause and Backonja, 2003). We employed this scoring system on the present data to compare scoring results with the LR and ANN results. Additionally, we created an ad hoc scoring system for the Neuropathic Pain Scale (NPS) as described by Argoff et al. (2004). According to Argoff et al., the NPS mean composite scores from all 10 items were used to evaluate pain reduction. We used the standard deviations they reported as a cutoff point to create neuropathic (greater than 42) and non-neuropathic (42 or less) groups.

2.3. Univariate analysis

Ordinal variables, e.g., those from the NPQ and NPS, were treated as linear variables. A linear fit is regarded as the natural approach when expecting a monotone effect of the ordinal covariate (Agresti, 1990). Thus, only evaluating continuous variables, the two-tailed Wilcoxon test for two independent samples (Mann–Whitney U test) was employed to reveal variables that possibly enable a helpful differentiation between NP and nonNP.

2.4. Multivariate analysis

The multivariate analysis was performed by a logistic regression (LR) model, the classical approach for a binary outcome, as well as by an artificial neural network (ANN), a method that is assumed to detect not only linear but also nonlinear relationships between the predicting variables and the dependent variable (NP, nonNP) as well as relationships between the predicting variables. Comparable results of both methods would indicate that nonlinear relationships are of no importance for the present classification problem. Otherwise – in case the ANN should outperform the LR approach – nonlinearities must be considered.

2.5. Artificial neural networks

ANNs are computer-based techniques which have been frequently applied for classifying clinical data and patients (Dybowski, 2001; Malmgren et al., 2000). ANNs use nonlinear mathematical equations to succes-sively develop meaningful relationships between input and output variables through a learning process. They have a "learning" or "training phase" and a "recall phase". In the learning phase, the relationships between the different input and output variables are established by adaptations of the weight factors assigned to the connections between the layers of artificial neurons. This adaptation is based on rules that are set in the learning algorithm. At the end of the learning process, the weight factors are fixed. In the recall phase, data from cases not previously interpreted by the network are entered, and an output is calculated based on the above-mentioned, and now fixed, weight factors. Diagrammatic representation of a standard feedforward network is given in Fig. 1. Data are entered at the input neurons and further processed in the hidden layer and output layer.

Features were fed into a feedforward network implemented in a prototype called ACMD, standing for "Approximation and Classification of Medical Data" (Microsoft Visual C++, Office Software International, USA) (Linder and Pöppl, 2001), which is in principle based on a multilayer perceptron (Rumelhart et al., 1986), using adaptive propagation (Linder et al., 2000) for automatic learning. Each of five networks were arranged as an ensemble, which makes them more robust than single networks and more suitable for learning with only a small amount of data (Bishop, 1995). To make the learning less susceptible to so-called overfitting (Geman et al., 1992), a strategy known as "Early Stopping" was used (Prechelt, 1998). ACMD comprises a number of further strategies both improving the generalization performance and accelerating the convergence speed. Several benchmarks give evidence that ACMD provides an excellent tool for learning feedforward networks. For details please see "ACMD: A practical tool for automatic neural net based learning" (Linder and Pöppl, 2001). A sequential backward feature selection
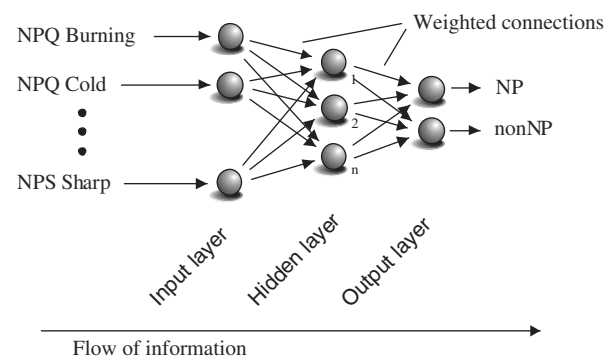

Fig. 1. Structure of ANN used for the differentiation of NP and nonNP patients: as an input pain descriptors are entered that result from the Neuropathic Pain Questionnaire (NPQ) and the Neuropathic Pain Scale (NPS), respectively. This information is forwarded to a maximum of n hidden neurons. The intermediate results of the hidden neurons are forwarded to output neurons representing neuropathic pain (NP) and non-neuropathic pain (nonNP).

was performed by the neural net clamping technique proposed by Wang (Wang et al., 1998).

2.6. Logistic regression

Binary logistic regression inclusive feature selection was calculated by the software SPSS V.11.5.1 (SPSS Inc., Chicago, USA). Exclusively default settings were chosen. In a resulting model, the predicted value $\widehat{P}(y)$ of a patient was calculated by

$$\widehat{P}(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_p \cdot x_p)}}$$

where β_i denotes the parameters estimate for the ith variable. As illustrated by the above equation, the LR approach assumes no interdependencies between the predicting variables and only linear relationships.

2.7. Validation

For both methods, validation was done by 5-fold cross-validation using 4/5 of all cases for specifying each classification method and testing with the remaining cases. By performing five rounds of allocation, every case is used once for the purpose of validation.

3. Results

3.1. Demographics

The patients' mean age was 46.4 years with a range 21–79 years. A total of 304 patients entered the study. Of the 165 who were enrolled complete data was available for 155 and they were classified as either predominantly neuropathic pain ($n = 91$) or non-neuropathic ($n = 64$). For 114 patients for whom duration of pain was established on the basis of a specific date of pain onset, there was an average duration of 78.9 (+64.2) months for patients with neuropathic and 131.6 (+113.4) months for patients with non-neuropathic pain.

Our clinics see many patients with neuropathic pain and the subjects' diagnoses reflected our patient populations: post-traumatic neuralgias including complex regional pain syndrome type I and II, radiculopathies, a wide variety of peripheral neuropathies, post-herpetic neuralgia, trigeminal neuropathy, and many etiologies of central pain syndrome. Non-neuropathic pain included many different diagnoses and etiologies such as arthritis and other musculoskeletal conditions, abdominal and other visceral pain, and headaches.

3.2. Scoring systems

The NPQ scoring system performed similarly to the LR and ANN overall. However, it performed better

Table 1
Results of scoring by NPQ and NPS

Score	NP correctly classified	NonNP correctly classified	Accuracy
NPQ	65.9% (60/91)	71.9% (46/64)	68.4% (106/155)
NPS	67.0% (61/91)	39.1% (25/64)	55.5% (86/155)

with nonNP cases than with NP cases (Table 1). The NPS scoring system did not perform well. Since classifying by chance achieves an accuracy of 58.7% (91/155), i.e., assigning all patients who suffer from NP, NPS performed worse than chance.

3.3. Univariate analysis

Evaluating the predictive power of each variable, the Mann–Whitney U test revealed a statistically significant difference regarding NP and nonNP in 12 of the 32 variables. Box-and-Whisker plots of the 12 variables in question are shown in Fig. 2 demonstrating a wide overlap between the boxes for NP and nonNP, respectively. As a consequence, none of the variables alone enables a sufficient distinction of the type of pain.

3.4. Multivariate analysis

Applying the ANN, NP and nonNP was correctly predicted in 69.0% of all cases (NP, 75.8%; nonNP, 59.4%). Applying the LR, NP and nonNP was correctly predicted in 67.7% of all cases (NP, 73.6%; nonNP, 59.4%). Thus, both methods achieved a statistically significant result (χ^2 test: p-value < 0.001). The difference between the accuracies of ANN and LR is only 1.3%. A two-tailed McNemar test showed that there is no significance between the ANN and LR results (p-value = 0.832) (Table 2).

Evaluating subsets of the patients comprising at least 13 cases, the predictive power greatly varies between the diagnostic sub-entities (Fig. 3). The degree of certainty with which the ANN is able to classify the diagnostic sub-entities is consistent with clinical experience.

4. Discussion

This study of non-selected chronic pain patients demonstrated that advanced methods of analysis, ANN and LR, performed slightly better in classifying patients with neuropathic pain than they did of those with non-neuropathic pain, when compared to the standard scoring of the NPQ. Characterizing a patient's pain symptoms and descriptors is always the starting point in the diagnosis and assessment of pain. Our results are consistent with the notion that neuropathic pain symptoms could provide a reasonable level of certainty in the diagnostic
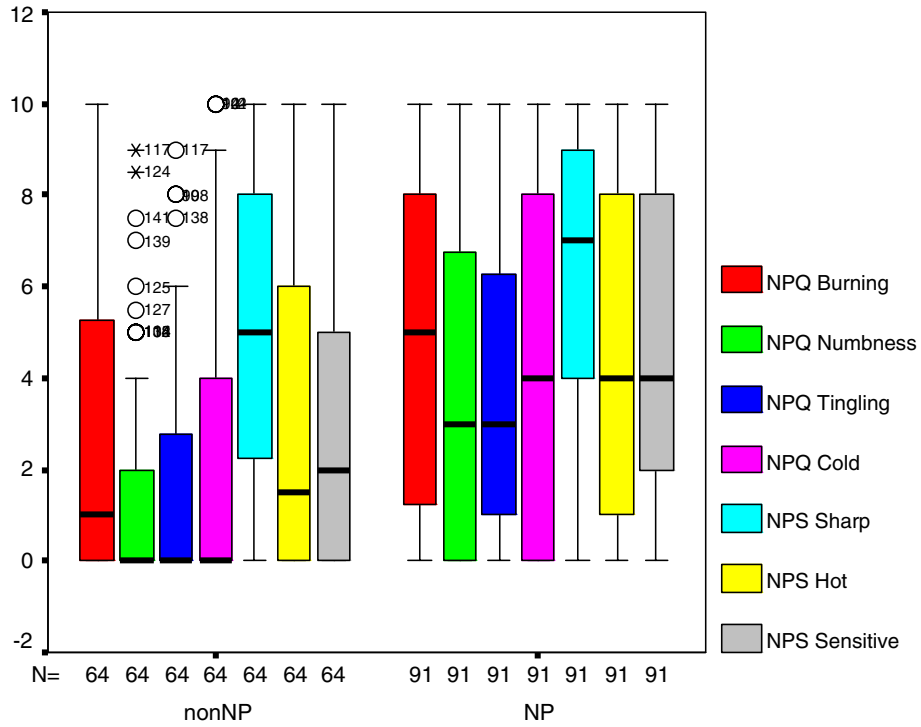
Fig. 2. Box-and-Whisker Plot. Variables shown had $p < 0.01$ (Mann–Whitney U test) and for this reason they have the greatest differentiating power. To make the different scaling of NPQ and NPS comparable, all variables are normalized into the interval [0, 10]. The only reason to scale the y-axis from −2 to 12 is not to mask the ends of the whiskers of the Box-and-Whisker Plots. Outliers (∗) and extreme values (○) have been marked.

Table 2
Classification of cases by ANN and LR

	ANN: cases wrongly classified	ANN: cases correctly classified
LR: cases wrongly classified	38	12
LR: cases correctly classified	10	95

process and that many pain descriptors are necessary since no single descriptor is pathognomonic for neuropathic pain. These observations are in accord with those that lead to development of neuropathic pain tools, including the NPQ, the Leeds assessment of neuropathic symptoms and signs (LANSS pain scale) (Bennett, 2001)
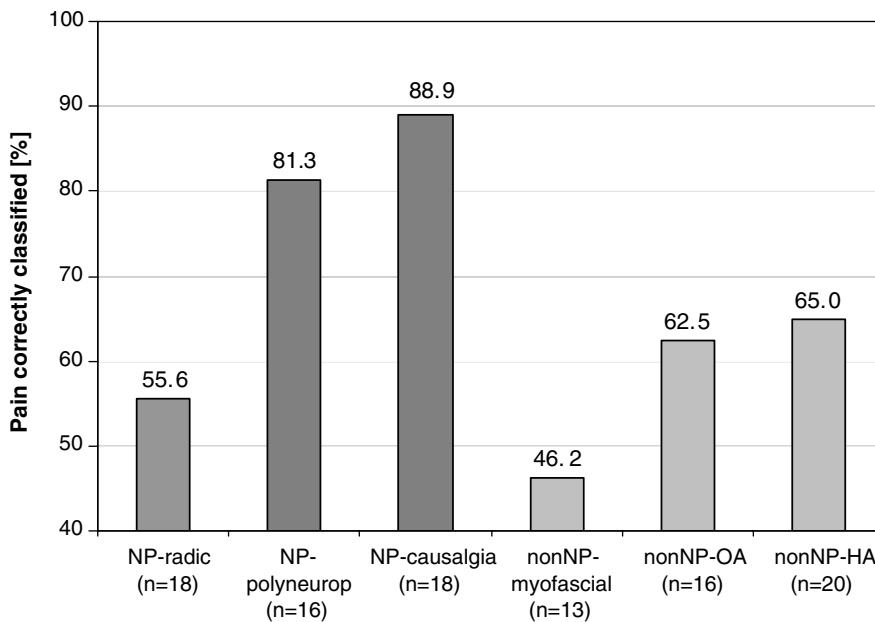

Fig. 3. Pain prediction for diagnostic sub-entities with $n \geqslant 13$ using the ANN results. *Abbreviations*. NP, neuropathic pain; nonNP, non-neuropathic pain; radic, radiculopathy; OA, osteoarthritis; HA, headaches; polyneurop, polyneuropathy.

and the Neuropathic Pain Symptoms Inventory (NPSI) (Bouhassira et al., 2005), all of which were developed with the goal of differentiating neuropathic pain from non-neuropathic pain.

The good performance of the NPQ scoring system reinforces previous studies showing this to be a valuable tool in identifying neuropathic pain (Krause and Backonja, 2003). Accuracy obtained by the NPQ scoring system (68.4%) is very similar to the 69.0% achieved by the ANN. Since the NPQ scoring system was based on 382 patients and therefore on a larger database than the ANN (155 patients), one may speculate that the ANN could have been somewhat more accurate using the same large database for training. This assumption is supported by the results of the LR approach based on 155 patients that is worse than the NPQ. However, a direct comparison of ANN and LR does not show a significant advantage of the ANN technique. This indicates that nonlinear relationships between pain descriptors do not play a significant role in classifying NP and nonNP patients for this dataset. NPS was not designed to be a discriminative tool and consequently it is not surprising that NPS did not perform well in classifying NP and nonNP patients.

The degree of certainty with which the ANN is able to classify the diagnostic sub-entities is consistent with clinical experience. For example, with neuropathic pain due to polyneuropathy (NP polyneurop) or causalgia (NP causalgia), where primary pathologic and pain mechanisms are due to direct neurological involvement, classification is more likely to be correct even if only on the basis of symptoms severity. In contrast, neuropathic pain due to radiculopathy (NP-radic) which is frequently associated with low back pain, is less likely to be classified correctly on the basis of the symptoms severity rating alone, in spite of patients' best efforts to rate only that particular pain. The presence of non-neuropathic pain certainly increases the challenge of accurately classifying the neuropathic pain based solely on symptoms.

The main methodological finding of our study is that nonlinearities within the data obtained at one time point are not significant since the ANN and LR results are roughly equivalent. This finding is reinforced by the rough equivalence of the ANN and LR results with the NPQ scoring system. Future work will evaluate if this is also true for data recorded over a period of time. Also, the diagnostic sub-entity classification results (Fig. 3) show that patients who exhibit mixed NP and nonNP traits are difficult to classify with data from NPQ and NPS. In comparison, cases with diagnoses which are "classical" are relatively easily classified. This likely indicates that more data for the difficult-to-classify cases are needed, as well as new methods for identifying whether a given patient is in a difficult-to-classify category need to be employed. We have also shown that modern methods, such as the ACMD ANN package,

handle pain classification problems as well as or better than methods established in pain research, such as LR and the NPQ scoring system.

Neuropathic pain is a new and still evolving area of pain research and classification of patients based on clinical presentation and according to diagnostic criteria is undergoing debate (Backonja, 2003). Consequently, at times questions are raised regarding the classification of pain disorders such as CRPS type I. In this study, CRPS type I was included into neuropathic pain group on the basis of sensory abnormalities of not only positive but also negative sensory signs, such as loss of sensation to one or more sensory stimulation modalities, as well as motor findings of weakness and movement disorders, such as tremor (Stanton-Hicks, 2003).

The clinical diagnostic process is complex and draws from multiple levels of clinical evaluation. For routine clinical use a diagnostic accuracy of 68% (LR) or 69% (ANN) may not be acceptable. However, this degree of accuracy derived from pain descriptors and pain qualities alone is statistically significant and points to an important area for further exploration. It is our belief that continuing to employ methods such as ANN, which are able to take advantage of nonlinearities within data, is important to further research, especially as more complete patient profiles are collected. Traditionally there has been a hesitation if not a strong resistance to obtaining large amounts of data over periods of time because of concerns about methods of analysis. We are confident that the methods presented here are appropriately suited for that type of data.

References

Argoff C, Galer B, Jensen M, Oleka N, Gammaitoni A. Effectiveness of the lidocaine patch 5% on pain qualities in three chronic pain states: assessment with the Neuropathic Pain Scale. Curr Med Res Opin 2004;20(Suppl. 2):S21–8.

Agresti A. Categorical data analysis. New York: Wiley; 1990.

Arbib MA. Handbook of brain theory and neural networks. 2nd ed. Cambridge (MA): Bradford Books/The MIT Press; 2002.

Backonja M. Defining neuropathic pain. Anesth Analg 2003;97(3):785–90.

Backonja M, Stacey B. Neuropathic pain symptoms in relation to overall pain. J Pain 2004;5(9):491–7.

Bennett M. The LANSS Pain Scale: the Leeds assessment of neuropathic symptoms and signs. Pain 2001;92:147–57.

Bishop CM. Neural networks for pattern recognition. Oxford: Clarendon Press; 1995.

Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, Cunin G, Fermanian J, Ginies P, Grun-Overdyking A, Jafari-Schluep H, Lanteri-Minet M, Laurent B, Mick G, Serrie A, Valade D, Vicaut E. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). Pain 2005;114(1–2):29–36.

Devor M. In: Arbib MA, editor. Pain networks. 2nd ed. Cambridge (MA): Bradford Books/The MIT Press; 2002.

Dybowski R. Clinical applications of artificial neural networks. Cambridge: Cambridge University Press; 2001.

Fallahati D, Backonja M, Eghbalnia H, Assadi A. Informatics of pain: applications to basic and clinical research. In: Proceedings of the first international symposium in neuroinformatics, Vienna, Austria; 2001.

Fallahati D, Backonja M, Eghbalnia H, Assadi A. Dynamic PCA for network feature extraction in multi-electrode recording of neurophysiological data in cortical substrate of pain. Neurocomputing 2002;4446:401–5.

Geman S, Bienenstock E, Doursat R. Neural networks and the bias/variance dilemma. Neural Comput 1992;4:1–58.

Haykin SS. Neural networks: a comprehensive foundation. 2nd ed. Upper Saddle River (NJ): Prentice Hall; 1999.

Krause S, Backonja M. Development of a Neuropathic Paine Questionnaire. Clin J Pain 2003;19:306–14.

Linder R, Pöppl SJ. ACMD: A practical tool for automatic neural net based learning.. Lect Notes Comp Sci 2001;2199:168–73.

Linder R, Wirtz S, Pöppl SJ. Speeding up backpropagation learning by the APROP algorithm. Millet (Alta.) Canada: ICSC Academic Press; 2000.

Malmgren H, Borga M, Niklasson L, editors. Artificial neural networks in medicine and biology. Proceedings of the Annimab-1 conference 2000. Telos: Springer; 2000.

Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. Neural Networks 1998;11:761–77.

Príncipe JC, Euliano NR, Lefebvre WC. Neural and adaptive systems: fundamentals through simulations. New York: Wiley; 2000.

Rasmussen PV, Sindrup SH, Jensen TS, Bach FW. Symptoms and signs in patients with suspected neuropathic pain. Pain 2004;110(1–2):461–9.

Rumelhart DE, Hinton GE, Williams RJ. Learning Representations by Back-Propagating Errors. Nature 1986;323:533–6.

Stanton-Hicks M. Complex regional pain syndrome. Anesthesiol Clin North Am 2003;21(4):733–44.

Wang W, Jones P, Partridge D. Ranking pattern recognition features for neural networks. Berlin: Springer; 1998.

Williams ACD, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. Pain 2000;85(3):457–63.