TOWARDS AN AUTOMATIC DIAGNOSIS SYSTEM FOR ACUTE ABDOMINAL PAIN

Support Vector Machines for the Diagnosis of Diverticulitis and Non-specific Abdominal Pain

Malin Björnsdotter Åberg
Institute of Neuroscience and Physiology, University of Gothenburg, Box 432, SE-405 30 Göteborg, Sweden

Kajsa Nalin
Centre of Interdisciplinary Research/Cognition/Information, SE-405 30 Göteborg, Sweden

Lars-Erik Hansson
Department of Surgery, Sahlgrenska University Hospital/Östra, SE-416 85 Göteborg, Sweden

Helge Malmgren
Department of Philosophy, University of Gothenburg, SE-405 30 Göteborg, Sweden

Keywords: Support vector machines, computer-aided diagnostics, acute abdominal pain.

Abstract: The process of medical diagnosis is highly complex, and automatic decision support systems are appealing. In this study we investigate the feasibility of automating one such decision-making process, namely the diagnosis of patients seeking care for acute abdominal pain, and, specifically the diagnosis of acute diverticulitis. We used a linear support vector machine (SVM) to classify diverticulitis from all other reported cases of abdominal pain and from the important differential diagnosis non-specific abdominal pain (NSAP). Using a database containing 3 337 patients, the SVM obtained results comparable to those of the doctors. The distinction between diverticulitis and non-specific pain was substantially better for the SVM. Here the doctor achieved a sensitivity of 0.714 and a specificity of 0.963. When adjusted to the physicians results, the SVM sensitivity/specificity was higher at 0.714/0.985 and 0.786/0.963 respectively. Age was found as the most important factor for diagnosis, closely followed by C-reactive protein level and various pain indicators on the left hand side. Thus, the support vector machine is a promising tool in the diagnosis of acute abdominal pain.

1 INTRODUCTION

The process of medical diagnosis and decision-making, that is, the classification of patients into disease groups based on various symptoms, is a highly complex problem – as evidenced by the near decade-long training required by specialist physicians. Computer-based decision support systems are therefore appealing tools in medical diagnostics, and in this study we investigate one such decision-making process, namely the diagnosis of patients seeking care for acute abdominal pain (AAP).

Emergency ward doctors face a highly demanding situation, where medical decisions with substantial impact on patients must be made under time pressure. The large number of potentially relevant physical measurements, from blood factors to face color, in combination with a stressful situation yields a challenging decision-making process. Moreover, physicians have reported lack of relevant experience and continuous feedback among other factors that affect the decision-making process negatively (Nalin, 2006). In addition, disease symptoms are highly variable between individuals, leaving the doctor to rely heavily on experience (Hansson, 2002).

Standardized, computer-based decision support systems, automatically identifying typical disease patterns in patient data, are thus appealing as a complement to the trained physician. These systems generally consist of computer models, or classifiers, which are trained to discover patterns related to a given disease in supplied patient data where the final

Björnsdotter Åberg M., Nalin K., Hansson L. and Malmgren H. (2009).
TOWARDS AN AUTOMATIC DIAGNOSIS SYSTEM FOR ACUTE ABDOMINAL PAIN - Support Vector Machines for the Diagnosis of Diverticulitis and Non-specific Abdominal Pain.
In *Proceedings of the International Conference on Health Informatics*, pages 51-57
DOI: 10.5220/0001546200510057

diagnosis is known. The classifiers are then applied to new patients, where an instantaneous diagnosis is made in order to assist the doctor.

Results on computer-aided diagnosis of abdominal pain were reported as early as in 1972, where de Dombal reported a surprisingly high diagnostic accuracy (91.8% vs. 79.6%) using decision support compared with the unaided examination (de Dombal et al., 1972). Moreover, a large British multi-center study with more than 16.000 patients confirmed the utility of computer aided diagnostic with accuracies of 65% vs. 46% (Adams et al., 1986).

A common acute abdominal disease, often a reason for emergency hospital admission – especially in elderly patients — is diverticulitis of the colon (Ambrosetti et al., 1994; Ferzoco et al., 1998; Young-Fadok et al., 2000; Laurell et al., 2006). The diagnosis is typically made at the emergency department, based on both medical history and clinical indications. The clinical presentation of acute diverticulitis was recently described by Laurell and colleagues, as well as the natural short-term development of the disease (Laurell et al., 2007). Primary diagnosis sensitivity, by the physician, was reported to be 64%, with a specificity of 97%. Moreover, Laurell et. al. identified non-specific abdominal pain (NSAP) as one of the most important differential diagnoses. For NSAP, the primary diagnosis sensitivity was reported to be 43%, with a specificity of 90%.

In the current study, we investigate the feasibility of using a decision support system for the automatic diagnosis of acute diverticulitis, contrasted with all other reported cases of abdominal pain and from the diagnosis category non-specific abdominal pain. Using a state-of-the-art classifier, namely (linear) support vector machines and feature selection, we also attempt to understand the underlying factors that are key to identifying diverticulitis.

2 METHODS

2.1 Data Acquisition

Mora Hospital in northern Sweden is a district hospital serving a population of 87 000 individuals, providing full emergency services. During the period of February 1997 to June 2000, all patients older than one years of age admitted to the hospital with abdominal pain of duration of up to 7 days were registered in a database. Details were registered according to a standardized form for history, clinical indications and laboratory results. The attending physician suggested a diagnosis, and a final diagnosis was given when the patient left the hospital. A definitive diagnosis was later established by a follow-up study of the patient's journal. Data for 3 337 patients was thus acquired.

A non-reported value can be assumed irrelevant and within the normal range, and missing data was, therefore substituted by estimated normal values. Moreover, normal values can be used in practical applications of decision support systems, without any knowledge of the statistics of the present sample.

Thus, a data-set consisting of 3 337 patients with 117 measured variables and an initial diagnosis by a trained physician was obtained. Out of the 3 337 patients, 148 obtained diverticulitis as a definitive (retrospective) diagnosis, whereas 1340 were diagnosed as having non-specific abdominal pain. In the training of the automatic system and in the performance analysis, these definitive diagnoses were the desired output of the system.

2.2 Support Vector Machines

Support vector machines (SVMs) is a type of classification algorithm which maximizes the geometric margin between the data classes and the separating hyperplane (Suykens et al., 2002).
Given our training data:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_{i=1}^{n} \qquad (1)$$

where y_i is the disease category (-1 or 1) to which patient $\mathbf{x}_i$ belongs, the hyperplane that maximally separates the data points must fulfill the following inequalities:

$$\omega \cdot \mathbf{x}_i + b \geq d \text{ for all } i \text{ where } y_i = 1 \qquad (2)$$

$$\omega \cdot \mathbf{x}_i + b \leq -d \text{ for all } i \text{ where } y_i = -1 \qquad (3)$$

where ω is the weight vector, b is the bias and d is the separating margin.

The SVM model is trained by adapting it's weights to the data at hand, using an algorithm that finds the optimal hyperplane that maximizes the margin d. Here, the matlab toolbox LS-SVMlab, developed by the group SCD/sista in the department ESAT at the KULeuven, Belgium (Suykens et al., 2002), available at *http://www.esat.kuleuven.be/sista/lssvmlab/*, was used.

2.3 Performance Measure

As a fitness measure indicative of classification performance, the receiver operating characteristic curve (ROC; a plot of the sensitivity versus 1-specificity for varying classifier thresholds) was computed and the area under the curve (AUC) is obtained. Larger AUC values indicate better classifier performance. Moreover, the specificity and sensitivity of the results were also computed for comparison with the physician's initial diagnosis.

In the Results section, histograms representing the discrimination ability of the classifier are presented. These are produced by plotting the estimated classifier output (aiming towards -1 or 1, that is, class 1 or class 2) frequencies color-coded according to the true class, that is, the better the classification performance, the more clustered is either color and the more distinctly separated are the clusters. The ROC is plotted in these figures as well, further illustrating classifier performance.

2.4 Variable Ranking

Although the database is substantial and there is a satisfactory number of instances (patients) compared to the number of available variables, feature selection, that is, the identification of a lower number of highly discriminatory features, can boost classification performance (Bellman, 1961; Blum and Langley, 1997).

A simple method was therefore implemented for variable ranking and subsequent selection as follows:

$$v_i = \mathrm{abs}(\frac{\mu_0 - \mu_1}{\sigma_0 + \sigma_1}) \tag{4}$$

where μ_0 and μ_1 represent the mean value of variable i over the patterns (patient data) belonging to class 0 and 1 respectively, and σ_0 and σ_1 are the standard deviations within each class. The variable ranking value is thus a measure of variable stability, over the patterns, as well as how well each variable taken by itself separates the data classes. For subsequent variable selection, the variables were thus ranked and a given number was selected accordingly (see the results section, figure 2).

2.5 Class Imbalance Correction

Compared to the remaining diseases, diverticulitis is heavily under-represented (148 instances out of 3 337), and this also holds in relation to the category of non-specific abdominal pain (1340). In order to lessen the effect of the imbalance a simple under-sampling scheme was used, where a number of instances (patients) of the over-represented category was removed until both categories were equally represented.

3 RESULTS

The patients were divided into a training (90%) and a validation data set (10%), after which the training data was adjusted for class imbalance. All results refer to the validation data set, unless otherwise specified. The support vector machine (SVM) classifier was trained on the training data, and subsequently applied to the validation data, for all variable subsets formed from the feature ranking, ranging from 1 through all 117 variables. The performance of the physician's initial diagnosis was also computed on the same datasets.

3.1 Classification Performance

The classification performance results are summarized in table 1. First, the SVM was applied to attempt discrimination between diverticulitis and the pool of all other diseases for a varying number of included variables. A maximum AUC of 0.95 was found for 64 variables, and a histogram illustrating the resulting discrimination ability of the classifier is shown in figure 1A. At this optimal point, the SVM obtained a substantially higher sensitivity of 1.00 – that is, correctly identified all of the diverticulitis cases – than the physician at 0.571. However, as is obvious from figure 1A, given this high sensitivity the specificity suffers, and a high amount of false positives are inevitable: the SVM obtained a specificity of 0.823, as opposed to the doctor who produced a much higher specificity of 0.987. When adjusted to the physician's values, the SVM achieved a sensitivity/specificity of 0.571/0.981 (at 105 variables) and 0.5/0.987 (at 111 variables) respectively – lower than the physician in both cases.

Non-specific abdominal pain (NSAP) appears more difficult to distinguish from the pool of all other diseases than diverticulitis. Not even the SVM training data (figure 2B) effectively achieves separation between the classes, and the highest SVM validation sensitivity and specificity are low at 0.687 and 0.721 respectively. Similarly, the doctor's diagnosis achieves a very low sensitivity of 0.455 but, again, a high specificity of 0.909. Also here, when adjusted to the physician's results, the sensitivity/specificity obtained was lower for the SVM than the physician at 0.455/0.878 (at 16 variables) and 0.41/0.909 (at 16 variables).

The discrimination between diverticulitis and non-specific pain, however, was substantially better. The doctor achieved a sensitivity of 0.714 and a specificity of 0.963, whereas the best SVM resulted in a substantially higher sensitivity of 1 and a satisfactory

Table 1: Summary of performance results for the diagnosis of diverticulitis and non-specific abdominal pain (NSAP): sensitivity/specificity.

	diverticulitis vs. others	NSAP vs. others	diverticulitis vs. NSAP
Physician	0.571/0.987	0.455/0.909	0.714/0.963
SVM at maximum AUC	1/0.823	0.687/0.721	1/0.858
SVM at physician's sensitivity	0.571/0.981	0.455/0.878	0.714/0.985
SVM at physician's specificity	0.5/0.987	0.41/0.909	0.786/0.963

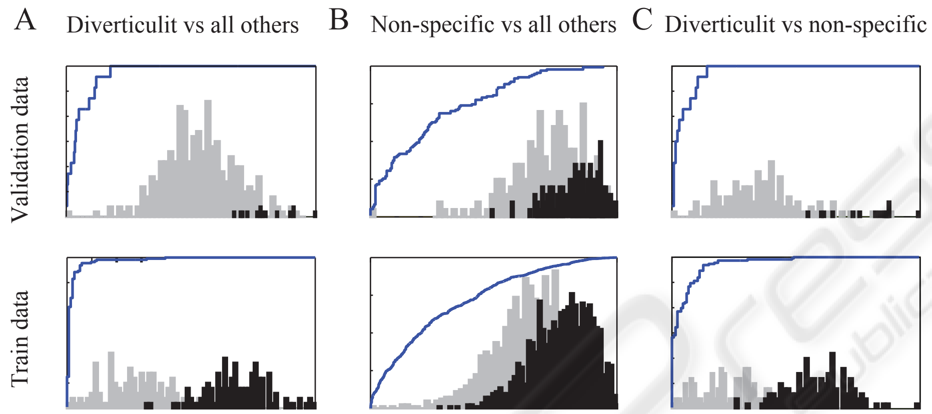

Figure 1: Histogram illustrating the separability of A) Diverticulitis (dark) vs. all other diseases (light), B) Non-specific abdominal pain (dark) vs. all other diseases (light) C) Diverticulitis (dark) vs. non-specific abdominal pain (light). The blue line represents the receiver operating characteristic curve.

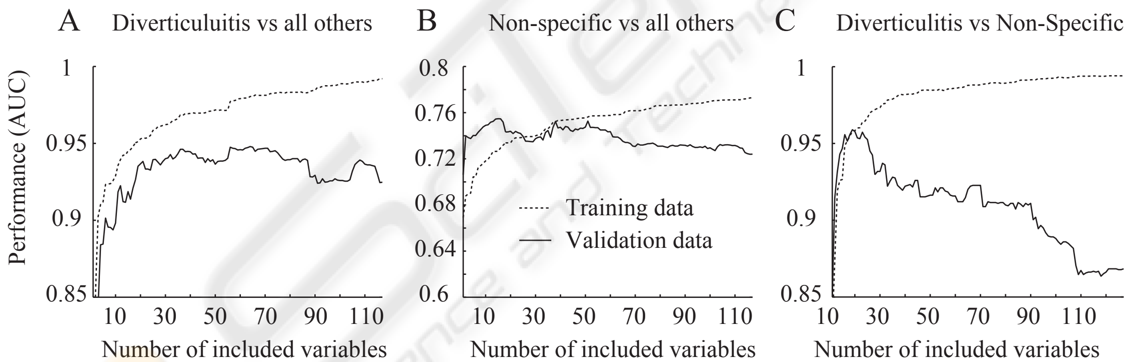

Figure 2: Performance as a function of the number of included variables for A) Diverticulitis vs. all other diseases, B) Non-specific abdominal pain vs. all other diseases C) Diverticulitis vs. non-specific abdominal pain.

specificity of 0.858 (AUC: 0.959). As can been seen in figure 1C, the validation data is distinctly separable. Also, when adjusted to the physicians levels, the SVM sensitivity/specificity was higher in both cases at 0.714/0.985 (at 44 variables) and 0.786/0.963 (at 19 variables), respectively.

3.2 Variable Selection

The variable selection proved to have substantial impact on all data sets (figure 2). For the discrimination between diverticulitis and all other diseases, the addition of variables from 1 through 30 had a large effect on classifier performance, after which it declined. Similarly, on the non-specific abdominal pain vs. all other diseases task, going from one to two features shows a dramatic increase in performance, whereas further addition does not have a large effect. On the other hand, for the more specific case of diverticulitis vs. non-specific abdominal pain, it is obvious that some variables contain large amounts of information regarding the categories – there is a sharp increase in performance up to nine variables, and an equally sharp decrease after the addition of 15 more variables.

A closer inspection of the highly rated features (see table 2-4) reveals that, for any data set combi-

Table 2: The top 10 ranked variables for diverticulitis vs. all other diseases.

Variable	Weight
Age	1.22099
C-reactive protein level	1.05363
Initial pain localization; left lower quadrant	0.917118
Tenderness on palpation	0.901306
Current pain localization; lower left quadrant	0.89903
Current pain localization; right upper quadrant	0.634565
Initial pain localization; right upper quadrant	0.604262
Vomiting	0.588099
Previous abdominal surgery	0.562905
Abdominal scars	0.551697

Table 3: The top 10 ranked variables for non-specific abdominal pain (NSAP) vs. all other diseases.

Variable	Weight
Age	0.605346
C-reactive protein level	0.369473
Serum bilirubin level	0.350293
Systolic blood pressure	0.303903
Decrease/absence of bowel movements	0.286566
Visible bowel movements	0.28654
Development of pain intensity; increase	0.277085
Localized swelling	0.267
S-amylase level	0.264904
Serum alanine aminotransferase level	0.251346

nation, age is the most important factor for discrimination, closely followed by C-reactive protein level. As can be expected, similar variables are important for the discrimination between diverticulitis and all other diseases and diverticulitis vs. non-specific abdominal pain, namely: initial pain localization (both in the left lower quadrant and the right upper quadrant), current pain localization (in the lower left quadrant) and tenderness on palpation (in the left lower quadrant). The focus on left side pain in the diagnosis of diverticulitis agrees with previous research (Laurell et al., 2006). However, in the case of non-specific abdominal pain vs. all other diseases, other variables, predominantly various fluid measurements, are highly rated. Moreover, the resulting rating coefficients are much smaller, thus indicating lower differentiation between variables.

4 DISCUSSION

We have investigated the utility of using a decision support system for the computer aided diagnosis of acute diverticulitis and non-specific abdominal pain (NSAP), as well as for the discrimination between the two, using the Mora acute abdominal pain database.

The general performance of the SVM was comparable to that of the doctor. Moreover, both sensitivity and specificity were higher than those of the physician in the distinction between diverticulitis and non-specific abdominal pain.

The discrimination between diverticulitis and the pool of other diseases, as well as that between non-specific pain and the other diseases, was substantially worse than the differentiation between the two disease categories. This suggests that incorporating known information about the other disease categories, including their respective distribution, in the training of the classifier model can aid in the subsequent diagnosis of new cases. This can, for example, be achieved using an ensemble of classifiers. For the case of multi-class data, where the diagnosis of all patients and thus all diseases is desired, this is, moreover, required for inherently binary classifiers such as SVMs. Standard schemes for ensemble encoding include the one-against-all and one-against-one approach. The latter is more computer-intensive than the former, but typically yields better results. Importantly, it also provides insight into the distinction between diseases. Moreover, the one-against-one is a more well-defined problem, contrasting data cate-

Table 4: The top 10 ranked variables for diverticulitis vs. non-specific abdominal pain.

Variable	Weight
Age	1.51921
C-reactive protein level	1.12222
Initial pain localization; left lower quadrant	1.03731
Current pain localization; lower left quadrant	0.944773
Tenderness on palpation; left lower quadrant	0.885332
Development of pain intensity; increase	0.762057
Current pain localization; right upper quadrant	0.677884
Local muscular defence	0.674493
Leukocyte level	0.664661
Initial pain localization; right upper quadrant	0.657553

gories with inherent similarities and differences, as was evidenced by the findings in our study. The one-against-all scheme, on the other hand, is more relevant to the problem of diagnosis.

The class imbalance was adjusted by simple under-sampling of the majority class. More sophisticate methods could be employed to this end, such as the SMOTE algorithm (Chawla et al., 2002).

The simple feature ranking and subsequent selection utilized in this study proved to be effective in boosting classifier performance. However, more suitable approaches can be used to obtain optimal variable subsets, such as evolutionary algorithms (Marchiori et al., 2007; Åberg et al., 2008). Moreover, there is reason to believe that non-linear relationships pertaining to the disease category exist between some parameters, and reducing the complexity of the data structures can potentially allow for better performance with non-linear classifiers (Åberg and Wessberg, 2007).

5 CONCLUSIONS

Automatic computer-based disease classification is a promising tool for the diagnosis of acute abdominal pain, but requires substantial research before a clinical implementation is feasible. The support vector machine is highly suitable for the discrimination between binary disease categories, and achieved results comparable to the medical doctor. Moreover, the classifier obtained higher sensitivity and specificity than the physician in the distinction between diverticulitis and non-specific abdominal pain. Age and C-reactive protein level, as well as left-hand side pain sensations, were identified as important factors for the classification of diverticulitis.

REFERENCES

Åberg, M. C., Löken, L., and Wessberg, J. (2008). An evolutionary approach to multivariate feature selection for fMRI pattern analysis. *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*.

Åberg, M. C. and Wessberg, J. (2007). Evolutionary optimization of classifiers and features for single trial EEG discrimination. *BioMedical Engineering Online*, 6(32).

Adams, I. D., Chan, M., Clifford, P. C., Cooke, W. M., Dallos, V., de Dombal, F. T., Edwards, M. H., Hancock, D. M., Hewett, D. J., and McIntyre, N. (1986). Computer aided diagnosis of acute abdominal pain: a multicentre study. *British Medical Journal (Clinical research ed.)*, 293(6550):800–804.

Ambrosetti, P., Robert, J., Witzig, J., Mirescu, D., Mathey, P., Borst, F., and Rohner, A. (1994). Acute left colonic diverticulitis: a prospective analysis of 226 consecutive cases. *Surgery*, 115(5):546–50.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton University Press, Princeton, NJ.

Blum, A. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2):245–271.

Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W. (2002). Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321357.

de Dombal, F., Leaper, D., Staniland, J., McCann, A., and Horrocks, J. (1972). Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, 2(5804):9–13.

Ferzoco, L., Raptopoulos, V., and Silen, W. (1998). Acute diverticulitis. *The New England Journal of Medicine*, 338(21):1521–6.

Hansson, L.-E. (2002). *Akut Buk*. Lund: Studentlitteratur.

Laurell, H., Hansson, L., and Gunnarsson, U. (2006). Acute abdominal pain among elderly patients. *Gerontology*, 52(6):339–44.

Laurell, H., Hansson, L., and Gunnarsson, U. (2007). Acute diverticulitis clinical presentation and differential diagnostics. *Colorectal Disease*, 6(9):496–501.

Marchiori, E., Moore, J. H., and Rajapakse, J. C., editors (2007). *A Genetic Embedded Approach for Gene Selection and Classification of Microarray Data*, volume 4447 of *Lecture Notes in Computer Science*. Springer.

Nalin, K. (2006). Den ideala kliniska beslutsprocessen. en studie av arbetsprocessen p en kirurgisk akutmottagning/The ideal clinical decison process. a study of the work process in an acute surgical ward. in swedish. masters thesis. Masters thesis in Cognitive Science, University of Gothenburg.

Suykens, J., Gestel, T. V., Brabanter, J. D., Moor, B. D., and Vandewalle, J. (2002). *Least Squares Support Vector Machines*. World Scientific.

Young-Fadok, T., Roberts, P., Spencer, M., and BG, W. (2000). Colonic diverticular disease. *Current Problems in Surgery*, (37):459514.