Computational diagnosis and risk evaluation for canine lymphoma

E.M. Mirkes [a], I. Alexandrakis [b], K. Slater [b], R. Tuli [b], A.N. Gorban [a],*

[a] Department of Mathematics, University of Leicester, Leicester LE1 7RH, UK
[b] Avacta Animal Health, Unit 706, Avenue E, Thorp Arch Estate, Wetherby LS23 7GA, UK

ABSTRACT

The canine lymphoma blood test detects the levels of two biomarkers, the acute phase proteins (C-Reactive Protein and Haptoglobin). This test can be used for diagnostics, for screening, and for remission monitoring as well. We analyze clinical data, test various machine learning methods and select the best approach to these oblems. Three families of methods, decision trees, kNN (including advanced and adaptive kNN) and probability density evaluation with radial basis functions, are used for classification and risk estimation. Several pre-processing approaches were implemented and compared. The best of them are used to create the diagnostic system. For the differential diagnosis the best solution gives the sensitivity and specificity of 83.5% and 77%, respectively (using three input features, CRP, Haptoglobin and standard clinical symptom). For the screening task, the decision tree method provides the best result, with sensitivity and specificity of 81.4% and $> 99\%$, respectively (using the same input features). If the clinical symptoms (Lymphadenopathy) are considered as unknown then a decision tree with CRP and Hapt only provides sensitivity 69% and specificity 83.5%. The lymphoma risk evaluation problem is formulated and solved. The best models are selected as the system for computational lymphoma diagnosis and evaluation of the risk of lymphoma as well. These methods are implemented into a special web-accessed software and are applied to the problem of monitoring dogs with lymphoma after treatment. It detects recurrence of lymphoma up to two months prior to the appearance of clinical signs. The risk map visualization provides a friendly tool for exploratory data analysis.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Biomarkers for canine lymphoma

Approximately 20% of all canine tumors are lymphoma [81]. The typical age of a dog with lymphoma is 6–9 years although dogs of any age can be affected. The biggest problem with cancer treatment in dogs or humans is the earlier diagnostics. Routine screening can improve cancer care by helping pick up tumors that might otherwise be missed.

The minimally invasive tests are needed for screening and differential diagnosis as precursors to histological analysis. It is also necessary to monitor the late effects of treatment, to identify or explain trends and to watch the lymphoma return. The modern development of veterinary biomarker technology aims to answer these challenges. In the discovery of cancer biomarkers the veterinary medicine follows human oncology with some delay. The controversies, potentials biases, and other concern related to the clinical application of biomarker assays for cancer screening are discussed in [32]. There is increasing interest in the study of

prognostic and diagnostic biomarker proteins for canine lymphoma [58].

Identification of several biomarkers for canine lymphoma has been reported during the last decade:

- The proteomic evaluation of lymph nodes from dogs with B-cell lymphoma (11 cases) was compared to those from unaffected controls (13 cases). The expression of prolidase (proline dipeptidase), triosephosphate isomerase and glutathione S-transferase was decreased in the samples from the lymphoma cases and the expression of macrophage capping protein was increased [52].
- The surface-enhanced laser desorption-ionization time-of-flight mass spectrometry (SELDI-TOF-MS) was used to identify biomarker proteins for B-cell lymphoma in canine serum. 29 dogs with B-cell lymphoma and 87 control dogs were involved in the study. Several biomarker protein peaks in canine serum were identified, and a classification tree was built on the basis of 3 biomarker protein peaks. It was reported that with 10-fold cross-validation of the sample set, the best individual serum biomarker peak had 75% sensitivity and 86% specificity and the classification tree had 97% sensitivity and 91% specificity for the classification of B-cell lymphoma [21].
- A commercially available canine lymphoma screening test was developed by PetScreen Ltd [69]. Serum samples were collected

from 87 dogs with malignant lymphoma and 92 control cases and subjected to ion exchange chromatography and SELDI-TOF-MS analysis. Nineteen serum protein peaks differed significantly ($p < 0.05$) between the two groups based on normalized ion intensities. From these 19 peaks, two differentiating biomarkers emerged with a positive predictive value (PPV) of 82%. These biomarkers were used in a clinical study of 96 dogs suspected of having malignant lymphoma. A specificity of 91% and sensitivity of 75% was determined, with a PPV of 80% and negative predictive value (NPV) of 88%. Later on, these peaks were identified as two acute phase proteins: Haptoglobin (Hapt) and C-Reactive Protein (CRP) [2].

- Some qualitative alterations were identified in dogs with lymphoma in the proteomic study [5]; 21 dogs included in the study had high grade lymphoma confirmed cytologically (16 cases) or histologically (five cases). The increased concentrations of haptoglobin in the sera of dogs with lymphoma could account for increased levels of α 2 globulins, α 2 macroglobulin, α-anti-chymotrypsin and inter-α-trypsin inhibitor, which were identified concurrently.
- Vascular endothelial growth factor (VEGF), metalloproteinase (MMP) 2 and 9 transforming growth factor beta (TGF-β) were tested in 37 dogs with lymphoma, 13 of which were also monitored during chemotherapy. Ten healthy dogs served as control. Lymphoma dogs showed higher activity of MMP-9 ($p < 0.01$) and VEGF ($p < 0.05$), and lower TGF-β than controls, and a positive correlation between act-MMP-9 and VEGF ($p < 0.001$). During chemotherapy, activity MMP-9 and VEGF decreased in B-cell lymphomas ($p < 0.01$), suggesting a possible predictive role in this group of dogs [3].

For use in clinics, the biomarkers should be identified and validated in preclinical settings and then validated and standardized using real clinical samples [59]. Intensive search of biomarkers requires standardization of this technology [51]. Proteins discovered in the research phase may not necessarily be the best diagnostic or therapeutic biomarkers. Therefore, after identification of a biomarker (Phase 1), the clinical assays are necessary to investigate if the biomarker can truly distinguish between disease versus control subjects (Phase 2). Then special retrospective and prospective research is needed for sensitivity and specificity analysis (Phases 3 and 4). Finally, the cancer control phase is needed (Phase 5) to "evaluate role of biomarker for screening and detection of cancer in large population" [51]. Discovery and identification of a promising biomarker does not mean that it will successfully go through the whole standardized procedure of testing and evaluation.

1.2. Acute phase proteins as lymphoma biomarkers

Acute phase proteins are now understood to be an integral part of the acute phase response which is the cornerstone of innate immunity [17]. They have been shown to be valuable biomarkers as increases can occur with inflammation, infection, neoplasia, stress, and trauma. All animals have acute phase proteins, but the major proteins of this type differ by species. Acute phase proteins have been well documented in laboratory, companion, and large animals. After standardized assays, these biomarkers are available for use in all fields of veterinary medicine as well as basic and clinical research [17].

Acute phase proteins, including alpha 1-acid glycoprotein [63,30,77], C-Reactive Protein (CRP) [55,57,69,2], and Haptoglobin (Hapt) [57,69,2], have been evaluated as tumor markers. Nevertheless, as is mentioned in review [32], it is still necessary to prove that these biomarkers are clinically useful in cancer diagnosis. Some authors even suggest that the non-specific serum biomarkers indicate inflammatory response rather than cancer [38].

In our research we evaluate the role of two biomarkers, CRP and Hapt, for screening and detection of lymphoma, for differential diagnosis of lymphoma and for monitoring of lymphoma return after treatment. Our research is based on the PetScreen Canine Lymphoma Blood Test (cLBT). This is advanced technology to detect lymphoma biomarkers present in a patient's serum [2]. The cLBT evaluates the concentration of two acute phase proteins: Hapt and CRP. High levels of these biomarkers indicate a high likelihood that the patient has lymphoma. The cLBT provides a minimally invasive alternative to a fine needle aspirate as a precursor to histological diagnosis of the disease. The cLBT should be used for differential diagnosis when a patient is suspected of having lymphoma by showing classical symptoms such as generalized lymphadenopathy, PU/PD and lethargy (we call all such cases the *clinically suspected* ones). It may be also useful in the monitoring of lymphoma return. In summary, the test provides:

- A simple blood test requiring only 2 ml of blood taken as part of existing biochemistry/haematology work up. Results are available the same day.
- A minimally invasive procedure.
- An alternative to taking an FNA sample and the associated risks of failing to retrieve sufficient lymphoid cells or encountering poor preservation of the cells.
- A monitoring tool to assess treatment progression and to detect recurrence.

Some of our previous results of canine lymphoma diagnosis are announced in [2,56].

1.3. The structure of the paper

The description of the database and statement of the problems are represented in Section 2. Two cohorts are isolated in the database and two problems are formulated: (i) differential diagnostic in clinically suspected cases and (ii) screening. The isolation of the clinically suspected cohort is necessary for formulation of the problem of differential diagnostics and selection of the appropriate methods. The healthy cohort and formulation of the screening problem demands the use of a prior probability of lymphoma and forbids the use of class weights as a parameter to select the best solution. This means that the weights of classes are determined by the prior probability. Both problems (differential diagnostics and screening) are formulated as problems of probabilistic risk evaluation [10]. Usual classifiers provide a decision rule and give the answer in the form "Yes" or "No" (cancer or not cancer, for example). We almost never can be sure that this "Yes" or "No" answer is correct. Therefore the evaluation of probability may be more useful than just a binary answer. If we evaluate the posterior probability of lymphoma under given values of features then we can take the decision about the next step of medical investigation or treatment. Probabilistic risk evaluation supports decision making and allows to evaluate the consequences of the decisions (risk management [10]).

Section 3 presents a brief review of the data mining methods employed in biomarker cancer diagnosis. We introduce the methods used in our work for the analysis of canine lymphoma. The detailed description of these methods is given in Appendix. Three used methods are described:

- *Decision trees* with three different impure-based criteria: information gain, Gini gain and DKM [70].
- *K nearest neighbors* method (KNN). Three versions of KNN methods are used: KNN with Euclidean distance [16], KNN with Fisher's distance transformation, and the advanced adaptive KNN [31]. All the three methods use statistical kernels to weight an influence of each of the k nearest neighbors to