Fractal Analysis and Neural Networks Used for Survival Time Estimation in Cancer Patients

Received for publication, June 15, 2014  
Accepted, September 23, 2014

Liviu GAIŢĂ, Claudiu GAL, Diana PĂRVAN, Manuella MILITARU  
University of Agronomic Sciences and Veterinary Medicine of Bucharest, Romania,  
Faculty of Veterinary Medicine  
Corresponding author: Manuella Militaru, Faculty of Veterinary Medicine, Bucharest,  
Splaiul Independenţei no. 105, sect. 5, 050097, Romania, Phone:+4021 318 0469,  
E-mail: militmanuella@yahoo.com

Abstract  
A model for evaluating the survival time for dogs and cats diagnosed with mammary gland carcinoma was developed. First results are reported, based on data recorded for 74 patients, all subject to mastectomy and some of them to adjuvant chemotherapy as well. Both statistical analysis and an artificial neural network were used, integrating over 20 assumed prognostic factors, provided by clinical, pathology, and other laboratory exams. The effective operation of the statistical tools is illustrated, along with the relevant results emerging from training and validating the neural network. Fractal dimension of chromatin regions in the histology pictures from the lesion is identified among the top prognostic factors, along with tumour dimension, age, alanine transaminase and creatinine serum levels, and white blood cells count at diagnostic time. A significant impact on the survival time was also revealed for adjuvant chemotherapy, NSAID treatment, and neutering.

Key words: fractal analysis, neural networks, breast cancer, mammary gland tumours, survival time

1. Introduction

What influences the survival time of breast cancer patients has been a constant and keen concern in medicine, for ages: hints emerge even from antiquity and methodical studies include cases dating back to 1805, from the Middlesex hospital in London [1]. Since then, the approach to estimating the outcome of this disease kept pace with the progress of diagnostic procedures, therapy means, and information processing tools. Experience and exigency on methodological requirements for trustworthy results of research are piling up. With cancer incidence growing, two factors shape the research: (i) the Evidence Based Medicine (EBM) approach is turning classic - only documented results get accepted for substantiating major therapy decisions, and (ii) molecular biology and genomic medicine provide constantly, at an ever growing pace, new potential cancer markers – with new therapy approaches associated to them.

UICC-Union Internationale Contre le Cancer acknowledged in 1995 no less than 76 prognostic factors for breast cancer in humans [2]. American Society of Clinical Oncology, in its 2007 guide for the EBM use of breast cancer markers organized markers in 13 categories. Only few of them were backed by compelling positive clinical evidence. Among those: tumour size, histological type and malignity grade, lymph node status, lympho-vascular invasion, density of receptors for estrogen, progesterone, and HER2, of p53 protein, along with several genes expression, some related to the proteins/receptors above. Practitioners currently differentiate [3] prognostic factors – with impact on the survival time regardless the
therapy paths (including no-therapy) and prediction factors – used for assessing the potential for benefits from a specific therapy. Concurrent classification criteria [4] list essential, additional, and new and promising factors, or factors related to the environment, to the host (patient), and to the lesion (illness). Obviously, therapy protocols also modulate prognostic.

Progress in statistics, information theory, and computers brought new tools in the hands of the oncologist. An example is the Adjuvant! programme, that can be interrogated free at www.adjuvantonline.com, to provide information on the potential benefits of adjuvant chemotherapy in a specific, new, real case. This programme is based [5] on the SEER (Surveillance, Epidemiology, and End Results Program) database of the National Cancer Institute and uses statistic analysis. An alternative approach is the use of artificial neural networks, proposed as early as 1999 for multifactor estimation of survival time in breast cancer [6].

The pathology lab is the place where many of those factors are assessed [7]: tumour size (nowadays corroborated with readings from MRI, radiology/CT, and ultrasonography), lymph node involvement, histotype type and grade, lympho-vascular invasion, necrotic or fibrotic areas. Immunohistochemistry highlights significant density levels or patterns for proteins/receptors known or alleged as prognostic or prediction factors. The fractal analysis of histology pictures provides valuable information for breast cancer prognostic, with notable advantages on possible automation, objectivity, and quick, high-volume processing of samples [8].

In veterinary medicine, mammary gland carcinoma for dogs is in top ranking positions as frequency. In female dogs, the prevalence of mammary gland tumours is three times bigger than in women [9], half of them being carcinomas [10]. Although for cats the incidence of mammary tumours is estimated to half of that in dogs, they stand for 17% of all tumours in female cats and over 85% of them are malignant [11]. Estimating the survival time in small animal veterinary practice is a growing concern as well. First dogs received cancer chemotherapy almost at the same time with the first humans to get it [12], in the 4th decade of the last century. Decision making in cancer therapy for dogs and cats requires adequate predictive factors. Following limited scale statistic studies [13, 11], in several Western countries (U.S.A., Canada, Norway, Denmark) veterinary oncology registers operate, gathering data very useful for prognostic studies, inter alia [14]. Several types of cancer have already been identified [14, 15] for sharing common histopathology and epidemiology features among humans, dogs, and cats: lymphomas, leukaemias, soft-tissue sarcomas, melanomas, mammary gland tumours, urinary bladder transitional carcinoma, prostate carcinoma, osteosarcoma.

The diagnosis and treatment of mammary gland tumours in dogs and cats have some specific limitations when compared with the corresponding human medicine practice: most often less financial resources are made available for the matter; cooperating with the patient is difficult both for clinical data collection and for therapy; a consistent deontology is difficult to set due to very uneven rapports between the patient and the owner or caretaker - from service animal, to virtual member of the family. In spite of such difficulties, in 1976 a first inventory of prognostic factors was published [16] for 2 year survival time of dogs diagnosed with mammary gland tumours: histology type, infiltrative vs. expansive growth, clinical stage, tumour size, and, with a lesser importance, histology malignity grade. Further studies changed the list and current oncology practice [17] relies for dogs and cats on: tumour size, lymph node involvement, presence of metastasis, histology malignity grade, completeness of exeresis, stage, and, only for dogs: age at diagnosis, protein level in diet (the one before diagnosis, along with the one during treatment), ulceration on tumour, infiltrative growth, invasion into lymphatics or blood vessels, differentiation of the tumour, presence/density of
estrogen and progesterone receptors, of Ki-67 protein (as a marker for cellular proliferation). Several tests for proposed cancer markers are being considered to be licensed for commercial use [18]: OncoPet RECAF – detecting a protein in the blood that is, or is similar to the alpha-fetoprotein receptor, VDXI-TK – detecting thymidine kinase, VDXI-CRP – detecting the canine-specific C-reactive protein.

This paper presents a model set up to provide an estimation of survival time for dogs and cats diagnosed with mammary gland carcinoma.

2. Materials and methods

The model we developed, named OncoVetNeuralNet, systematically records information on previous cancer patients and then uses it to provide support for estimating the survival time for any new case, on alternative therapy paths. The information stored on past cases is processed and made useful by two complementary and relatively independent modules:

1. A statistic module, based on Kaplan-Meier analysis and Cox regression.
2. An artificial neural network. Once trained and validated, the network can be interrogated for an estimation of survival time on any new case, choosing a proposed therapy path. Eventually, when the outcome is known, the case is added to the training set and the prediction performance of the neural network is improved.

The data fed in the model were selected from those being available and being proven or suggested as relevant for the survival life:

- general patient data: species, age at diagnosis, age when tumour became apparent
- clinical data: ovariectomy status (actually, ovariohysterectomy in all cases), lymph node macroscopic status, metastasis status, recurrence status
- pathology data: histology type and grade (according to WHO classification updated in [19]), lymph node involvement, fractal dimension of chromatin regions
- paraclinical data (measurements at diagnosis time): number/level of white blood cells, neutrophiles, monocytes, eosinophiles, serum alkaline phosphatase, alanine transaminase, aspartate transaminase, creatinine, urea, glucose
- therapy data: chemotherapy, surgery, NSAID treatment, graded on a 3 level scale: 0- absent, 1- partial protocol, 2- complete protocol.

Data was collected for oncology patients of ORTOVET clinic, diagnosed in the pathology laboratories of FMV-USAMV, HISTOVET, and „Prof. Dr. Dimitrie Gerota” Emergency Hospital, all in Bucharest. The pool of recorded cases included, at the time of this report, 74 patients, 41 dogs and 33 cats, diagnosed with mammary gland carcinoma by histology exam following mastectomy, after 2002, with the latest update on 15 of April 2014.

The digital histology images were captured on Olympus BX41 microscope with its built-in camera assisted by Olympus Cell®B software; further image processing in preparation of fractal analysis was carried out with Corel® Photo-Paint. The fractal analysis has been done with FracLab 2.05 developed by Research Centre INRIA Saclay - Île-de-France. The database is being hosted in Epi Info™ 7, made available by CDC – Centers for Disease Control and Prevention in U.S.A.. Statistical analysis was carried out using StatsDirect© 3.0 software, from StatsDirect Ltd. in U.K.. The neural network was grown on EasyNN© plus 14.0g platform, licensed by Neural Planner Software Ltd. in U.K..

The fractal dimension

The fractal dimension of histology (and even cytology) samples from the tumour site was found to be useful for the diagnostic and prognostic of mammary gland tumours, in humans,
Fractal Analysis and Neural Networks Used for Survival Time Estimation in Cancer Patients

dogs, and cats [8, 20, 21]. Fractal dimension is a real number measuring the complexity (irregular, non-periodic), the self-similarity over multiple scales, and the degree of filling the space hosting the fractal object. Biological objects were indicated [22] as fractals by Benoît Mandelbrot himself, the founding father of fractal analysis, only to be confirmed repeatedly later [23]. A project was started in 2009 in the Pathology Laboratory of FMV-USAMV to develop practical procedures based on fractal analysis of histology pictures. The project went through progressive phases of taking stock of possibilities, optimising work techniques, and applying tools on several groups of cases in the laboratory [24, 25, 26, 27, 28, 29].

![Figure 1](image1.jpg) Histology picture processing: (a) original picture (HE stain, x400 magnification), (b) after balance (hue, contrast, brightness, saturation) and directional sharpen, (c) after segmentation by colour mask and conversion to black and white

Briefly, microscopic pictures (HE, x400) are processed and reduced to chromatin regions (Figure 1) and on that form the fractal dimension is computed with the box method (Figure 2), known to provide, in such cases, a good approximation for the generalised Hausdorff dimension [20].

\[ D_H = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log \frac{1}{\varepsilon}} \]

\( \varepsilon \): size of cell in the net
\( N \): number of positive cells

![Figure 2](image2.jpg) Computation of the fractal dimension using FracLab [27]

The artificial neural network

The neural network is a computer based model simulating a complex set of interconnected information processing units, set which provides output results based on input values. Inside the network information propagates through neurons, each with a transfer function, connected among them by synapses, each with a weight or transfer function of its own. The network can be very complex and is usually structured in layers of neurons, parallel or hierarchical (pyramidal), all, except for the first (input) and last (output) neurons, being called hidden (Figure 3).
The plasticity of the neural network is the basis for its “learning” capacity. The cases for which all inputs and outputs are known can be used for training the network, by optimisation algorithms which change internal, local transfer functions and weights, even modify architecture, aiming at minimising the difference between the calculated outputs and the „correct” answers. Once that difference is within acceptable limits, the network can be interrogated with a new case, for which a prediction is needed. When time closes the case and the actual outcome is known, that case is added to the training pool and the performance of the network is expected to improve.

Rooted in the perceptron proposed by Rosenblatt in 1958, after a temporary blunt rejection from artificial intelligence experts, the seminal 1986 work of David Rumelhart, Geoffrey Hinton and Ronald Williams propelled the neural networks in the position of major working tools for a wide range of applications in automation, aerospace engineering, in search, monitoring and surveillance for military and public safety purposes, in monitoring and diagnostic in medicine [6, 31, 32, 33] as well as in civil and mechanical engineering, in big data mining, stock and market predictions, weather forecasts, social, educational, epidemiology modelling and many others. Masters summarises [34] when neural networks perform better than alternative tools:

- when the input data are, to a significant degree, „fuzzy” or subject to possibly large error: human opinions, ill-defined categories
- when patterns searched for are deeply hidden in a large amount of data, or noise dominates the signal
- when unpredictable nonlinearity is present
- when chaotic (in the mathematical sense) dynamics is affecting to a significant extend the analysed data.

The OncoVetNeuralNet model yields on the work of the authors since 2010 to develop means for superior use of quantitative data such as the fractal dimension of chromatin regions, in correlation with other information available in the pathology lab. After the initial testing by retrospective studies [35, 36], efforts were made to acquire the needed experience in handling
the neural networks, particularly the facilities provided by EasyNN© plus for gauging the training of the network to overcome the dead-end of overfitting/overtraining situation [30].

The statistic analysis
Statistic analysis is the regular tool for studies that summarise data from past cases and suggest means to issue a prediction for the survival time in new cases [37, 38]. The OncoVetNeuralNet model relies for that on the Kaplan-Meier analysis and on the Cox regression [39, 40]. The statistical significance of the impact of various prognostic/predictive factors on the survival time is evaluated by the non-parametric tests log-rank and Wilcoxon.

Here are the main assumptions used by these methods:
- censored cases bear the same survival probability as the rest of the cases,
- at equal values for the prognostic factors, the survival probability is the same, regardless the delay of joining the study for each case.
- additionally, the Cox regression and the log-rank test rely on the fact that the predictor variables are constant in time for a case and have cumulative effects.

On the statistical soundness of data included in OncoVetNeuralNet model, several weaknesses can be listed:
- diagnostic and malignity grade were set in three different pathology laboratories, over many years. Technical variations and subjectivity might have impacted the data;
- veterinary oncology therapies (except for surgery) are not standardised, as are those in human oncology;
- without a veterinary oncology register in Romania, the relation between the set of cases included in the OncoVetNeuralNet model and the overall population of dogs and cats is difficult to assess.

3. Results and comments
The limited space allows only for some examples to illustrate the potential use of OncoVetNeuralNet. In Figure 4, Kaplan-Meier survival plots are presented for all cases.

![Survival Plot](image)

**Figure 4** Kaplan-Meier survival plots for all cases to date in OncoVetNeuralNet. Mastectomy was performed at least once for all patients.

One can see that 50% of dogs lived more than 3 years, while 50% of cats survived less than 1.5 years, the median survival time values being estimated with wide confidence intervals.
The way the model reveals the impact of a prognostic factor is shown in Figures 5, 6.
Figure 5 Age at diagnosis time as a prognostic factor in dogs and cats

For the Cox regression plot, the vertical axes measures the individual, cumulated (from start to Time) probability of survival for a patient.

Figure 6 The impact of age (at diagnosis time) on the survival time is confirmed by statistic analysis for both dogs and cats, to a larger extend in dogs (parallel groups if hazards are proportional)

The difference of survival time between the two groups segregated by the age at the time of diagnosis and treatment inception was confirmed by log-rank and Wilcoxon tests. The difference is greater and proportional in dogs.

The importance of the statistic assumption that the effects of a prognostic factor do not vary with time is highlighted by looking at the impact of chemotherapy on survival time. In Figure 7 one can see a plot for the survival time up to 2 years, distinct for groups organised by this criterion, and by its side, a similar plot but without any time limit. Time limit is not reflected by a simple crop of the global plot: all events post-2 year limit turn into censored events, thus changing the probability levels. From another perspective, there are medical reasons to assume that the effect of a status descriptor on the evolution of the organism has a limited time span, after which the dynamics of patient-environment interactions and the inner biological developments overwhelm any such effect (if death has not occurred). Two years is an arbitrary but easy to accept time limit of the kind, and first studies [16] on prognostic factors for this disease used that limit.
Fractal Analysis and Neural Networks Used for Survival Time Estimation in Cancer Patients

It is also a fact that some prognostic factors, especially related to therapy, are not measured, identified, or present at a single moment in time, quite the opposite, their presence is associated with a time interval. The assumed duration of their (constant) influence appears reasonable to be extended for a period starting when that treatment ended. For patients currently included in our model, the adjuvant chemotherapy protocol had a 35 to 105 days duration. Finally, the multifactor complexity of the real situation has to be kept in mind. For instance, chemotherapy was more keenly proposed and more often accepted in high grade tumour cases. We do not have yet a data base large enough to carry out a sound multifactor (stratified) statistic analysis.

**Figure 7** Adjuvant chemotherapy as a prognostic factor for survival time. Impact of a time limit set for the model: time limit set to 800 days (left) vs. no time limit (right)

**Figure 7** also reflects the importance of the censored data for the statistical analysis. An event is censored if it satisfies only partially the criteria for selection:

- the patient is still alive at the end of the study: an event with the consumed survival duration is marked, but censored
- the patient dies or leaves the studied group for causes unrelated to the topic of the study: all the same, an event with the (in the study) consumed survival duration is marked, but censored.

This censorship, known also as „censorship to the right” (the time value of the event is actually larger than the one that we record, but we do not know how large) is not an approximation, and should be differentiated from *capping* and *rounding*. Occasional difficulties in tagging cancer as direct or indirect cause of death have also to be acknowledged.

Other prognostic factors were similarly evaluated and for some of them noticeable impact on survival time was revealed: lymph node involvement, malignity grade, NSAID treatment. While statistic plots give some insights on the impact of each prognostic factor on the survival time in the population, the neural network in OncoVetNeuralNet aims at providing an estimation for the survival time for an individual, a new patient, based on all prognostic factors evaluated for that patient and on the known data from all previous cases in the model. **Figure 8** presents the progress of the iterative training process for the network, described by the error plots. Maximal and average error can be seen during training on completely known old cases, as well as average validation error on several other cases, randomly selected from the same set of available old cases. In the process, along with classic gradient optimization algorithms, the tools specific for handling neural networks are put at work [30]: random noise, cloning, pruning, jitter, jogging, freezing, resampling by cross-validation or bootstrapping.
Figure 8 Artificial neural network OncoVet NeuralNet: error plots during training and validation

The architecture of the grown and trained network can be seen in Figure 9.

Figure 9 Artificial neural network OncoVet NeuralNet, layout after training: 22 prognostic variables, one output – survival time, 21 neurons in two hidden layers, 341 synapses

The network can be now interrogated on the predicted survival time in a new case. How good its predictions are can be seen on the square diagrams in Figure 10, drawn for training and validation cases. When the predicted value coincides with the real value, the dot lies on the diagonal of the square; the error is proportional with the deviation from the diagonal.
Fractal Analysis and Neural Networks Used for Survival Time Estimation in Cancer Patients

Figure 10 Artificial neural network OncoVet NeuralNet: Relative importance of prediction factors (left); Quality of prediction (right)

Further valuable information is made available (Figures 10, 11) by the network: the importance of each prediction factor (its weight in determining the survival time) and the sensitivity analysis (relative impact of unit variations in the prognostic factors, on the survival time).

Figure 11 Artificial neural network OncoVet NeuralNet: Sensitivity analysis

We learn from it that, at least for the group of patients in the study:

- The fractal dimension of chromatin regions (DF max in Figures 10, 11) is an important prognostic factor.
- Age at diagnosis, tumour size, levels/numbers for alanine transaminase, white cells, neutrophiles, creatinine in the blood at diagnosis time are relevant for estimating the survival time. It is worth to stress that these are easy to measure, bias-free variables.
- NSAID treatment and ovariectomy have a positive impact on survival time.
The number of cases included in the study has not allowed a comparison between the effect size of each prognostic factor on the survival time, as suggested by the neural network vs. by the statistic analysis. It is a task for a future phase of the project. Time will reveal the usefulness of the OncoVetNeuralNet model. It is in our plan to assess it at least on some aspects:

- Therapy plans put forward by the veterinary doctor will be better substantiated, on an Evidence Based Medicine approach.
- The owner/caretaker of the oncology patient will be better informed when facing decisions on engaging therapy.
- Some of the insights on prognostic and prediction factors might prove relevance for human oncology as well [15].
- Such models highlight the span and outcomes of impressive, enduring efforts of veterinary professionals for the benefit of non-human cancer patients. One can foresee increasing valuation of the life of the non-human individual, not only from an anthropocentric perspective, but from that of his/her own interests and aspirations.

Several topics have been identified for further work on developing and improving OncoVetNeuralNet:

- Fractal analysis can be extended by considering the dispersion or span of the fractal dimension for the same patient. Along with its maximal value, confirmed also as a diagnostic factor [28], several other values of the fractal dimension on mammary gland tissue will be assessed for prognostic usefulness: the average and minimal values on tissue free of lesions, the average one and the standard deviation on patient. Minimal values for fractal dimension are difficult to assess, as small values can be measured on histology frames with occasional scarce cell population, irrelevant for the pathology under scrutiny. A validation key based on minimal number of nuclei (or of compact regions of chromatin) might be a solution.

- A more rigorous assessment of the histology grade is desirable, with the systematic scoring of each defining component, and that is also an ongoing ambition for veterinary pathology laboratories in Romania.

- The list of prognostic and prediction factors can be extended. We consider for that: delayed healing of surgery wound, recurrence and metastatic status at 3 months from surgery, paraclinical data at defined milestones during treatment, EMT (Epithelial-Mesenchymal Transition) indicators.

- The neural network can be improved by a more rational inclusion of censored events. The current approach is to include them as uncensored, a simplification used by other researchers as well [32].

- The model is being improved by extending the pool of cases on which data is collected. Along with oncology cases presented at ORTOVET, online, internet website collection of data on patients of other veterinary practices is considered. Costs are an issue, but voluntary participation is considered for trading case information for survival time estimation. Well defined, standardised procedures for data collection are prerequisites for such an approach.

- The experience gathered in the project can be put at work for other types of cancer in dogs and cats as well. Lymphomas, sarcomas, and mastocitomas are the strongest candidates.
4. Conclusions

We presented a model, based on statistic analysis and on an artificial neural network, that allows the estimation of probable survival time for dogs and cats with mammary gland carcinoma, diagnosed on mastectomy collected samples. Using data from the first 74 cases included in the OncoVetNeuralNet model, the capabilities of its statistical and neural network tools were illustrated. Several prognostic factors were highlighted as relevant for the survival time, among which: the fractal dimension of the chromatin region, the age, tumour size, and several plasma and blood cell parameters at diagnosis time. Adjuvant chemotherapy, NSAID treatment and neutering were also revealed as beneficial. The study confirms and opens important ways of using advanced information technology to support the small animal pathology work.

REFERENCES

22. B. MANDELBROT, The fractal geometry of the nature, Freeman, San Francisco, 1993
24. B. MANDELBROT, The fractal geometry of the nature, Freeman, San Francisco, 1993
27. L. GAIȚĂ, M. MILITARU, Performance Optimization of Fractal Analysis Used for the Microscopic Pathology Diagnostic, Contribuții ale cercetării științifice la progresul medicinii veterinare, Facultatea de Medicină Veterinară, USAMV, București 17-18 noiembrie 2011
29. L. GAIȚĂ , M. MILITARU, G. POPESCU, Fractal dimension of chromatin regions in histological pictures reveals the presence of epithelial tumours, 5th International Conference „Computational Mechanics and Virtual Engineering” COMEC 2013, Brașov, Romania, pg. 209-213
31. L. GAIȚĂ, M. MILITARU, Fractal analysis used for identifying the impact of drugs on the hepatic tissue, Journal of Comparative Pathology, Vol. 148:1, January 2013
32. L. GAIȚĂ, M. MILITARU, Optimization of an artificial neural network used for the prognostic of cancer patients, 5th International Conference „Computational Mechanics and Virtual Engineering” COMEC 2013, Brașov, Romania, pg. 214-219
33. M. SORDO, Introduction to Neural Networks in Healthcare, OpenClinical, 2002