APPLICATIONS OF FRACTAL ANALYSIS IN THE PATHOLOGICAL EXAMINATION OF TUMOURS

SUMMARY

Key words: fractal analysis, fractal dimension, digital image processing, virtual microscopy, oncological diagnosis, prognosis, artificial neural networks, evidence based medicine

Oncological research is on the conjunction of several fields that are presently undergoing spectacular changes. The clinical medical practice is in the process of progressive reorganization in accordance with the requirements falling under the concept of Evidence Based Medicine. Genomics and molecular biology are constantly generating new explanatory models for various stages of carcinogenesis, thus impacting therapeutic and diagnosis approaches. Examination techniques, from noninvasive imagery to deciphering biological samples down to a molecular and genomic level offer such strong instruments for research that it is technically possible to rapidly advance in almost any direction and the only problem remaining is choosing the direction. Information technology is undergoing a change in the heuristic and operational paradigm, as progress in microelectronics has made it possible to store and process an enormous volume of data. However, even though cancer has been studied intensively in the last 60 years, it remains a subject in which every confirmed progress in research raises at least as many new questions as it answers. A perspective that has been gaining ground lately approaches oncology from the perspective of nonlinear dynamic systems that can have spontaneous self-organization manifestations. Cancer would thus reflect a manifestation of the perturbation of swarm intelligence, which in the state that we refer to as *normal* makes the enormous population of somatic cells work together in the multicellular organism. This unifying vision marks an essential link between **carcinogenesis**, **fractal analysis** – which mainly developed as a modeling instrument for nonlinear dynamic systems and **artificial neural networks** – which are one of the best documented examples of man-made *swarm intelligence.*

Fractal analysis facilitates the obtaining of measurements, of numbers that quantify the non-periodical complexity that is so characteristic of biological structures, on all size scales in which life is manifested. This research has investigated the possibility of including fractal analysis applied to histopathological or cytopathological microscopic images amongst the instruments used for diagnosing tumours in animals. I then tested the introduction of the fractal dimension amongst factors of oncological prognosis. For this purpose, I developed statistical models and artificial neural network (ANN) type models for the survival duration, including the fractal dimension in the set of predictors on which these models estimate the prognosis for cancer patients.

Structure of the thesis. The doctoral thesis „Applications of fractal analysis in the anatomopathological examination of tumour processes" is organized into two main parts. ***Section I,*** **Bibliographical Study,** briefly presents the main

results published in previous years with regard to the application of fractal analysis to anatomical pathology in other countries, but in Romania as well, especially in human medicine. It also summarizes the performances reported for some fractal parameters, particularly for the fractal dimension of some elements of histologic or cytologic images, of representing prognosis factors for cancer patients. I highlighted models that have been used internationally to formulate such predictions – mainly statistical models and models in the class of artificial intelligence, respectively artificial neural networks. ***Section II, Original Research*** starts off by presenting the objectives and organization of the research, the theoretical background, the software instruments and technical means used, the populations of cases included and the populations of microscopic images analyzed in the six studies. The main software packages that were used are: Olympus Cell^B for capturing images, CorelPHOTO-PAINT® for the preliminary processing of images, FracLab 2.05 for fractal analysis, StatsDirect 3.0 for the statistical analysis and graphical representation of the results, Epi Info 7™ for the inventory of patients and information, EasyNN© plus 14.0g for the development of artificial neural networks ANN. Capturing histologic and cytologic images, as well as anatomopathological diagnosis were done in most cases in the Anatomical Pathology Laboratory of FMV-USAMV Bucharest. For very few cases, some pathological exams were carried out at S.C. Histovet s.r.l. and the anatomical pathology laboratory of the „Prof. Dr. Dimitrie Gerota" Emergency Hospital in Bucharest. The thesis continues by presenting, in separate chapters, six studies which represented procedural and chronologic steps in the development of the research.

Study 1 **Preliminary research: Fractal Analysis Compared with Classical Morphometry In Histological and Cytological Exams** tested the performance of fractal analysis in comparison with classic nuclear morphometrics in providing quantitative parameters correlated with the presence/absence of pathological modifications on histologic and cytologic images. In this part of the research, I also verified the possibility of using the selected software packages for the fractal analysis of some microscopic histologic or cytologic images. I made a retrospective analysis of 88 histopathological images (magnification ×40 and ×100, Hematoxylin and Eosin HE stain), originating from a previous study executed at FMV-USAMV on Walker 256 rats with iatrogenic cancer, following the reflection in the images of the microscopic morphologic modifications of the liver under the action of the toxic agents (cytostatics) and substances with a liver protecting role (phenolic extracts from plants). I also analyzed 17 cytopathological images (magnification ×100, May-Grünwald-Giemsa stain), of samples obtained through aspiration with a fine needle from dog lymph nodes, which were normal (4 images) or affected by malign lymphoma. The classical morphometric examination determined values for the nuclear perimeter and nuclear area established through the manual contouring of the nuclei on the digital microscopic images. Through fractal analysis, I used the box counting method to calculate the fractal dimension of the integral images – **FDII**, and the fractal dimensions of chromatin regions **FDCR**. For images from histologic samples of liver tissue, I established that the fractal dimension – both of the chromatin regions **FDCR,** as well as of the integral images **FDII**, has levels of performance that are clearly superior in comparison with classical nuclear morphometrics with regard to correlation with the impact of administering toxic agents and even liver protecting agents: lower value dispersions, leading to trust intervals of 95% for narrower averages and very small probabilities ($P < 0.0001$, in some cases) for the differences in averages between

the analyzed groups to be due to hazard. I established that the level of performance of fractal analysis as an instrument for the identification of lesions on a histologic image depends critically on the choice of morphologic elements in the image extracted to be subjected to the evaluation of the fractal dimension: on liver tissue, some toxic factors generate lesions that are optimally revealed through the fractal dimension of the chromatin regions **FDCR**, other factors produce lesions that are better highlighted by the fractal dimension of the integral histologic image **FDII**. Therefore, even for the same organ, one must evaluate several image segmentation possibilities in order to establish the optimum one is for evaluating a certain type of lesion through fractal analysis. In the case of the images from cytologic samples, the methods of classic morphometrics proved superior to fractal analysis in providing parameters that were well correlated with the diagnosis. This conclusion can be considered preliminary, due to the relatively small number of cytologic images included in the analysis.

Study 2 **Preliminary research: Optimization of Procedures and Parameters for Applying Fractal Analysis in Pathological Diagnostic** was carried out on the same population of images used in *Study 1*. I identified the options for operating programs, the file formats and the file naming conventions that ensure the compatibility upon data transfer between the programs involved in the successive image processing phases, fractal analysis and statistical analysis. The sequence of digital images processing operations was translated into a program (*Corel script*) for the optimization of quality, extraction of regions of interest (chromatin) and chromatic reduction (shades of grey) and dimensional reduction (800 x 600 pixels) in the proper format for calculating fractal dimension: this program allowed the automatic processing of a large number of microscopic images in later studies. I established, through successive attempts and evaluations using the ANOVA method, *the work parameters* in the software packages for image processing and fractal analysis leading to *the best correlation* between the fractal dimension calculated for processed microscopic images and the clinical and anatomopathological data associated with the samples from which these images were taken.

Study 3 **Fractal Dimension Used for Revealing Epithelial Tumours** was executed on 194 histologic images (magnification ×40, HE stain) from 19 dogs and 5 cats diagnosed with benign and malign epithelial tumours: mammary gland carcinoma, other carcinomas, mammary gland adenoma, epithelioma, seminoma, mammary fibroadenoma, trichoblastoma, trichoepithelioma. I established that the fractal dimension of chromatin regions - **FDCR** increases significantly when there are elements characteristic of epithelial tumours on the analyzed image. The increase is large enough to allow the formulation of a test based on **FDCR** that differentiates between images containing epithelial tumours and those representing healthy tissue. The performance level of the test (sensitivity and specificity) increases if the population of cases is segregated into sex groups and age groups. In addition, of the types of epithelial tumours included in the study, the best correlation between the presence of tumour modifications with **FDCR** was noted for mammary gland carcinoma, the area under the ROC *Receiver Operating Characteristic* curve being 0.869.

Study 4 **Fractal Analysis Used for The Identification of Tumoral Changes in the Mammary Gland** followed the confirmation of the hypothesis that **FDCR** – the fractal dimension of chromatin regions – is statistically correlated with the presence or absence of a tumour in a mammary gland histologic image.

For this purpose, the research compared **FDCR** for histopathological images of mammary tissues previously anatomopathologically characterized as representing *physiologically normal tissue / adenoma / carcinoma* – for dogs and for cats *physiologically normal tissue/ adenoma / fibroadenoma / carcinoma*. The study included 1237 histologic images (magnification ×40, HE stain) from 62 patients (30 dogs and 32 cats) subjected to a mastectomy as a result of the clinical diagnosis of suspicion of neoplasia at the level of the mammary gland. All of the images selected for the study were of normal mammary tissue or mammary tissue that was tumorally modified exclusively at the level of the epithelial components. The study did not include images that had a mixed content of anatomopathological characters, with the presence in significant weights of several types of lesions. I confirmed that the **fractal dimension of chromatin regions FDCR shows a significant statistical variation (P < 0.001) depending on the presence or absence in the images of mammary gland carcinoma.** Both the values of the fractal dimensions and the high level of their statistical correlation with the presence of malign lesions show a good concordance with results reported in human oncology for mammary gland, prostate and oral mucosa carcinomas. In the case of mammary gland images, for dogs **FDCR** – the fractal dimension of chromatin regions increases progressively between the image groups *with no lesions – adenoma – malign tumours*, whereas for cats, the benign modifications (adenoma, fibroadenoma) decrease **FDCR**, and the malign ones are accompanied by an increase in **FDCR**. As a result of this unsystematic means of **FDCR** variation, the statistical model necessary for testing malignity based on **FDCR** is more complex for cats. By evaluating the ROC - *Receiver Operating Characteristic,* the malignity test based on **FDCR** for dogs that distinguishes whether a histologic image is part of the „*carcinoma*" group compared to the combined group „*no lesions* or *adenoma*", has an area under the ROC curve of 0.7725 (very good performance), and with a balanced choice of the limit value **FDCR** = 1.8214, the test has a chance of correct diagnostic of 5 : 1, a sensitivity of 0.73 and a specificity of 0.66. Similarly, the malignity test for cats, that distinguishes between images in the group „*carcinoma*" and those in the combined group „*no lesions* or *fibroadenoma* or *adenoma*", has the area under the ROC curve of 0.8513 (excellent performance), and for the limit value **FDCR**= 1.7891 the test has a correct diagnostic chance of 16.8 : 1, a sensitivity of 0.88 and a specificity of 0.70. I outlined the possibility of directly translating these conclusions into a procedure that, through the fractal analysis of a histologic image, evaluates the probability of the image containing tumour modifications. This probability, through the Bayesian inference with the statistics data of the studied population, can provide an estimation of the chances of a positive tumour diagnosis for the patient that the sample originates from. For the dog mammary gland, a parameter that can be directly used as an evaluation of the probability of the presence of malign lesions in an image for which **FDCR** is determined is the predictive value for the presence of the lesion in the case of a positive test result (*post-test likelihood of lesion*), as provided by the ROC model if one introduces the test value of **FDCR** into it as a limit value. I created a **fractal index for lesion highlighting – FILH**, derived from the ROC statistic model that I demonstrated offers a robust orientation with regard to the presence/absence in a histologic image of malignity characters, both in the case of dogs and in the case of cats. **FILH** can be used directly in *screening* procedures in order to select regions of interest in the preparations or select images to be sent to the pathologist for evaluation and comprehensive diagnosis. I also argued the possibility of applying

the procedures in this study in order to evaluate images of immunohistochemical preparations and images resulted through the RCM (*Reflectance Confocal Microscopy*) technique through fractal analysis. The study did not identify any advantage resulting from the distinct use of the *structure* and *texture* components of **FDCR**. I confirmed the correlation between the *tumour invasion of lymph capillaries* and the median **FDCR** in the case of the cat mammary gland. I investigated and confirmed the correlation between the *degree of malignity* of the mammary carcinoma lesions and the median **FDCR** in the group of histologic images based on the respective lesions, both for dogs and for cats. Given the recognized value as a prognostic factor of the degree of malignity, by corroborating with similar conclusions in human oncology with regard to fractal analysis, the results of the study bring positive arguments in favor of the usefulness of **FDCR** as a prognosis factor for dog and cat patients with mammary carcinoma.

Study 5 **Assessment of The Possibility to Use Artificial Neural Networks to Issue A Prognostic Based On An Extended Set of Clinical and Pathological Data** developed a model for the prognostic estimation based on artificial neural networks ANN, and was also an opportunity to successfully test the EasyNN© plus 14.0g programming environment, identifying the operationalization problems of ANN and the practical means of overcoming these problems. Data from 39 dogs and cats diagnosed with mammary gland tumours (18 cases), osteosarcoamas, cutaneous lymphomas, hemangiomas, ovarian tumours, malign melanomas, histiocytomas, various types of carcinomas with various locations were used. 27 predictors were quantified on a scale from 0 to 4 for every patient: *clinical and paraclinical criteria* (speed of tumour growth, post-surgical tumour-free period, deviation from normal values of blood tests, surgical wound healing time, age, TNM score at the start of treatment), *macroscopic anatomopathological criteria* (the size of the tumour at the moment of diagnosis, its consistency – hardness, vascularization, ulcerations), *histologic criteria* (degree of malignity, lymph node implication, *therapy related criteria* (surgical treatment, adjuvant chemotherapy, immunity stimulation, hormonal therapy before diagnosis, therapy with nonsteroidal anti-inflammatory drugs NSAIDs). The survival duration (in days) from the moment of diagnosis was the only interrogation parameter of the neural network. I established that the most efficient means of overcoming the blockage situation represented by *overfitting* is cloning the high-performance hidden nodes combined with the adaptive freezing of weights and pruning the low-performance nodes and synapses.

Study 6 **Fractal Analysis and Artificial Neural Networks Applied in Assessing the Prognostic for Cancer Patients** led to the development of the **OncoVetNeuralNet** model that allows for the estimation of the survival duration for dogs and cat with mammary carcinoma, both based on statistic modeling (Kaplan-Meier analysis and Cox regression) and with the help of an artificial neural network - ANN. The study included 74 patients of the ORTOVET clinic, 41 dogs and 33 cats diagnosed with mammary gland carcinoma between 2002 and 2014. 22 factors of prognostic/prediction were quantified and introduced in the model: general patient data (species, age at the moment of diagnosis, age when the tumour was observed), clinical data (macroscopic lymph node status, presence of metastases, presence of relapses), anatomopathological data (histologic type and degree, lymph node involvement, the fractal dimension of chromatin regions – **FDCR** – the maximum value for histologic images from patient samples), blood parameters at the date of diagnosis (number/level for leucocytes, neutrophils, monocytes,

eosinophils, alkaline phosphatase, alanine-transaminase, aspartate-transaminase, creatinine, urea, glucose), therapy related data (gonad status, chemotherapy, excision surgery, NSAIDs). The ANN component of the **OncoVetNeuralNet** model allowed for the conclusion that **FDCR – the fractal dimension of chromatin regions is an important prognosis factor** (in the first 3 positions of the 22 predictors) for dogs and cats diagnosed with cancer. The study also established that the age at the moment of diagnosis, the size of the tumour, the number/level of serum leucocytes, neutrophils, alanine-transaminase, and creatinine at the moment of diagnosis are relevant *prognosis factors* for the estimation of survival time. Treatment with NSAIDs and ovariectomy are also *prediction factors* with a favorable impact on survival time.

The general conclusions and recommendations are included in the final chapter of the thesis, which also offers a brief presentation of some extended applications of the results of this research, given that the procedures for the processing of images and their fractal analysis developed in *Study 1* and *Study 2* have since become operational research instruments in the Anatomical pathology laboratory of FMV-USAMV Bucharest; as a result, in the last few years, from the level of this laboratory, results have been communicated of applying fractal analysis in the study of liver fibrosis, of adipocytes and adipose tissue, in identifying the means used for the thermal preparation of meat destined for human consumption, in investigating new options for studying epithelial-mesenchymal transition.

Section I **Bibliographic study** is 28 pages long, representing 15.5 % of the thesis.

Section II **Original research** comprises 153 pages, representing 84.5 % of the thesis. It contains 8 chapters: ***General Materials and methods***, 6 chapters comprising the 6 research studies and the ***Conclusions, Applications and Recommendations*** chapter.
The two parts of the thesis include 82 figures (of which 7 are reproduced from the bibliography works) and 30 tables.

The bibliography comprises 216 titles of works.

The annexes comprise details with regard to the reproduction rights for the 7 figures taken from works in the bibliography, the list of works published by the PhD candidate, his *Curriculum vitae* and activity report.

The thesis has a total of 250 pages (***Summary*** in Romanian, English, and French**,** ***Contents*** in Romanian, English, and French**,** ***Introduction*** in Romanian and English**,** ***Section I, Section II, Bibliography, Annexes)***.