

UNIVERSITY OF BRISTOL
Department of Engineering Mathematics

Final year Project

**AN INVESTIGATION INTO
FEATURE SELECTION
AND
MACHINE LEARNING
FOR DETECTING DISEASE
FROM LUNG SCANS**

Author: Manuel Algar Gonzalez

Supervisor: Jonathan Rossiter

I wish to thank Jorge Galan Vioque for affording me the opportunity to experience a year in Bristol, and for his generous support back in Spain. I thank Dr. Jonathan Rossiter for his continuous guidance, inspiration and enthusiasm, always striking the perfect balance between providing direction and encouraging independence.

I am grateful to all my friends, especially my classmates, for all the good and bad moments we shared together. And especial thanks to my parents, my brother and all my family, who have always encouraged me, never trying to limit my aspirations. I am grateful to them and amazed at their generosity once more. To them I dedicate this project.

Bristol, April 2003

PRESENTATION

This project has been proposed in the context of a combined collaboration between the Bristol Royal Infirmary's Department of Radiology and the University's Faculty of Engineering to develop programs for the automatic detection of pulmonary diseases.

The development of new techniques and tools in computer-aided diagnostic systems has proved to be a powerful and encouraging tool for the future of the diagnostic field, and many investigations are being carried out nowadays in this sense.

The aim of this project is to merge medical digital imaging and computer processing capabilities into an automatic diagnosis program able to detect potential abnormalities in the lungs, using artificial intelligence techniques.

1. INTRODUCTION.....	1
1.1. Artificial intelligence and Medicine.....	1
1.2. Motivations and Objectives.....	4
2. TEXTURE ANALYSIS.....	6
2.1. Introduction.....	6
2.2. Texture models.....	8
2.2.1. Surveys.....	9
2.2.2. Model-based features.....	10
2.2.2.1.Fractal models.....	10
2.2.3. Non-model-based features.....	12
2.2.3.1.Co-occurrence matrices and other statistical features.....	12
2.2.3.2.Frequency domain methods.....	14
2.3. Texture analysis problems	15
2.4. Applications in medical image processing	17
2.4.1. Applications in the detection of lung diseases	18
2.5. Summary and Conclusions	20
3. LUNG FUNCTION AND PATHOLOGY	21
3.1. Introduction	21
3.2. Lung function	22
3.2.1. The muscles of respiration and the chest wall	22
3.2.2. The airways	23
3.2.3. The alveolar-capillary unit	23
3.3. Lung diseases	25
3.3.1. Overall view.....	25
3.3.2. Classification	26
3.4. Computed Tomography (CT)	31
3.4.1. CT and Magnetic Resonance Imaging (MRI)	31
3.4.2. General procedure	33
3.4.3. Benefits vs. risks	35
4. DEVELOPMENT OF THE DIAGNOSTIC TOOLS	37
4.1. Previous considerations. Chosen language and environment	37
4.2. Developed methodology and program structure	39
4.3. 2D Interface	42
4.4. Summary and Conclusions	57

5. FEATURE SELECTION	59
5.1. Introduction	59
5.2. Selection of a set of features to evaluate	63
5.2.1. Statistical features	63
5.2.2. Co-occurrence matrix features	66
5.2.3. Fourier features.....	70
5.2.4. Grey-level image fractal features	71
5.2.5. Binary features	73
5.3. Initial trials	80
5.3.1. Abnormally low parenchymal density: <i>Emphysema</i>	82
5.3.2. Abnormally high parenchymal density: <i>Fibrosis</i>	85
5.4. A special case: features in edges	88
5.5. Summary and Conclusions	89
6. NEURAL NETWORK. THE LEARNING PROCESS	90
6.1. Introduction	90
6.1.1. Neural networks architectures	91
6.1.2. Cascade-forward networks	93
6.1.3. Learning process. Supervised learning & Back-propagation algorithm	94
6.1.4. Matlab functions	98
6.1.5. Medical applications	101
6.2. A second approach on features selection	102
6.2.1. Methodology	102
6.2.2. Training/Testing process for emphysema	104
6.2.3. Training/Testing process for fibrosis	114
6.2.4. Other points of interest	118
6.3. Summary and Conclusions	121
7. A 3D APPROACH	122
7.1. Motivations	122
7.2. 3D Interface	123
7.3. Comparison between 2D and 3D Diagnosis. Advantages and disadvantages	129
7.4. Summary and Conclusions	136

8. CONCLUSIONS AND FUTURE WORK	137
8.1. Conclusions	137
8.2. Future work	138
9. REFERENCES	139