TECHNICAL UNIVERSITY OF CRETE

Department of Electronic & Computer Engineering (ECE)

IDENTIFICATION OF SIGNIFICANT BIOLOGICAL MARKERS AT METABOLIC & GENOMIC LEVEL FOR NON-INVASIVE DISCRIMINATION OF BRAIN TUMORS

Supervisor Committee

Professor ZERVAKIS MICHALIS (Chairperson)
Professor KALAITZAKIS KONSTANTINOS
Professor GAROFALAKIS MINOS
Professor LIAVAS ATHANASIOS
Associate Professor PETRAKIS EURIPIDES
Associate Professor BALAS KONSTANTINOS
Principal Researcher KAFETZOPOULOS DIMITRIOS

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philoshophy, PhD

by **MICHAIL G. KOUNELAKIS**

CHANIA, GREECE, December 2011

TECHNICAL UNIVERSITY OF CRETE (TUC)

Department of Electronic & Computer Engineering (ECE)

IDENTIFICATION OF SIGNIFICANT BIOLOGICAL MARKERS AT METABOLIC & GENOMIC LEVEL FOR NON-INVASIVE DISCRIMINATION OF BRAIN TUMORS

by **MICHAIL G. KOUNELAKIS**

A thesis submitted as partial fulfillment
of the requirements for the degree of
Doctor of Philosophy, PhD

Supervisor Committee:
Professor ZERVAKIS MICHALIS (Chairperson) – *T.U.C (ECE)*
Professor KALAITZAKIS KONSTANTINOS – *T.U.C (ECE)*
Professor GAROFALAKIS MINOS – *T.U.C (ECE)*
Professor LIAVAS ATHANASIOS - *T.U.C (ECE)*
Associate Professor PETRAKIS EURIPIDES - *T.U.C (ECE)*
Associate Professor BALAS KONSTANTINOS - *T.U.C (ECE)*
Principal Researcher KAFETZOPOULOS DIMITRIOS – *FORTH (IMBB)*

CHANIA, GREECE, December 2011

…to my beloved wife Sofia
and my Parents

Η γνώση δεν είναι στο μέρος,

είναι στο όλον.

(Σωκράτης, Έλληνας Αρχαίος Φιλόσοφος – 470 π.Χ.)

CONTENTS OF THE THESIS

ΠΡΟΛΟΓΟΣ

Με κίνητρο την αγάπη μου για την έρευνα αλλά και την προσδοκία μου να προσθέσω και εγώ ένα λιθαράκι στις προσπάθειες που καταβάλλονται σε Παγκόσμιο επίπεδο για την καταπολέμηση του καρκίνου του εγκεφάλου, ξεκίνησα το 2003 τη Διδακτορική μου διατριβή στο Πολυτεχνείο Κρήτης και στο Τμήμα Ηλεκτρονικών Μηχανικών & Μηχανικών Ηλεκτρονικών Υπολογιστών. Υπήρξε ιδιαίτερα σημαντική για εμένα η ημέρα όπου ο επιβλέπων Καθηγητής μου κ. Ζερβάκης Μιχάλης μου πρότεινε να εντάξω τη Διδακτορική μου έρευνα στο πλαίσιο ενός Ευρωπαϊκού προγράμματος (Biopattern IST EU) που θα ξεκινούσε και στο οποίο θα συμμετείχαν πολλά Πανεπιστημιακά Ιδρύματα σε Ευρωπαϊκό και Διεθνές επίπεδο με βασικό στόχο την κατανόηση, εφαρμοσμένη ανάλυση και εξερεύνηση διαφόρων τύπων καρκίνου. Αυτό ήταν για εμένα μια μεγάλη ευκαιρία αφού μου έδινε τη δυνατότητα να συνδυάσω δύο πεδία επιστημών, την πληροφορική και τη βιολογία, που πραγματικά αγαπώ.

Το σύγγραμμα αυτό είναι αποτέλεσμα μίας εναγώνιας προσωπικής προσπάθειας αλλά και της συνεργασίας μου με πολλούς ανθρώπους και Οργανισμούς. Την αξία της δουλειάς αυτής θα την κρίνετε εσείς αλλά και οι μελλοντικοί ερευνητές όμως θα ήθελα να τονίσω ότι πραγματικά αποτελεί μια ειλικρινή και αναλυτική παρουσίαση αποτελεσμάτων, μεθόδων και παρατηρήσεων που καταγράφηκαν καθ' όλη τη διάρκεια της έρευνας μου. Θα ήθελα λοιπόν να ευχαριστήσω όλους τους Καθηγητές της συμβουλευτικής επιτροπής για το χρόνο που θα αφιερώσετε να μελετήσετε και να αξιολογήσετε τα αποτελέσματα της προσπάθειάς μου αυτής.

Τίποτα από όλα αυτά όμως δε θα ήταν δυνατό να πραγματοποιηθούν χωρίς την αμέριστη συμπαράσταση του επιβλέποντα Καθηγητή μου κ. Ζερβάκη Μιχάλη τον οποίο και ευχαριστώ θερμά. Η συνεχής παρακολούθηση της πορείας μου, οι εύστοχες παρατηρήσεις του, η βαθιά και τεκμηριωμένη γνώση του αντικειμένου του αλλά πάνω από όλα η ανθρωπιά του με οδήγησαν στην επιτυχημένη ολοκλήρωση της διατριβής μου. Είναι σίγουρο ότι μέσα από την προσπάθεια αυτή γνώρισα ένα πραγματικό φίλο και αυτό είναι πολυτιμότερο από οτιδήποτε άλλο.

Θα ήθελα επίσης να ευχαριστήσω θερμά τους φίλους, συναδέλφους-ερευνητές Δρ. Μπλαζαντωνάκη Μιχάλη και Μανίκη Γιώργο (MSc) του Εργαστηρίου Ψηφιακής Επεξεργασίας Σημάτων & Εικόνας (DISPLAY) για τις πολύ χρήσιμες παρατηρήσεις τους, τις αμέτρητες ώρες εποικοδομητικών συζητήσεων, τη βοήθειά τους στην ανάπτυξη μεθόδων επιλογής χαρακτηριστικών (feature selection methods) και το πραγματικό τους ενδιαφέρον όλα αυτά τα χρόνια.

Ευχαριστώ επίσης τον φίλο Ιατρό-Νευροχειρούργο Δρ. Κοτσιάκη Ξενοφώντα (Δ/ντή Νευροχειρουργικής Κλινικής Γενικού Νοσοκομείου Χανίων) για την ουσιαστική συμβολή του στα ιατρικά-βιολογικά και όχι μόνο θέματα της διατριβής αυτής.

Σημαντική κατά διάρκεια της Διδακτορικής μου διατριβής υπήρξε και η συμβολή του φίλου μου ερευνητή Dr. Geert Postma (Department of Analytical Chemistry, Institute of Molecules & Materials, Radboud University, the Netherlands) αλλά και των Καθηγητών Lutgarde Buydens

(Department of Analytical Chemistry, Institute of Molecules & Materials, Radboud University, the Netherlands) και Arend Heerschap (Department of Radiology, University Medical Center, Radboud University, the Netherlands) στην επεξήγηση των ιατρικών απεικονιστικών διαγνωστικών μεθόδων που χρησιμοποιήθηκαν στη διατριβή μου. Τους ευχαριστώ επίσης για την παραχώρηση της βάσης βιοϊατρικών δεδομένων ασθενών του Πανεπιστημιακού Νοσοκομείου του Radboud, που αποτέλεσε το βασικό εργαλείο στην έρευνα μου.

Ευχαριστώ επίσης θερμά τον Καθηγητή και IEEE Fellow, George Giakos (Department of Electrical & Computer Engineering, University of Akron, USA) για τη συμμετοχή του και τις πολύ σημαντικές παρατηρήσεις του στην έρευνά μου που βελτίωσαν σημαντικά την συνολική εικόνα της διδακτορικής μου διατριβής.

Βασικοί υποστηρικτές των προσπαθειών μου ήταν και πάντα θα είναι οι Γονείς μου. Υπήρξαν πάντα απρόσκοπτοι εμψυχωτές μου ακόμα και κατά τη διάρκεια πολύ δύσκολων προσωπικών τους στιγμών λόγω ασθένειας και των δύο από την επάρατη νόσο και γι αυτό τους ευχαριστώ μέσα από τα βάθη της ψυχής μου. Η δοκιμασία που πέρασαν αποτέλεσε και αποτελεί έναν ακόμα λόγο για την ενασχόληση μου με θέματα που αφορούν στον καρκίνο. Ευχαριστώ το Θεό αλλά και τους επιστήμονες που τους έσωσαν και τους έδωσαν ελπίδα στη ζωή τους.

Αφήνω για το τέλος τις πιο θερμές μου ευχαριστίες στην αγαπημένη μου σύζυγο και σύντροφο της ζωή μου, Σοφία. Η αμέριστη συμπαράστασή της σε στιγμές απογοήτευσης, η διαρκής ψυχολογική υποστήριξή της αλλά πάνω από όλα η πραγματική της αγάπη και ο σεβασμός της προς εμένα, υπήρξαν καταλυτικά καθ' όλη τη διάρκεια της προσπάθειας αυτής.

Μιχαήλ Γ. Κουνελάκης

Χανιά,
Δεκέμβριος 2011

Thesis work was supported by Biopattern, IST EU (FP6) funded project, Proposal/Contract no: 508803 ("Grand Vision" was to develop a pan-European, coherent and intelligent analysis of a citizen's bioprofile; to make the analysis of this bioprofile remotely accessible to patients and clinicians; and to exploit bioprofile to combat major diseases such as cancer and brain diseases.)

ABSTRACT

Cancer is generally considered as the public nuisance of our century since it is one of the most complex diseases that the medical community must face. Its undetermined pathological origins, its unpredicted biological behavior and its lethal, most of the times, outcome are some of its main characteristics that experts have to deal with.

Among the most lethal types of cancer is the *Brain cancer* which is characterized by the formation of one or more solid tumors within the brain parenchyma. The Brain tumors have the ability to rapidly progress form low malignancy to high malignancy, restricting thus the oncologist's ability to accurately evaluate their behavior and design an effective treatment in order to improve the patient's clinical image.

The recent release of the human genome enabled experts to understand that abnormal genetic mutations are the basis for cancer genesis. Furthermore, the introduction of other "omics" fields such as transcriptomics, proteomics and metabolomics, descendants of genomics, revolutionized the way experts analyze Brain tumors today. State of the art "omics" technologies and pattern recognition methods managed to revealed useful information regarding Brain tumors' pathology that has been unknown for many decades.

Although a lot has already been done in the field of Brain cancer diagnosis, prognosis and treatment, a lot more must be achieved. Most of the patients with high grade brain malignancy die within 24 months from initial diagnosis. The need to design new and more effective treatments that will prolong patients' life expectancy is overwhelming.

Motivated by this need under the major hypothesis that the selection and the effectiveness of the therapy to be followed is primarily based on the estimation of the histopathological profile of the tumor at diagnosis stage, we attempt to identify novel and reliable biological features (markers) sets that can be adopted to accurately discriminate the type and grade of a brain tumor for a new patient. The selection of significant features, which describe the tumors' type and grade, is the foundation for the design of novel non-invasive patient specific therapies. This is actually an open challenge in this continuous fight against Brain as well as other types of cancer.

To accomplish this goal, the data of Brain cancer patients provided from two *"omics" technologies*, named *Magnetic Resonance Spectroscopy (MRS)* and *DNA Microarrays*, were utilized. MRS technology reveals the metabolic profile of a Brain tumor while DNA Microarrays provides its genetic identity. Analyzing the information provided from MRS spectra, we identified *novel metabolic marker sets* that can be used to classify the type and grade of a new patient with high accuracy. On the other hand our genetic analysis was based on the Otto Warburg's hypothesis in 1956 who observed that tumorous cells exhibit increased rates of glycolysis (sugar splitting process for cellular energy production). Examining the glycolytic profile of Brain tumor patients we managed to discover that, apart from the well known from bibliography genetic markers (genes), *glycolycis related genetic markers* play a very significant role in Brain tumor's behavior. Based on this two-fold analysis a novel *medical Decision Support System (DSS)*, which bridges the knowledge extracted from two different "omics" modalities, i.e. genomics and metabolomics, is proposed. As a primary result, we

verify the importance of metabolites in cancer-type and grade discrimination and validated their metabolic and genetic association in cancer progression, through the glycolysis process.

In order to implement the analysis of the data at genomic and metabolomic levels, modern pattern recognition methods were applied. Two well known classifiers named Support Vector Machines (SVM) and the Least Squares-SVM (LS-SVM), widely used in biomedical problems, were used exploiting their unique property to cope quite well with complex data as occurs in brain cancer. Based on these classifiers we managed to develop a *reliable feature selection and classification system* that embeds the intrinsic characteristic of patients' data into the classification process resulting to high classification accuracy rates and identification of significant metabolic and genetic marker sets. This was a secondary accomplishment of this thesis.

GENERAL INTRODUCTION

Brain tumors and especially the *brain gliomas* are among the most aggressive and lethal types of cancer. Their ability to rapidly infiltrate into the brain tissue causing irreversible damages requires quick and accurate clinical response both at diagnostic and therapeutic stages.

Nowadays, the diagnostic protocol followed begins with the application of diagnostic imaging techniques such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) in order to obtain information concerning the size, the location, the extent, the type and grade of the tumorous lesion. Most of the times though, this is not enough and a biopsy is necessary for the evaluation of some cancer biomarkers, already known from literature, in order to determine the type and grade of tumor's malignancy. Biopsy however, forces an oncologist to proceed to a brain surgery for the resection of a small tumorous part for further examination. Although necessary, many times this operation hides significant risks for the patient. The undetermined borders of the tumorous area involve the danger of removing, apart from tumorous, healthy brain tissue and destroying in this way vital functions of the patient's brain. In addition, it is also often decided, as a last choice for therapy and usually after the application of chemotherapy and/or radiotherapy, to proceed to a surgical operation through craniotomy in order to remove the observable tumorous area. Remember however that brain is the organ that maintains and controls the functions of our body, so even when a small part of the tissue is removed, there will be definitely a cost for the patient. The bet therefore is to limit this cost by inventing alternative methods, less invasive than biopsy and surgery, to accurately diagnose and eventually treat these tumors.

During the last decades, efforts to develop new non-invasive imaging diagnostic methods or even improve the existing ones have been generated. The need to acquire more sensitive diagnostic tools which will provide a clearer view and understanding of the intrinsic characteristics of such complex tumors resulting to the design of new more efficient treatment protocols, is urgent.

Among the most promising imaging diagnostic techniques applied today, is the *Magnetic Resonance Spectroscopy (MRS)*. This technique, which is a descendant of Magnetic Resonance Imaging (MRI), is capable to monitor the biochemical/metabolic activity of brain gliomas by analyzing the MRS spectra acquired from patient's brain. Studying the metabolism of a tumor has been recently introduced into the oncology science. Many researchers have recently shown that cancer's metabolic activity can be a field where significant information regarding cancer's behavior can be extracted *[Seyfried T.N. et al 2005 - Griffin J.L. et al 2006 - Spratlin J.L. et al 2009 - Seyfried T.N. et al 2011]*.

Although the metabolic information of tumors enable experts to study their behavior, classify them based on their grade and decide the treatment that is most suitable for each patient, today's diagnostic practices mostly examine a predefined set of known metabolic markers in order to classify a glioma patient into one of the known, according to World health Organization (WHO), types and grades *[Callot V. et al 2008 - Galanaud D. et al 2006 -*

Heerschap A. 2007 - Howe F.A. et al 2003(a, b) - Majos C. et al 2004 - Sjobakk T.E. et al 2006]. Therefore, the need to discover new metabolic markers and test whether they can help experts to increase the diagnostic outcome and improve the therapeutic process is immediate.

Furthermore, the release of the human genome working draft marked the biomedical discipline opening a new era in the fields of biology and medicine with the use of bioinformatics. In combination with the advent of microarray technology, scientists can now derive a vast amount of valuable information but the need still remains to understand and exploit it. *DNA microarray technology* allows researchers to study the behavior of thousands of genes in a single experiment, exploring and monitoring their expression in various diseases with the aim of understanding or discovering the biological mechanisms involved *[Shalon D. et al 1996 - Golub T.R. et al 1999 - Ramaswamy S. et al 2002]*. Today, specific gene alterations are identified using DNA microarrays in order to determine the aggressiveness of the tumor. But again are these genetic markers enough to obtain the best possible diagnosis?

Both of these two *"omics" technologies* (MRS and DNA Microarrays) offer clinicians a non-invasive way to search for unique cancer characteristics which will enable them to achieve better diagnosis and treatment results. However another issue must be addressed. Today's clinical methods applied for the therapy of brain tumors, most of the times, follow a similar pattern. Patients that have been diagnosed with the same grade of brain glioma undergo a similar treatment plan (chemotherapy, radiotherapy and/or surgery through a craniotomy). But is this always the correct or the only way to face this disease? And if this is true, why then the same therapy has not the same impact on different patients suffering from the same type of brain tumor?

This type of syllogism has generated the idea that new-generation medicine, which moves from disease-based to patient-based treatment, is necessary. *Patient-based or personalized treatment* considers each patient as a unique pathological case. Based on this consideration new drugs can be designed to focus on specific metabolic and genetic characteristics of the patient which will hopefully restrict or even eliminate the brain tumor *[Van't Veer L.J. et al 2008 - Lesniak M.S. et al 2004]*.

What research has offered so far?

At genomic level most of the studies accomplished in the area of brain cancer involve the evaluation of specific tumor suppressors (PTEN, Rb, TP53) and oncogenes (CDK4, EGFR, PDGF) and their impact on tumorigenesis. The examination of their genetic mutations is now considered the gold standard for diagnostic, prognostic and treatment purposes. Their interrelations are described by genetic networks (also known as pathways) which are available in large databases on the Internet (KEGG, HGP) for further study and use *[Shiraishi T. et al 2003 - Furnari F.B. et al 2007 - Ohgaki H. et al 2009 - Jones R.G. et al 2009]*. However less work has been done on the evaluation of the impact of other biological pathways on brain tumor pathology and behavior that could possibly reveal new potential

biomarkers which will open new ways towards a more reliable non-invasive management of this disease.

At metabolic level now the use of spectroscopic data obtained from patients' MRS brain tumor spectra, enabled the determination of specific metabolites that have been found to play a significant role mainly in the diagnosis of the type and grade of brain tumors. The estimation of the spectroscopic intensities of NAA (N-acetyl-aspartate), Cre (Creatine), Cho (Choline) and others metabolites offer to clinicians a clear image of the metabolic activity within the brain which is a useful diagnostic tool *[Howe F.A. et al 2003(a, b) - Simonetti A.W. et al 2005 – Lehnhardt F.G. et al 2005 - Devos A. et al 2005 - Galanaud D. et al 2006 -Luts J et al 2007 - Heerschap A. 2007 - Callot V. et al 2008 - Postma G.J et al 2011]*. The problem however is that most of the studies done in this area propose a predefined set of metabolites that is used as a common base for the discrimination of any brain tumor type and grade. In other words they do not examine each brain tumor separately so they can determine specific sets of metabolic markers that influence their type and grade. This has a negative impact on the prognosis and therapeutic estimates for two different tumors since their metabolic profile is not extensively investigated.

DEFINITION OF THE PROBLEM
GOALS OF THE THESIS

Definition of the problem

Brain is perhaps the most complex organ of human body. It actually controls our body's functions and therefore any problem arising in its tissue must be carefully managed by experts. Brain tumors and especially gliomas are among the most aggressive neoplasms which must be immediately faced since their prognosis estimate is very poor (less than 24 months).

During this short period of time several critical decisions must be taken regarding the treatment protocols to be followed (chemotherapy, radiotherapy and/or surgical operation). However, even with the most effective therapies the patients will, most of the times, not make it at the end. The selection of the proper therapy is always based on the proper diagnosis of the disease. Today when two patients are diagnosed to suffer from the same brain tumor, both will follow the same or similar treatment pathways. What quite often happens though is that the two patients respond quite different from the expected.

Based on these observations we directed our research efforts on two clinical areas, i.e. *the diagnosis and the therapy of brain tumors*. The problems that this thesis tried to face are:

- The weaknesses of today's conventional imaging diagnostic methods such as CT (Computer Tomography) and MRI (Magnetic Resonance Tomography) to accurately discriminate brain tumor types and grades. These methods provide only morphological information for the brain tumor, i.e. they do not investigate its biological profile. This inevitably leads to biopsy (surgical operation) which alters new dangers for patient's life.
- The inability of today's conventional diseased-based treatments to effectively manage brain tumors. The toxicity of the drugs applied today many time affect even the healthy tissue of the brain causing irreversible problems in patient's functionality.
- The lack of reliable brain cancer biomarkers that can reveal the tumor's biological behavior which will provide more reliable diagnosis and set the foundations for new non-invasive treatments (reduce surgical operations and therefore also reduce the cost of treatments for the hospitals).
- Finally, the inexistence of a new medical Decision Support System in the clinical practice that could establish a new non-invasive diagnostic protocol based on these biomarkers and focuses primarily on the patient not the disease only (personalized treatment).

Goals of the Thesis

The aim of this thesis is three-fold. The first major goal is to exploit the information extracted from the metabolic (MRS spectra) and genetic (DNA Microarrays) analysis of brain tumors (mainly brain gliomas) in order to identify unique sets of metabolic and genetic markers that will accurately discriminate their type and grade but also open new ways for their non-invasive treatment.

These sets of biomarkers will offer to oncologists the ability to find answers into two crucial matters regarding the diagnosis and treatment of a new patient. These are:

- to accurately classify a new patient into a specific type and grade, based on the metabolic and genetic characteristics (markers) of his brain tumor, not only on the clinical (histopathological) characteristics of his tumor.
- to manage each new patient as a unique case and decide for his treatment in an individualized manner designing and administrating targeted therapies based on these markers' expression profiles.

Second, to provide a robust pattern recognition method able to identify significant characteristics of brain tumors, both at metabolic and genomic level, for diagnostic purposes.

Finally, the last goal is to demonstrate the significance of developing new state-of-the-art medical Decision Support Systems (DSS), based on reliable pattern recognition methods and bioinformatics, which will provide clinicians new analytic tools to implement alternative non-invasive methods for the diagnosis, prognosis and treatment of brain tumors.

CONTRIBUTION OF THE THESIS

Although a more detailed presentation of the achievements of this work is provided in Chapter 6, we would like to briefly address the most important here. Based on the analysis of the main problems involved in the management (diagnosis and treatment) of this disease, as explained in the General Introduction, and the main goals set in the section above, the contribution of our work is inevitably related to diagnosis, treatment and the design of medical DSS.

At clinical level

More specifically at diagnostic level our research has identified reliable sets of metabolic and genetic biomarkers which determine the metabolic and genetic profile of brain gliomas and can be integrated into the clinical practices to further improve the diagnostic accuracy through a robust non-invasive way. In contrast to other researches where a predefined set of markers is used for the discrimination of the brain tumors, our research defined optimal sub-sets of metabolic and genomic markers (shown in Tables 6.1 and 6.2 respectively but also in Chapters 4 and 5) that can be used to investigate each type and grade separately. Furthermore these markers offer the opportunity to discover new types and/or sub-types of brain tumors since they provide an in depth examination of the metabolic and genomic differentiations of the cancerous areas even within the same tumor.

At treatment level our work has managed to prove that the genomic and metabolic analysis of the glycolysis pathway, a part of the cellular respiration pathway, must be considered in the design of new targeted therapies. The glycolysis-related genes discovered, shown in Table 6.2 of Chapter 6, suggest that brain tumors have altered glycolytic activity which is a characteristic that scientists can study in order to generate new less toxic but highly effective drugs. It must be noticed here that this is the first time that a specific set of glycolytic markers is proposed for brain tumors treatment and especially gliomas.

Finally our research proposes a new medical DSS (shown in Chapter 6, Figure 6.1) which presents an alternative non-invasive protocol for diagnosis, prognosis and therapeutic purposes. This DSS suggests that the metabolic and genetic markers found from our studies can be integrated into the diagnosis, prognosis and treatment phase in a new patient case. This can be done in combination with the conventional protocol followed today.

At theoretical level

At theoretical level our work exhibits the remarkable potential of maximal margin classifiers such as Support Vector Machines (SVM) and Least Squares - SVM in cancer discrimination problems. A feature selection and classification method which embeds filter methods such as, Fisher's criterion and Relief-F ranking, was designed in order to reveal significant characteristics of brain tumor both at metabolic and genetic level, for diagnosis and treatment purposes. Furthermore we managed to show that the selection and preprocessing of the initial brain tumor features, inputted into the classifier, significantly influence the identification of the final cancer markers' set and so the classification of a new patient.

OUTLINE OF THE THESIS

This section presents both analytically and schematically (Figure 1) the development and progression of this PhD thesis, aiming to help the potential readers understand the research that has been done.

The thesis consists of six main chapters, which demonstrate both the background theory and the applications (studies) in the field of Brain cancer diagnosis and treatment, using Magnetic Resonance Spectroscopy (metabolomics) and DNA Microarrays (genomics) information. More specifically:

Chapter 1 gives an overview of Brain cancer pathology presenting its incidence, types, grades and also the diagnosis, prognosis and treatment clinical practices followed nowadays. In addition this chapter exhibits the role of the so called “omics” (genomics – transcriptomics – proteomics - metabolomics) their interrelations and their contribution to Brain cancer analysis. Finally it reveals the need to design and apply non-invasive methods for the accurate evaluation and treatment of this disease.

Chapter 2 reviews the state of the art pattern recognition methods used in biomedicine. It describes two supervised kernel-based classification models, called Support Vector Machines (SVM) and Least Squares - Support Vector Machines (LS-SVM), commonly applied for classification and marker selection tasks. An explanation of the feature selection methods and the classifier performance evaluation techniques is also provided. Finally an example study which practically demonstrates the application of SVM and LS-SVM for the classification of Acute Myeloid Leukemia patients, using clinical, cytogenetic and molecular biology features, is presented.

Chapter 3 reveals the significant role of the Magnetic Resonance Spectroscopy (MRS) imaging technique and the metabolomics in Brain tumors diagnosis. It also describes the background theory and the main physical principles behind the MRS and its predecessor, the Magnetic Resonance Imaging (MRI). Furthermore it presents an application study on the strengths and weaknesses of MRS data obtained from two different MRS scanners (1.5 Tesla and 3 Tesla respectively) in the classification of Brain tumors (gliomas).

Chapter 4 presents a study integrating the areas of metabolomics and MRS in respect to Brain tumors diagnosis. Specifically, this study focus on the identification of novel sets of metabolic markers that can describe different types of brain gliomas and meningiomas providing in this way a classification tool for accurate clinical diagnosis. Furthermore, a statistical methodology which exploits the intrinsic properties of data by integrating the Fisher’s marker ranking criterion into the classification process is proposed for diagnostic purposes.

Chapter 5 presents a study in the area of metabolomics, genomics and cell glycolysis in respect to Brain tumors treatment. The proposed analysis considers aspects of both statistical and biological validation of glycolysis effects on brain gliomas, at both genomic and metabolic level. Furthermore, it discusses main issues of the cellular respiration and the role of glycolysis process in cancerous cells (Otto Warburg's hypothesis) and how these can be utilized for the design of new therapeutic protocols for Brain gliomas.

Chapter 6 concludes the Thesis, presenting the main achievements at both medical (diagnosis, treatment) and engineering (proposal of a medical Decision Support System) areas concerning Brain cancer. Finally it addresses some issues for future research in the field of Brain cancer management.

OUTLINE OF THE THESIS

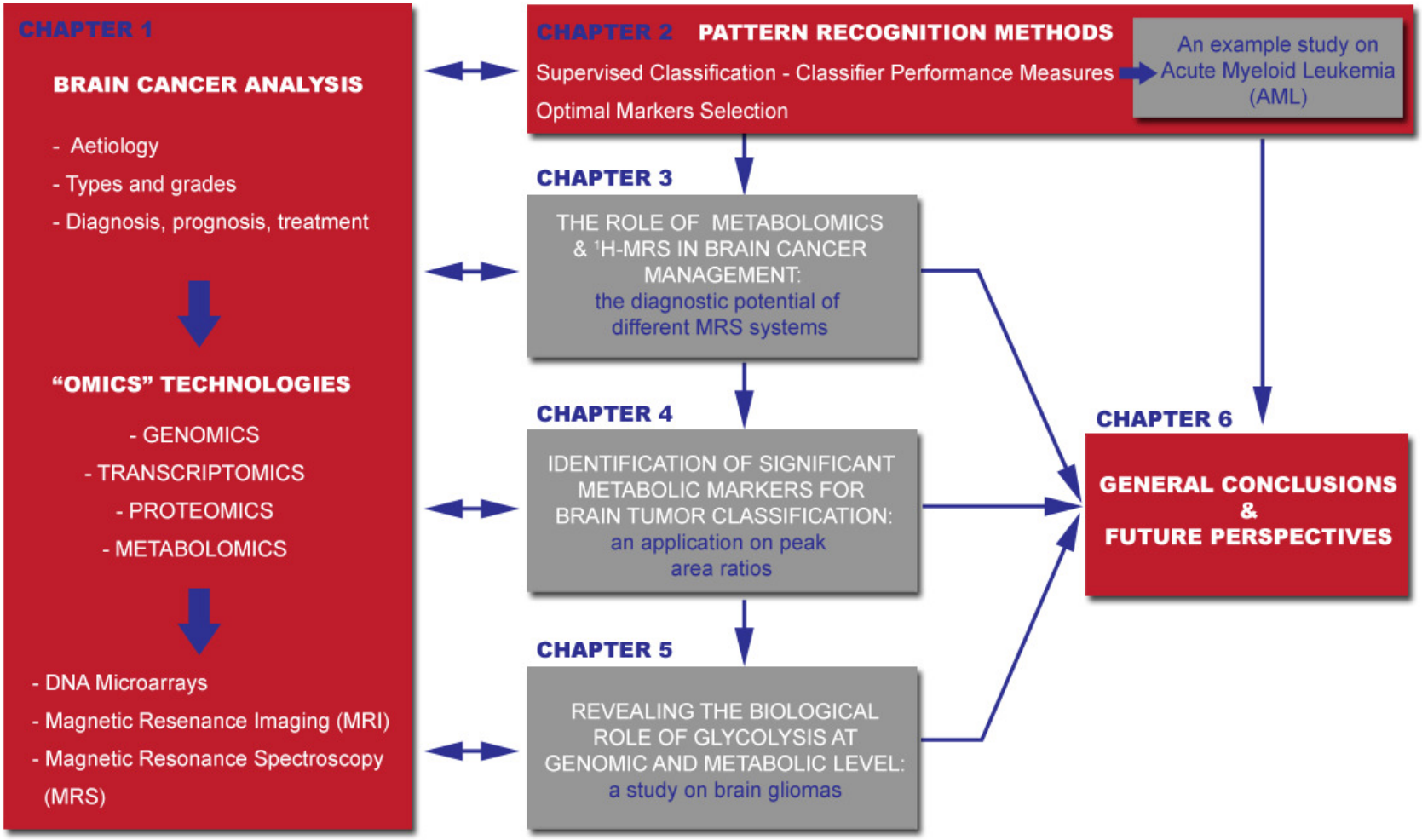


Figure 1 – Flow diagram of the Thesis – (The figure was designed with Adobe Photoshop ver. CS5)

References

Callot V. et al 2008, '^{1}H MR spectroscopy of human brain tumours: a practical approach', *European Journal of Radiology*, vol. 67, pp. 268-74.

Devos A. et al 2005, 'The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification', *Journal of Magnetic Resonance*, vol. 173, pp. 218-28.

Furnari F.B. et al 2007, 'Malignant astrocytic glioma: genetics, biology, and paths to treatment', *Genes and Development*, vol. 21, pp. 2683-2710.

Galanaud D. et al 2006, 'Non-invasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy', *Magnetic Resonance in Medicine*, vol. 55, pp.1236-1245.

Golub T.R. et al 1999, 'Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring', *Science*, vol. 286, no. 5439, pp. 531–537.

Griffin J.L. et al 2006, 'A metabolomics perspective of human brain tumours', *FEBS Journal*, vol. 274, pp. 1132-1139.

Heerschap A. 2007, 'In vivo magnetic resonance spectroscopy in clinical oncology. in Shields A.F. and Price P. (eds) Cancer Drug discovery and Development In Vivo Imaging of Cancer Therapy, *Humana Press Inc.,* Totowa, pp. 241-258.

Howe F.A. et al 2003a, '^{1}H MR spectroscopy of brain tumours and masses', *NMR in Biomedicine*, vol. 16, pp. 123-131.

Howe F.A. et al 2003b, 'Metabolic Profiles of Human Brain Tumors Using Quantitative In Vivo ^{1}H Magnetic Resonance Spectroscopy', *Magnetic Resonance in Medicine*, vol. 49, pp. 223-232.

Jones R.G. et al 2009, 'Tumor suppressors and cell metabolism: a recipe for cancer growth', *Genes and Development*, vol. 23, pp. 537-548.

Lehnhardt F.G. et al 2005, 'Metabolic differences between primary and recurrent human brain tumors: a ^{1}H NMR a ^{1}H NMR spectroscopic investigation', *NMR in Biomedicine*, vol. 18, pp. 371-382.

Lesniak M.S. et al 2004, 'Targeted therapy for brain tumours', *Nature Reviews*, vol. 3, pp. 499-508.

Luts J. et al 2007, 'A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection', *Artificial Intelligence in Medicine*, vol. 40, pp. 87-102.

Majos C. et al 2004, 'Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE', *American Journal of Neuroradiology*, vol. 25, pp. 1696–1704.

Postma G.J. et al 2011, 'On the relevance of automatically selected single-voxel MRS and multimodal MRI and MRSI features for brain tumour differentiation', *Computers in Biology and Medicine*, vol. 41, pp. 87-97.

Ohgaki H. et al 2009, 'Genetic alterations and signalling pathways in the evolution of gliomas', *Cancer Science*, vol. 100, pp. 2235-2241.

Ramaswamy S. et al 2002, 'DNA Microarrays in Clinical Oncology', *Journal of Clinical Oncology*, vol. 20, no. 7, pp. 1932-1941.

Seyfried N.T. et al 2005, 'Targeting energy metabolism in brain cancer: review and hypothesis', *Nutrition & Metabolism*, vol. 2, pp(no). 30

Seyfried N.T. et al 2011, 'Metabolic management of brain cancer', *Biochemica et Biophysica Acta*, vol. 1807, pp. 577-594.

Shalon D. et al 1996, 'A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization', *Genome Research*, vol. 6, pp. 639-645.

Shiraishi T. et al 2003, 'Genetic alterations of human brain tumors as molecular prognostic factors', *Neuropathology*, vol. 23, pp. 95-108.

Simonetti A.W. et al 2005, 'Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification', *NMR in Biomedicine*, vol. 18, pp. 34-43.

Sjobakk T.E. et al 2006, 'Clinical 1H magnetic resonance spectroscopy of brain metastases at 1.5T and 3T', *Acta Radiologica*, vol. 47, pp. 501-508.

Spratlin J.L. et al 2009, 'Clinical Applications of Metabolomics in Oncology: A Review', *Clinical Cancer Research*, vol. 15, pp. 431-440.

Van't Veer L.J. et al 2008, 'Enabling personalized cancer medicine through analysis of gene-expression patterns', *Nature,* vol. 452, pp. 564–570.

Chapter 1

THE "OMICS" IN BRAIN CANCER ANALYSIS

Based on Author's published studies:

- ***Kounelakis M***., Zervakis M., Kotsiakis X., The impact of Microarray Technology in Brain Cancer, Outcome Prediction in Cancer, A. F. G. Taktak & A. C. Fisher (eds): A Multi-Perspective View on Outcome Prediction in Cancer, *Elsevier Publishing Company, pp. 339-388, 2006*

- ***M. Kounelakis***, M. Blazantonakis, M. Zervakis and X. Kotsiakis, The Impact of Bioinformatics In Clinical Practices On Human Cancer, *European Conference on Emergent Aspects in Clinical Data Analysis, Pisa, 2005*

1.1 Introduction: What is Cancer?

Cancer is a malignant neoplasm that develops when cells in a tissue of the body begin to grow in an anarchic manner. Although there are many types of cancer, they all start because of out-of-control growth of abnormal cells. Normal body cells grow, divide (mitosis), and die (apoptosis) in an orderly fashion. During the early years of a person's life, normal cells divide more rapidly until the person becomes an adult. After that, cells in most parts of the body divide only to replace worn-out or dying cells and to repair injuries. Because cancer cells continue to grow and divide, they are different from normal cells. Instead of dying, they outlive normal cells and continue to form new abnormal cells *[Internet Sources: American Cancer Society - National Cancer Institute of US]*.

Cancer cells develop because of damage to DNA (Deoxyribonucleic Acid) also known as the genome, which contains all the genetic instructions used in the development and functioning of all known living organisms, packed in genes as shown in Figure 1.1 *[Gray J.W. et al 2000]*.

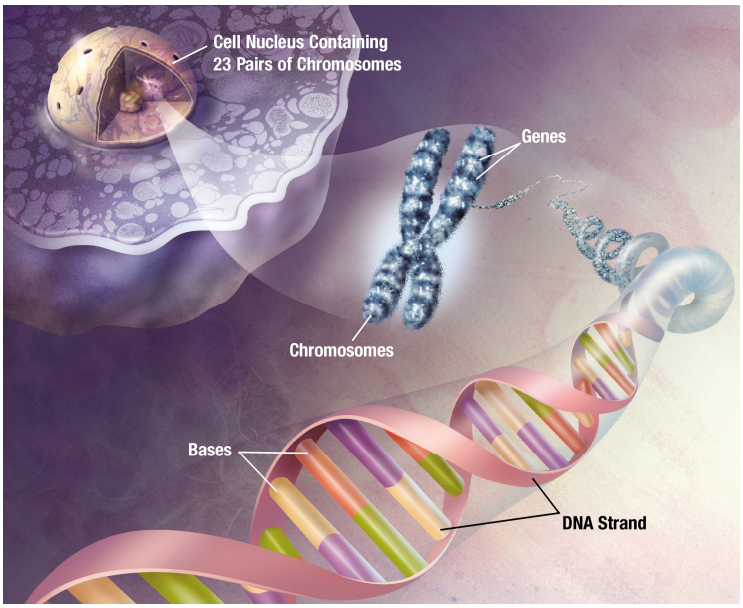


Figure 1.1 - The DNA expressed in chromosomes and genes – [Internet Source: U.S. National Institute of Health]

A gene is a DNA segment that codes for a type of protein that has a function in the organism. Within a cell's nucleus, DNA is organised into long structures called chromosomes whose analysis is also known as the "karyotype". DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases (known as adenine - cytosine - thymine – guanine or generally ACTG). It is the sequence of these four bases along the backbone that encodes information. This information is read using the

genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA (single DNA strand), in a process called transcription.

Most of the times when DNA becomes damaged the body is able to repair it. In cancer cells, the damaged DNA is not repaired. People can inherit damaged DNA, which accounts for inherited cancers. Many times though, a person's DNA becomes damaged by several other factors like, such as environmental factors (smoking, radiation exposure) and immune system disorders.

Cancer usually forms as a tumor, as shown in Figure 1.2. Some cancers, like leukaemia, do not form tumors. Instead, these cancer cells involve the blood and blood-forming organs and circulate through other tissues where they grow. Often, cancer cells travel to other parts of the body, where they begin to grow and replace normal tissue. This process is called metastasis. Regardless of where a cancer may spread, it is always named from the place it began. For instance, lung cancer that spreads to the brain is still called lung cancer, not brain cancer. Not all tumors are cancerous (malignant). Benign (non-cancerous) tumors do not spread (metastasize) to other parts of the body and, with very rare exceptions, are not life threatening. Different types of cancer can behave very differently.

For example, brain cancer and breast cancer are very different diseases *[Internet Source: Cancer Research in UK]*. They grow at different rates and respond to different treatments. That is why people with cancer need treatment that is aimed at their particular kind of cancer.

Today, millions of people are living, or have lived, with cancer. The risk of developing most types of cancer can be reduced by changes in a person's lifestyle. The sooner a cancer is found and treatment begun, the better the chances of survival.

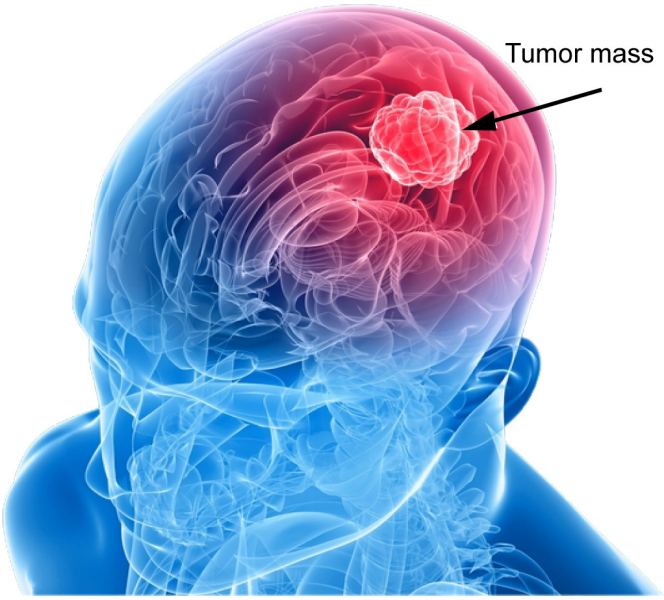


Figure 1.2 - A brain tumor mass – [Internet Source: National Brain Tumor Society]

1.2 The “Omics” in Cancer

“Omics” is an emerging and exciting area in the field of science and medicine. Numerous promising developments have been elucidated using “omics” (genomics, transcriptomics, proteomics and metabolomics), in cancer research, as shown in Figure 1.3. The development of high-throughput technologies that permit the solution of deciphering cancer from higher dimensionality will provide a knowledge base which changes the face of cancer understanding and therapy *[Cho W.C.S 2010]*.

After the first draft sequence of the human genome was announced in 2001 by the International Human Genome Sequencing Consortium *[Internet Source: National Human Genome Research Institute]* and the beginning of the design of The Cancer Genome Atlas *[Internet Source: The Cancer Genome Atlas]*, the scientific era of “omics” has emerged to revolutionize our way of studying and understanding cancer. The term “omics”, derived from the Greek suffix “ome” meaning “collection or body”, represents the rigorous study of various collections of molecules, biological processes, physiologic functions and structures as systems *[Cho W.C.S 2010 - Keusch G.T. 2006]*.

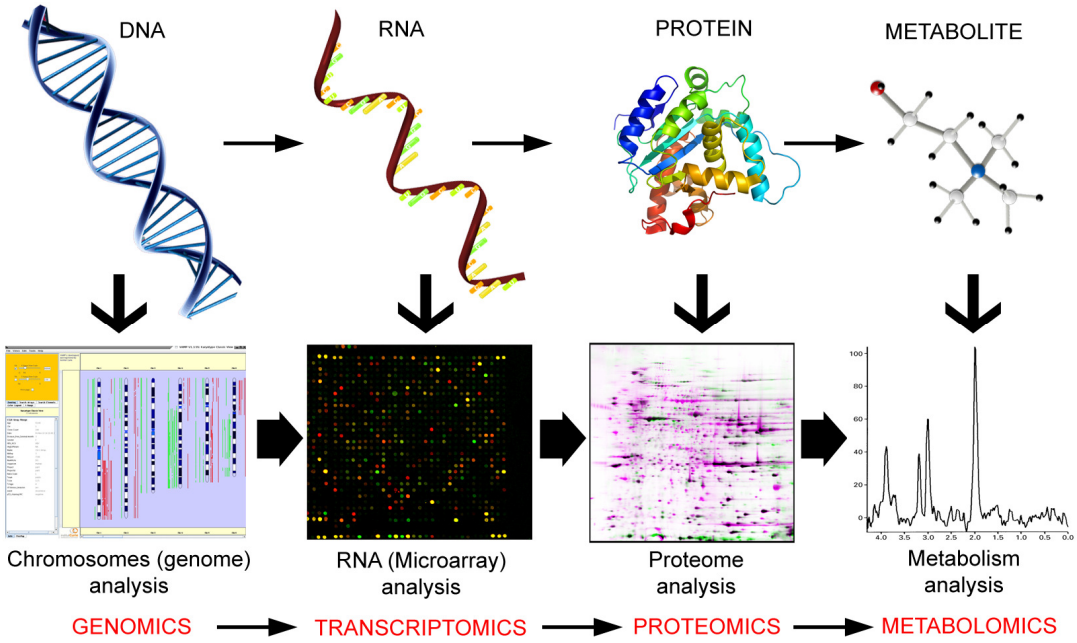


Figure 1.3 -The “omics” stages in cancer research. From Genomics to Metabolomics – [Internet Sources: Curie Institute – Stanford Functional Genomic Facility – University of York, UK – British Medical Journal – Wikipedia] - (The figure was designed with Adobe Photoshop ver.CS5)

Although it is now known that carcinogenesis is the result of abnormal genetic_mutations that result in uncontrolled cellular proliferation, cell cycle deregulation and a decrease in cell death, experts throughout the world have also focused on transcriptomics, proteomics and metabolomics, hoping for a better monitoring of cancers' biological behaviour. Application of "omics" in cancer research provides multi-dimensional analytical methods that reveal cancer's biological profile. The study of "omics" in cancer provides experts the ability to discover new biomarkers which can explain carcinogenesis, metastasis and grade progression and even identify new types of cancer. In other words they facilitate an insight into genes, proteins and metabolites of cancer cells, at various stages of the disease, enabling a better diagnosis, prognosis and treatment management.

Genomics is a discipline in genetics concerning the analysis and study of the genomes of organisms. Their application in cancer research can lead to the discovery of novel oncogenes and tumor suppressors which will have a great impact in our understanding of cancers' pathological mechanisms that cause carcinogenesis. The analysis of the genomic profile of a patient with respect to the outcome of a chemotherapy and/or radiotherapy followed, determines the efficacy of the treatments applied and provides significant knowledge about the grade and state of the tumor. Furthermore, genomics can be used to identify molecular pathways which describe the functions and the interrelations of genes, enabling the design of new classes of drugs and treatments *[Furge K.A. et al 2007]*.

The human genome contains the complete set of genes required to build a functional human being. However, the genome is only a source of information. In order to function, it must be expressed. The transcription of genes to produce RNA is the first stage of gene expression. The transcriptome is the complete set of RNA transcripts produced by the genome at any moment.

Unlike the genome, the transcriptome is extremely dynamic. Most of our cells contain the same genome regardless of the type of cell, stage of development or environmental conditions. Conversely, the transcriptome varies considerably in these differing circumstances due to different patterns of gene expression. *Transcriptomics*, the study of the transcriptome, is therefore a global way of looking at gene expression patterns. Its main applications involve cancer diagnostics and prognostics based on the gene expression profile of RNA, as well as in drug development *[He Y.D. et al 2006]*.

Proteomics enable the quantitative investigation of both cellular protein expression levels and protein-protein interactions involved in tumors. Monitoring the protein expression patterns, the proteome, in tumor cells offers the opportunity to discover potential cancer biomarkers and also detect carcinogenesis in early stages. Targeting the protein-network of the tumor enables the assessment of the therapeutic efficacy and toxicity of the drugs selected for a specific patient *[Cho W.C.S et al 2007(a)]*.

Metabolomics is a dynamic portrait of the metabolic status of a living system, also called the metabolome. This new "omic" approach reveals the metabolic profile of a cell, tissue, body fluids or the whole body at any moment in time. By analyzing the relative concentrations

of specific small molecules, called metabolites, we can determine the metabolic activity of a tumor which is significant for tumor grade progression evaluation. The diagnostic, prognostic and therapeutic value of metabolomics has been further enhanced since 1956, when Otto Warburg stated that cancers present altered glycolysis which is a fundamental metabolic process in all living organisms *[Warburg O. 1956]*. For human studies, the Human Metabolome Project (HMP) attempts to identify and record all the metabolites found in our body, aiming to complete a metabolic map which will provide useful information about the linkage of the human metabolites and proteins, genes and pathways with which they are involved *[Wishart D.S. 2007]*.

The introduction of the “omics” technologies into the cancer research exhibited the need to find a reliable and understandable way to visualize the interrelations between the different modalities. For this purpose the concept of the *pathway* was generated. A pathway is a biological network which illustrates the gene-gene (in genomics), protein-protein (in proteomics) and metabolite-metabolite (in metabolomics) interactions. Based on these pathways the experts now can also understand more complicated interactions such as gene-protein or gene-enzymes interactions. In Chapter 5 two pathways are used (Figures 5.6 and 5.7) one at genetic level (showing the gene interactions of tumor suppressors and oncogenes) and a second one at metabolic level (showing the metabolic interactions occurring in the cellular glycolysis pathway). Today, popular published on-line databases, such as the Kyoto Encyclopedia of Genes and Genome (KEGG) and the Human Metabolome Database (HMDB) offer access to these pathways so experts can extract useful information for brain cancer pathology.

The grand vision of all experts is to integrate the knowledge derived from these “omics” modalities under a biomedical manner in order to design patient specific diagnostic, prognostic and therapeutic protocols that will eventually be adopted by daily clinical practices. However, the major challenge is how to bring the best outcome from the “omics” research into clinical practice as accurate and reliable standardised procedures. Each “omic” approach has its strengths and weaknesses and these must be carefully addressed if we really want to achieve the best results in cancer management.

1.2.1 The “omics” laboratory tools for cancer management

It is widely believed that thousands of genes and their products (i.e., RNA and proteins) residing in a given living organism function in a complicated and orchestrated way that creates the mystery of life. However, traditional methods in molecular biology generally work on a "one gene in one experiment" basis, which means that the output is very limited and the "whole picture" of gene function is hard to obtain.

In the past several years, a new technology called DNA microarray, Figure 1.4, has attracted tremendous interest among biologists *[Yang Y.H. et al 2002 - Golub T.R. et al 1999 - Ramaswamy S. et al 2002]*. This technology promises to monitor the whole genome on a single chip so that researchers can have a better picture of the interactions among thousands

of genes simultaneously. These techniques include various microarray-based approaches that allow for global, systematic, high-throughput comparisons of the gene expression differences between normal and cancerous tissues. In the field of cancer research, the most commonly used microarray techniques for the molecular profiling of human tumors have been cDNA and oligonucleotide microarrays *[Pomeroy S.L. 2002 – Kounelakis M.G. et al 2006 - Kounelakis M.G. et al 2005]*.

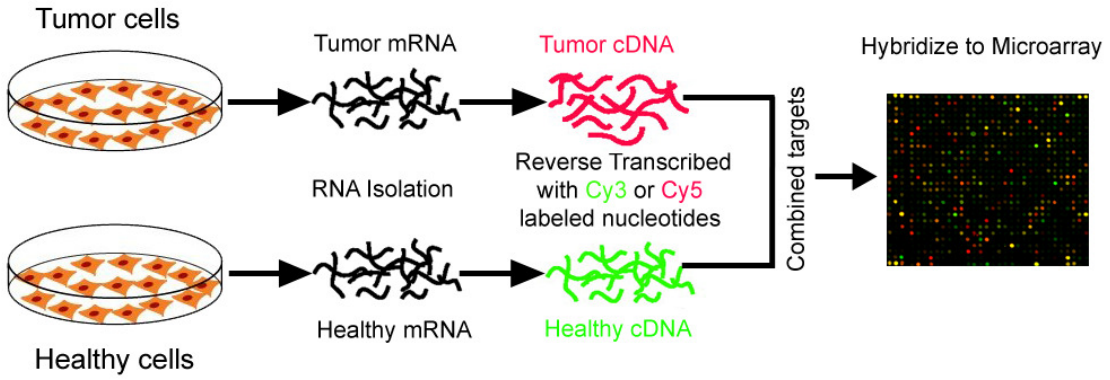


Figure 1.4 -The DNA Microarray hybridization process – [Internet Source: DNA Sequencing] – (The figure was designed with Adobe Photoshop ver.CS5)

Significant progress in clinical proteomics has been observed due to the contribution of state-of-the art technologies for proteome analysis, Figure 1.3. Several sophisticated tools such as 2D Difference Gel Electrophoresis (2-DE), Mass Spectrometry (MS) and Electrospray Ionization (ESI) have been used for differential evaluation of biological samples. The adoption of such high-throughput methods of molecular analysis, applied in vivo or in vitro, can comprehensively survey the proteomic profile of tumors and identify patterns that are associated with a particular malignancy and clinical behavior *[Cho W.C.S. et al 2007(b)]*.

It is essential to point out that there are substantial differences in the nature of the information between genomics, proteomics and metabolomics. Although genomics and proteomics yield information regarding the presence or absence of a genetic expression in the form of RNA or protein, the information obtained from the metabolomics differs in that it provides quantitative and dynamic screening of the behavior of the metabolic network of an organism. Metabolomics research focuses on functional metabolites that can be measured quantitatively so that net effects of genetic and environmental influences on disease-related cellular processes can be elucidated *[Kim Y.S. et al 2008]*.

Metabolite detection and quantification is usually carried out by Magnetic Resonance Spectroscopy (MRS) which make use of the magnetic properties of certain atomic nuclei with an odd number of protons and/or neutrons (e.g. ^{1}H, ^{31}P and ^{13}C). These nuclei possess a magnetic moment and when placed in a strong magnetic field will resonate at a particular radiofrequency that subtly depends upon the chemical environment. In MR spectroscopy, the

frequencies and intensities of these resonances are measured and represented graphically in an MR spectrum, Figure 1.3.

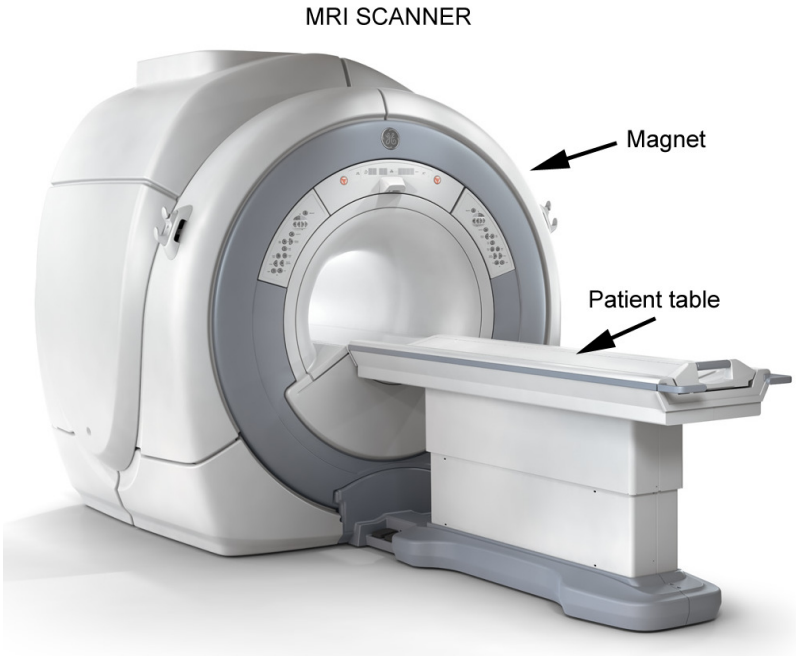


Figure 1.5 - An MRI scanner (Sigma HDe 1.5 Tesla) – [Internet Source: GE Healthcare]

The MRS technology originates from the well known Magnetic Resonance Imaging (MRI) widely applied today for diagnostic purposes. Magnetic Resonance Image generally refers to the spatial distribution of the ^{1}H (hydrogen protons). Magnetic Resonance Imaging is a non-invasive technique which does not involve any radiation risk for the patient. Using an MRI scanner, shown in Figure 1.5, images of cross-sections of the diseased/tumorous area can be acquired with several contrasts, depending on the acquisition parameters. Images can also be acquired after intravenous administration of a contrast agent. Together, they provide the clinician with high resolution images on which most tissue types and their morphology are clearly displayed.

However, even with the precise morphologic information, it is not always able to distinguish between different tumor types or to indicate the spatial extent of the tumor. In such cases the application of MRS fills the gap. While MRI uses the signal from hydrogen (^{1}H) protons to form anatomic images, proton MRS (^{1}H-MRS) uses this information to determine the concentration of metabolites in the tissue examined.

1.3 Oncology and Brain Tumors

Oncology is certainly one of the most demanding fields of medical science. Accurate diagnosis, prognosis and application of effective therapeutic methods is a genuine challenge for oncologists, biologists, radiologists and bioinformaticians worldwide, who daily unite their efforts daily to face such a complex disease as cancer. Although it is now clear that the understanding of the genetic abnormalities evolving in human cells could be the “Achilles heel” of carcinogenesis, the fact that clinical practices and treatment protocols daily followed often fail to save a patient’s life, verify that the discovery of the real causes for carcinogenesis is a true labyrinth. This is even harder both psychologically and scientifically when the survival time due to a poor prognosis is short which is common in brain tumor patients.

Brain tumors are among the most complex and hard-to-tackle cancers. The fact that they develop within the brain tissue demands immediate and accurate management in diagnostic, prognostic and treatment stage.

The brain is made from two types of cells: neurons or nerve cells and supporting or glial cells, as shown in Figure 1.4. Neurons are cells that send and receive electro-chemical signals to and from the brain and nervous system. There are about 100 billion neurons in the brain. There are many more glial cells; they provide support functions for the neurons, and are far more numerous than neurons. The neuron consists of a cell body (or soma) with branching dendrites (signal receivers) and a projection called an axon, which conducts the nerve signal. At the other end of the axon, the axon terminals transmit the electro-chemical signal across a synapse (the gap between the axon terminal and the receiving cell). The word "neuron" was coined by the German scientist Heinrich Wilhelm Gottfried von Waldeyer-Hartz in 1891 (he also coined the term "chromosome"). Myelin coats and insulates the axon (except for periodic breaks called nodes of Ranvier), increasing transmission speed along the axon. Myelin is manufactured by Schwann's cells, and consists of 70-80% lipids (fat) and 20-30% protein. The cell body (soma) contains the neuron's nucleus (with DNA and typical nuclear organelles). A typical neuron has about 1,000 to 10,000 synapses (that is, it communicates with 1,000-10,000 other neurons, muscle cells, glands, etc.).

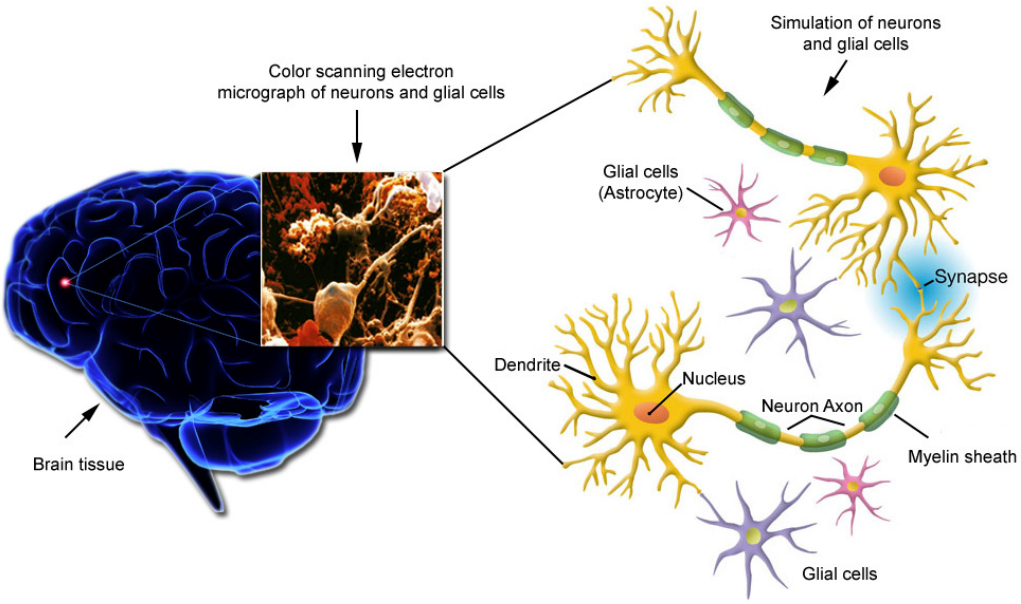


Figure 1.6 - Brain neurons and glial cells – [Internet Sources: National Institute of Child Health & Human Development – Science Clarified – Brain Supplements] - (The figure was designed with Adobe Photoshop ver.CS5)

A brain tumor is a mass of unnecessary and abnormal cells growing in the brain. The cell of origin determines the type of the brain tumor. A tumor that starts in the brain is a primary brain tumor. If cells have spread to the brain from a cancerous tumor in another part of the body, they are known as secondary brain tumors or metastatic brain tumors *[DeAngelis L.M. 2001]*.

Tumors can be benign and are usually but not necessarily localized in a small area, or they can be malignant and invasive (spreading to neighbouring areas). Brain cells can be damaged by tumor cells:

(i) directly when they are pushed or displaced due to growth of the tumor,

(ii) indirectly when affected through ongoing inflammation in and around the tumor mass,

(iii) via brain edema (swelling) or

(iv) increased pressure in the skull (due to brain edema or to the blockage of cerebrospinal fluid (CSF) circulation).

Brain tumors are usually located in the posterior third of the brain in childhood and in the anterior two-thirds of the brain in adulthood. Some benign tumors grow for many years and reach a large size before being detected. Malignant tumors tend to grow more rapidly and will have been present for a shorter period when they are discovered. The onset of symptoms cannot be used to detect how long the tumor has been present or indicate whether it is cancerous or non-cancerous.

The most common types of brain tumor are those which originate from malignant primary tumors arising somewhere else in the body and spreading out to the brain. Those secondary or metastatic brain tumors are always malignant. Approximately one quarter of metastatic

cancers spread to the brain. The lungs and breasts are the most common locations from which secondary brain tumors originate. Tumor cells travel to the brain via blood vessels. Since the brain has no lymphatic drainage system like other organs (the cerebrospinal fluid system acts as a lymphatic system in the brain), the spreading of tumor cells by the lymphatic route (which is very typical for cancers of other organs) is impossible. In contrast to primary brain tumors, metastatic tumor masses may occur in various remote locations in the brain. Symptoms, diagnosis and treatment are quite similar to those of primary tumors. However in the case of secondary tumors the initial location of the tumor cells must be identified and treated as well. Primary or secondary brain tumors may cause herniation of the brain (displacement of one part of the brain tissue due to mass effect of a lesion, usually causing the compression of the neurons controlling the respiratory system in the brainstem and eventually death) and permanent neurologic changes including intellectual decline.

1.4 Classification and Grading of Brain Tumors (WHO)

Histological grading, also called differentiation, refers to how much the tumor cells resemble normal cells of the same tissue type. More specifically, tumor grade is a system used to classify cancer cells in terms of how abnormal they look under a microscope and how quickly the tumor is likely to grow and spread. Many factors are considered when determining tumor grade, including the structure and growth pattern of the cells. The specific factors used to determine tumor grade vary with each type of cancer. In clinical practice, the tumor grade is a key factor influencing the choice of therapies, particularly determining the use of adjuvant radiation and specific chemotherapy protocols.

The World Health Organization (WHO) classification of the Central Nervous System (CNS) has included a grading scheme that is a "malignancy scale" ranging across a variety of neoplasms rather than a strict histological grading. The WHO grading is widely used, having incorporated or largely replaced other previously published grading systems *[Louis D.N et al 2007]*. Tumors are graded based on their microscopic appearances. The grade indicates the level of malignancy. Tumors are graded on their mitotic index (growth rate), vascularity (blood supply), presence of a necrotic (dead cells) center, invasive potential (border distinctness) and similarity to normal cells (atypia). Malignant tumors may contain several grades of cells. The most malignant grade of cell found determines the grade for the entire tumor, even if most of the tumor is a lower grade. Based on the microscopic appearance of cancer cells, experts commonly describe tumor grade by four degrees of severity: grades I, II, III, and IV. Grades I and II are generally characterised as low grade whereas III and IV as intermediate and high grade respectively.

Grade I lesions generally include tumors with low proliferative potential. These tumors grow slowly and microscopically appear almost normal. Surgery alone may be effective for grade I tumors. However, even a grade I tumor may be life-threatening if it is inaccessible for surgery. Grade I tumors are often associated with long-term survival. Lesions designated grade II are generally infiltrative in nature and, despite low level proliferative activity, often recur. Furthermore, they may invade surrounding healthy tissue. Grade III lesions are malignant. Most of the times these lesions present nuclear atypia and brisk mitotic activity while an invasion to surrounding healthy tissue is very common. Finally grade IV characterized lesions are the most malignant and invade wide areas of surrounding healthy tissue. These tumors reproduce rapidly, appear very unusual under the microscope and are necrotic in the center. Grade IV tumors cause new blood vessels to form, to help maintain their rapid growth. Most of the times, despite therapy, they have a fatal outcome.

1.4.1 Types of brain tumors

Brain tumors, as mentioned, are generally distinguished in primary (originate within the brain) and the secondary (metastasize to brain tissue from other parts of the body). A tumor that spreads to the brain is the same disease and has the same name as the original

(primary) tumor. Brain metastases outnumber primary neoplasm by at least 10 to 1, and they occur in 20% to 40% of cancer patients. The most common primary tumors metastasizing to the brain are lung cancer (50%), breast cancer (15%-20%), unknown primary cancer (10%-15%), melanoma (10%), and colon cancer (5%). Eighty percent of brain metastases occur in the cerebral hemispheres, 15% occur in the cerebellum, and 5% occur in the brain stem *[Levin V.A. et al 2001]*.

Metastases to the brain are multiple in more than 70% of cases, but solitary metastases also occur. Brain involvement can occur with tumors of the nasopharyngeal region by direct extension along the cranial nerves or through the foramina at the base of the skull. Dural metastases may constitute up to 9% of total CNS metastases *[Kounelakis M.G. et al 2006]*. Due to the fact that secondary brain tumors do not have their origin in the brain tissue, their further analysis is out of the scope of this thesis.

Primary brain tumors are classified by light microscopy according to their predominant cell type and graded based upon the presence or absence of standard pathologic features. Histological evaluations have shown that there are more than 120 primary brain tumors types. According to WHO, these are categorised into two broad categories, named neuroepithelial and non-neuroepithelial tumors. The most known and frequently diagnosed tumors belonging to these two classes are shown in Table 1.1 *[Louis D.N. et al 2007]*.

Table 1.1 – Classification of primary brain tumors by histology

Neuroepithelial Tumors	***Non-Neuroepithelial Tumors***
Astrocytic tumors	Meningial tumors
Oligodendroglial tumors	Pituitary tumors
Ependymal tumors	Nerve sheath tumors
Embryonal tumors	Germ cell tumors
Other Neuroepithelial tumors	CNS Lymphomas

1.4.1.1 Neuroepithelial brain tumors

The most known neuroepithelial brain tumors are the *Astrocytic*, *Oligodendroglial*, *Ependymal* and *Embryonal* tumors.

Astrocytic tumors, commonly called Astrocytomas, arise from small, star-shaped cells called astrocytes, shown in Figure 1.6. Astrocytes regulate and support the foundation of the functions of the brain. These tumors may grow anywhere in the brain or spinal cord. In adults, astrocytomas most often arise in the cerebrum. In children, they occur in the brain stem, the cerebrum, and the cerebellum, shown in Figure 1.7. The most malignant astrocytoma is the Glioblastoma multiforme (GBM) which is of grade IV.

Oligodendroglial tumors, with Oligodendrogliomas most often diagnosed, arise in the cells that produce myelin, the fatty covering that protects nerves, shown in Figure 1.6. These tumors usually arise in the cerebrum. They grow slowly and usually do not spread into surrounding brain tissue. Oligodendrogliomas are rare. They occur most often in middle-aged adults but have been found in people of all ages.

Ependymal tumors, with most common the Ependymomas, usually develop in the lining of the ventricles. They may also occur in the spinal cord. Although these tumors can develop at any age, they are most common in childhood and adolescence.

Embryonal tumors, with most common the Medulloblastomas, were once thought to develop from glial cells. However, recent research suggests that these tumors develop from primitive (developing) nerve cells that normally do not remain in the body after birth. For this reason, medulloblastomas are sometimes called primitive neuroectodermal tumors (PNET). Most medulloblastomas arise in the cerebellum; however, they may occur in other areas as well. These tumors occur most often in children and are more common in boys than in girls.

1.4.1.2 Non-Neuroepithelial brain tumors

The non-neuroepithelial brain tumors class includes the Meningial, Pituitary, Nerve sheath, Germ cell and CNS Lymphoma tumors.

Meningial tumors, with most common the Meningiomas, grow from the meninges, shown in Figure 1.7. They are usually benign. Because these tumors grow very slowly, the brain may be able to adjust to their presence. Meningiomas often grow quite large before they cause symptoms. They occur most often in women between 30 and 50 years of age.

Pituitary tumors, with Craniopharyngiomas mostly diagnosed, develop in the region of the pituitary gland near the hypothalamus, shown in Figure 1.7. They are usually benign; however, they are sometimes considered malignant because they can press on or damage the hypothalamus and affect vital functions. These tumors occur most often in children and adolescents.

Nerve sheath tumors, with Schwannomas are one of the most common. These are benign tumors that begin in Schwann cells, which produce the myelin that protects the acoustic nerve-the hearing nerve. Acoustic neuromas are a type of schwannoma. They occur mainly in adults. These tumors affect women twice as often as men.

Germ cell tumors arise from primitive (developing) sex cells, or germ cells. The most frequent type of germ cell tumor in the brain is the germinoma.

CNS Lymphomas start in lymphocytes (the main cell type of the immune system). They typically remain confined in the brain (leptomeninges, spinal cord, or eyes) and rarely spread outside the central nervous system.

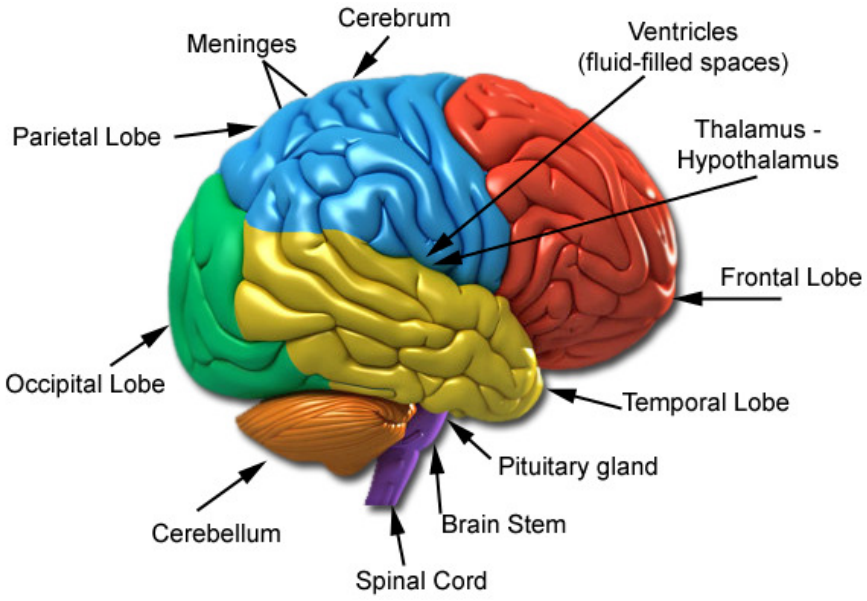


Figure 1.7 – Human brain parts - [Internet Source: Creative Crash] - (The figure was designed with Adobe Photoshop ver.CS5)

1.4.2 The case of gliomas

Among the five classes of neuroepithelial tumors shown in Table 1.1, the first three (Astrocytic, Oligodendroglial and Ependymal) comprise a well known class called *gliomas*. Gliomas are the most common type of primary brain tumors arising from the glial cells. There are four types of glial cells each one with different function.

- *Astrocytes* – which transport nutrients and energy to neurons, hold them in place and maintain the blood-brain barrier.
- *Oligodendrocytes* – which provide insulation (myelin) to neurons
- *Microglia* – digest dead neurons and pathogens
- *Ependymal cells* – line the ventricles and secrete cerebral spinal fluid (CSF)

According to the 2011 Report of the Central Brain Tumor Registry of the United States (CBTRUS), gliomas account for approximately 30% of all tumors, as shown in Figure 1.8.

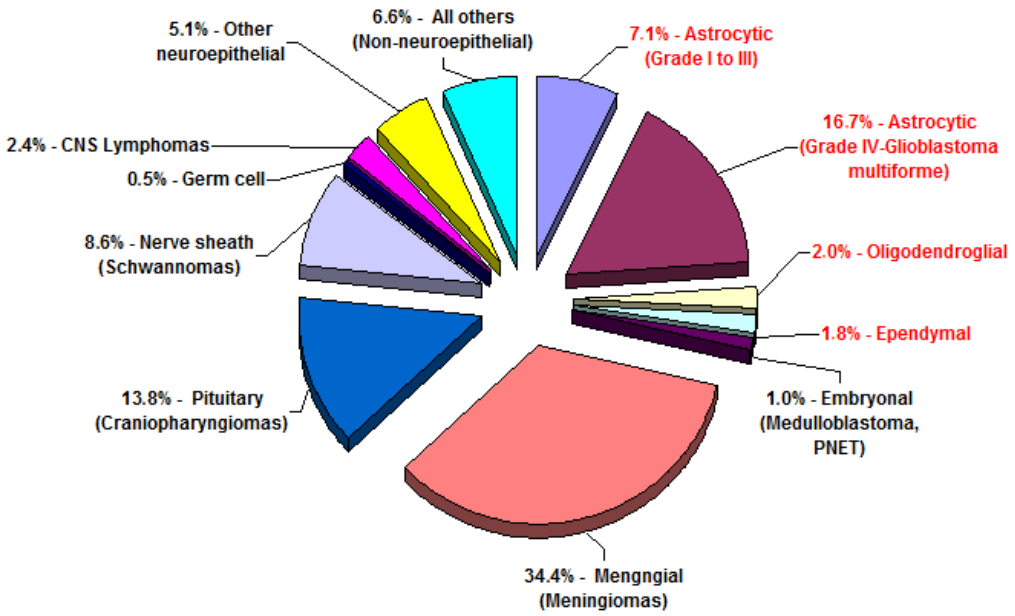


Figure 1.8 – Distribution of all primary brain tumors by histology – Gliomas distribution is with red color – [Internet Source: Report of Central Brain Tumor Registry of the United States 2011 (CBTRUS)]

Accounting for approximately 76% of all gliomas (where 53.7% accounts for Glioblastoma multiforme i.e. grade IV astrocytoma), astrocytomas are the most common glioma brain tumors, as shown in Figure 1.9. Astrocytomas are slow growing tumors that originate from astrocytes, most commonly found in the cerebrum. Based on the grade of their malignancy these tumors are generally distinguished into four sub-types as presented in Table 1.2.

Oligodendroglial tumors or oligodendrogliomas originate in oligodendrocytes; cells that help insulate the nerve fibers that transmit nerve impulses. Oligodendrogliomas are frequently located within the frontal, temporal or parietal lobes of the brain, causing seizures in the majority of patients. These tumors account for 6.5% of all gliomas, as shown in Figure 1.9. Most of the oligodendrogliomas are of grade II. When grade III is diagnosed these are called anaplastic oligodendrogliomas.

Ependymal tumors or ependymomas are rare types of glioma brain tumors that occur more often in children. They arise from ependymal cells, cells lining cavities of the brain where production and storage of Cerebral Spinal Fluid (CSF) occurs and account for 6.0% of all gliomas, as shown in Figure 1.9. The most frequently diagnosed ependymomas are of grade II. Grade III ependymomas are called anaplastic ependymomas.

Table 1.2 – Histological grading of gliomas sub-types

Glioma Sub-types	Grade I	Grade II	Grade III	Grade IV
Pilocytic Astrocytoma	■			
Diffuse Astrocytoma		■		
Anaplastic Astrocytoma			■	
Glioblastoma multiforme				■
Oligodendrogliomas		■		
Anaplastic Oligodendrogliomas			■	
Ependymomas		■		
Anaplastic Ependymomas			■	

[Internet Source: World Health Organisation (WHO)]

Apart from the three general classes aforementioned, others less frequently diagnosed have been also recorded according to WHO. These are the Oligoastrocytomas, Mixed gliomas, Optic gliomas and Gliomatosis Cerebri (GC). These cases are grouped in the "All other gliomas" class presented in Figure 1.9.

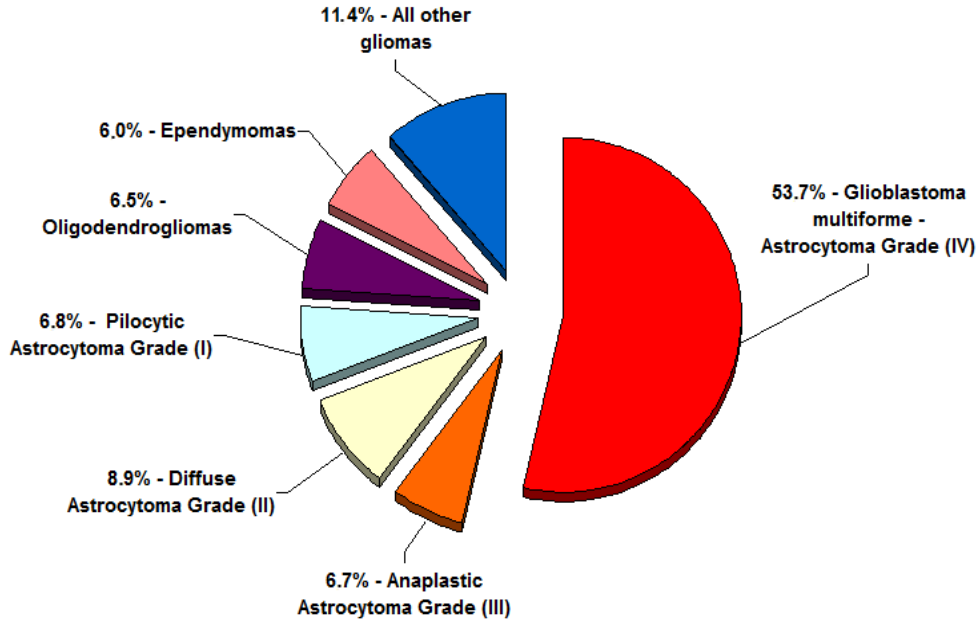


Figure 1.9 – Distribution of all brain gliomas sub-types by histology – [Internet Source: Report of Central Brain Tumor Registry of the United States 2011 (CBTRUS)]

The research efforts currently applied on gliomas brain tumors worldwide are exhaustive. Medical doctors monitor glioma patients daily to measure the efficacy of the therapies applied; biologists, genetists and chemical engineers try to identify the factors for gliomagenesis to generate new drugs and finally bioinformaticians develop new statistical models used for even more accurate diagnosis and prognosis. The main motivation for all these efforts is the fact

that, despite optimal treatment, the median survival for glioma patients is short, as shown in Table 1.3. Other reasons that justify this urgency lie in the nature of these tumors. Gliomas have no clear boundaries, which is why they tend to regrow after removal. In fact, it is unusual to be able to cure these tumors only through surgery because of the tendency of the tumor cells to extend well beyond where they can be seen on x-rays or at surgery.

Therefore there is a great need to discover new biomedical methods to support early and accurate diagnosis, prognosis and efficient therapies that will extend the patient’s survival. Towards this direction the “omics” analysis of gliomas has been adopted. The application of genomics, transcriptomics, proteomics and metabolomics on gliomas, opens new ways to fight this cancer since they can be used to identify new biological markers to help us understand their behaviour.

1.5 Current Brain Tumor Diagnostic Approaches

Diagnosis holds a key role in understanding the type and the characteristics of a brain tumor. False diagnosis can often lead to a false prognosis and treatment path. Current diagnostic approaches offer the ability to observe the morphology, size and stage of the tumor and decide on the therapy that has to be followed. The most recent diagnostic approaches are presented in the following with a brief analysis and discussion *[Fuller G.N. 2006]*.

Brain tumors are usually found because of signs or symptoms a person feels. The brain tumor symptoms can be fairly general, or they may be more specific depending on the tumor's location within the brain. Symptoms may occur gradually and become worse over time or they can happen suddenly, such as with a seizure or a coma. The pressure within the brain skull caused by a tumor can lead to general symptoms such as headache, nausea, vomiting, blurred vision, balance and movement problems, drowsiness and personality changes. Almost 50% of all patients have complained of severe and persistent headaches. Of course, having a headache does not necessarily mean that a brain tumor is present. Most headaches are caused by other factors. It is therefore a combination of these symptoms that can lead to the conclusion that a brain tumor has developed. Furthermore, due to the fact that the brain controls functions of other organs in the human body, such as the production of specific hormones, other symptoms, not listed here, may be present.

1.5.1 Neurological examination

If symptoms suggest a brain tumor may be present, the clinician will take a complete medical history and do a neurological examination to evaluate brain functions. A neurological examination is usually the first test given when a patient complains of symptoms that imply a brain tumor. The examination includes checking eye movements, hearing, sensation, muscle movement, sense of smell, and balance and coordination. The clinician will also test mental state and memory. If the results of the neurological examination are abnormal then the clinician will order further tests to better evaluate the status of the patient. The first tests to be applied are the imaging tests *[Abdullah N.D. et al 1999 - Barkovich A.J. 1992]*.

1.5.2 Non-invasive imaging modalities

Imaging tests are non-invasive examinations that use x-rays, strong magnets or radioactive substances to produce images of the tumorous brain area. The outcomes of these studies are examined and interpreted by radiologists who are specialized on imaging diagnostic techniques. The most current and state-of-the-art imaging diagnostic approaches applied today are analysed below *[Wen P. et al 2001 – Kounelakis M.G. et al 2006]*.

Skull x-rays were once standard diagnostic tools but are now performed only when more advanced procedures, for example MRI, are not available. Occasionally they may be useful in demonstrating calcification usually present in relatively slow growing brain tumors.

Astrocytomas are the most common calcifying tumor. Although calcifications occur only in approximately 20% of astrocytomas, their overall frequency more compensates for this percentage. Calcification also occurs in 50–60% of oligodendrogliomas, 70–80% of craniopharyngiomas, 50% of ependymomas, 35% of gangliogliomas, and about 10% of meningiomas *[Segall H.D. et al 1990]*.

Computed Tomography (CT) scan is an x-ray test that produces detailed cross- sectional images of the brain tissue. Instead of taking one image, like a regular x-ray, a CT scanner takes many images of the brain tissue, as it rotates around the patient's head. A computer then combines these images producing a number of slices of the brain tissue. Unlike a regular x-ray, a CT scan has the ability to create detailed images of the soft tissues in the body, like the brain tissue. CT scans tend to be better tolerated than MRI because of their shorter scanning time and are more sensitive for detecting acute haemorrhage, calcifications, and bony involvement. CT angiography or CTA is another procedure that is also used in combination with the CT. During this test a contrast material is intravenously injected into the patient. The CT scanner creates detailed images of the blood vessels in the brain, which can help clinicians to plan a surgery.

Magnetic Resonance Imaging (MRI) scans are particularly helpful in examining the brain and spinal cord and are considered the best way to look for tumors in these areas. The acquired images are usually more detailed than those provided from the CT scanner. A major drawback though is the fact that the MRI scanner, shown in Figure 1.5, does not display the bones of the skull as well as CT scans making difficult for the radiologist to see the effects of tumors on the skull. MRI scans use radio waves and strong magnets instead of x-rays. The energy from the radio waves is absorbed and then released in a pattern formed by the type of body tissue and by certain diseases. Dedicated computer software translates the pattern into a very detailed image of parts of the body. In most patient cases a contrast material called gadolinium (Gd) may be injected into a vein before the scanning to obtain clearer images.

Magnetic Resonance Spectroscopy (MRS) also known as MR spectroscopy, is like an MRI, except that it provides the ability to measure the interactions of the radio waves with different atoms, such as hydrogen (^{1}H), as mentioned above. MRS images highlight some substances of brain tumors, called metabolites, which are not clearly seen by MRI. The study of these metabolic substances within the brain tumorous are provides the metabolic profile, called metabolism, of the tumor which greatly supports experts treatment decisions. MRS can also be used after treatment to help determine if an abnormal area is a remaining tumor or if it is more likely to be scar tissue. CT angiography can provide better details of the blood vessels in and around a tumor than MR angiography (analysed next) in selected cases *[Doherty G.M. 2010]*.

Magnetic Resonance Angiography (MRA) is a special form of MRI, which is also applied to observe the structure of the blood vessels in the brain, as CT angiography. Like CT angiography, MRA is very useful before a surgery.

Magnetic Resonance Perfusion or Perfusion Weighted Imaging (PWI) is another significant imaging test where a contrast material is injected quickly into the patient's vein. Then a special type of MR image is obtained to look at the amount of blood going through different parts of the brain. This test is based on the observation that brain tumors need a greater blood supply than normal areas of the brain. This means that the faster a tumor is growing, the more blood it needs. Therefore, Perfusion MRI can give clinicians an idea of how quickly a tumor is growing or help show them the best place to take a biopsy.

Positron Emission Tomography (PET) is a special test based on the brain tumor's property to absorb larger amounts of glucose (a form of sugar) than the healthy tissue, as they grow. For this reason glucose that contains a radioactive atom is injected into the patient's blood. A special camera can then create an image of areas of radioactivity in the body. This image is not finely detailed like a CT or MRI scan, but it can provide helpful information about whether abnormal areas seen on other tests (such as MRIs) are likely to be cancerous or not. This test is also useful after treatment, as it can help tell whether the tumor cells have been killed since dead cells do not use glucose. Abnormal areas may still show up on an MRI scan. PET scans can help determine if the abnormal area is a remaining tumor or if it is more likely to just be scar tissue *[Hustinx R. et al 1999]*.

Chest x-ray is a plain x-ray of the patient's chest, which is sometimes applied by clinicians. The purpose of this test is to identify possible tumors in other organs of the body which have metastasized to the brain. Such an organ could be the lung which most of the times give metastases to brain.

1.5.3 Histological evaluation through biopsy

Although diagnosis using imaging techniques can give clinicians information about the tumor's location, size, morphology and tumorous borders, a histological evaluation through biopsy is, oftentimes, inevitable. A definite evaluation and classification of a tumor can only be accomplished by the histological analysis under a microscope of a small tissue of the tumor. The removal of this tissue from the brain is called a biopsy. The main biopsy procedure followed nowadays is the guided or needle biopsy, as shown in Figure 1.10.

Guided biopsy is done using a needle. The needle is guided by a CT or MRI scan. The scan helps to ensure that the neurosurgeon can move the tip of the needle into exactly the right place to take a sample from the tumor. There are two ways of doing this - stereotactic biopsy or neuronavigation. Surgeons most often use guided biopsy for tumors that are very deep inside the brain or for tumors that are widely spread throughout an area of the brain.

For *stereotactic biopsy*, a head frame is fitted on the patient. Once the scan is obtained, the neurosurgeon uses the scan and the reference points from the head frame to work out exactly where to guide the needle. In this type of biopsy a general anesthetic is used. The neurosurgeon makes a very small hole in the skull with a drill. Then the frame is set to guide a fine needle into exactly the right position to take the tissue sample.

For *neuronavigation*, the neurosurgeon takes the biopsy with a fine needle in much the same way. In this procedure though the head frame is not needed. The neurosurgeon looks at the scan while guiding the needle into position. Sometimes neurosurgeons use the natural landmarks of patient’s nose, eyes and ears to help them guide the needle into position.

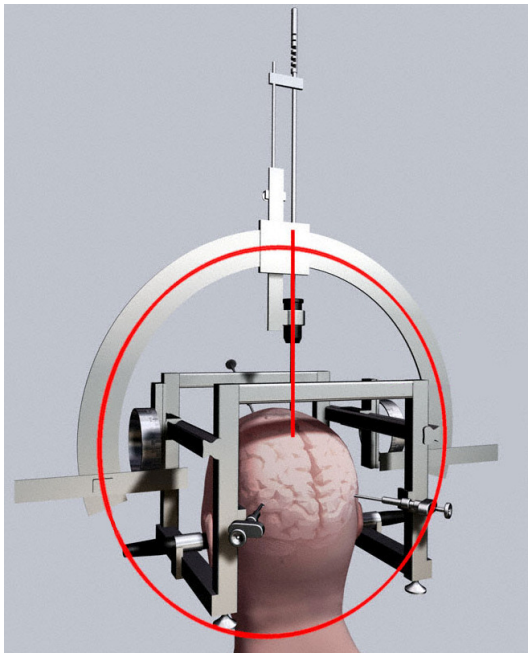


Figure 1.10 – The head frame used for brain stereotactic (guided) biopsy – [Internet Source: Med Art]

Biopsy is, as mentioned, both inevitable and necessary in most of the brain tumor diagnosis cases. Nowadays though, experts’ efforts focus on the development of new diagnostic techniques that will allow an accurate evaluation of the tumor circumventing the surgical operation of a biopsy. It is a common belief that such tools and methods can only be generated from the “omics” area. The evolution and improvement of “omics”-related technologies such as DNA Microarrays, MRS metabolic profiles will provide a deeper insight into a tumor’s intrinsic characteristics.

1.6 The Prognosis Stage

The prognosis of a brain tumor usually refers to the likely outcome of the disease. The prognosis phase considers the survival rates, death rates, prospects for patient's recovery after treatment, recovery period for patient and chances of brain tumor to recurring. Naturally, such forecast issues are by their nature unpredictable. The survival rates in people with brain tumors depend on many different factors *[Kounelakis M.G. et al 2006]*. The most important are shown below.

- Whether the tumor is malignant or benign.
- The size of the cancerous tumor in the brain impacts the prognosis significantly. There is limited growth space in the brain, so as a tumor grows in the brain it severely affects the surrounding brain tissue and its function. This can result in speech, sight, and other abilities being adversely affected when the tumor pushes into other brain areas because of its size. Additionally, the size of a tumor will dictate treatment options pursued (and how quickly).
- The tumor's cell type and location. The location of a tumor is important, since some tumors that would normally be considered benign are instead labeled malignant due to their complicated location in the brain, preventing the ability to be safely removed during surgery without risking life or quality of life, or quality of.
- Tumor grade. According to WHO there are four grades of malignancy (I to IV), aforementioned. The grade also shows the possible tendency to have a tumor spread.
- Patient's age. The outlook is poorer in the very youngest (<20 years old) and very oldest patients (>65 years old), although younger patients who survive two years after diagnosis have a much better outlook than older patients.
- Patient's ability to function and duration of symptoms.

Survival statistics show what is likely to happen in large groups of people. They can sometimes be useful as a general guide, but each person's situation is unique. Despite the factors mentioned, others, like the amount of tumor that can be surgically removed, can also affect the patient's outlook. The usually estimated *5-year survival rate* refers to the percentage of patients who live at least 5 years after being diagnosed. Some of these patients, though, live longer than 5 years and others die during this time period. Another term also used is relative survival rates, as in Table 1.3. The relative survival rates differ in that they do not include patients who die from other causes during the 5-year period.

In Table 1.3 it can be observed that among gliomas sub-types, Anaplastic astrocytoma and Glioblastoma multiforme (in red color) have the worst survival rates. In addition, it can be seen that, in all cases, younger people have better outlooks than older ones.

Table 1.3 – 5 Year relative survival rates

Glioma Sub-types	Ages 20-44	Ages 45-54	Ages 55-64
Pilocytic and Diffuse Astrocytomas	91%	82%	81%
Anaplastic Astrocytoma	48%	28%	8%
Glioblastoma multiforme	16%	5%	3%
Oligodendrogliomas	84%	76%	65%
Anaplastic Oligodendrogliomas	65%	52%	32%
Ependymomas and Anaplastic Ependymomas	90%	84%	84%

[Internet Sources: Report of Central Brain Tumor Registry of United States 2011 (CBTRUS) – Surveillance Epidemiology and End Results (SEER) of National Cancer Institute of US]

1.7 Current Brain Tumor Treatment Strategies

1.7.1 Today's treatment methods

When a patient is diagnosed with a malignant brain tumor, the treatment protocol applied, in all oncologic clinics, involves the following therapeutic approaches *[Doherty G.M. 2010 - Castro M.G. et al 2003]*.

- *Surgery*
- *Radiation treatment (also called radiotherapy)*
- *Chemotherapy*

The intensity, the combination and time of application of these treatments depends on the tumor's type, size and location and patient's age, health status and medical history. Unfortunately, even when treated with aggressive combined therapies, most of the times, the high grade glioma patients do not have the desired clinical outcome. Glioblastoma multiforme (GMB) often recurs between 6 and 12 months while anaplastic astrocytoma within 18–36 months.

Recent advances though in surgical, radiation and chemotherapeutic treatments hope to have significantly extended average survival times compared to those of standard therapy. Investigative treatments, such as monoclonal antibodies and gene therapies also show promise of a great future.

1.7.1.1 Removing the tumor through surgery

Surgery plays a crucial role in the management of brain tumors and especially gliomas. Low grade tumors should be resected to the limitation of clinical deficits. On the other hand, high grade tumors and particularly GBM can be difficult to resect due to the undefined tumor's edges; the tumor may extend into healthy-looking brain tissue, and/or can be localized near critical areas for the brain's function. Before surgery, the size and number of tumors must be considered, along with the general health status of the patient.

Tumors can be either completely or partially resected, or biopsied. Tumor resection is beneficial to alleviate mass effect, such as pressure. It is thought that even partial resections are beneficial to the patients, as it improves body functions, relieves pressure in the brain, and disrupts the blood-brain barrier [*Balmaceda C. 2000*]. This allows for enhanced exposure to chemotherapeutic drugs. It also provides space for the tumor to grow, and pushing the tumor mass into a growing cell cycle appears to achieve better responses to radiation and chemotherapy. At diagnosis, tumors are usually localized and are less than 5 cm in diameter [*Eck S.L. et al 1996*]. It is thought that surgery for the treatment of GBM can prolong the life of the patient for up to 6 months *[Shand N. et al 1999]*.

The standard surgical procedure followed is craniotomy or open surgery. If the brain tumor appears to be treatable with surgery, the neurosurgeon will cut out an area from the patient's skull to search for the tumorous tissue. In cases where the tumor is reachable all of its mass can be removed. Unfortunately this is not always the case and some small, unapparent tumor tissue will remain within the brain resulting in tumor re-growth.

The surgeon has various surgical options for breaking down and removing the tumor. These include:

- *Standard surgical procedures.*
- *Laser microsurgery (which produces great heat and vaporizes tumor cells).*
- *Ultrasonic aspiration (which uses ultrasound to break the glioma tumor into small pieces, which are then suctioned out).*

Most malignant tumors require additional treatments, including repeated surgery. Additional procedures to enhance brain surgery have been developed in order to allow maximum removal of the cancerous tissue. Some of them are: *Stereotaxy, Cortical Localization, Image-Guided Surgery, and Magnetic-Tipped Catheters.*

1.7.1.2 Applying radiotherapy

Radiotherapy plays a central role in the treatment of most brain tumors, whether benign or malignant. There are different phases where radiotherapy could be applied.

- *Radiotherapy after Surgery.* Even when it appears that the entire tumor has been surgically removed, microscopic cancer cells often remain in the surrounding brain tissue.
- *Radiotherapy when Surgery is not appropriate.* Radiotherapy may be used instead of surgery for inaccessible tumors or for tumors that have properties that are particularly responsive to radiotherapy.
- *Radiotherapy and Chemotherapy.* Combining chemotherapy with radiotherapy is beneficial in some patients with high grade tumors.
- *Stereotactic Radiosurgery.* It has been developed to allow highly targeted radiation to be delivered directly to the small tumors while avoiding healthy brain tissue. The term radiosurgery is used because the destruction is so precise that it acts almost like a surgical knife. Some studies are finding that stereotactic radiosurgery improves survival, even in patients with the highly aggressive glioblastoma multiforme brain cancer. The procedure is being tested to boost standard radiotherapy.

1.7.1.3 Applying chemotherapy

Chemotherapy can be used as a primary therapy or an additional therapy following surgery and/or radiation therapy. Chemotherapy involves the use of toxic drugs to kill cancer

cells. They may be given orally, intravenously, or administered directly into the central nervous system. Chemotherapy is not an effective initial treatment for low grade brain tumors, mostly because standard drugs cannot pass through the blood brain barrier.

Chemotherapy prolongs survival, especially in patients with anaplastic gliomas, oligodendrogliomas, medulloblastoma, primitive neuroectodermal tumors (PNET), germ cell tumors and primary CNS lymphoma. Glioblastoma multiforme though tends to become chemoresistant. As is also the case with many systemic cancers, chemotherapy of brain tumors is not curative, and the goals of the treatment are mainly to control the growth of the tumor and to maintain good performance and quality of life for the patient for as long as possible *[Burton E.C et al 2000 – Beauchesne P. 2002 - Kleinberg L. et al 2002 -Tentori L. et al 2002 -Trent S. et al 2002 -Watling C.J. et al 2002]*.

The most common drugs (agents) are the nitrosoureas (BCNU, CCNU), platinum-based drugs (Cisplatin, Cisplatinum, Carboplatin), Temozolomide, Procarbazine, and natural-occurring compounds (Taxol). Each one of these drugs is also used separately or in other combinations.

1.7.2 Novel treatment approaches

The main drawback of chemotherapy is that not all tumors are responsive to chemotherapeutic drugs, and treatment may cause damage to the bone marrow of the patient. Actively dividing cells, as in high grade tumors are most susceptible to this form of treatment. Recently, however, researchers have identified certain genetic arrangements in specific brain tumors that make them sensitive to the effects of chemotherapy. Several other promising treatment paths have been introduced. Some of them are: Immunotherapy, Gene Therapy, Angiogenesis Inhibitors, Transplantation Procedures and High-Dose Chemotherapy, and Photodynamic Therapy. Among them, gene therapy seems to be the most promising.

Gene therapy *[Lam P.Y.P. et al 2001 - Castro M.G. et al 2003]* refers to the introduction of genes into a person's DNA in order to treat brain tumors. Gene therapy is an emerging medical technique that involves the addition of DNA to the human genome in order to replace a defective gene or to provide a gene that the body can use to fight disease. Several strategies as well as other novel approaches have been attempted in the treatment of malignant gliomas *[Williams D.A. et al 2003]*.

- *delivery of prodrug-activating genes that confer sensitivity to toxic metabolites.*
- *replacement of tumor suppressor genes known to be deficient in gliomas usually resulting in tumor apoptosis.*
- *delivery of genes resulting in suppression of angiogenesis.*
- *delivery of genes resulting in activation of host antitumor immune responses.*
- *antisense cDNA delivery to negatively regulate tumor-related protein.*
- *conditionally replicating viruses that selectively infect and destroy tumor cells.*

Although these approaches significantly vary in strategy, they all share a common goal: *to deliver the therapeutic gene or virus efficiently and specifically to the targeted tissue.*

Gene therapy can be distinguished into two categories: *ex vivo*, in which cells are modified outside the body and then transplanted back, and *in vivo*, in which genes are changed in cells that are still in the body. The ex vivo approach was the first to be applied. In this approach, cells are removed from a patient's tumorous area and incubated with vectors (carriers) to introduce genes. For in vivo techniques the challenge of inserting genes is greater. Here, vectors have a more difficult task to complete. They must deliver genes to enough cells so as to have an effect. They have to remain undetected by the body's immune system and they must deliver genes into a precise spot on the genome for the body to properly produce desired proteins *[Lam P.Y.P. et al 2001]*.

Vectors *[Basillion J.P et al 2000 – Martinez A.R. et al 2002]* are mechanisms that allow genes to be carried into the genome. Modified cells are then transplanted back into their host, where it is hoped that they will replace defective genes to correct protein problems. There are two main types of vectors, namely viral and non-viral vectors as shown in Figure 1.11.

The most important viral vectors are the retroviruses and the adenoviruses. Retroviruses are small RNA-based viruses. They reproduce by integrating their RNA into a host's DNA. For gene therapy, scientists modify these viruses' genetic code so that none of their natural proteins are produced, meaning that they cannot replicate and damage a host. Because retroviruses target fast-growing cells, they are especially promising for possible cancer treatments. Adenoviruses are larger DNA-based viruses. They can hold more genes and are not limited to just targeting fast-dividing cells. However, their larger size makes them more difficult to manipulate. The viral vectors contain specifically designed viruses to carry DNA for gene therapy. Common viral vectors are the adenovirus, adeno-associated virus, Epstein–Barr virus, HSV, papova virus, vaccinia virus, retrovirus, lenti virus and hybrid virous.

A significant problem affecting all virus-based vectors relates to the recognition by the immune system. When familiar viruses are detected in the bloodstream, the body sends antibodies to bind to and consume them. A second problem relates to the unpredictability of where viruses will insert genes into a person's DNA. If genes are inserted in the wrong place, then they may not be expressed. Additionally, gene insertion could cause diseases, such as cancer, by adversely affecting the function of nearby genes. Thus, the insertion of genes using viral vectors can cause cells to behave irregularly and dangerously *[Basilion J.P. et al 2000]*.

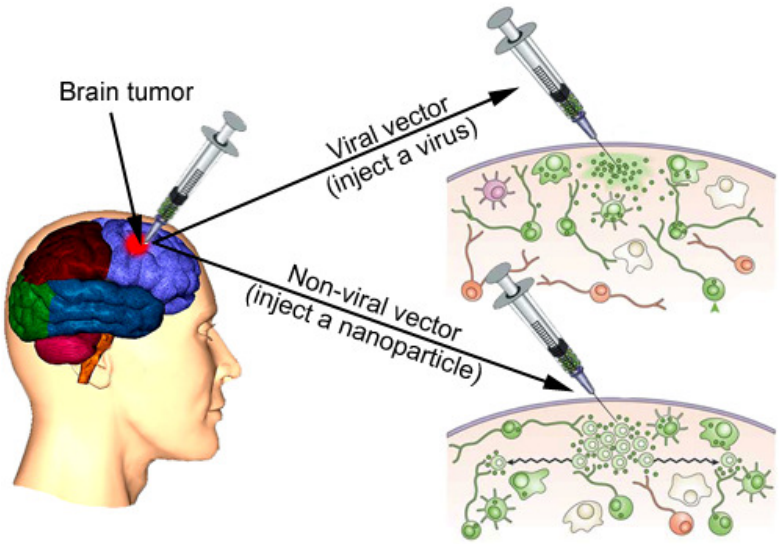


Figure 1.11 Viral (up) and non-viral (down) vectors for the treatment of a brain tumor – [Internet Source: Nature] - (The figure was designed with Adobe Photoshop ver.CS5)

Researchers are also examining non-viral vectors such as nanoparticles that can deliver therapeutic genes. Scientists are also considering introducing an extra chromosome into cells. Alongside existing DNA, this additional chromosome could contain therapeutic genes. Introduced into the body as a large vector, it should not be targeted by the immune system. Commonly used non-viral vectors are the DNA–polylysine complexes, liposomes, lipofectin, ligand-targeted liposomes and hybrid nonviral vectors.

As mentioned above, there are several strategies used in gene therapy of brain tumors. These belong to the general principles that scientists follow and shown below:

- *Immunomodulation by gene therapy*
- *Apoptosis-inducing genes*
- *Blocking angiogenesis*
- *Oncolytic viruses*

Therapeutic genes for brain tumor treatment based on the approaches above are summarized in Table 1.4 *[Kounelakis M. et al 2005]*.

Table 1.4 – Gene therapy strategies and gene targets

Strategy	***Therapeutic genes***
Immunomodulation	Cytokine genes such as *IL-2, Il-4, IL-12, GM-CSF*
Induction of Apoptosis	*TP53, CDK4,Rb, PTEN* *BAX* *hREC2* Caspase-8
Blocking Angiogenesis	Angiostatin, Endostatin
Oncolytic Viruses	HSV γ34.5 minus RR-minus Ad E1B-minus

Gene therapy cannot be considered as the “magic wand” approach in brain tumor treatment. There is a lot that has yet to be done. One of the most promising issues is the generation of even more accurate gene carriers, but also the invention of specific therapeutic genes for individualized treatment. These two issues become crucial due to the complexity of the brain compared to other tissues of the human body *[Martinez A.R. et al 2002]*.

1.8 Conclusions

The fight against brain cancer is a continuous process. Despite the fact that the brain is one of the most important organs in the human body for the sustainability of life, its structure and functionality is far more complex compared to other organs.

When a malignant brain tumor is present these considerations are always in the clinicians’ minds. Accurate and sophisticated manipulations are required to face this disease. The rapid growth of this tumor and its ability to spread is a serious obstacle to manage since it shortens the patient’s survival. Although state-of-the-art treatments are today applied to prolong patients’ survival, at the end of the day many of them will not make it.

Nowadays, most of the experts in brain oncology believe that, despite the recent advances in brain tumor treatment methods, the key issue is to see and study each patient as a unique case. This statement opens the way to a personalised treatment management where hopefully each patient will be treated according to his/her tumor’s characteristics and not following a general therapeutic protocol, as is done today. Most experts also agree that this can be achieved only when more accurate and robust diagnostic methods and protocols are developed and eventually integrated into the clinical practices. In order to proceed with an individualised therapy, reliable sets of biomarkers extracted from accurate diagnostic procedures must be defined to better describe the tumor’s type, grade and metastatic trend. Based on the expression of these biomarkers, clinicians will be able to determine a patient-specific healthcare.

The “omics” technologies and methods can be a promising way towards this direction. It is now known that brain cancer’s genesis is due to abnormal genetic mutations happening in its cells. These mutations and their impacts are mirrored in several other “omics” areas, such as transcriptomics, proteomics and metabolomics. Among them, the genomics and the metabolomics, the two ends of the “omics” era, provide the ability to obtain a clear picture of a tumor’s biological behaviour. This is because the diagnostic tools relevant to these two “omics” modalities, such as DNA microarrays and MR Spectroscopy respectively, have great potential.

The integration of the knowledge derived from these two modalities can define a more effective manner to fight this disease. The scope of this thesis is to exhibit their significant diagnostic role both for brain tumor discrimination and treatment design purposes.

References

Abdullah N.D. et al 1999, 'Contrast issues in brain tumor imaging', *Neuroimaging Clinics of North America*, vol. 9, pp. 733–749.

Balmaceda C. 2000, 'General considerations', in Rowland L.P (ed.) Merritt's Neurology, Section VI, Tumors, *Lippincott Williams & Wilkins*, Philadelphia.

Barkovich A.J. 1992, 'Neuroimaging of pediatric brain tumors', *Neurosurgery Clinical*, vol. 3, pp. 739–769.

Basilion J.P. et al 2000, 'Gene therapy of brain tumors: problems presented by physiological barriers', *Neurosurgical Focus*, vol. 8, pp. 1-7.

Beauchesne P. 2002, 'Promising survival and concomitant radiation plus temozolomide followed by adjuvant temozolomide', *Journal of Clinical Oncology*, vol. 20, pp. 3180– 3181.

Burton E.C. et al 2000, 'Malignant gliomas', *Current Treatment Options in Oncology*, vol. 1, pp. 459– 468.

Castro M.G. et al 2003, 'Current and future strategies for the treatment of malignant brain tumors', *Pharmacology & Therapeutics*, vol. 98, pp. 71– 108.

Cho W.C.S. 2010, 'An Omics Perspective on Cancer Research', *Springer-Verlag*, New York.

Cho W.C.S. et al 2007a, 'Oncoproteomics: current trends and future perspectives', *Expert Review of Proteomics*, vol. 4, pp.401-410.

Cho W.C.S. et al 2007b, 'Proteomic approaches to cancer target identification', *Drug Discovery*, *Today: Therapeutic Strategies*, vol. 4, pp. 245-250.

DeAngelis L.M. 2001, 'Brain Tumors', *New England Journal of Medicine*, vol. 344, no. 2, pp. 114-123.

Doherty G.M. 2010, 'Current diagnosis and treatment: surgery', *The McGraw-Hill Companies*, New York.

Eck S.L. et al 1996, 'Treatment of advanced CNS malignancies with the recombinant adenovirus H5.010RSVTK: a phase I trial', *Human Gene Therapy*, vol. 7, pp. 1465– 1482.

Fuller G.N. 2006, 'Molecular classifications', in Barnett G.H, (ed.) Current Clinical Oncology: High-Grade Gliomas: Diagnosis and Treatment., *Humana Press*, Totawa NJ, pp. 37-42.

Furge K.A. et al 2007, 'Identification of deregulating oncogenic pathways in renal cell carcinoma: an integrated oncogenomic approach based on gene expression-profile', *Oncogene*, vol. 26, pp.1346-1350.

He Y.D. et al 2006, 'Genomic approach to biomarker identification and its recent applications', *Cancer Biomarker*, vol. 2, pp. 103-133.

Hustinx R. et al 1999, 'SPECT and PET imaging of brain tumors', *Neuroimaging Clinics of North America*, vol. 9, pp. 751–766.

Golub T.R. et al 1999, 'Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring', *Science*, vol. 286, no. 5439, pp. 531–537.

Gray J.W. et al 2000, 'Genome changes and gene expression in human solid tumors', *Carcinogenesis*, vol. 21, no. 3, pp. 443-452.

Keusch G.T. 2006, 'What do -omics means for the science and policy of the nutritional sciences?', *American Journal of Clinical Nutrition* vol. 83, pp.5205-5225.

Kim Y.S. et al 2008, 'Frontiers in metabolomics for cancer research: Proceedings of a National Cancer Institute Workshop', *Metabolomics*, vol. 4, pp. 105-113.

Kleinberg L. et al 2002, 'Survival of patients with newly diagnosed glioblastoma multiforme treated with RSR13 and radiotherapy: results of a phase II new approaches to brain tumor therapy CNS consortium safety and efficacy study', *Journal of Clinical Oncology*, vol. 20, pp. 3149– 3155.

Kounelakis M.G. et al 2006, 'The impact of Microarray Technology in Brain Cancer', in A. F. G. Taktak A.F.G & Fisher A.C. (eds.) Outcome Prediction in Cancer, *Elsevier Publishing Company*, London, pp. 339-388.

Kounelakis M.G. et al 2005, 'The Impact of Bioinformatics in Clinical Practices on Human Cancer', *European Conference on Emergent Aspects in Clinical Data Analysis*, Pisa.

Lam P.Y.P. et al 2001, 'Potential of gene therapy for brain tumors', *Human Molecular Genetics*, vol. 10, pp. 777-787.

Levin V.A. et al 2001, 'Neoplasms of the central nervous system', in DeVita V.T. Jr, Hellman S, & Rosenberg S.A. (eds.) Cancer: Principles and Practice of Oncology, *Pa: Lippincott Williams & Wilkins*, Philadelphia, pp. 2100-2160.

Louis D.N. et al 2007, 'WHO Classification of tumors of the central nervous system', *Springer-Verlag*, Lyon.

Martinez, A.R. 2002, 'Gene therapy for cancer', *Revista de Investigacion Clinica*, vol. 54, pp.56-67.

Pomeroy S.L 2002, 'Prediction of central nervous system embryonal tumor outcome based on gene expression', *Nature*, vol. 415, no. 6870, pp. 436-42.

Ramaswamy S. et al 2002, 'DNA Microarrays in Clinical Oncology', *Journal of Clinical Oncology*, vol. 20, no. 7, pp. 1932-1941.

Segall H.D. et al 1990, 'CT and MR imaging in malignant gliomas', in: Apuzzo M.L.J (ed.) Malignant Cerebral Glioma, *American Association of Neurologic Surgeons*, Park Ridge IL, pp. 63–78.

Shand N. et al 1999, 'A phase 1 –2 clinical trial of gene therapy for recurrent glioblastoma multiforme by tumor transduction with the herpes simplex thymidine kinase gene followed by ganciclovir. GLI328 European-Canadian Study Group', *Human Gene Therapy*, vol 10, pp. 2325–2335.

Tentori L. et al. 2002, 'Pharmacological strategies to increase the antitumor activity of methylating agents', *Current Medicinal Chemistry*, vol. 9, pp. 1285– 1301.

Trent S. et al 2002, 'Temozolomide as second-line chemotherapy for relapsed gliomas', *Journal of Neurooncology*, vol. 57, pp. 247– 251.

Warburg O. 1956 'On the origin of cancer cells', *Science*, vol.123, pp. 309–314.

Watling C. J et al 2002, 'Acetazolamide therapy for symptomatic plateau waves in patients with brain tumors. Report of three cases', *Journal of Neurosurgery*, vol. 97, pp. 224– 226.

Wen P. et al 2001, 'Clinical imaging, and laboratory diagnosis of brain tumors', in: Laws E, Kaye A, (eds.) Brain tumors, *Jones and Bartlett*, New York, pp. 24–50.

Williams D.A. et al 2003, 'Gene therapy – new challenges ahead', *Science*, vol. 302, pp. 400-401.

Wishart D.S. 2007, 'Proteomics and the human metabolome project', *Expert Review of Proteomics*, vol. 4, pp. 333-335.

Yang Y.H. et al 2002, 'Design issues for cDNA Microarray experiments', *Nature Reviews Genetics*, vol. 3, no. 8, pp. 579-588.

Internet sources

American Cancer Society, http://ww.cancer.org

British Medical Journal, http://jnnp.bmj.com

Brain Supplements, http://ebrainsupplements.com

Cancer Research in UK, http://www.cancerresearchuk.org

Central Brain Tumor Registry of the United States 2011 (CBTRUS), http://www.cbtrus.org

Creative Crash, http://www.creativecrash.com

Curie Institute, http://bioinfo-out.curie.fr

DNA Sequencing, http://www.dnassequencing.com

GE Healthcare, http://www.gehealthcare.com

Med Art, http://www.med-ars.it

National Brain Tumor Society, http://www.braintumor.org

National Cancer Institute of US, http://www.cancer.gov

National Human Genome Research Institute, http://www.genome.gov

National Institute of Child Health & Human Development, http://www.nichd.nih.gov

National Instirute of Health of the US, http://www.nia.nih.gov

Nature, http://www.nature.com

Surveillance Epidemiology and End Results of National Cancer Institute of US, http://seer. cancer.gov (SEER)
Science Clarified, http://www.scienceclarified.com

Stanford Functional Genomic Facility, http://www.microarray.org

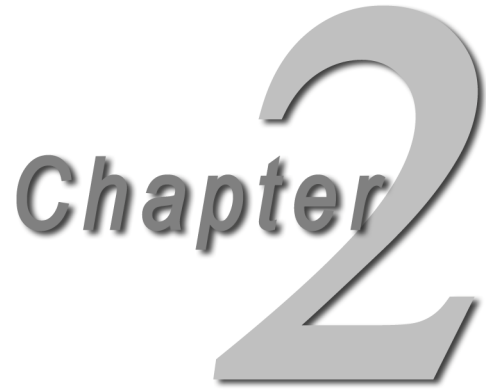
The Cancer Genome Atlas, http://cancergenome.nih.gov

The Human Metabolome Project, http://www.metabolomics.ca

World Health Organization (WHO), http://www.who.int

Wikipedia, http://en.wikipedia.org

York University, http://www.york.ac.uk



PATTERN RECOGNITION METHODS FOR BIOMEDICAL DATA MINING: an example study on Acute Myeloid Leukemia cancer

Based on Author's published studies:

- G. C. Manikis, **M. G. Kounelakis** and M. E. Zervakis, Pattern Recognition Methods for AML-Induction- Treatment Analysis, in Computational Intelligence and Bioengineering: Essays in Memory of Antonina Starita, *Frontiers in Artificial Intelligence and Applications, F. Masulli, A. Micheli & A. Sperduti (eds) – IOS Press, vol. 196, pp. 69-80, 2009*

- I. Ardoino, F. Ambrogi, S. Iacobelli, P. Boracchi, G.C. Manikis, **M.G. Kounelakis**, M.E. Zervakis, P.J.G. Lisboa, P. Fazi, M. Vignetti, E.M. Biganzoli and A. Starita, Evaluation of short- and long-term response to treatment in GIMEMA protocol for Acute Myeloid Leukaemia, *International Journal of Biomedical Engineering & Technology, vol. 3, pp. 329-348, 2010*

- G. C. Manikis, **M. G. Kounelakis** and M. E. Zervakis, Pattern Recognition Methods for Acute Myeloid Leukemia (AML) Induction-Treatment Analysis, *International Workshop on Computational Intelligence and Bioengineering, Pisa, 2009*

2.1 Introduction: The Role of Pattern Recognition in Cancer Research

Nowadays, the exhaustive research efforts being applied on cancer have produced a vast amount of patients' data. The "omics" technologies, mainly developed during the last two decades, enabled the design of large databases containing complex data of the genomic, transcriptomic, proteomic and metabolomic profiles of cancer patients. This information though is most of the times valueless if it is not further processed by statisticians and bioinformaticians who aim to extract useful and valuable knowledge which will be integrated into the diagnostic, prognostic and treatment cancer practices.

The principle of extraction or discovery of new knowledge from raw data is called data mining. Towards this direction, experts today join their efforts to develop new methods/tools that will enable oncologists and neurosurgeons to observe and study the cancer's genesis, pathology and behaviour and decide the type of treatment to follow. This is exactly the point where pattern recognition is introduced.

2.2 Fundamentals of Pattern Recognition

2.2.1 Patterns and features

Patterns are the means by which we interpret the world. Pattern recognition involves the identification of *patterns* from raw data that can reliably describe a situation. More specifically a pattern is a set of measurements or observations that characterize a situation and is usually represented in a vector or matrix form, as shown in Table 2.1. For example, a pattern in cancer field could be a patient with specific attributes such as abnormally altered gene, protein or metabolite expressions, which determine a situation like a cancer type or grade. In the same manner even a group of patients with similar attributes or clinical profile constitute a pattern of a pathological state. These measurements or attributes are often called as *features*.

Table 2.1 – A dataset of ten cancer patients distinguished into two separate classes. Each patient is described by twenty features

		FEATURES									
		Feat1	**Feat2**	**Feat3**	**……**	**…….**	**…….**	**…….**	**…….**	**…….**	**Feat20**
Class1	**Patient1**										
	Patient2										
	Patient3										
	Patient4	x	x	x	x	x	x	x	x	x	x
	Patient5										
Class2	**Patient6**										
	Patient7										
	Patient8										
	Patient9										
	Patient10										

The shaded area represents a patient vector which is a pattern of class 1. The symbol (x) corresponds to the values of the features

Pattern recognition is actually information mapping as shown, in an abstract view, in Figure 2.1 *[Schalkoff R. 1992]*. Each one of the classes in the classes's membership space C is mapped $(P\,mapping)$ to a subset of patterns in the pattern space P. These subspaces however may overlap, allowing patterns from different classes to share characteristics. Furthermore, the pattern space P is mapped $(F\,mapping)$ to the features space F or in other words, each pattern vector contains features from the F space. As it can be seen though, the same feature can belong to different pattern subspaces which can also correspond to different classes.

This is even more obvious in real life problems of pattern recognition. A trivial example is the case were two patients suffering from two different cancer types, i.e. belong to two different classes, have common characteristics (features) such as the same tumor location, or morphology, age, sex etc. It is clear that the identification of those optimal features or feature

sets that could uniquely describe a specific cancer type or grade could be the best solution to such problems.

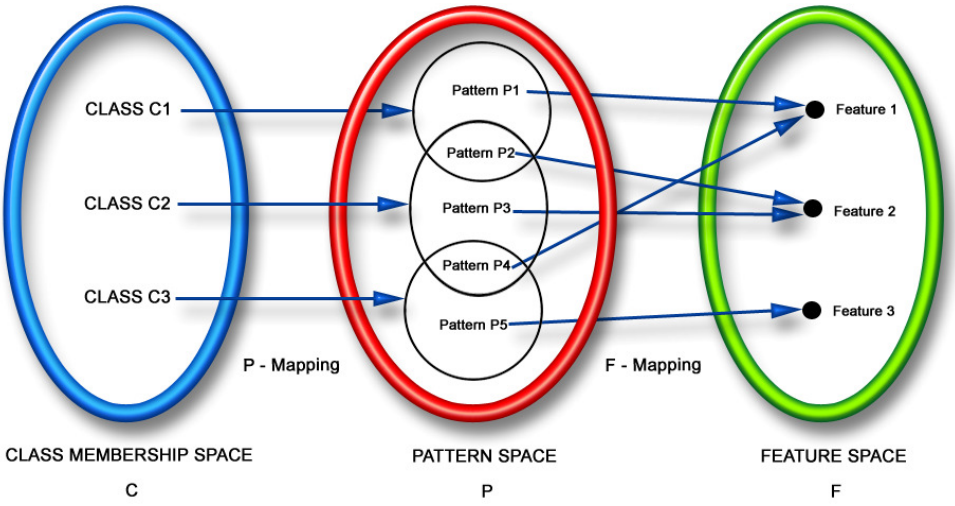


Figure 2.1 – The process of pattern recognition – (The figure was designed with Adobe Photoshop ver.CS5)

In the field of pattern recognition the aim is to learn a computer by examples to recognize patterns in datasets. The recognition process performed by a computer is also called classification. In fact this is the process where the computer categorises a new instance, for example a new patient, into a specific class (group) of patients with similar features which correspond to a specific pattern that the computer has already learned to recognise.

2.2.2 Feature selection: why is it so important in cancer research?

Feature selection, also called attribute or variable selection, is the technique of selecting a subset of relevant features for building robust learning models (classifiers). Selecting significant features is a very important issue in every data mining problem. The most obvious reason for that is the fact that often a dataset contains more information than is needed to build a robust model. Using all features involves the danger of building models that cannot recognise meaningful patterns from data and therefore are unable to accurately classify new instances. By removing the most irrelevant and redundant features from the data, the performance of the learning model is improved because of the alleviation of the effect of the curse of dimensionality which results to a much quicker learning process and enhanced generalization capability and interpretability *[Internet Source: Wikipedia]*.

Especially in cancer research the need to define optimal sets of cancer features (today called as cancer markers) is crucial both for patients and clinicians. Having reliable cancer markers as a tool, clinicians can diagnose a cancer at early stages which has a great impact on patient's follow up. Furthermore they can generate new drugs and define specific

treatment protocols which actually constitute the foundations of most medical Decision Support Systems (DSS) applied today.

2.2.2.1 Feature selection methods

Most of the methods used for feature selection were developed for classification purposes *[Geng X. et al 2007]*. Basically, feature selection methods in classification fall into three categories *[Dietterich T.G. 1998 - Guyon I. et al 2003]* referred to as *filter*, *wrapper* and *embedded* methods.

Filter feature selection is defined as a pre-processing step and can be independent from learning. Filter approaches focus on the intrinsic properties of data in each feature direction. A filter method computes a score for each feature and then selects features according to the scores *[Mladenic D. et al 1999]*. More specifically a feature ranking is applied which ranks the features based on a stochastic metric and eliminates all features that do not achieve an adequate score. The metrics typically used *[Dash M. et al 1997]* to score the features in filter selection are the,

- *distance measure,*
- *the information measure,*
- *the dependence or correlation measure,*
- *and the consistency measure*

In distance measure we are computing the physical distance between instances. Features that can support instances/records of the class to stay together are selected. The key concept is the assumption that instances of the same class must be closer than those in different class. For a two-class problem, a feature $feat_1$ is preferred to another feature $feat_2$ if $feat_1$ induces a greater difference between the two-class conditional probabilities than $feat_2$. An example is the Euclidean distance measure.

The information measure determines the information gain from a feature. The information gain from a feature $feat_1$ is defined as the difference between the prior uncertainty and expected posterior uncertainty using $feat_1$. Feature $feat_1$ is preferred to feature $feat_2$ if the information gain from feature $feat_1$ is greater than that from feature $feat_2$.

Dependence measures or correlation measures qualify the ability to predict the value of one variable from the value of another. The coefficient is a classical dependence measure and can be used to find the correlation between a feature and a class. If the correlation of feature $feat_1$ with class C is higher than the correlation of feature $feat_2$ with C, then feature $feat_1$ is preferred to $feat_2$.

Finally the consistency measures are rather new and have been in much focus recently. These measures find out the minimally sized subset that satisfies the acceptable inconsistency rate that is usually set by the user.

Characteristic tests based on the stochastic metrics mentioned, implemented by modern statistical tools, are the Fisher's discriminant criterion, Relief-f, T-statistic (Student's t-test), Chi-square statistic (X^2 statistic), Information gain, Cross-entropy measure, Person Correlation coefficient, Kruskall-Wallis test, Analysis of variance (ANOVA), Mann-Whitney U test, Liu's consistency test and others.

In contrast to filter methods which ignore the impact of the learning algorithm the second category, referred to as *wrapper*, utilizes the learning system as a black box to score subsets of features *[Kohavi R. et al 1997]*. The disadvantages of wrapper methods relate to the high computational cost of the search and their inability to take advantage of intrinsic data structures.

The third category, called the *embedded method [Breiman L. et al 1984 – Blum A. et al 1997]*, performs feature selection within the process of training. Embedded methods aim to immediately integrate the feature selection or weighting procedure into the learning algorithm of the classifier succeeding thus to retain the intrinsic characteristics of the data in the classification process. Comparing with the wrapper type, the embedded feature selection methods are usually more efficient, since they look into the structure of the involved learning model and use its properties to guide feature evaluation and search. In recent years, the embedded methods are gaining increasing interests in feature selection research due to their superior performance *[Liu H. et al 2010]*. The embedded type of feature selection is adopted in two published studies of our team presented in Chapters 4 and 5 with remarkable results in feature selection and classification of brain tumors.

In general feature selection is performed before the model is trained (learning or training phase), to automatically choose the features in a dataset that are most likely to be used in the model. A schematic representation of the feature selection and classification process is provided in Figure 2.2.

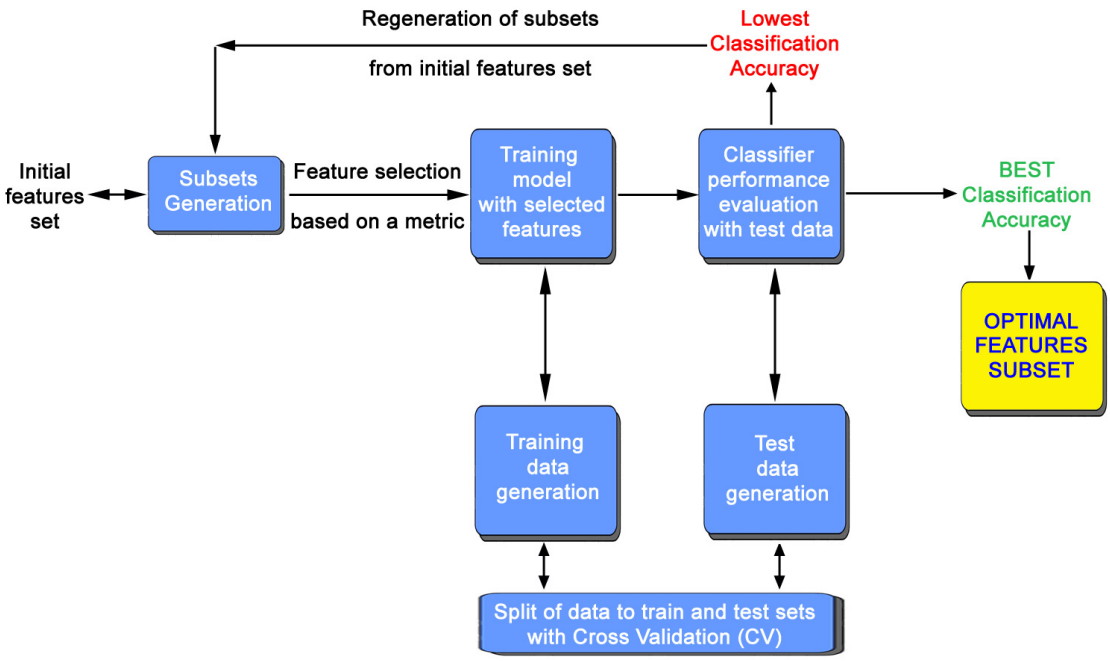


Figure 2.2 – The Feature Selection and Classification process that leads to the identification of the Optimal Features Subset – (The figure was designed with Adobe Photoshop ver.CS5)

2.2.3 Classification: linear and nonlinear classifiers

The statistical analyses applied on cancer type and/or grade discrimination studies mostly involve binary classifiers. Binary classification is dealing with the categorization of a new instance, for example a new patient, into one of two known labelled classes. The correct classification is based on classifier's prior knowledge obtained through the training phase, as shown in Figure 2.2.

Typical binary classification cases in today's oncology are the:

- *healthy vs pathological,*
- *one type of cancer vs another type of cancer* (for example Acute Myeloid Leukemia vs Acute Lymphocytic Leukemia or Brain glioma vs brain meningioma)
- *low grade cancer vs high grade cancer* (for example Brain glioma grade I vs brain glioma vs II or III or IV)

The last case of binary classification is appeared to be the most challenging one since the discrimination of subtypes of the same type of cancer is complex. Small differentiations in a patient's cancerous characteristics (features) can confuse clinicians whether the patient belongs to one subtype or to another and therefore lead to a wrong decision regarding the treatment that must be followed. For these reasons many cancer researchers during the last decade have been trying to identify significant features that are able to accurately distinguish cancer subtypes. Towards this direction two studies, published by our team, are presented in Chapters 4 and 5.

Concerning the classification of data another issue must also be addressed. That is the ability of the binary classifier to recognise whether the data provided is linearly or nonlinearly separable, as shown in Figure 2.3. The goal of statistical classification is to use an instance's characteristics to identify which class it belongs to. A linear classifier achieves this by making a classification decision based on the value of a linear combination of the characteristics. An instance's characteristics are typically presented to the classifier in a vector form called a feature vector.

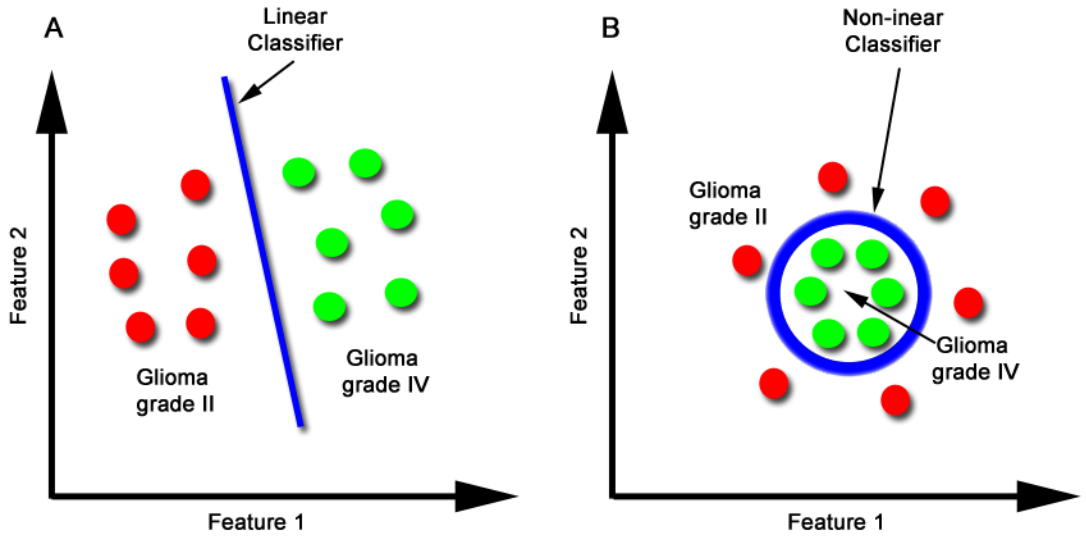


Figure 2.3 – A: A linear separable data of two glioma classes in 2D space and B: A nonlinear separable case of the same classes in 2D space – (The figure was designed with Adobe Photoshop ver.CS5)

However, in reality it is very difficult and uncommon to find data that is linearly separable. Especially in the case of cancer, the enormous amount of data produced and stored in large databases, contain many patient features whose interrelation can only be defined through a nonlinear way. Typical cases of high complexity can be found in glioma grade II vs grade III and also grade III vs grade IV discriminations where patients can present minor changes in their features. Therefore powerful nonlinear classifiers must be designed to cope with such complex cases.

2.2.3.1 Support Vector Machines (SVM)

The selection of the proper classifier is always an important issue in biomedical cancer data mining, for two main reasons:

- the great complexity and the high dimensionality (high number of features) observed in such datasets and

- the urgency to discover reliable cancer markers which will eventually be used in cancer diagnostic, prognostic and treatment procedures at clinical level.

Among the classifiers developed and applied during the last decade on cancer discrimination problems, *Support Vector Machines (SVM)* have been widely recognized as one of the most promising. SVM that first introduced Vapnik N. V. et al in 1963, belong to the maximal margin classifiers that have been found to cope quite well with complex nonlinear datasets by *[Vapnik N. V et al 1963 - Vapnik N. V. 1999]*. SVM have been an attractive choice to bioinformaticians due to their three main properties:

- the discrimination (decision) function is determined by that hyperplane which optimally separates the two classes. The optimal hyperplane is found when the boundaries of the two classes are as far as possible (maximal margin) from each other, and is placed in the middle of these boundaries, as shown in Figure 2.4.
- the decision function is defined only by the instances (vectors) and not the whole data. These vectors, which are placed on the boundaries of the two classes and therefore define these boundaries (margins), are called Support Vectors (SV), as shown in Figure 2.4.
- the nonlinearity of data can be faced through the application of the *'Kernel trick'* which maps the input features into a higher dimensional feature space where data is linearly separable, as shown in Figure 2.6.

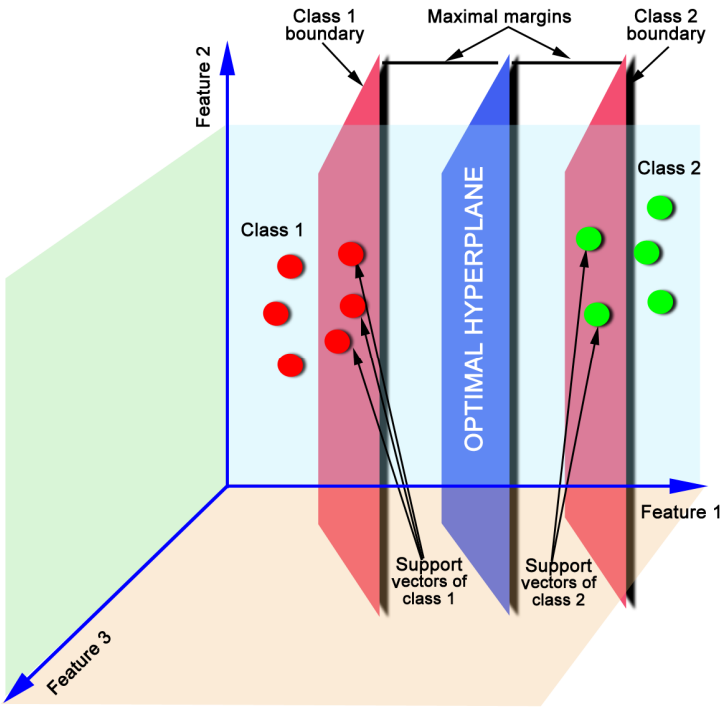


Figure 2.4 – The SVM classifier showing the optimal hyperplane (in blue), maximal margin of the two classes and the support vectors of each class, in 3D space – (The figure was designed with Adobe Photoshop ver.CS5)

Binary classification can be viewed as the task of separating two classes, usually defined as positive $(+1)$ and negative (-1) as shown in Figure 2.5(A). At diagnosis phase, when a new patient (x_{new}) has to be categorised into one of the two classes (e.g. tumor type I or tumor type II), the decision is determined from the $sign$ of the function given in Eq. (1).

$$f(x_{new}) = sign((w \cdot x_{new} + b) \quad (1)$$

where $(\cdot)$ denotes the dot product, w is the direction vector of the hyperplane and $\frac{b}{\|w\|}$ is the offset of the hyperplane from the axes' origin along the vectror w.

However, as it can be seen from Figure 2.5(B), there are many hyperplanes (w,b) that can separate the two classes and therefore the question that is generated is: *which is the optimal one?*

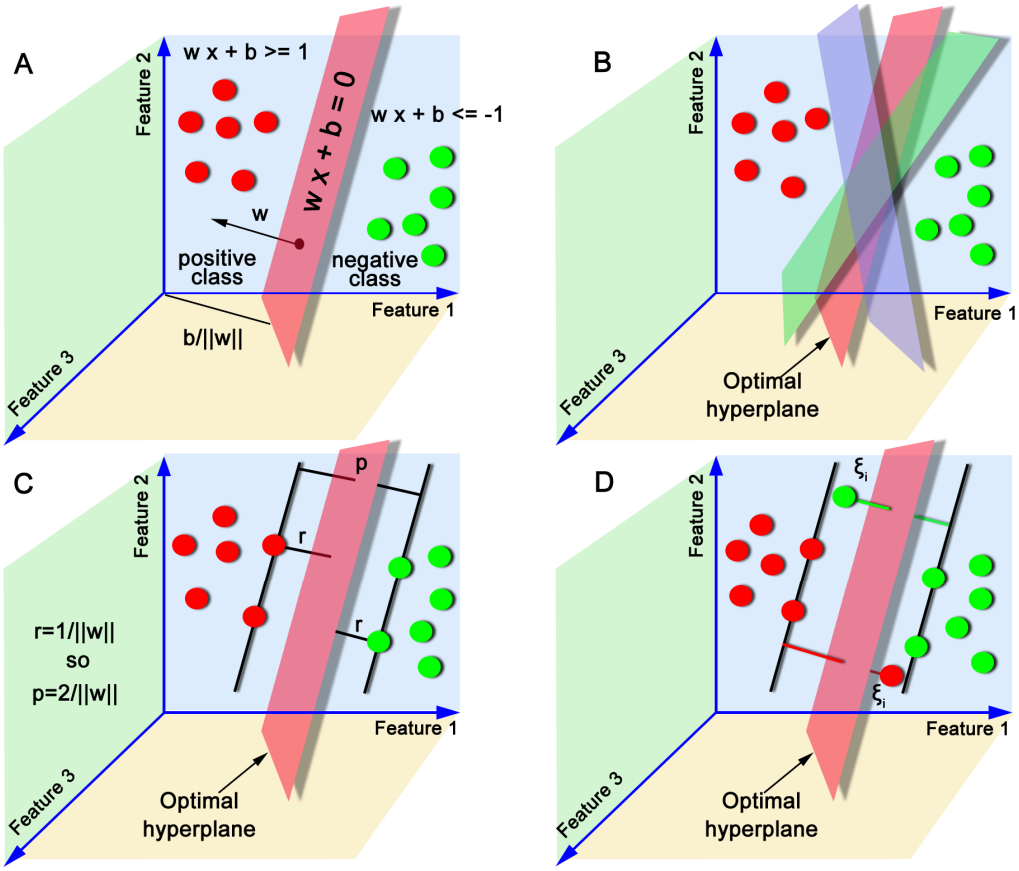


Figure 2.5 – A: Two classes (binary) classification and the separating hyperplane showing the decision function (w x + b = 0), B: Three possible hyperplanes that separate the two classes. The optimal one is in light red color, C: The distance (r) of the support vectors of the both classes from the optimal hyperplane (p=2r), D: The case of two misclassifications and the slack variable (ξi) – (The figure was designed with Adobe Photoshop ver.CS5)

The solution to that question comes from the SVM's theory which supports that the optimal hyperplane can be found by maximizing the distance between the two classes' boundaries

(margins). Then the optimal (or maximum margin) hyperplane is that one which passes through the middle of the maximum marginal distance $(p = 2r)$, as shown in Figure 2.5(C). This fact also implies that the optimal hyperplane (so optimal w, b as explained below) is defined by the support vectors. In other words, only the support vectors count in the decision function and the rest of the training data can be ignored.

To better understand this, let a training set $\{(x_i, y_i)\}$ where $i = 1...n,\ x_i \in R^d, y_i \in \{-1, +1\}$, be separated by the hyperplane with marginal distance p. Then for each training example x_i:

$$w \cdot x_i + b \geq \frac{p}{2} \text{ , if } y_i = +1 \tag{2}$$

$$\text{and } w \cdot x_i + b \leq -\frac{p}{2} \text{ if } y_i = -1 \tag{3}$$

which leads to the Eq. (5)

$$y_i (w \cdot x_i + b) \geq \frac{p}{2} \text{ for all } \{(x_i, y_i)\},\ i = 1...n, \tag{4}$$

As we also have to prevent data points from falling into the margin, we reformulate the Eq. (2), (3) and (4) as follows:

$$w \cdot x_i + b \geq +1 \text{ if } y_i = +1 \tag{5}$$

$$w \cdot x_i + b \leq -1 \text{ if } y_i = -1 \tag{6}$$

or more compactly

$$y_i (w \cdot x_i + b) \geq 1 \text{ for all } \{(x_i, y_i)\},\ i = 1...n, \tag{7}$$

In fact we require that at least one example on both sides has a distance of exactly 1.

Now let x_i^{sv} be a support vector of a class (i.e. lays on the margin of the class, so $y_i^{sv}(w \cdot x_i^{sv} + b) = 1$). Using geometry we can find that the distance from a support vector x_{sv} to the hyperplane, shown in Figure 2.5(C), is given in Eq. (8).

$$r = \frac{w \cdot x_{sv} + b}{\|w\|} = \frac{1}{\|w\|} \tag{8}$$

and therefore the marginal distance of the two classes from the hyperplane is $p = 2r = \frac{2}{\|w\|}$. The goal of SVM is to maximize p, so minimize $\|w\|$ or equivalently $\frac{1}{2}\|w\|^2$ in the equation above. In other words,

$$\min \frac{1}{2}\|w\|^2 \tag{9}$$

subject to $y_i(w \cdot x_i + b) \geq 1$ for all $\{(x_i, y_i)\}$, $i = 1...n$,

which is actually a convex Quadratic Programming (QP) optimization problem. Fortunately the original or primal optimization problem can be solved in the dual form by introducing the non-negative Lagrange multipliers $\lambda_i, i = 1...n$. Using the Lagrange multiplier the problem can be transformed as in Eq. (10).

$$\max_{\lambda} \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \lambda_i \lambda_j (x_i \cdot x_j) \tag{10}$$

subject to: $\sum_{i=1}^{n} \lambda_i y_i = 0$ for all $\lambda_i \geq 0$

From the solution of these equations it can be shown that the optimal (maximum margin) hyperplane is described by Eq. (11) and Eq. (12), only if the training examples are support vectors. This is derived from the fact that if $\lambda_i = 0$ then the training example x_i has no influence on the hyperplane (or the decision function). So, each $\lambda_i > 0$ indicates that the training example x_i is a support vector.

$$w = \sum_{i=1}^{\#sv} \lambda_i y_i x_i^{sv} \tag{11}$$

and because for any support vector $y_i^{sv}(w \cdot x_i^{sv} + b) = 1$

$$b = y_i^{sv} - w \cdot x_i^{sv} \tag{12}$$

Reformulating the Eq. (1), the decision function for a new instance (patient) in terms of Lagrange multipliers λ_i, becomes,

$$f(x_{new}) = sign\left(\sum_{i=1}^{\#sv} \lambda_i y_i x_i^{sv} \cdot x_{new} + b\right) \tag{13}$$

Notice that the final decision (or classification) function is defined in terms of the dot product $(\cdot)$ which is a very useful property of SVM as we will explain below where the *'Kernel trick'* is introduced.

The margin error

The constraints in Eq. (9) above ensure that the maximum margin classifier classifies each training example correctly, which is possible when the data is linearly separable (hard margin case). In practice, data is often not linearly separable, as shown in Figure 2.5(D); and even if it is, a greater margin can be achieved by allowing the classifier to misclassify some points.

To allow some errors, which is know as the soft margin case first introduced in 1995 *[Cortes C. et al 1995]*, we reformulate the Eq. (9) as follows,

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{14}$$

$$\text{subject to } y_i(w\cdot x_i + b) \geq 1 - \xi_i \text{ for all } i = 1...n \text{ and } \xi_i > 0$$

where ξ_i are slack variables (margin error) that allow an example to be in the margin $(0 \leq \xi_i \leq 1)$ or to be misclassified $(\xi_i > 1)$.

Again, the goal is to minimize $\frac{1}{2}\|w\|^2$ but now also penalize the possible misclassifications and margin errors. The constant $C > 0$ is a tuneable hyperplane parameter. Large values of this parameter decrease the margin of error, while small values increase it.

The Kernel trick: from a nonlinear to a linear solution

As previously mentioned, in practice the cancerous patients' datasets provided from cancer research, are most of the times not linearly separable. As explained, the main reason for that are both the complexity and the high dimensionality of the data provided from biomedical and clinical experiments. A representative case where this becomes clear is in the discrimination of complex types of cancer, such as brain tumors, with negative impact on diagnosis and treatment decision. In such cases the experts look for classifiers where they can face the problem of non linearity.

The original optimal hyperplane SVM solution described in previous section, proposed by Vapnik N. V. et al in 1963, was a linear classifier. Later though, in 1992, Boser B., Guyon I. and Vapnik N. V. suggested a way to design nonlinear classifiers by applying the *'Kernel trick'* to maximum-margin hyperplanes *[Boser B.E. et al 1992]*. The basic idea was generated from the fact that the dot product $(\cdot)$ in Eq. (13) can be replaced by a nonlinear kernel function. Using the kernel function the original input feature space where data is not linearly separable is transformed into another high dimensional feature space where a linear solution can be obtained, as shown in Figure 2.6.

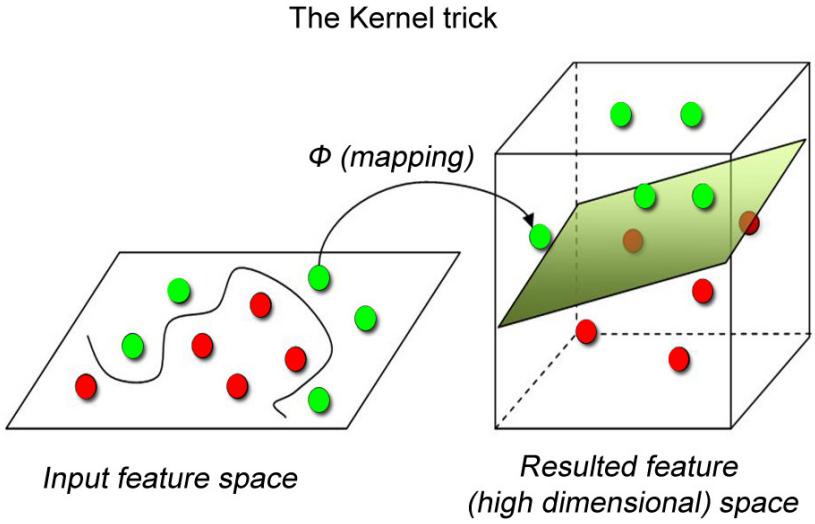


Figure 2.6 – The transformation from the input nonlinearly separable space to a higher dimensional linearly separable feature space – [Internet Source: Institute of Microbial Technology] - (The figure was designed with Adobe Photoshop ver.CS5)

A kernel function is a function that is equivalent to the dot product $(\cdot)$ in the some feature space. Therefore if every instance x_i is mapped into a high dimensional space via some transformation $\Phi : x_i \to \phi(x_i)$, the dot product becomes,

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{15}$$

Based on Eq. (15) the Eq. (13) becomes,

$$f(x_{new}) = sign(\sum_{i=1}^{\#sv} \lambda_i y_i k(x_i^{sv}, x_{new}) + b) \tag{16}$$

which is the decision function in the high dimensional feature space.

Some common Kernels are,

- *Polynomial of power* p : $K(x_i, x_j) = (x_i \cdot x_j)^p$
- *Gaussian Radial Basis Function (RBF)*: $K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2)$ *for* $\gamma > 0$
- *Sigmoid*: $K(x_i, x_j) = \tanh(kx_i \cdot x_j + c)$ *for* $k > 0$ *and* $c < 0$

In cancer discrimination tasks, the RBF kernel is commonly used due to its ability to construct classifiers that can quite satisfactory discriminate complex and heterogeneous data. Furthermore, with careful tuning of its parameter γ and the soft margin SVM's parameter C (in case of inseparable data) allows to avoid overfitting of the classifier *[Scholkopf B. et al 2002 - Wang W. 2003]*.

2.2.3.2 Least Squares Support Vector Machines (LS-SVM)

Two of the main advantages of nonlinear SVM classifiers are the ability to solve classification problems by means of convex Quadratic Programming (QP) optimization as well as the sparseness of the data as a result of this QP problem. Suykens J.A.K. et al in 1999 proposed the idea of modifying the Vapnik's SVM formulation by adding a least squares term in the cost function, as shown in Eq. (17), which transformed the problem from solving a QP problem to solving a set of linear equations *[Suykens J.A.K et al 1999]*.

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} e_i^2 \tag{17}$$

$$\text{subject to } y_i(w \cdot x_i + b) = 1 - e_i \text{ where } i = 1...n$$

As one can observe the Vapnik's initial Eq. (14) is modified in Eq. (17) in two points. First the inequality constraints are substituted by equality constraints, where the value 1 (distance fro the optimal hyperplane) at the right hand side is considered as a target value instead of a threshold value. The error variable e_i allows some tolerance of misclassification in the case of overlapping distributions. This error variables function is similar to the slack variable ξ_i in SVM formulation. Second, a squared loss function is taken for this error variable *[Selvaraj H. et al 2007]*. This approach significantly reduces the cost in complexity and computation time for solving the problem.

Applying again the Lagrange multipliers method, the solution derived is described by Eq. (18).

$$w = \sum_{i=1}^{n} \lambda_i y_i \phi(x_i) \tag{18}$$

subject to

$$\sum_{i=1}^{n} \lambda_i y_i = 0$$

$$\lambda_i = Ce_i \text{ where } C > 0$$

$$y_i(w \cdot \phi(x_i) + b) - 1 + e_i = 0$$

As explained above, an interesting property of SVM is that many of the resulting λ_i values are equal to zero. Hence the obtained solution is sparse. This means that in the resulting classifier in Eq. (13) the sum should be taken only over non-zero λ_i values, i.e. the support vectors, instead of all training data points.

In the LS-SVM's however the decision function is based on all training examples. That is because each training example is considered a support vector since λ_i is always non-zero,

based on the constraint $\lambda_i = Ce_i$ in Eq. (18). This is a major drawback of LS-SVM especially when the datasets provided are large since the solution obtained is not sparse as in SVM.

2.2.4 Classifier performance measures

A crucial point in every classification task is to validate the classifier's overall performance. For this purpose several classifier performance measures have been developed and utilized over the last decades. A performance measure estimates the classification accuracy which must objectively reflect the predictive quality of the classifier in assigning a new unseen instance to the correct class.

It is also clear that the performance of a classifier on the unseen data is directly related to the training data. That is because the classifier searches the training data to discover patterns of data which will be used as the basis for the correct classification of the new unseen instance. Therefore the selection of a suitable training set is very important for the classification process of a new unseen instance.

2.2.4.1 Selection of the training and test sets

The ultimate goal in a classification process is to get as high accuracy as possible but at the same time ensure that overfitting has been avoided. Overfitting occurs when the classifier fits too much to the training data, but cannot generalize well to new unseen data. This leads to unrealistic accuracy values i.e. low performance of the classifier.

A common way to fairly measure classification performance is by dividing the initial data into a) a training set, b) a validation set and c) a test set.

The classifier is then trained with the training set, while the validation set is used to decide when the training process must stop (i.e. when a minimal error on this validation set is reached). Finally the test data are only used after the training process has been terminated and the classification accuracy is measured.

Unfortunately in real life problems, it is quite often very difficult to have a large dataset. Especially in real life cases, such as in clinical medicine there is virtually no perfect test. With the limited available number of data, a trade-off should be taken to divide the dataset into a training set and a test set, e.g. 2/3 of the dataset are used as training data while the rest is used later as test data.

There are several statistical methods that can be utilized to help us select the proper training and test sets for the classifier. Most of them follow a Cross Validation (CV) motif. The most known are the K *-fold CV* and the *Leave-one-Out CV (LOOCV) [Kohavi R. 1995]*.

In K-fold CV the initial dataset is randomly partitioned into K folds (also called K subsets). Of the K subsets, one is retained for testing purposes while the rest $K-1$ subsets are inputted into the classifier for training. This process is repeated K times, so all K subsets are used for testing at least once. 10-fold CV is most commonly used in classification tasks.

In LOOCV method, as the name suggests, each instance of the dataset is left out for testing and the remaining instances are used to train the classifier. This process is repeated until all instances are tested at least once. LOOCV is usually very expensive from a computational point of view because of the large number of times the training process is repeated *[Internet Source: Wikipedia]*.

2.2.4.2 Correct classification rate

The correct classification rate, sometimes called accuracy, shows the proportion of correct classifications to the total number of classification tests. In medicine, this is described by four terms, the true and false positive and true and false negative. More specifically,

- *True Positive (TP)*: the test result is *positive* in the *presence* of a disease (i.e. the test is positive for the disease because the person has the disease).
- *True Negative (TN)*: the test result is *negative* in the *absence* of a disease (i.e. the test is negative for the disease because the person is healthy)
- *False Positive (FP)*: the test is *positive* in the *absence* of a disease (i.e. the test is positive for the disease but the person is healthy)
- *False Negative (FN)*: the test is *negative* in the *presence* of a disease (i.e. the test is negative for the disease but the person has the disease)

Apart from the first two terms (TP, TN) that determine the accuracy of the classifier as shown in Table 2.3, the FP and FN terms have also great impact on medical decision making. Telling to a person which is healthy that is sick (i.e. FP) has a cost, but on the other hand telling to a person that is healthy while it has a disease (i.e. FN), is life threatening. These kinds of classification errors must be eliminated.

In a tabular format, these four terms are shown in Table 2.2. Furthermore, based on these four terms some formulae have been defined to express the classification accuracy and the efficacy of the classifier. These are shown in Table 2.3.

Table 2.2 – The contingency table commonly used in medical classification tasks

	Predicted group	
Actual group	Normal	Disease
Normal	TN	FP
Disease	FN	TP

Table 2.3 – The contingency table commonly used in medical classification tasks

Term	Formulae
Sensitivity (or True positive rate)	$= \frac{TP}{TP + FN}$
Specificity (or True negative rate)	$= \frac{TN}{TN + FP}$
Likelihood ratio	$= \frac{Sensitivity}{1 - Specificity}$
Accuracy	$= \frac{TP + TN}{Total}$

2.2.4.3 Receiver operating characteristic (ROC) curve

The Receiver Operating Characteristic (ROC) analysis [*Swets J.A. 1979*] is considered an objective and highly effective technique for assessing the performance of a classifier in a binary classification task. The ROC curve is a graphical representation or better a plot of the sensitivity, also called as True positive rate (TPR) versus the 1-specificity, also called as False positive rate (FPR = 1-True negative rate). More specifically the TPR determines the proportion of the positively classified instances among all positive instances available during the test. Accordingly, the FPR defines the number of incorrectly classified instances as positive among all negative instances available during the test.

In order to draw an ROC curve all we need is these two rates. The ROC space is defined by FPR and TPR as x and y axes respectively, which actually depicts relative trade-offs between true positive and false positive. Therefore each prediction result is represented by a point on this ROC space.

The best possible prediction result (i.e. the one that gives 100% classification accuracy) is represented by a point in the upper left corner (coordinates 0,1) of the ROC space, as shown in Figure 2.7(A). This is also called a *perfect classification*.

On the other hand, a completely random guess would give a point along a diagonal line which divides the ROC space in two subspaces above and below this line. This is called the *no-discrimination line*. Points above this line are considered good classification results while points below poor results.

Based on the observations aforementioned a classifier performance index, called the area under the ROC curve (AUROC) has been introduced. In medical discrimination problems AUROC serves as a well established index of diagnostic accuracy. For example as shown on Figure 2.7(B) a test with an AUROC of 1.0 is perfectly accurate as the sensitivity is 1.0 when the specificity is 1.0 (perfect test). In contrast, a test with an AUROC of 0.0 is perfectly

inaccurate. The line segment from (0,0) to (1,1) has an area of 0.5 and is called the chance diagonal. Classification tests with an AUROC value larger than 0.5 have at least some discrimination ability. In other words the closer the AUROC reaches 1.0, the better the diagnostic test. Hence, the AUROC is independent of the cut-off points used or the prevalence of disease and is therefore a good summary measure of test accuracy.

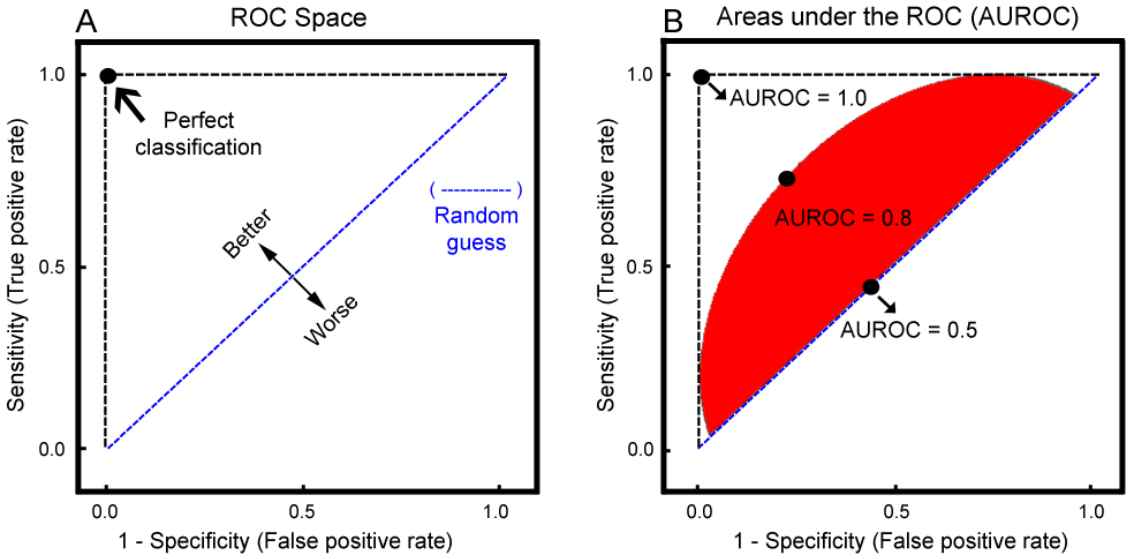


Figure 2.7 – A: The ROC space showing how classification accuracy varies B: Three different AUROC values for three different classification tests (black dots) – [Internet Source: Wikipedia] - (The figure was designed with Adobe Photoshop ver.CS5)

2.3 An Application of Pattern Recognition Methods on Acute Myeloid Leukemia: Example Study

2.3.1 Abstract

An example study which attempts to evaluate the potential of SVM and LS-SVM classifiers to discriminate Acute Myeloid Leukemia (AML) patients who a follow a treatment scheme according to GIMEMA LAM-99 protocol, has been accomplished *[Manikis G.C, Kounelakis M.G, Zervakis 2009(a) - Manikis G.C, Kounelakis M.G, Zervakis 2009(b) - Ardoino I., Manikis G.C, Kounelakis M.G, Zervakis et al 2010]*. Importance is given to feature selection and classification potential using SVMs.

The feature selection and classification process was applied using patients' features measured at the evaluation stages of two significant clinical endpoints, known as the first (Short Term) and the second (Long Term) induction treatments, (i.e. at the end of the first and second induction treatments) as shown in Figure 2.8. The results have shown that classification accuracies, in terms of AUROC, provided by both SVM and LS-SVM were quite satisfactory with LS-SVM performing slightly better than SVM. Finally the optimal feature set selected, i.e. that providing the best AUROCs, contains 9 markers (clinical, cytogenetics and molecular) and it was selected among 3 feature sets all tested for classification purposes.

This example study demonstrates the aspects of pattern recognition in terms of the final goal and the alternative decision paths that may exist in clinical diagnosis.

2.3.2 Introduction

GIMEMA (Gruppo Italiano Malattie Ematologiche dell'Adulto) is an Italian group which is expertly involved to Adult Hematological Diseases with coordination among more than 100 centers for uniform treatment protocols and data collection. One of these protocols is the LAM-99 presented in Figure 2.8 which describes the steps followed to treat the AML.

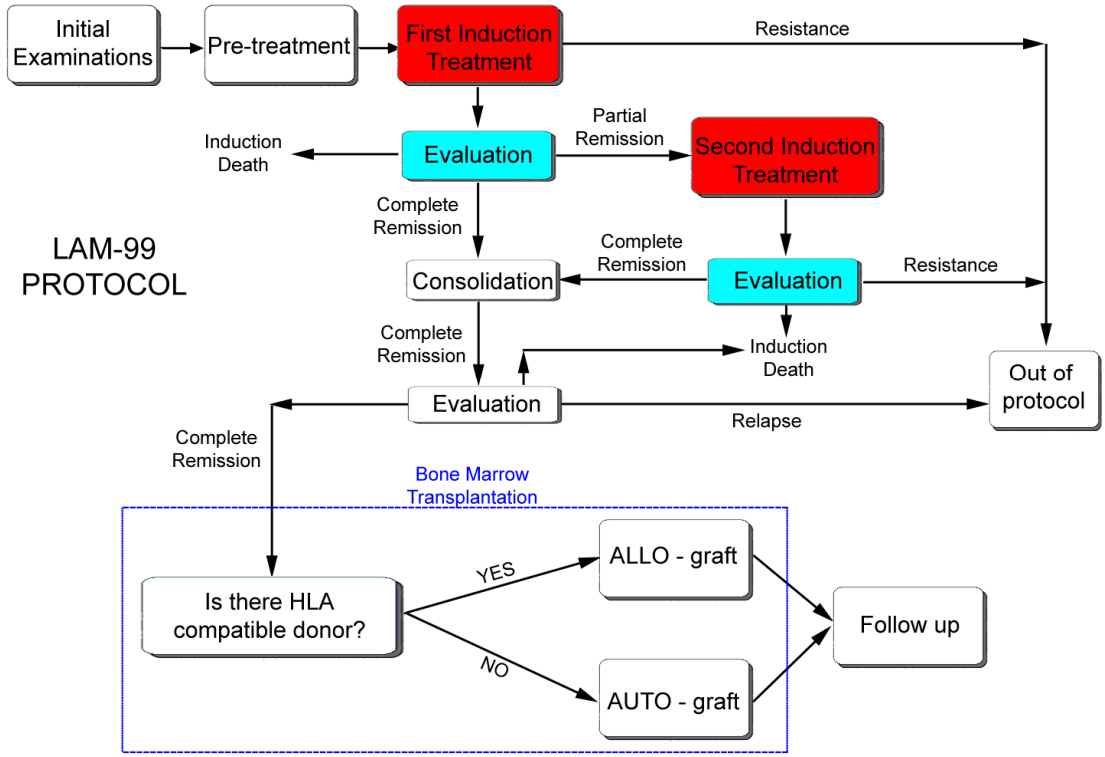


Figure 2.8 – The LAM-99 clinical protocol for AML treatment proposed by GIMEMA group. The two (first and second) induction treatments are presented in red color. The corresponding evaluations in blue color – [Source: Ardoino I. et al 2010] - (The figure was designed with Adobe Photoshop ver.CS5)

Myeloid Leukemias (ML) belong to a heterogeneous group of diseases (also known as blood cancers) characterized by uncontrolled proliferation of immature (neoplastic) white blood cells (known as blasts) which result to a serious malfunction of the hematopoietic system which is controlled by the bone marrow. The incidence of ML has been increasing over the last decades. On the basis of their clinical and pathological course, they are typically split into *Acute* or *Chronic*.

The Acute Myeloid Leukemia is considered the most aggressive form and if untreated, patients die of infection or bleeding usually in a matter of weeks. Some older adults may have a slower progressive clinical course. The symptoms of AML are caused by replacement of normal bone marrow with leukemic cells, which causes a drop in red blood cells, platelets,

and normal white blood cells. Several risk factors and chromosomal abnormalities have been identified, but the specific cause is not clear.

Historically, AML were classified based on blasts morphology and cytochemistry, that is to say the type of cell from which leukaemia developed and its degree of maturity. Classification according to WHO (World Health Organisation) and FAB (French–American–British), identifies eight major subtypes denoted *M0* through to *M7*. However, recent findings indicated that genetic abnormalities, detected by classic cytogenetics (flow cytometry), are fundamentally implicated in the leukogenesis processes. These features, associated with clinical and pathological ones, may provide useful diagnostic and prognostic information for this disease. Moreover, new integrated techniques of molecular biology allow for the evaluation of an increasing number of patients with apparently normal karyotype or where usual cytogenetics examination failed, and for the detection of further aberrations *[Ardoino I. et al 2010]*. During the last four decades, research efforts have investigated a wide variety of cytotoxic antileukemic agents. Most recently, insights into the molecular pathogenesis of AML have led to the development of the more specific targeted therapies.

GIMEMA group has managed to create a database of 509 AML patients, implementing the protocol LAM-99, whose cytogenetic and molecular data at diagnosis and, moreover, clinical and outcome data has been stored. Based on those clinical, molecular and genetic features, the aim of this study is to identify the optimal set of features that best classifies those patients after the first (*Short Term*) and the second (*Long Term*) induction treatment. The reason that our analysis focused on the Short and Long Term evaluations relies on the fact that these two stages of treatment are very crucial if we want to follow bone marrow transplantation in the case where an HLA donor is available. According to LAM-99 protocol, only if a complete remission (CR) of the disease, either at the end of first or second induction treatments is achieved, the patient will be scheduled for bone marrow transplantation. In any other case, i.e. if the outcome is partial remission (PR), resistance (Res) or induction death (ID) of the patient, this is not an option. Another reason was to assess the efficacy of the therapies applied (first and second induction treatments).

2.3.3 Materials and methods

2.3.3.1 LAM-99 protocol description

Patients were all admitted to the LAM99P trial and treated in several centres in Italy. All patients are supposed to receive for a short period a pre-treatment intended to decrease the number of abnormal white blood cells. The main purpose of the administration of the pre-treatment in the protocol was to keep the disease under control to allow a delay in the beginning of the treatment course (chemotherapy) permitting cytogenetic and molecular biology proceeding. The next stage of therapy was an intensive chemotherapy induction treatment with the primary purpose of achieving Complete Remission (CR). Patients having only a Partial Remission (PR) response were given a second induction treatment: patients

who failed both induction treatments were no longer eligible for continuing on the protocol and underwent savage therapy.

The evaluation of the response to the first induction treatment for Acute Myeloid Leukemia (AML) is scheduled between 31 and 38 days from the beginning of the induction treatment. At this stage patients are classified as CR, PR, Resistant, and dead in induction. Then, patients in PR enter a second induction treatment. The definitive evaluation of the response to the induction treatment should happen at about 80 days. This term should (could) be procrastinated at 90 days for accounting of a further shift induced by a possible cardiac or haematological toxicity.

All patients in CR, either at the end of the first or second induction treatment, start, as soon as possible, a Consolidation Treatment, necessary to eliminate non-detectable disease and prevent relapse – that is, to achieve a cure. *Allogeneic* or *Autologous Bone Morrow Transplantation* according to the availability of an Human Leukocyte Antigens (HLA) identical sibling donor were recommended for all patients following the consolidation treatment. Patients were, then, followed and events of special interest – i.e., relapse or death in remission for any cause – were monitored *[Ardoino I. et al 2010]*.

2.3.3.2 Dataset description and binary classification design

The core dataset consisted of 509 AML patients enrolled in different Italian centers partaking in the GIMEMA group. These patients were admitted to the LAM-99 protocol trial on the basis of eligibility and exclusion criteria detailed in the "*Guidelines for the treatment of AML in adults drawn on the basis of more recent European clinical studies EORTC-GIMEMA*".

A survival analysis was set up on a discrete time basis by our coordinators in Milan, following a partition of the time axis, using as a rationale for the discretisation the time points indicated by the protocol for the clinical evaluation of the response ([0, 30), [30, 60), [60, 90), [90, ..)) and then studying the hazard and/or the survival function within each interval (Figure 2.9). The curves represent the probability of the response as the first occurring event in presence of the possibility of the occurrence of the other events. The largest numbers of non-ID events are recorded between 30 and 60 days from diagnosis (at the end of the first induction treatment) and successively between 60 and 90 days (at the end of the second induction treatment). The survival analysis was performed with SPSS ver. 19.0 software package *[Internet Source: http://www.spss.com]*.

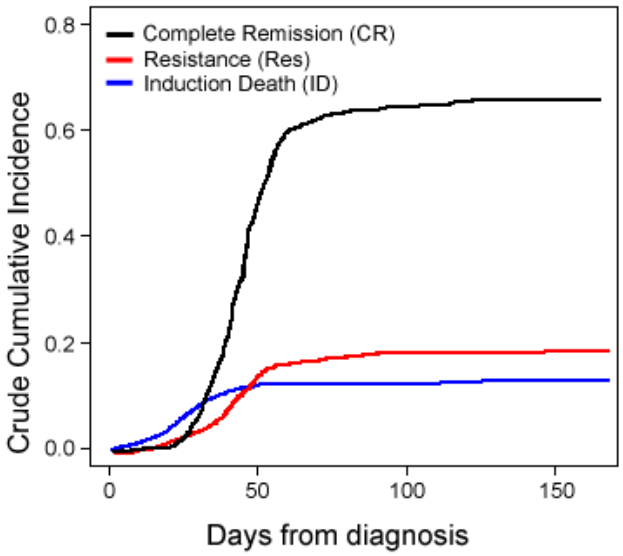


Figure 2.9 – Assessment of the response to the induction treatment – Source: [Ardoino I. et al 2010]

The survival analysis provided us with significant information about the possible classification outcome and the classes that should be chosen. Our research was based on the evaluation of the response of every patient after the first induction and second induction treatments. According to Figure 2.9, in discrete time interval [0, 30) three possible responses were recorded. These were the "Complete Remission", "Induction Death" and "Resistance". The class "Resistance" is actually the union of "Partial Remission" and "Resistant" AML patients.

The binary classification schemes decided for the Short Term Analysis was the "Induction Death" vs "all others", where "all others" is the union of responses "Complete Remission" and "Resistant". On the other hand, Long Term Analysis is the analysis when the two-cycle, when needed, induction treatment was completed (after 60 days of treatment). It is obvious from the survival curve that the two classes under examination in this term are classes "Complete Remission" and "all others", where now "all others" contains classes "Induction Death" and "Resistant".

The schemes designed for the binary classification are shown in Table 2.4 for Short Term and Long Term Analyses.

Table 2.4 – Binary classification schemes for Short Term Analyses

Analysis	***Binary schemes***	***No of Patients***
Short Term	Induction Death vs All others (CR+Res)	67 vs 442
Long Term	Complete Remission vs All others (ID+Res)	347 vs 162

2.3.3.3 Initial features selection

During initial examinations (stage of diagnosis prior to pre-treatment) a set of clinical features were measured as presented in Table 2.5.

Table 2.5 – Clinical, cytogenetic and molecular biology features at initial examinations stage

No	Features		Types of data
1	***Clinical***	WBC (White Blood Cells)	Numerical
2		PS (Performance Status)	Categorical
3		Bl_bm (no of Blasts in Bone Marrow)	Numerical
4		Hb (Hemoglobin)	Numerical
5		Plts (Blood Platelets)	Numerical
6		Exm (Extramedullary infiltration)	Categorical
7	***Cytogenetics (Cyto)***	Distinguished into three risk groups: Low risk: t(8;21), inv(16) Intermediate risk: 12p, -7/del(7q), del(9q), t(3 ;5), t(8 ;16), t(15;17), –5/del(5q), del(13q), mark, -5; -7 High risk: 3q, t(6;9), t(9;22), 11q23	Categorical
8	***Molecular***	Flt3/ITD	Binary
9		Flt3/D835	Binary
10		NPM	Binary
11		AML_ETO	Binary
12		INV16	Binary
13		FinalMLL	Binary
14		BCR_ABL	Binary
15		DEK_CAN	Binary

Cytogenetics exams were synthesised by currently defining three risk classes (lower – LR, intermediate – IR and High risk – HR) relying on Karyotype profile and prevalent abnormalities, by physicians of the GIMEMA group on the basis of literature. Molecular biology studies were also carried out, with the aim of identifying the presence of molecular abnormalities, thus increasing the sensitivity of the biological assessments.

Normalization of the features was also applied prior to classification. As it can be observed in Table 2.6 the features selected have different data types. These were normalised in the range from –1 to 1 ($mean = 0,\ \mathrm{var}iance = 1$) to improve classifier performance.

In order to investigate the discriminative power of these features under a combined scheme, three different input feature sets were designed and validated by GIMEMA's biologists and literature. These are shown in Table 2.6.

Table 2.6 – Feature sets selected for classification at Short and Long Term Analysis

1	WBC	PS	Bl_bm	Hb	Plts	Exm		
2	WBC	PS	Bl_bm	Hb	Plts	Exm	Cyto	
3	WBC	PS	Bl_bm	Hb	Plts	Exm	Cyto	Molecular

As it can be observed these feature sets were designed on an additive mode. In other words, the first set contains only the clinical features while in the second one the cytogenetic features are also added. Finally the third set contains all three types, i.e. clinical, cytogenetic and molecular.

2.3.3.4 Statistical analysis

Our study focuses on classification and optimal feature selection based on both SVM and LS-SVM. For this purpose two classification models were designed in Matlab ver. R2010 software package *[Internet Source: http://www.mathworks.com]*.

For each binary classification scheme, the entire dataset was separated into two subsets, where 80% of the data was randomly selected for training and the remaining 20% for testing the overall classification process, as shown in Figure 2.10. This procedure was repeated for 20 iterations in a 10-fold cross validation scheme and the final result of the classification was the average accuracy of all the iterations.

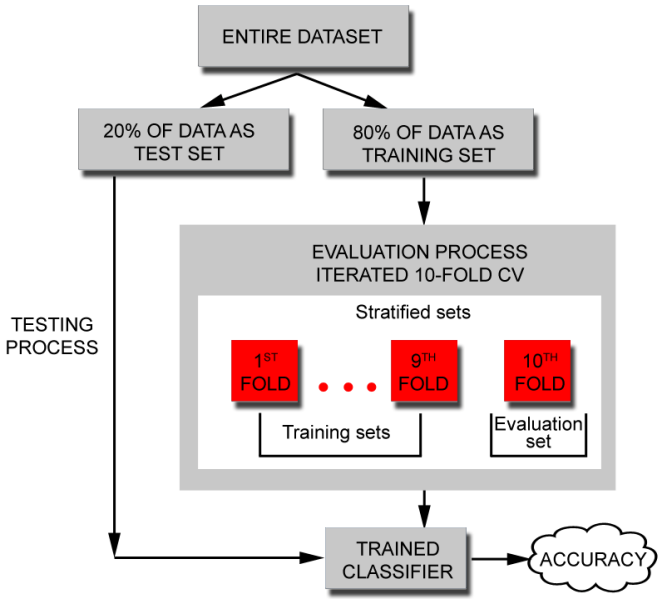


Figure 2.10 – The classification process using 10-fold cross validation - (The figure was designed with Adobe Photoshop ver.CS5)

During the evaluation process the training set (80% of the data) functioned as an evaluation criterion for adjusting the parameters of the algorithms and accessing their effectiveness to the classification. Specifically, through the 10-fold cross validation, the 9 folds trained the classifier, and the remaining fold evaluated its performance keeping out of the procedure the independent test set (20% of the data) which was used afterwards for the testing process. Furthermore, the randomly partitioned folds were also stratified so that they contain approximately the same proportions of labels as the original dataset. The Radial Basis Function (RBF) Kernel was adopted as the most suitable for our classification process. A graphical representation of the design of the training, evaluation and test sets is given in Figure 2.10.

2.3.4 Experimental results

The results from the classification approach are categorized into Short Term and Long Term Analysis results. Tables 2.7 and 2.8 show the classification accuracies obtained at Short and Long Term Analyses, from the application of both SVM and LS-SVM classifiers, in terms of sensitivity (TP), specificity (TN), false positives (FP), false negatives (FN) and AUROC values. Furthermore, Table 2.9 presents the most significant features set (or markers set), i.e. those features provided the highest average AUROC values at Short and Long Term. These features were derived from the third set of features, shown in Table 2.6, where clinical, cytogenetic and molecular values are included. Their significance was estimated in terms of their frequency of appearance at the highest AUROC values measured at both short and long term classification process.

Table 2.7 – Feature sets selected for classification at Short Term Analysis

Classifier	Binary scheme	TP	FP	TN	FN	AUROC
SVM	*Induction Death vs All others (CR+Res)*	*70%*	*45%*	*55%*	*30%*	*66%* (± 0.2)
LS-SVM		*74%*	*39%*	*61%*	*26%*	*68%* (± 0.1)

The numbers in the brackets represent the Confidence Intervals measured for each AUROC value

Table 2.8 – Feature sets selected for classification at Long Term Analysis

Classifier	Binary scheme	TP	FP	TN	FN	AUROC
SVM	*Complete Remission vs All others (ID+Res)*	*69%*	*32%*	*68%*	*31%*	*70%* (± 0.3)
LS-SVM		*73%*	*31%*	*69%*	*27%*	*71%* (± 0.2)

The numbers in the brackets represent the Confidence Intervals measured for each AUROC value

Table 2.9 – The optimal features set (markers) at Short Term Analysis

Clinical						***Cyto***	***Molecular***	
WBC	PS	Bl_bm	Hb	Plts	Exm	(High risk group genes)	Flt3/ITD	Flt3/D835

2.3.5 Discussion

Observing the classification results (Tables 2.7 and 2.8) we can clearly see that LS-SVM performs slightly better than the SVM classifier in both Short and Long Term Analyses.

As far as it concerns the optimal features identified (Table 2.9), the 5 clinical factors WBC, PS, Plts, Bl_bm and Exm were found to have a significant impact on both Short and Long Term classification outcome. In particular, the discriminative potential of WBC is greater in Short Term than in Long Term and that is because WBC is strongly associated with Resistance and Induction Death. On the other hand the Hb feature appears as less important.

Finally the contribution of Cytogenetics (especially those genes involved in the High risk group) has been crucial. Among the two molecular biology features, Flt3/ITD, Flt3/D385 the Flt3/ITD feature improves the classification outcome in the Short Term Analysis.

2.3.6 Conclusions

This study exhibits the necessity to involve modern pattern recognition methods for the evaluation of the therapeutic protocols applied today on aggressive blood cancers, such as Acute Myeloid Leukemia.

Based on the well known GIMEMA LAM-99 protocol, modern classifiers such as SVM and LS-SVM, widely adopted for binary classification problems, have been proved to be an effective tool for diagnostic and treatment purposes. Their potential to identify optimal sets of clinical, cytogenetic and molecular features that can be used to decide whether an AML patient has been benefitted from specific treatments is a key issue.

Furthermore, such supervised classification methods can be enriched in order to be used in the investigation of the behaviour of other types of solid tumors such as brain gliomas and meningiomas, as we will see at Chapters 3, 4 and 5.

References

Ardoino I., Manikis G.C, Kounelakis M.G., Zervakis M.E. et al 2010, 'Evaluation of short- and long-term response to treatment in GIMEMA protocol for Acute Myeloid Leukaemia', *International Journal of Biomedical Engineering & Technology*, vol. 3, pp. 329-348.

Blum A. et al 1997, 'Selection of relevant features and examples in machine learning', *Artificial Intelligence*, vol. 97, pp. 245–271.

Boser B.E. et al 1992, 'A training algorithm for optimal margin classifiers', *5th Annual ACM Workshop on COLT*, pp. 144–152.

Breiman L. et al 1984, 'Classification and regression trees', *Wadsworth and Brooks*, Belmont, CA.

Cortes C. et al 1995, 'Support-Vector Networks', *Machine Learning*, vol. 20, pp. 273-297.

Dash M. et al 1997, 'Feature selection for classification', *Intelligent Data Analysis*, vol. 1, pp. 131-156.

Dietterich T.G. 1998, 'Machine learning research: Four current directions', *The Artificial Intelligence Magazine*, vol. 18, pp. 97–136.

Geng X. et al 2007, 'Feature selection for ranking', *Proceedings of the 30th Annual International ACM SIGIR Conference*, pp. 407-414.

Guyon I. et al 2003, 'An introduction to variable and feature selection, *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182.

Kohavi R. 1995, 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *Proceedings of 14th International Joint Conference on Artificial Intelligence*, pp 1137-1143.

Kohavi R. et al 1997, 'Wrappers for feature subset selection', *Artificial Intelligence*, vol. 97, pp. 273–324.

Liu H. et al 2010, 'Feature selection: an ever evolving frontier in data mining, *Proceedings of the 4th International Workshop on Feature Selection in Data Mining*, vol. 10, pp. 4-13.

Manikis G.C., Kounelakis M.G and Zervakis M.E. 2009a, 'Pattern Recognition Methods for AML-Induction-Treatment Analysis', *Computational Intelligence and Bioengineering: Essays in Memory of Antonina Starita, Frontiers in Artificial Intelligence and Applications, F. Masulli, A. Micheli & A. Sperduti (eds)*, vol. 196, pp. 69-80.

Manikis G.C., Kounelakis M.G and Zervakis M.E. 2009b, Pattern Recognition Methods for Acute Myeloid Leukemia (AML) Induction-Treatment Analysis, *International Workshop on Computational Intelligence and Bioengineering*, Pisa.

Mladenic D. et al 1999, 'Feature selection for unbalanced class distribution and naïve bayes', *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 258-267.

Schalkoff R. 1992, 'Pattern recognition: statistical, structural and neural approaches,' John Wiley & Sons, New York, US.

Scholkopf B. et al 2002, Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond, *MIT Press*, Cambridge, MA.

Selvaraj H. et al 2007, 'Brain MRI Slices Classification Using Least Squares Support Vector Machines', *International Journal of Intelligent Computing*, vol.1, pp. 21-33.

Suykens J.A.K et al 1999, 'Least Squares Support Vector Machine Classifiers', *Neural Processing Letters*, vol 9, pp. 293-300.

Swets J.A. 1979, 'ROC analysis applied to the evaluation of medical imaging techniques', *Investigative Radiology*, vol. 14, pp. 109-121.

Vapnik N.V. et al 1963, 'Pattern Recognition using Generalized Portrait Method', *Automation and Remote Control*, vol. 24, pp. 774-780.

Vapnik N.V. 1999, 'The Nature of Statistical Learning Theory', *Springer –Verlag*, New York.

Wang W. 2003, 'Determination of the spread parameter in the Gaussian kernel for classification and regression, *Neurocomputing*, vol. 55, pp. 643– 663.

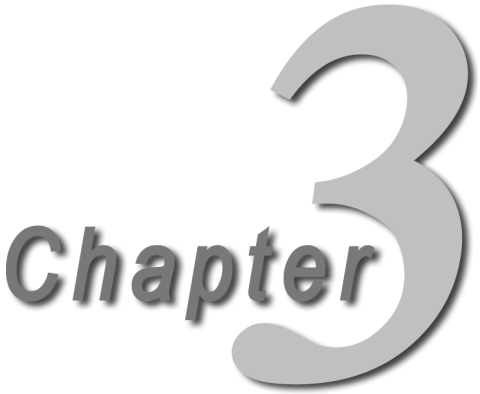
Internet sources

Institute of Microbial Technology, http://www.imtech.res.in

Matlab, http://www.mathworks.com

SPSS, http://www.spss.com

The Wikipedia, http://www.wikipedia.org



THE ROLE OF METABOLOMICS & ^{1}H (PROTON) MAGNETIC RESONANCE SPECTROSCOPY IN BRAIN CANCER MANAGEMENT: the diagnostic potential of different MRS systems

Based on Author's published studies:

- **M. G. Kounelakis**, I. N. Dimou, M. E. Zervakis, I. Tsougos, E. Tsolaki, E. Kousi, E. Kapsalaki and K. Theodorou, Strengths and Weaknesses of 1.5T and 3T MRS Data in Brain Glioma Classification*, IEEE Transactions on Information Technology in Biomedicine, vol. 15, pp. 647-654, 2011*

- I.N. Dimou, I. Tsougos, E. Tsolaki, E. Kousi, E. Kapsalaki, K. Theodorou, **M. Kounelakis** and M. Zervakis, Brain lession classification using 3T MRS spectra and paired SVM kernels, *Elsevier Biomedical Signal Processing and Control, vol. 6, pp. 314-320, 2011*

3.1 Introduction: The Promise of Metabolomics

Metabolism, (which comes from the Greek word 'metabolismos'), is the set of biochemical reactions that occur in living organisms to maintain life. These processes allow organisms to grow and reproduce, maintain their structures, and respond to their environments. Metabolism is distinguished to anabolism and catabolism (the build-up and breakdown of substances, respectively). The biochemical reactions in the human body form a metabolic network or map consisting of metabolic pathways which involve enzymes that either breaking down or building up a new chemical substance. The human's body metabolic map is shown in Figure 3.1 *[Internet Source: Wikipedia]*.

*Enzymes***,** are a type of proteins that catalyze i.e. increase the rates of chemical reactions. For example, enzymes in our digestive system break the food elements down into sugars and acids which consists our body's fuel or energy. Our body can use this energy right away, or it can store it in our body tissues, such as our liver, muscles and body fat *[Kell D.B. 2004]*.

Enzymes are crucial to metabolism because they allow organisms to drive desirable reactions that require energy. Enzymes also allow the regulation of metabolic pathways in response to changes in the cell's environment or signals from other cells. An example of a metabolic pathway, crucial for life maintenance, is the cellular respiration or cell's breathing pathway, shown in the center of Figure 3.1. This concept is analytically described in fourth chapter.

The main advantage of studying the metabolome of a living organism can be revealed through a better clarification of the difference between the genome, proteome and metabolome. As mentioned in first chapter too, the genome is the complete genetic sequence of an organism and also the blueprint for its cellular proteome, since proteome is the full set of proteins produced by a particular genome.

The genome of an organism is essentially static. The instructions for making all of an organism's cellular proteins are always there. It only changes when a genetic mutation occurs. Genes will be expressed to make proteins only when the organism or better its cells, require them. In contrast to genome the proteome continually changes in response to external and internal events.

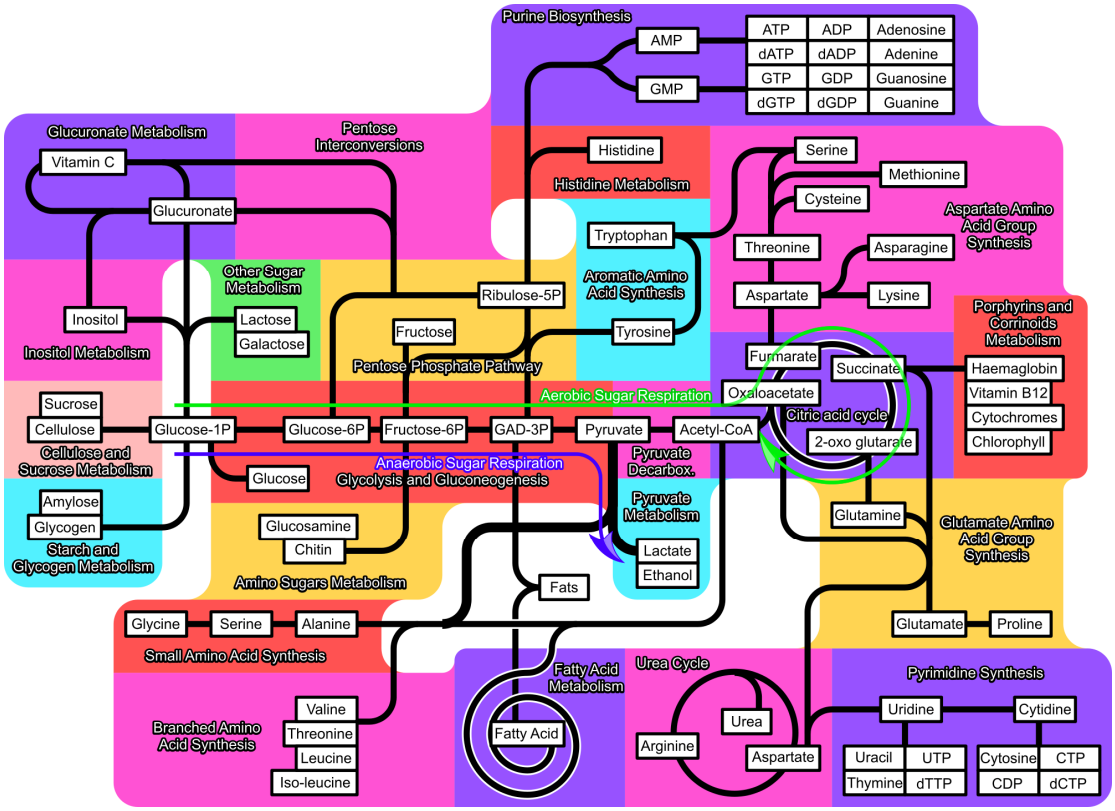


Figure 3.1 – The metabolic map of the human. The metabolic pathway of the cellular respiration is presented in the middle. The aerobic respiration (green pathway) and the anaerobic respiration (blue pathway) are the two main processes of cell's breathing – [Internet Source: Encognitive]

On the other hand, metabolome is the complete set of the metabolites produced by a single organism. Metabolites are endogenous small-molecule substances produced by the biochemical reactions of metabolism necessary for the cells of a living organism. In other words metabolites are the products of metabolism *[Internet Source: Wikipedia]*.

Like the proteome, the metabolome is closely tied to an organism's genome, but is also influenced by the genes which are expressed as well as materials that the cell can obtain from its environment, for example glucose through food. The study of metabolome enables experts to look at the relationship between an organism's genotype (genetic makeup) and phenotype (physical observable characteristics) and also the relationship between its genotype and environment *[Dunn W.B. et al 2005]*.

Metabolomics, one of the "omics" in systems biology, is the study of metabolism. More specifically it is the global assessment and validation of metabolites within a biologic system. With the advent of bioinformatics that are capable of detecting and interpreting globally the metabolites present in a biological system, biochemistry has re-emerged as a primary tool for research and discovery.

Another advantage of the study metabolome comes from the fact that the number of metabolites in human body is significant smaller than the number of genes, as shown in

Figure 3.2. *[Internet Sources: Kyoto Encyclopedia of Genes and Genome – Human Metabolome Database – Expasy – The Comprehensive Enzyme Information System]*. Therefore the statistical methods applied on metabolome for diagnostic, prognostic and treatment purposes can be implemented easier and quicker due to lower dimensionality.

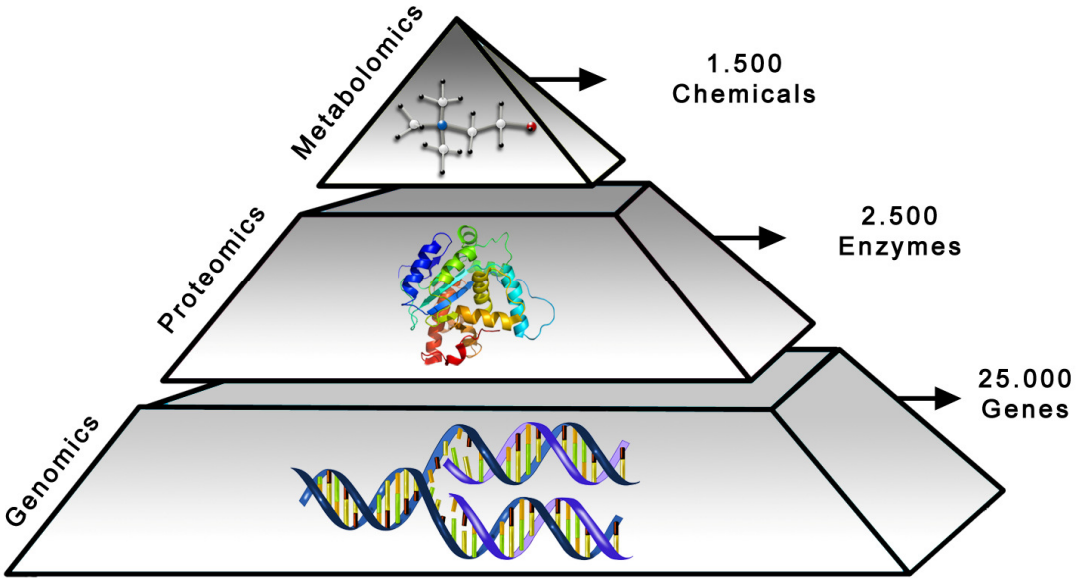


Figure 3.2 – The pyramid of life, from genomics to metabolomics – [Internet Sources: Medscape – Wikimedia - US National Institute of Health] - (The figure was designed with Adobe Photoshop ver. CS5)

Due to the dominating influence of genomic, transcriptomic and proteomic technologies on modern biological research, it is natural for research groups to approach this new field of metabolomics from a topdown point of view (i.e. gene–transcript–protein–metabolite), as explained in first chapter.

Although the term of “metabolic profile” was first introduced by Horning et al in 1971, the concept that all individuals have a metabolic profile or pattern was informally introduced by Roger Williams in the late 1940s that used paper chromatography to suggest characteristic metabolic patterns in urine and saliva were associated with diseases such as schizophrenia. However, it was only through technological advancements in the 1970s that it became feasible to quantitatively measure metabolic profiles *[Gates S.C. et al 1978]*.

Among the latest technologies, such as Mass Spectroscopy (MS) and *Nuclear Magnetic Resonance (NMR) Spectroscopy*, developed to monitor the metabolic profile of a human tissue or organ, NMR promises to revolutionize the way clinicians examine diseases, such as brain cancer. Today, Magnetic Resonance Imaging (MRI), based on the principles of the NMR technology, has become a sophisticated and powerful analytical technology that has found a variety of applications in many disciplines of scientific research, medicine and various industries.

Recent clinical applications *[Internet Source: World of Teaching, Banu H.B]* involve NMR in diagnosis and drug development for diseases arising in:

- *Brain* (for detection and discrimination of tumors, hemorrhages, infarctions)
- *Muscular skeletal system* (for demonstration of Osteomyelitis, tumor metastasis in bones and imaging of muscles, tendons and ligaments)
- *Heart* (for tomographic imaging of heart muscle, chambers and vascular structures)
- *Breast* (for detecting breast abnormalities)

3.2 Fundamentals of the MRI

3.2.1 The operation of an MRI scanner

An *MRI scan* is a radiology technique that uses magnetism, radio waves (oscillating electromagnetic fields), and a computer to produce images of body parts/organs, such as the brain, as illustrated in Figure 3.3. The technique is now called magnetic resonance imaging (MRI) rather than nuclear magnetic resonance imaging (NMRI) because of the negative connotations associated with the word '*nuclear' [Rochester Institute of Technology, Hornak J.B]*.

The basic stages of an MRI scanning are simple. First the patient is placed in a strong constant magnetic field and is surrounded by several coils. Radiofrequency (RF) radiation is then applied to the system, causing certain atoms within the patient's body to resonate. When the RF radiation is turned off, the atoms continue to resonate. Eventually, the resonating atoms return (relax) to their natural state and emit a radiofrequency radiation that is the NMR signal. This radiation is picked up by the radiofrequency coils transforming it into electrical current which is then processed through a computer and converted into a visual image of patient.

The image and resolution produced by MRI is quite detailed and can detect tiny changes of structures within the body. For some procedures though, contrast agents, such as gadolinium (Gd), are used to increase the accuracy of the images.

Generally, an MR system consists of the following components:

- *a large magnet* to generate the magnetic field,
- *a radiofrequency (RF) coil* to transmit a radio signal into the body part being imaged and a receiver coil to detect the returning radio signals,
- *gradient coils* to provide spatial localization of the signals, and
- *a computer* to reconstruct the radio signals into the final image.

[University of California - Center for Functional MRI, Hesselink J.R]

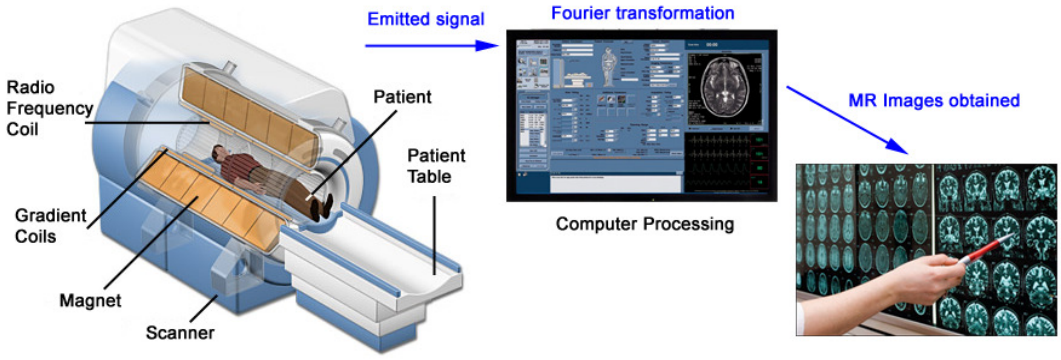


Figure 3.3 – The components in a typical MRI system and the brain MR images obtained - [Internet Sources: NASA– Mark's Psychiatry] - (The figure was designed with Adobe Photoshop ver.CS5)

3.2.2 The magnetic property of atoms

The operation of the MRI technique is based on the magnetic property of the atomic nuclei, especially the hydrogen (H) nuclei, existing within our body tissues.

It is widely known in Physics that all matter is composed of molecules, which in turn are composed of atoms. These atoms are constituted of a positively charged nucleus, made up of protons and neutrons, surrounded by negatively charged electrons, as depicted in Figure 3.4.

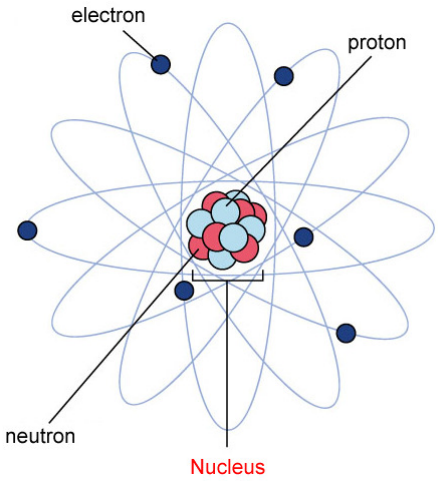


Figure 3.4 – The atomic structure - [Internet Source: PF Science] - (The figure was designed with Adobe Photoshop ver.CS5)

All elements containing an odd number of protons or neutrons, such as hydrogen, have an inherent net spin, so they possess an angular momentum. Due to the fact they are charged particles, spinning nuclei generate a small magnetic field. In other words they behave like tiny magnets. In Physics this small magnetic field is called the nuclear magnetic moment. These nuclei are usually arranged in a completely random orientation. When placed in a magnetic field such that in an MRI scanner, the nuclei (and so their protons) are forced to align with the applied magnetic field like a compass needle aligns with the Earth's magnetic field, as shown in Figure 3.5 *[Internet Source: Queensland Diagnostic Imaging]*. The magnetization of the hydrogen's nuclei which causes the alignment of the protons is fundamental property in today's MRI technology as explained next.

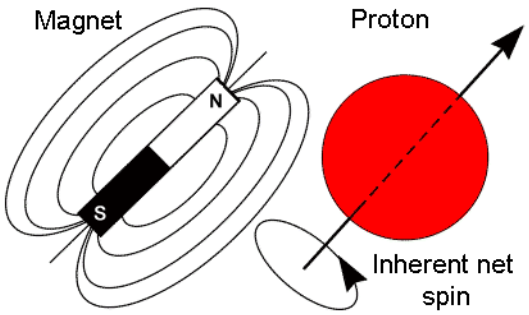


Figure 3.5 – A proton spinning about its own axis creates a magnetic charge similar to a bar magnet – [Internet Source: Wikidoc, Chao C.] - (The figure was designed with Adobe Photoshop ver.CS5)

3.2.3 The (proton) or ^{1}H-MRI

Hydrogen is a common element in nature and in human tissues due to its presence in water. Hydrogen is approximately 70% abundant in our body. For this reason but also because hydrogen nucleus contains only one proton, we use only the hydrogen proton in today's routine clinical imaging. Due to this fact the current MRI systems are called *proton or ^{1}H-MRI*.

When the body is placed in an MRI scanner of magnetic field B_0, the spinning protons of the hydrogen atoms, although arranged in random orientation (Figure 3.6(A)), will largely align along or against the field (Figure 3.6(B)). In addition to aligning with B_0, the protons will precess at some frequency called the "Larmor frequency" ω_0, Eq. (1), as shown in Figure 3.6(C). This results in a net magnetization M of the tissue, the magnitude of which is proportional to the magnitude of the external magnetic field B_0.

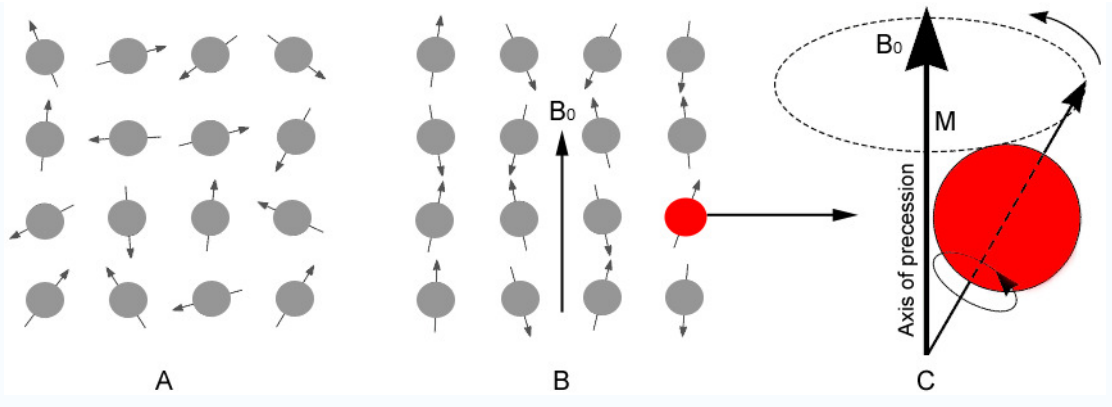


Figure 3.6 – A) The protons with random orientation – B,C) The protons are aligned to the external magnetic field B_0 and also precess about the axis of the magnetic field B_0 in a path of a cone – [Internet Source: Wikidoc, Chao C.] - (The figure was designed with Adobe Photoshop ver.CS5)

This phenomenon, called nuclear magnetic resonance (NMR), only occurs at the Larmor frequency, corresponding to the specific strength of the magnetic field. The Larmor equation tells us that the precessional frequency (or the rate of rotation about B_0) is equal to the strength of the external static magnetic field B_0 multiplied by the gyromagnetic ratio γ (ratio of proton's magnetic moment to its angular momentum). Increasing B_0 will increase the precessional frequency and conversely, decreasing B_0 will decrease the precessional frequency.

$$\omega_0 = B_0 \cdot \gamma \quad (1)$$

Then a radiofrequency pulse (which is actually a new magnetic field B_1) is transmitted onto a patient, being inside an MRI magnet, perpendicular to the direction of B_0. At that moment the hydrogen nuclei protons begin to spin 'in phase' with one another. In other words all the tiny bar magnets begin spinning together, pointing in the same direction. When the radiowave's frequency becomes equal to the Larmor frequency the hydrogen nuclei protons will absorb the maximum energy and excite.

The radiofrequency pulse is then switched off. The hydrogen nuclei protons spinning together in phase start losing their absorbed energy producing a radiowave signal of their own which is picked up by receiver coils placed around the patient. The generation of this signal is based on the well known Faraday's law of electromagnetism. The received signal, initially in time domain is then transformed (Fourier Transformation) to a signal in the frequency domain. This signal is composed of multiple frequencies, reflecting different positions along the magnetic field gradient. When the signal is broken into its component frequencies, the magnitude of the signal at each frequency is proportional to the hydrogen density at that location, thus allowing an image to be constructed *[Internet Source: E-Radiography, Bradley W.G.]*. Thus, spatial information in MRI is contained in the frequency of the signal. By using multiple pulse sequences and by varying the applied magnetic field strength by using 'gradient coils' an image can be created in two or three dimensions. A more detailed explanation of how the MR images are created is given next.

3.2.4 Creation of the MR images

As mentioned, the NMR signal is produced from the emitted energy absorbed by the hydrogen nuclei when the transmitted RF pulses (also called B_1 magnetic field), onto the tissue under examination, are turned off. The gradual release (decay) of their energy forces them to relax back to their initial state at specific relaxation rates. The initial signal produced is referred to as the Free Induction Decay (FID).

The size of the signal depends largely on four parameters. The first of these is proton density or in other words, the number of hydrogen nuclei per unit volume. It is self evident that

a magnetic resonance image will reflect the density of hydrogen in the section being displayed. The three other parameters are the spin-lattice (T1) and spin-spin (T2) relaxation times and the motion of protons *[Armstrong P. et al 1991]*.

T1 and T2 relaxation times depend on the physicochemical environment of the hydrogen protons. They actually reflect the rate at which the excited protons lose energy (their rate of relaxation). Protons lose energy by a variety of mechanisms, resulting in changes in the intensity of the signal that they produce. Every magnetic resonance image contains both T1 and T2 information, but by appropriate choice of the timing and length (scanning parameters) of the radiofrequency pulses (B_1) the image can be weighted to depend mainly on one or other of these relaxation times or to represent mainly proton density.

A *T1 relaxation (or longitudinal T1) curve* represents the time at which the excited protons realign with the initial magnetic field B_0 (Z axis) when the B_1 RF field (perpendicular to B_0) is turned off (Figure 3.7(A-D)). This curve is exponential in form, and the number T1 represents the time in milliseconds it takes for 63% of the magnetisation due to the excited protons to realign with the field, as shown in Figure 3.7(E). In images of complex structures such as the brain there are, of course, many different tissues, each with its own individual T1 curve. Differences in gray-scale in the final image reflect the difference between the heights of the T1 curves. These differences in signal intensity are referred to as T1 contrast *[Armstrong P. et al 1991]*.

T2 relaxation (or transverse T2) time reflects the rate of signal decay along the XY plane due to dephasing of the spinning protons, as shown in Figure 3.8(A-D). Once the pulse is switched off the protons start to dephase, and the rate at which they dephase is characterised by a time known as T2, representing the time it takes for 37% (100%-63%) of the magnetisation due to the spinning protons to decay due to dephasing, as shown in Figure 3.8(E). The contrast due to T2 decay in the final image depends on the difference in magnitude of the T2 curve at the point in time when the signal is collected *[Armstrong P. et al 1991]*.

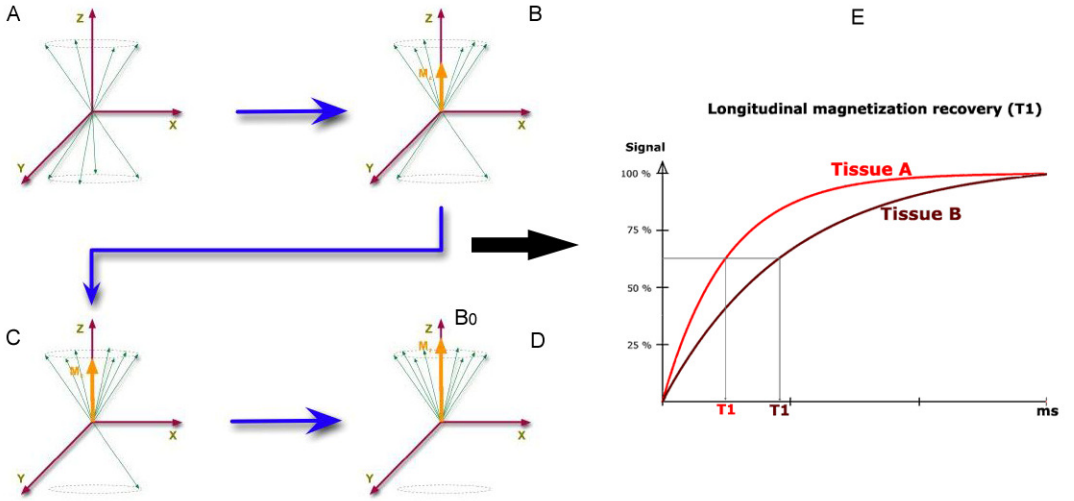


Figure 3.7 – A to D) The excited protons gradually relax to their initial magnetization B_0 along the Z axis after the RF B_1 field is switched off - E) The T1 time curve which reflects this gradual relaxation in two different tissues – [Internet Sources: Wikibooks – Imaios] - (The figure was designed with Adobe Photoshop ver.CS5)

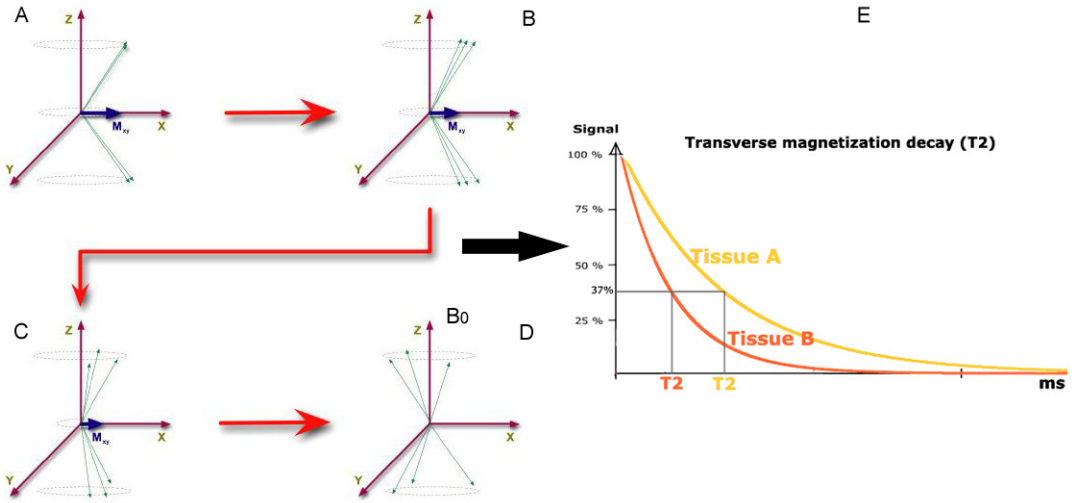


Figure 3.8 – A to D) The excited protons gradually dephase along the XY plane when the RF B_1 field is switched off - E) The T2 curve which reflects this gradual dephasing time in two different tissues - [Internet Sources: Wikibooks – Imaios] - (The figure was designed with Adobe Photoshop ver.CS5)

Different types of tissues will exhibit different T1 and T2 values, as shown in Figures 3.7(E) and 3.8(E). For example, the gray matter in the brain has a different T1 and T2 value than blood. Using the T1, T2 and proton density values, a highly resolved image can be constructed.

The *proton density (PD) image* is an image produced by controlling the selection of scanning parameters to minimize the effects of T1 and T2. This results in an image dependent primarily on the density of protons in the imaging volume. Proton density contrast

is a quantitative summary of the number of protons per unit tissue. The higher the number of protons in a given unit of tissue the brighter the signal on the proton density contrast image. Conversely, the lower the number of protons in a given unit of tissue the darker the signal on the proton density image *[Internet Source: Magnetic Resonance – Technology Information Portal]*.

Finally as mentioned, in some cases contrast agents are used to improve the magnetic resonance images obtained. Such an agent is the Gadolinium (Gd) also called diethylenetriaminepenta-acetic acid (DTPA) which is commonly used to alter the signal intensity of the organ examined. This contrast medium is either injected into the patient's vein or drunk by the patient. Although the use of gadolinium contrast agent provides better imaging result its application is not always possible. For example it is not appropriate to patients that face cardiovascular problems. Figure 3.9 illustrates the four images usually acquired during an MRI scanning procedure. In Figure 3.9(C) the borders of the main tumor area and also a new smaller tumor are clearly detected due to the use of gadolinium contrast agent.

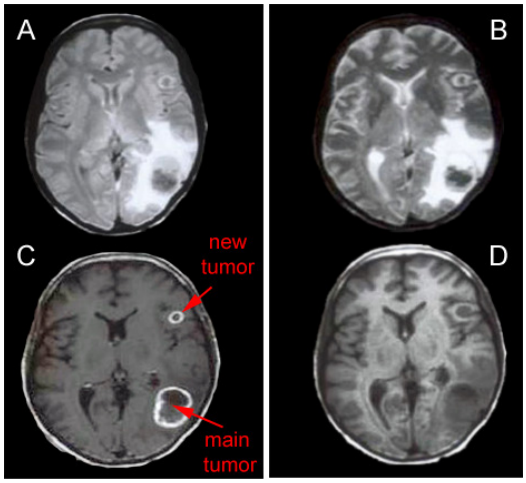


Figure 3.9 - Axial MR of a patient with glioblastoma multiforme. A) Proton density (PD) image, B) T2-weighted image, C) T1 with Gd contrast enhancement and D) T1-weighted image – [Internet Source: Michigan State University (Radiology)]

The fact that the potential information derived from the MRI signal is based on four parameters whereas that obtained from the computed tomography (CT) depends only on two (the number of atoms in a given volume and the atomic number of these atoms) renders the MRI technique more reliable in diagnosis, prognosis and treatment stages. Figure 3.10 below shows the contrast difference between two images of a glioma tumor obtained with CT (A) and MRI (B) scanning techniques respectively. As it can be observed the tumor area is much clearer in MRI image than in CT image.

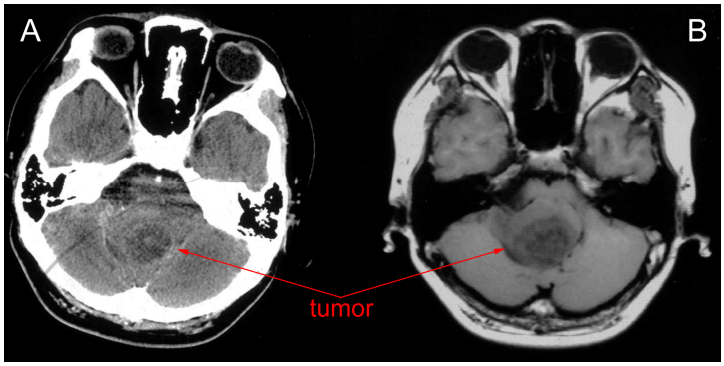


Figure 3.10 – A) A brain glioma tumor image obtained with a CT scanner B) An MRI T1-weighted image obtained from the same region – [Internet Source: Haodf]

3.2.5 Basic MRI scanning parameters

The biggest and most important component of an MRI system is the magnet. Although it is believed that strengthening the magnet will provide better quality MR images due to increased Signal-to-Noise Ratio (SNR), this is not always the case as we can see at section 3.4.The strength of a magnet in an MRI system is rated using a unit of measure known as a Tesla. Another unit of measure commonly used with magnets is the Gauss (1 Tesla = 10,000 Gauss). The magnets in use today in MRI systems create a magnetic field of 0.5-tesla to 7.0-tesla, or 5,000 to 70,000 gauss. To realize how powerful these magnets are we can just say that the Earth's magnetic field measures are 0.5 Gauss, we can understand how powerful these magnets are *[Internet Source: How MRI Works]*.

Inside the main magnet there are three gradient coils located, which produce the desired gradient (magnetic) fields. These fields are used to alter the influence of the static magnetic field B_0 on the imaged object by increasing or decreasing the field strength and changing the direction.

Apart form the magnet selection, other scanning parameters (sequence, timing and imaging parameters) must be adjusted in order to acquire the best possible MR images of the tissue under examination.

The most significant one is the RF pulse sequence (magnetic field B_1) which will be transmitted on the tissue to excite its hydrogen nuclei. A pulse sequence is a preselected set of defined RF and gradient pulses usually repeated many times during a scan, wherein the time interval between pulses and the amplitude and shape of the gradient waveforms will control NMR signal reception and affect the characteristics of the MR images. Three different types of pulse sequences can be applied in an MRI system. These are the Gradient-Echo (GE), the Spin-Echo (SE) and the Inversion Recovery (IR) sequences. The most common is the Gradient-Echo sequence which is the simplest too.

Along with the RF pulse sequence applied two timing parameters are also adjusted. The first timing parameter is called Repetition Time (TR) and the second one is called Echo Time

(TE). TR is the amount of time that exists between successive pulse sequences applied to the same region of the tissue (slice). It is delineated by initiating the first RF pulse of the sequence then repeating the same RF pulse at a time t. Variations in the value of TR have an important effect on the control of image contrast characteristics. Short values of TR (< 1000 ms) are common in images exhibiting T1 contrast, and long values of TR (> 1500 ms) are common in images exhibiting T2 contrast. TE is time represents the time in milliseconds between the application of the 90° pulse and the peak of the echo signal. In MRI, an echo is the emission of energy in form of an electromagnetic resonance signal of a nucleus after its excitation. At this point spins are back in phase again and the signal is measured. The desired number of echoes is selectable. As in TR both short (usual 20ms to 35ms) and long echo (usual 135ms to 290ms) times are used in MR image acquisition process. Selection of either short or long has a significant impact on the MR images obtained. More specifically long TE signals contain fewer components and are easier to process. However, they also provide limited information and are therefore restrictive in diagnosing some patient diseases. Hence, there is a growing interest for the processing of short echo time signals. Furthermore, in the case of the MRS signals, where the metabolic profile of the tissue is examined, analyzed in next sections, more metabolite information is contained in short echo time MRS signals. However they are also much more difficult to process.

Finally some additional imaging parameters must be determined such as the number of acquisitions or excitations (NEX), the field of view (FOV), the slice thickness and the MR image resolution (number of pixels). NEX is the number of times a sequence is repeated. The data is averaged together to create a single image with better SNR. Too many acquisitions can lead to long acquisition times and worse motion artifacts (noise) with a smaller than expected gain in SNR. The FOV specifies the area from which the MR signals are accurately sampled *[Internet Source: Magnetic Resonance - Technology Information Portal]*. An attempt to summarize the MRI parameters is given in Table 3.1 below.

Table 3.1 – Basic MRI scanning parameters

Parameters	***Types or measurement units***
Magnet's strength	Tesla (from 0.5 to even 7.0)
RF pulse sequences	Gradient-echo, Spin-echo, Inversion recovery
TR (Repetition time)	ms (short<1000ms and long>1500ms)
TE (Echo time)	ms (short: 20 to 35ms and long: 135 to 290ms)
NEX (no of excitations)	Integer number
FOV (Field of view)	mm
Slice thickness	mm
MR image resolution	pixels

When the Inversion recovery pulse sequence is used in the MRI, a new image can be obtained named Fluid Attenuated Inversion Recovery (FLAIR) image along with the conventional T1-weighted, T2-weighted, PD and Gd MR images. FLAIR is a special type of T1-wieghted image where any fluids, existing in the tissue under examination, can be eliminated. For example it can be used in brain imaging to suppress the cerebrospinal fluid (CSF) so as to bring out a better contrast of the interested regions of the tissue, such as tumor's borders, as shown in Figure 3.11.

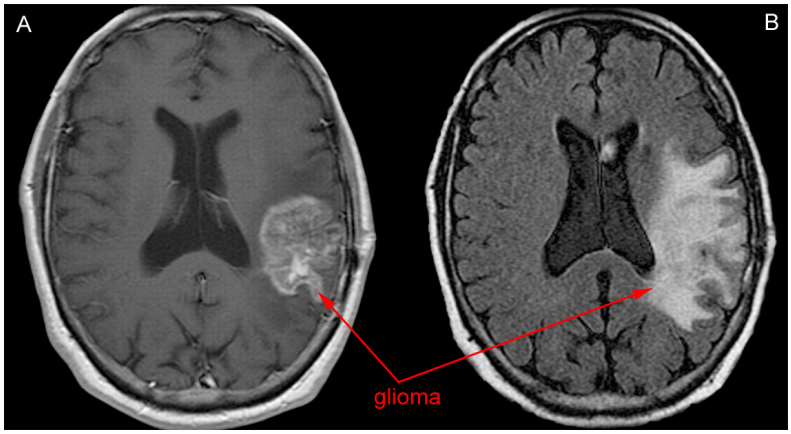


Figure 3.11 – A) A T1-weighted brain glioma MR image B) A FLAIR MR image of the same region – [Source: Heina P.A. et al 2004]

3.3 From MRI to MRS

3.3.1 What is the MR spectroscopy?

Nowadays MRI is often used for clinical diagnosis of brain tumours due to its high spatial resolution and signal-to-noise ratio of the images. This technique provides the clinician with high resolution images on which most tissue types and their morphology are clearly displayed. However, even with the precise morphologic information, it is not always able to distinguish between different tumour types or to indicate the spatial extent of the tumour *[Devos A. et al 2004]*.

A step beyond the MRI technique is the Magnetic Resonance Spectroscopy (MRS) technique which can be performed on the same MR scanner. MRS provides chemical information of specific molecules (metabolites) present in living tissue (*in vivo*) and has the potential to facilitate the characterization of tissue and in particular of human brain tumours *[Nelson S.J. 2003 – Smith I.C.P. et al 2002]*.

Although both MRI and MRS are based on the same principles in Physics, i.e. the magnetic resonance of the hydrogen nuclei protons (^{1}H-MRI and ^{1}H-MRS), the MRS can go a step further by revealing the chemical behaviour of significant metabolites whereas MRI focus only on the hydrogen molecules existing in the water within the tissue which is under examination. More specifically, MRS obtains resonance signals from molecules in the tissue and cells, whereas MRI obtains resonance signals limited to intracellular and extracellular water *[Cousins J.P. 1995]*.

As it is already known from previous sections, the signal from an MR procedure can be displayed either as a function of frequency (a spectrum) or converted from signal intensities to gray-scale values. In contrast to MRI where the evaluation of the resulted gray-scale images is mostly followed for diagnostic purposes, the MRS technique stands on the spectra obtained during the MR scanning process. In other words it is the spectra obtained from the tissue where the MRS focuses on.

3.3.2 MRS data acquisition

In MR spectroscopy, the strength of the MR signal obtained is proportional to the number of protons of the nuclei of the different molecules existing within the tissue under examination. For example in a brain tumor tissue there are many different molecules that contain different number of protons which resonate at different frequencies. Therefore the intensity of the signal at a specific frequency is proportional to the number of protons at that frequency. In other words, rather than displaying MRI proton signals on a gray scale as an image depending on its relative signal strength, MRS displays the quantities as a spectrum.

Once an MR image is obtained as a localizer image, a volume of interest (VOI) is selected. In order to acquire the spectra needed to evaluate a disease, such as a brain tumor, the radiologist is provided with two spectroscopic methods. These are the Single Voxel

Spectroscopy (SVS) and the Magnetic Resonance Spectroscopic Imaging (MRSI) also called as Chemical Shift Imaging (CSI) which is a multivoxel spectroscopic imaging. Figure 3.12 shows these two different approaches *[De Seze J. et al 2010]*.

SVS acquires one signal from a certain volume element (voxel), while MRSI acquires simultaneously signals from a two dimensional grid of voxels. In contrast to SVS, MRSI can facilitate the identification of heterogeneity of a tumourous region, since spatial variations of the tissue characteristics can be assessed at metabolite level. For each of the voxels, the intensity of the biochemically relevant metabolites can be determined *[Devos A. et al 2004]*.

As mentioned, the spectrum or spectra collected with the SVS or MRSI are based on the amount of protons in the voxel(s). The proton signals are detected and represented as a Free Induction Decay (FID). A Fourier transform is applied to the FID, converting the temporal information into frequency information. The resonant frequency is then plotted versus signal intensity on a spectrum, instead of the typical gray-scale image as is done in MRI *[Wirt M.D. 2003]*.

Therefore a spectrum is the Fourier-transformed information obtained from an MR spectroscopy study. It is presented as a series of peaks along an axis labelled in Hertz (Hz) or parts per million (ppm). The ppm scale describes the chemical shift in Hertz from a reference peak divided by the frequency of excitation *[Cousins J.P. 1995]*.

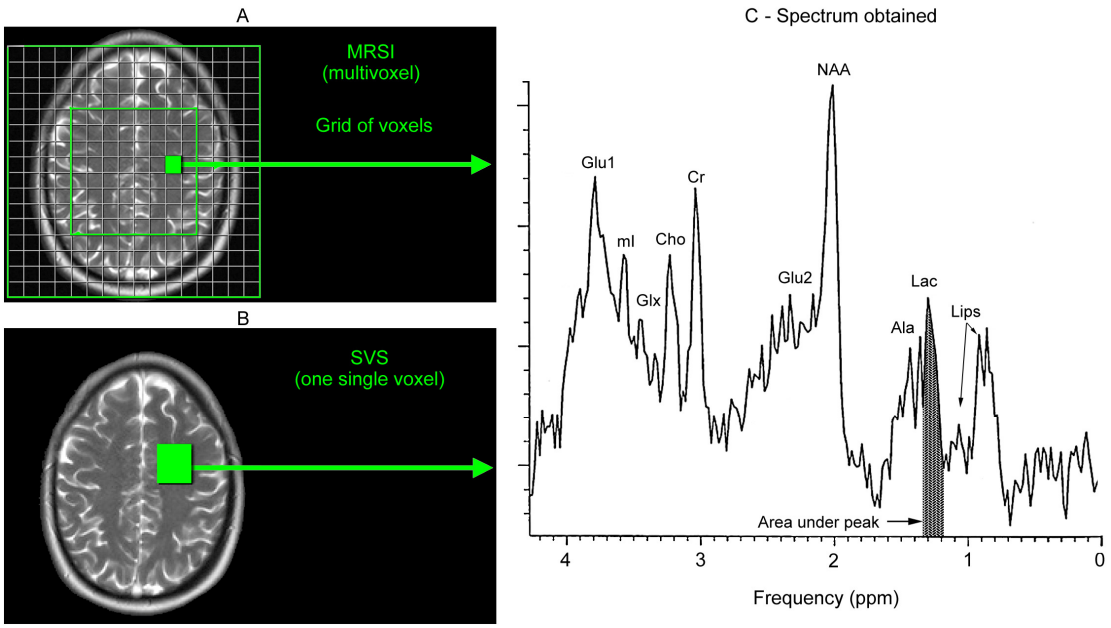


Figure 3.12 – A) The multivoxel (MRSI) technique and the grid of voxels. The inner grid shows the volume of interest (VOI) B) The single voxel (SVS) technique C) The spectrum acquired showing the main metabolites measured in the brain and the area under the peak of Lactate - [Source: De Seze J. et al 2010] - (The figure was designed with Adobe Photoshop ver.CS5)

3.3.2.1 ^{1}H-MRS pulse sequences

The pulse sequences used in the ^{1}H-MRS process are basically two. These are the Stimulated Echo Acquisition Mode (STEAM) and the Point Resolved Spectroscopy (PRESS) *[Bottomley P.A. 1987 – Frahm J. 1987]*. Nowadays these sequences can be found on most MR systems in both SVS and MRSI techniques. Both STEAM and PRESS have their advantages and limitations. Most MR studies use the STEAM sequence due to the fact that it is easier to use, it provides excellent voxel selection and can use shorter echo time (TE) values than PRESS which allows the observation of more metabolites in the spectra, as shown in Figure 3.13. Another useful characteristic of STEAM is the use of 90^{o} RF pulses for slice selection whereas PRESS uses 180^{o} RF pulses which are more difficult to implement. A disadvantage of STEAM though is the fact that it provides half the theoretical signal-to-noise ratio than PRESS *[Cousins J.P. 1995]*.

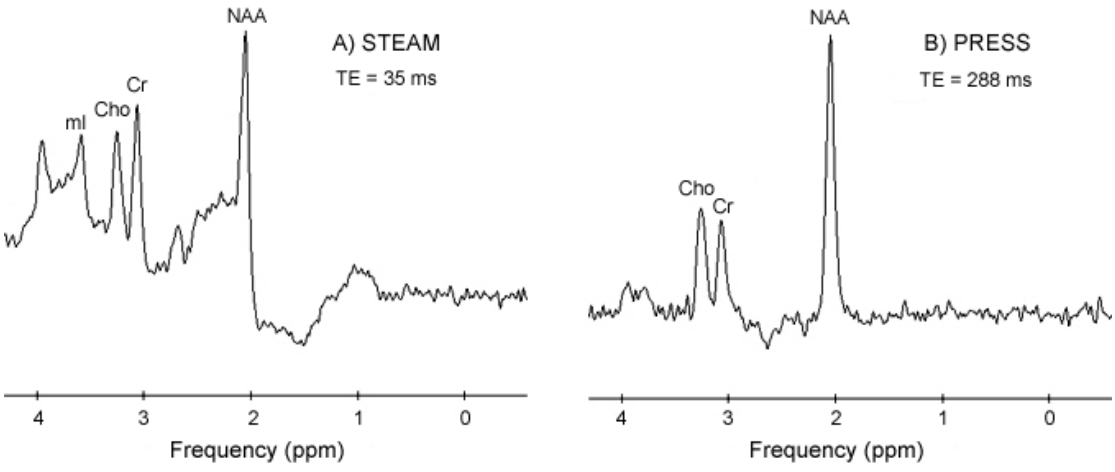


Figure 3.13 - Proton magnetic resonance spectroscopy in a normal patient. A) STEAM technique with short echo (35 ms) demonstrates normal Choline (Cho), Creatine (Cr), N-acetyl aspartate (NAA), and Myo-Inositol (mI) peaks. B) PRESS technique with a long echo (288 ms) in the same location demonstrates only Choline, Creatine, and NAA peaks. Due to the long echo time, the Myo-Inositol peak is not seen with PRESS – [Internet Source: Barrow Neurological Institute, Bohnert B.J.]

3.3.3 Significant brain metabolites

The most significant brain metabolites containing protons that can be measured during the MRS procedure are shown in Table 3.2 and in Figure 3.12(C).

N-acetyl aspartate (NAA) is a neuronal metabolite i.e. it is considered to be present only in neurons and dendrites and is reduced in brain tumors such as gliomas. Creatine (Cre) is an energy reservoir and is decreased in brain tumors. Choline (Cho), increased in tumors, is associated with glial cell membrane integrity. Glutamate (Glu1), Glutamine (Glu2) and/or Glx (Glu1/Glu2) are important in neurotransmission and are usually found increased in low grade

brain tumors. Myo-Inositol (mI) is important in cell growth and is possibly a glial cell marker which is increased in low grade brain tumors. Alanine (Ala) is elevated mostly in meningial tumors. Finally Lactate (Lac) is an indicative of anaerobic metabolism (explained in Chapter 5) and Lips (lipids) are fatty acids existing in the brain all increased in brain tumors *[Drost D.J. et al 2002 – Howe F.A. et al 2003(a) – Heerschap A. 2007]*.

Table 3.2 – The main brain metabolites measured in MRS along with their proton numbers

Metabolite's name and abbreviation	***Number of hydrogen protons***	***Resonance frequencies in descending order***
Glutamate (Glu1)	1	3.75 ppm (and in 2.20 to 2.40 ppm)
myo-Inositol (mI)	4	3.56 ppm
Glutamate/Glutamine (Glx)	1/2	3.44 ppm (and in 2.2 to 2.4 ppm)
Choline (Cho)	9	3.20 ppm
Creatine (Cre)	3	3.02 ppm
Glutamine (Glu2)	2	2.20 ppm (and in 3.75 ppm)
N-acetyl-aspartate (NAA)	3	2.02 ppm
Alanine (Ala)	3	1.48 ppm
Lactate (Lac)	3	1.33 ppm
Lips (mobile lipids)	5 (in total)	1.30 and 0.90 ppm

[Source: Govindaraju V. et al 2000] - ppm stands for parts per millions of frequency

3.3.4 Clinical and preclinical interpretation of MRS data

Today most of the ^{1}H-MRS systems' manufactures provide the MRS equipment embedded within the MR scanner facilitating the clinicians to investigate the biochemical/metabolic behaviour of brain tumors such as gliomas, meningiomas etc.

More specifically, clinicians and especially neurosurgeons are now able to combine the spectroscopic (MRS) and the MR imaging (MRI) information in order to decide the therapeutic protocol they have to follow.

At clinical diagnosis stage the combination of the MRI and MRS enables neurosurgeons to clearly identify the morphology, borders, type, grade and trend to infiltrate of the brain tumor. Based on this information they can decide whether they will proceed with a biopsy or directly plan a radiotherapy and/or surgery. Especially at this phase the MRS can be a significant tool since the true extent of the tumor tissue, not shown in conventional MRI, can be identified and possibly removed.

At preclinical stage biologists, bioinformaticians and radiologists join their efforts to provide clinicians with further information which will enrich their 'medical pharetra' with new tools to

fight this lethal disease. Towards this direction pattern recognition methods (supervised and unsupervised classification, feature selection etc) are applied on MRS data. The MRS data interpretation requires the application of state of the art pattern recognition methods in order to statistically validate the spectral information acquired and identify new sets of metabolic features (markers) which will enable experts to reveal new types, sub-types and metabolic profiles of brain tumors.

The steps followed to evaluate the diagnostic value of the MRS data obtained are described in Figure 3.14. A study which introduces pattern recognition methods to reveal the strengths and weaknesses of the spectroscopic data obtained from two different MRS systems (1.5Tesla and 3Tesla) used for brain tumor diagnostic purposes follows.

In Chapters 4 and 5, two more studies are presented showing the potential of modern feature selection and classification methods on metabolic and genomic brain tumor data.

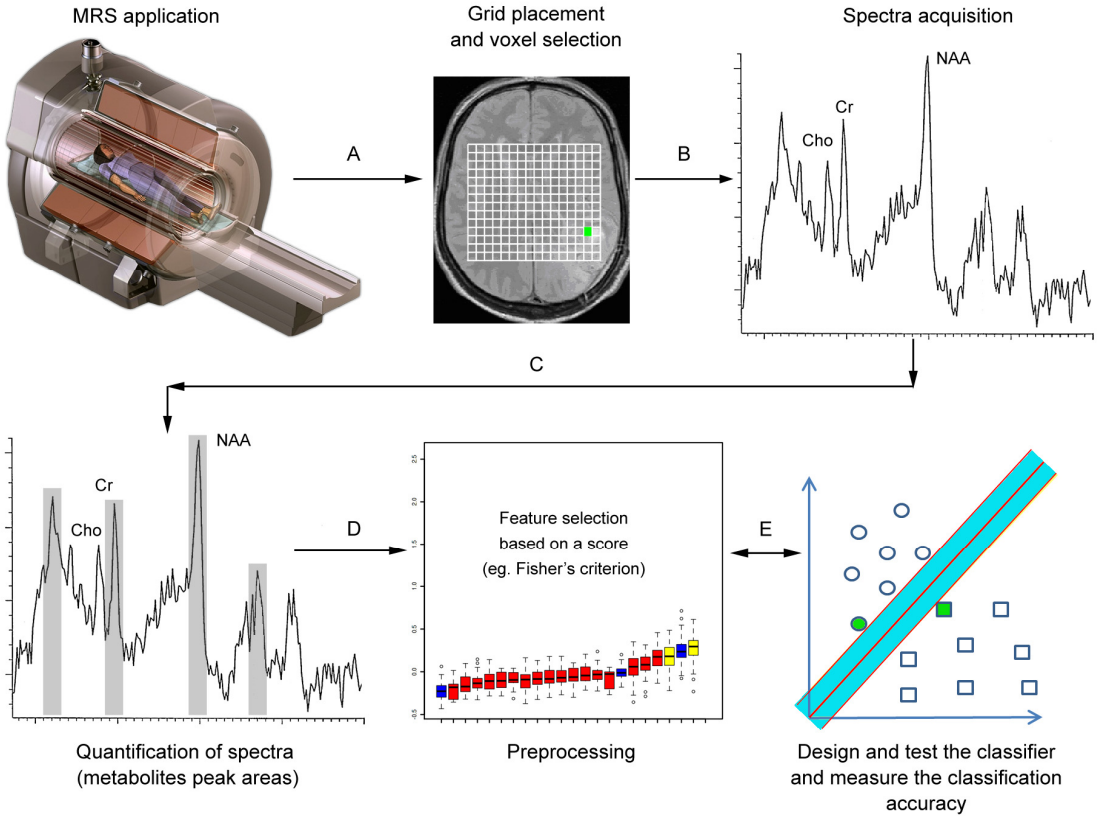


Figure 3.14 – Application of pattern recognition methods on MRS data. A) MRS application and voxel selection B) Spectra acquisition C) Quantification of spectral data D) Significant feature selection and E) Classification accuracy estimation (Note: the double arrow here denotes that feature selection can also be embedded within the classification process) - [Internet Sources: Griff Wason – Computer Science Source – Cybaea] - (The figure was designed with Adobe Photoshop ver.CS5)

3.4 Strengths and Weaknesses of 1.5T and 3T MRS Data in Brain Glioma Classification

3.4.1 Abstract

Although Magnetic Resonance Spectroscopy (MRS) methods of 1.5Tesla (T) and 3T have been widely applied during the last decade for non-invasive diagnostic purposes, only a few studies have been reported on the value of the information extracted in brain cancer discrimination. The purpose of this study is threefold. First, to show that the diagnostic value of the information extracted from two different MRS scanners of 1.5T and 3T, is significantly influenced in terms of brain gliomas discrimination. Second, to statistically evaluate the discriminative potential of publicly known metabolic ratio markers, obtained from these two types of scanners in classifying low, intermediate and high grade gliomas. Finally to, examine the diagnostic value of new metabolic ratios in the discrimination of complex glioma cases where the diagnosis is both challenging and critical.

Our analysis has shown that although the information extracted from 3T MRS scanner is expected to provide better brain gliomas discrimination, some factors like the features selected, the pulse-sequence parameters and the spectroscopic data acquisition methods, can influence the discrimination efficiency. Finally, it is shown that apart from the bibliographical known, new metabolic ratio features such as NAA/S, Cho/S, Cr/S and mI/S, play significant role in gliomas grade discrimination.

3.4.2 Introduction

Proton Magnetic Resonance Spectroscopy (^{1}H MRS) is a non-invasive technique that plays an important role in determining most brain tumor types and grade, as well as in monitoring disease progression and response to therapy *[Howe F.A. et al 2003(a) - Moller-Hartmann W. et al 2002 - Majos C. et al 2004 - Devos A. et al 2004 - Callot V. et al 2008]*. For these purposes it provides metabolic information such as cell proliferation and degradation, energy metabolism, neuronal integrity and necrotic transformation. It complements the anatomic information obtained with magnetic resonance imaging (MRI), computer tomography (CT) and angiography. One of the key issues in brain tumor discrimination though, is the diagnostic value of the information extracted from the MRS scanner which is influenced by several factors presented below.

With the advent and proliferation of strengthen MRS scanners, like a 3T one, into clinical practice, researchers have been motivated to examine whether a benefit to MRS can be achieved over those of 1.5T *[Gonen O. et al 2001 - Kim J.H. et al 2006(a) - Tanenbaum L.A. et al 2005 - Kim J.H. et al 2006(b)]*. The ability to scan with strength of 3T has multiple potential benefits. In theory, the signal-to-noise ratio (SNR) and chemical dispersion are almost doubled at 3T. The former improves spatial resolution, while the latter increases spectral resolution. The gain in SNR can be used to either improve image quality or decrease the scan time in contrast to 1.5T imaging. Scanning with 3T MRI also provides an increase in spatial and temporal resolution. This results in the ability to perform smaller field of views (FOVs) and thinner slices as a result of almost double the SNR compared to 1.5T MRI. Therefore, higher magnetic fields enable physicians to improve the accuracy of diagnosis and treatment for a broad category of diseases related to brain tumors. Findings like these are reflected in Figure 3.15(A, B) where finest metabolite structures are achieved at 3T, for example in Lactate-Lipids (LL) and myo-Inositol (mI) regions depicted by arrows. Furthermore, broad resonance dephasing in 3T results in flatter baseline aiding the more reliable estimation of major metabolite peaks.

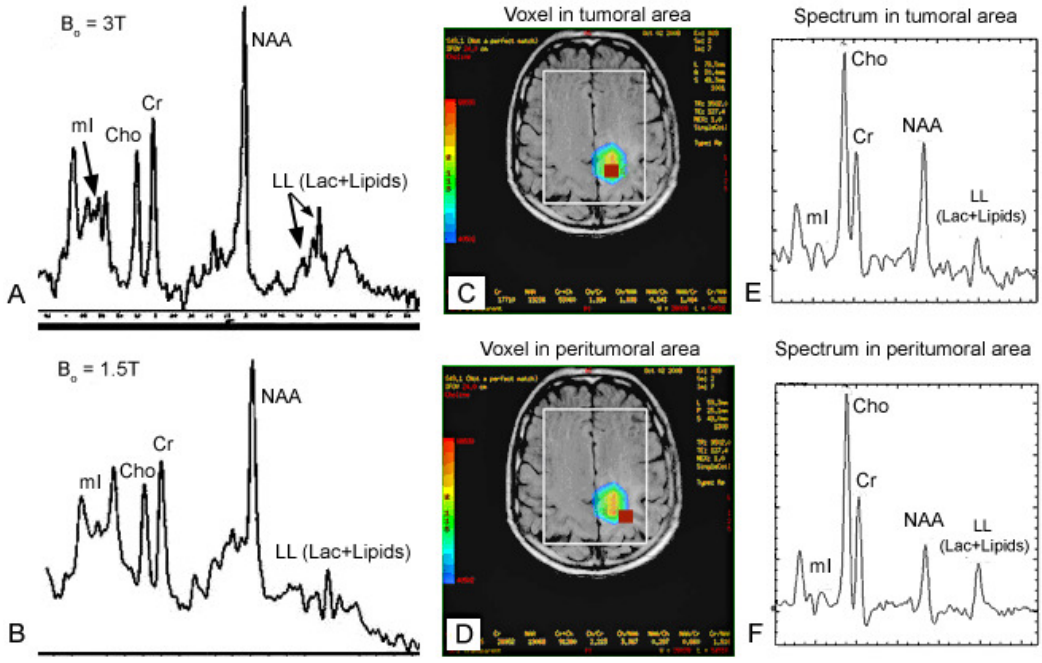


Figure 3.15 - A, B) Two spectra extracted from normal brain parenchyma at 3T and 1.5T respectively, using single-voxel technique. Note the improved spectral resolution of peaks such as mI and LL (Lactate and Lipids) (black arrows), better SNR and flatter baseline at 3T versus 1.5T. C, D) T2 FLAIR images (acquired at 3T) showing the position of voxel (gray square) placed in tumoral and peritumoral region respectively from a patient with glioblastoma multiforme. Spectra were acquired using intermediate TE =144ms, TR=1000ms and a voxel size of 225mm3. E) 1H-MRS in the tumoral region showing elevated Cho, Lac and lipid peaks and decreased NAA peak. F) 1H-MRS in the peritumoral region showing increased Cho, greater Lac and lipid peaks indicating tumor infiltration and greater decrease in the NAA peak than in tumoral region. The gray color transition inside and outside the tumoral area indicates Choline's color map. The transition from light gray (outer region of tumor) to dark gray color (central region of tumor) corresponds to the transition from low to high Choline's concentration

However, radiologists familiar with 3T MRI have cited several limitations to the increased field strength, such as a greater amount of noise, imaging contrast issues, and safety concerns. Gradient noise increases with magnetic field strength (B_0), and gradient noise at 3T can be twice that of 1.5T. This fact can lead to a fault determination of the grade and invasiveness of the tumor *[Tanenbaum L.A. et al 2005 - Kim J.H. et al 2006(b)]*. Lower-strength magnets closer to 1.5T are preferred in order to overcome this problem. Furthermore, the pulse-sequence parameters like the echo time and repetition time (TE, TR respectively) and the spectroscopic data acquisition methods also affect the power of discrimination.

Besides the MRS magnet strength to be used, another important aspect for efficient evaluation of tumour metabolic profiles and effective discrimination among tumour grades is

the selection of proper cerebral metabolites that can lead to both high classification rates and diagnostic confidence. By exploring the metabolic differences between grade I, grade II, grade III and grade IV of gliomas, it is inferred in *[Fountas K.N. et al 2004]* that grade I gliomas are characterized by a mild decrease of N-acetyl aspartate (NAA) and Creatine (Cr) and a mild increase of Choline (Cho). However, grade II gliomas present significant metabolic overlapping with grade I subtype *[Fountas K.N. et al 2004]*. Similarly, it has been suggested that the values of all metabolite ratios including Cho/Cr and Lipids/Cr overlap among each grade, particularly between grade II and III and between grade III and IV *[Kim J.H. et al 2006(a)]*. For this reason, low grades (I and II), as well as intermediate and high grades (III and IV), are often grouped to one grade for each case due to discrimination inability *[Kim J.H. et al 2006(a) - Meyerand M.E. et al 1999]*. The differences among glioma subtypes can therefore be extremely subtle, hampering the potential to effectively distinguish contiguous glioma grades as presented in Figure 3.16, especially for cases of grade I and II. Meanwhile, the metabolic profile of the peritumoral region can give valuable information about tumour type and grade *[Di Costanzo A. et al 2006 - Caprinelli G. et al 1996]*. Both high Cho and low NAA peaks in peritumoral regions are more valuable indicators, compared with those in tumoral regions, for high-grade gliomas with poor prognosis.

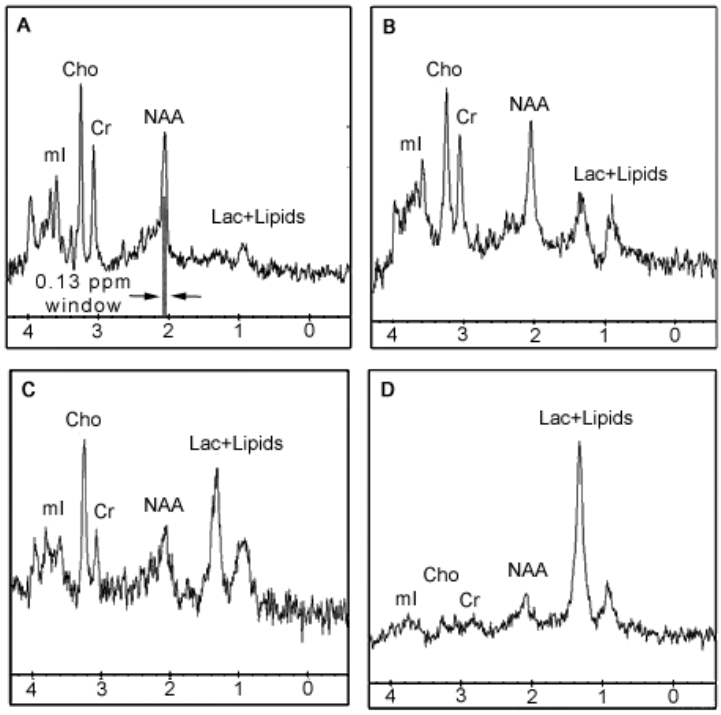


Figure 3.16 - ^{1}H-MR spectra of different glioma grades at 3T: A) Low grade astrocytoma (grade I), B) Low grade astrocytoma (grade II), C) Anaplastic astrocytoma (grade III) and D) Glioblastoma multiforme (grade IV). The shaded area in (A) represents the window width selected for the integration of the peak areas and it is used only for presentation purposes in this figure

3.4.3 Main goals and achievements of the study

Based on these observations, this study focuses on three main goals. The first is to reveal the diagnostic value of the data obtained from two different MRS scanners of 1.5T and 3T, in brain glioma discrimination *[Kounelakis M.G. et al 2011 – Dimou I.N. et al 2011]*. The second is to identify the statistical and biological significance of well known *[Kounelakis M.G. et al 2008 - Kounelakis M.G. et al 2009 - Howe F.A. et al 2003(b) - Ott D. et al 1993 - Kumar A. et al 2003 - Lehnhardt F.G. et al 2005 - Galanaud D. et al 2006 - Dimou I.N. et al 2009(a) - Dimou I.N. et al 2009(b) - Bulakbasi N. et al 2003 - Castillo M. et al 2000 - Likavcanová K. et al 2005]* ratio-type metabolic features extracted from two different datasets of 1.5T and 3T, respectively. Finally, to evaluate the diagnostic value of new metabolic ratio markers stemming from recent published studies *[Kounelakis M.G. et al 2008 - Kounelakis M.G. et al 2009]*.

The results of this study clearly emphasized the fact that despite the strength of the MR magnet applied, the spectral information obtained and therefore its discriminatory efficacy, is greatly influenced by several factors such as the volume of interest (VOI), the metabolic features selected from the spectra and also the pulse sequence parameters (TE, TR and acquisition methods) and noise. Furthermore this study identified four new metabolic markers significant in the classification of low (grade I and II), intermediate (grade III) and high grade (IV) gliomas.

3.4.4 Materials and methods

In order to compare the 1.5T and 3T MR scanners in terms of their discriminative ability, we analyze the information extracted from two different brain-glioma datasets.

3.4.4.1 Multi-center patients' demographics

The first MRS dataset from University Medical Center Nijmegen (UMCN - the Netherlands) was acquired using a 1.5T MR scanner and involved 21 glioma patients. The second dataset from Larissa University Hospital (LUH - Greece) was obtained with a 3T MR scanner and included 43 glioma patients. Details on the brain-glioma patients from both datasets are shown in Table 3.3. The 1.5T core dataset consisted of 303 pre-processed spectra sets from 303 voxels (volume elements). It included 10 patients of low grade (class II), 4 intermediate grade (class III) and 7 of high grade gliomas (class IV). The low grade gliomas consisted of diffuse astrocytomas and oligodendrogliomas, the intermediate ones of anaplastic oligodendrogliomas and the high-grade gliomas were glioblastomas multiforme. Each patient case had passed strict quality control and validation procedures, including consensus histopathologic determination. Per tissue type, the voxels are taken from homogeneous intratumoral regions.

The dataset collected from the 3T consisted of 43 glioma patients as shown in Table 3.3 where one voxel per patient from the intratumoral region was picked for further analysis. Particularly, 15 tumours were classified as low grade gliomas (classes I and II), 8 lesions as intermediate grade (class III) and 20 lesions were classified as high grade gliomas (class IV). Low grade gliomas (class I) consisted of 9 astrocytomas and those of class II consisted of 5 astrocytomas and 1 oligodendroglioma. The intermediate grade consisted of 8 anaplastic astrocytomas and the high grade included 20 glioblastomas. Glioma grades were verified in histopathological terms by means of either stereotactic biopsy or surgical resection on 30 out of 43 patients. When biopsy was performed, the location was chosen similar to the voxel placement in ^{1}H-MRS. The rest of gliomas were diagnosed by an experienced neuroradiologist (follow up). All patients gave a written informed consent to participate in the study.

The subjects enrolled in this study were examined before any surgical operation. During voxel localization, inclusion of obvious necrosis, cyst, haemorrhage, edema, calcification and normal appearing brain tissue in the voxel was omitted, to avoid lesion's underestimation. Thus volumes of interest (VOIs) with potential contamination with cerebrospinal fluid, subcutaneous fat, or eye motion were excluded from analysis. In order to facilitate the statistical analysis but also in accordance to the bibliography, the two grades I and II at 3T were integrated to one broader class, called grade II. This was decided with the experts' agreement. Another reason for this data integration was the fact that the 1.5T dataset did not contained patients of grade I class, which renders the comparison with the 3T impossible.

Table 3.3 – Patients demographics, MRS pulse-sequence parameters and techniques provided by each center

Parameters	***1.5T (UMCN)***	***3T (LUH)***
No. of Patients in Low grade glioma (class I)	0	9 (9 voxels)
No. of Patients in Low grade glioma (class II)	10 (176 voxels)	6 (6 voxels)
No. of Patients in Intermediate grade glioma (class III)	4 (57 voxels)	8 (8 voxels)
No. of Patients in High grade glioma (class IV)	7 (70 voxels)	20 (20 voxels)
Echo Time (TE)	20ms (short)	144ms (long)
Repetition Time (TR)	2000ms	1000ms
Voxel Size	$20mm^3$	$225mm^3$
MRS Acquisition Techniques	2D STEAM	2D PRESS

The numbers in the brackets represent the number of voxels measured from all patients of the class

3.4.4.2 MRSI data acquisition of 1.5T center

All measurements for the 1.5T system were performed on a (Siemens Vision) whole body system, using a circularly polarized (CP) transmitter/receiver head coil. The protocol consisted of the acquisition of conventional T1-weighted, T2-weighted and proton density weighted images (PD image). This was followed by a T1-weighted image after intravenous administration of Gadolinium-DTPA (Gd image) and concluded with water suppressed and unsuppressed proton MR Spectroscopic Imaging (^{1}H-MRSI). Eventually, only the four images (with different contrasts) acquired from the same location as the MRSI were retained. Images from areas just below or above were discarded, since they were not totally within the MRSI slice. The MRSI were acquired applying fat and water suppresses 2D Stimulated Echo Acquisition Mode (STEAM) ^{1}H-MRSI sequence with short echo time (TE). The STEAM box was positioned in a transversal plane through the brain showing the largest Gadolinium-DTPA enhanced tumor area in the Gd image. It was placed entirely in the brain parenchyma avoiding leakage of disturbing signals from fatty tissue surrounding the skull. One 2D MRSI slice was acquired per patient. The MRSI parameters were: 16 x 16 x 1024 samples, TR/TE (Repetition/Echo times) = 2000/20ms, slice thickness = 12.5 or 15 mm, Field of View (FOV) = 200 mm and number of excitations (NEX) = 2. Each voxel within the STEAM box was corrected for eddy current effects in the spectra using a method that prevents the occasional occurrence of eddy current correction induced artifacts. This was followed by Hankel-Lanczos Singular Value Decomposition (HLSVD) filtering to remove the residual water signal between 4.3 and 5.5 ppm. Next, a simple but efficient baseline correction with a filter width of 5 ms was applied to remove broad resonances. Finally, each time-domain signal was Fourier transformed to obtain a spectrum from which only the region between 0.5 to 4.0 ppm was retained.

3.4.4.3 MRSI data acquisition of 3T center

The dataset was collected from the 3T whole body MR unit (GE Healthcare Signa® HDx). The imaging protocol for voxel positioning was consisted of Fluid attenuated inversion recovery (FLAIR TR=9502ms, TE=128ms) or a home-designed T2-weighted fast spin echo (FSE, TR=4520ms, TE=102ms) sequence in axial, coronal and sagittal planes, which were performed using FOV = 260mm, slice thickness = 5mm and NS=1. Spectra from ^{1}H-MRSI were acquired by applying fat and water suppressed 2D-Point Resolved Pulse Sequence (2D-PRESS) before contrast administration using a 4-channel bird cage coil with frequency and phase resolution of 12 and 24 steps, respectively. ^{1}H-MRSI scan parameters were TE=144ms and TR=1000ms. Table 3.3 summarizes patient demographics and time parameters of MR spectroscopic pulse-sequences.

Spectroscopic data from patients were acquired from the following regions of interest: 1) inside the lesion, as in Figure 3.15(C, E), 2) contralateral side, 3) outer diameter of the lesion (if possible), as in Figure 3.15(D, F) and 4) normal appearing white matter. For the

classification purposes herein only region 1 was taken into account. Regions 2, 3 and 4 were used for reasons of comparison and to investigate the degree of tumor infiltration, which aided experts to evaluate tumor grade in lack of other histological diagnosis. A multi-voxel PRESS box was initially localized to the region of interest, which according to the number of phase and frequency directions was divided into subvoxels. Only the voxel from the intratumoral region was picked for further analysis and comparison purposes.

The MRS data collected from the MRS scanner were first preprocessed off line using the software provided by the manufacturer, to ensure that metabolite ratios were accurately calculated. Phase correction was implemented to remove baseline roll and restore pure shapes of metabolite peaks. The existence of baseline signal in the spectrum makes spectral analysis unreliable, since the estimation of peak areas in the presence of an unknown background suffers from baseline dependent bias. Baseline correction consisted of the subtraction of the function describing the course of the background signal.

Water normalization of the metabolites MRSI was performed by the division of each metabolite's free induction decay (FID) by its calculated mean water signal. In other words the mean water signal of all voxels of each subject was calculated and was used as an intra-patient normalization factor. Similarly, the calculated water signal of each voxel in a patient's data set was divided by the mean water signal, to obtain inter-patient normalized water signals for each voxel. The normalization process was applied before the metabolic ratio calculation, in both the 1.5T and 3T datasets.

3.4.5 Metabolic features selection

The metabolic features often acquired from the MRS spectra are either a) peak heights of known metabolites at specific resonances, or b) their integrated peak areas, or c) ratios of their integrated peak areas. Measuring peak-area ratios of metabolites has the advantage of cancelling out the effects of uncontrolled system and measurement variations on signal intensity. Ratio-type features have been successfully used in several studies over the last decade, mostly based on 1.5T MRS scanners, addressing the discrimination of brain-tumor type, grade and heterogeneity. Investigating the current bibliography shown in Table 3.4, involving ratio-type features, we observe that there exists a dominant set of ratio markers (NAA/Cho, NAA/Cr, Cho/Cr, mI/Cr and LL/Cr) that facilitates the non-invasive discrimination of different grades of gliomas, providing quite satisfactory classification rates.

Motivated from these observations, we designed a set of significant ratio markers, based on the peak areas of well known metabolites shown in Figure 3.15(E, F) and Figure 3.16(A to D). In addition, we considered five new metabolites (NAA/S, Cho/S, Cr/S, mI/S and Ala/S) found to be highly significant in recent studies of our team *[Kounelakis M.G. et al 2008 - Kounelakis M.G. et al 2009]*. In addition, the LL/S ratio feature was also included in this feature set to test its discriminative potential. The S variable denotes the sum of the peak areas of these metabolites. The whole set of metabolic ratio markers, is presented in Table 3.4.

In order to calculate the ratios shown in Table 3.4, integration of the peak areas of each metabolite was applied *[Simonetti A.W. et al 2003]*. The areas were estimated by integrating each metabolite's spectral intensity around its peak within a window of 0.13 ppm, as shown in Figure 3.16(A). The 0.13 ppm window was selected as a spectral width able to completely cover most peaks of interest, without being extended to the regions of neighboring peaks.

Table 3.4 - Significant Metabolic Ratio Markers

Bibliography	Markers used in this study									
	NAA/Cho	NAA/Cr	Cho/Cr	ml/Cr	LL/Cr	NAA/S	Cho/S	Cr/S	ml/S	LL/S	Ala/S
Ratio Markers know from Bibliography						Ratio Markers from recent studies of our team					
Kounelakis M.G. et al 2008	X	X	X		X	X	X	X	X	X	X
Kounelakis M.G. et al 2009	X	X	X			X	X	X	X	X	X
Howe F.A. et al 2003(b)			X	X							
Ott D. et al 1993	X		X								
Kumar A. et al 2003	X	X	X								
Lehnhardt F.G. et al 2005			X								
Galanaud D. et al 2006	X	X	X		X						
Dimou I.N. et al 2009(a)	X	X	X	X	X						
Dimou I.N. et al 2009(b)	X	X	X	X	X						
Bulakbasi N. et al 2003	X	X			X						
Castillo M. et al 2000		X	X	X							
Likavcanová K. et al 2005		X	X								

The second line of this table shows the metabolic ratio markers used in this study. NAA: N-acetyl aspartate, Cho: Choline, Cr: Creatine, Lac: Lactate, LL: Lipids and Lac, ml: myo-Inositol. The "S" variable denotes the sum of the peak areas of these metabolites. The "X" letter under each ratio marker is used to denote the studies in which this particular ratio was found significantly different (based on well known statistical tests like Independent Samples t-test, Mann-Whitney U-test and Analysis of Variance F-test applied in these studies) among gliomas grades and therefore important for discrimination purposes

3.4.6 Statistical analysis at 1.5T and 3T datasets

The diagnostic models used to classify tumor grades were developed using support vector machine (SVM) classifiers *[Cortes C. et al 1995]*. SVMs have been widely applied for binary (two-classes: positive (+1) and negative (-1)) classification problems. The learning scheme seeks for the optimal separating hyperplane where the margin of class separation is maximal. The SVM solution is based only on those data points that are at the margin of the decision boundary, called support vectors, as already explained in Chapter 2. The classifier model in primal space is given by Eq. (2),

$$y(x) = sign((w \cdot x) + b) \tag{2}$$

where $y(x)$ is the binary class indication encoded as -1 versus +1 of the x sample vector, w is a weighting vector, $(w \cdot x)$ denotes the dot product of the two vectors and b is a bias term (the offset of the hyperplane form the origin).

As described in Chapter 2, one of the main advantages of SVMs is the kernel trick. In other words the mapping of x in the input space, where the separation of the two classes is non-linear, to the feature space $\phi(x)$ i.e. a space of higher dimensionality where the separation can be linear. Based on this idea the dot product in Eq. (2) can be rewritten as:

$$y(x) = sign((w \cdot \phi(x)) + b) \tag{3}$$

We implicitly work in the feature space by applying a positive definite kernel as in Eq. (4) and Eq. (5).

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \tag{4}$$

In particular, we use the radial basis function (RBF) SVM kernels:

$$k(x_i, x_j) = \exp\left(- \frac{\left\| x_i - x_j \right\|_2^2}{\sigma^2} \right) \tag{5}$$

The SVM classification was applied under different binary tests. The schemes examined are glioma grade II vs. III, grade II vs. IV and grade III vs. IV, which are the most important according to clinicians and the bibliography. The Matlab ver. R2010 *[Internet Source: http://www.mathworks.com]* software package was used to build and test the classifiers.

3.4.7 Classifier performance evaluation

The evaluation of classifier performance was assessed through a 10-fold Cross Validation (CV) method and the classification accuracy was measured in terms of Area under the ROC curve (AUROC) values. Repeated stratified runs (10 folds X 10 times = 100 runs) were applied for each dataset (1.5T, 3T) which according to bibliography is, among others (Leave one out, Bootstrapping), the most suitable evaluation method for cases where data samples are few as in 3T dataset *[Kohavi R. 1995]*. For each binary classification scheme the train sets contained 90% of the voxels of each of the two classes and the remaining 10% of these two classes were contained in the test sets. Due to the fact that in 3T dataset the number of voxels was small compared to that of 1.5T, the training sets were carefully selected to avoid overlap. Confidence Intervals were also estimated for each classification scheme using the SPSS ver. 19.0 statistical software package *[Internet Source: http://www.spss.com]*.

3.4.8 Experimental results

The first step of our analysis was the application of the SVM classifier on both 1.5T and 3T datasets. The feature set used in the classification process was derived from the union of ratio features from recent bibliography expanded with those found significant by our team's recent studies *[Kounelakis M.G. et al 2008 - Kounelakis M.G. et al 2009]*, all presented in Table 3.4. Overall, we ended up with 11 ratio features used in the classifier. In Table III the classification outcome in terms of AUROC, for each binary scheme, is presented. Moreover the confidence intervals estimated for each scheme are shown.

Another aspect of this study was to estimate the contribution of each ratio feature on the classification success. To quantify this contribution, the frequency (%) of participation of each ratio feature at the highest AUROC values was recorded (using SPSS package). Figures 3.17 and 3.18 below represent in different colored bars the 11 metabolic ratios participating in the classification process for 1.5T and 3T datasets, respectively. In these two charts, a higher bar underlines a greater contribution of the metabolite in the classification procedure.

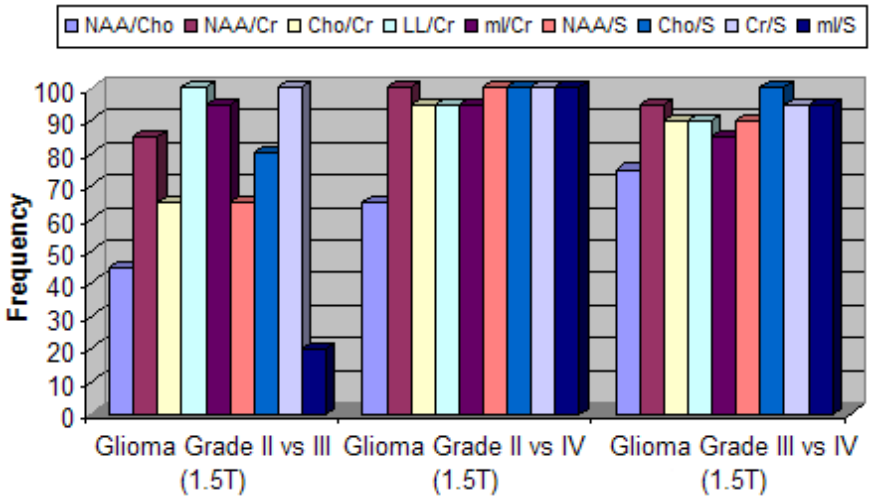


Figure 3.17 - Frequencies of features' participation in each binary classification scheme of the 1.5T dataset. The ratio markers, from left to right, correspond to the bars in the same order

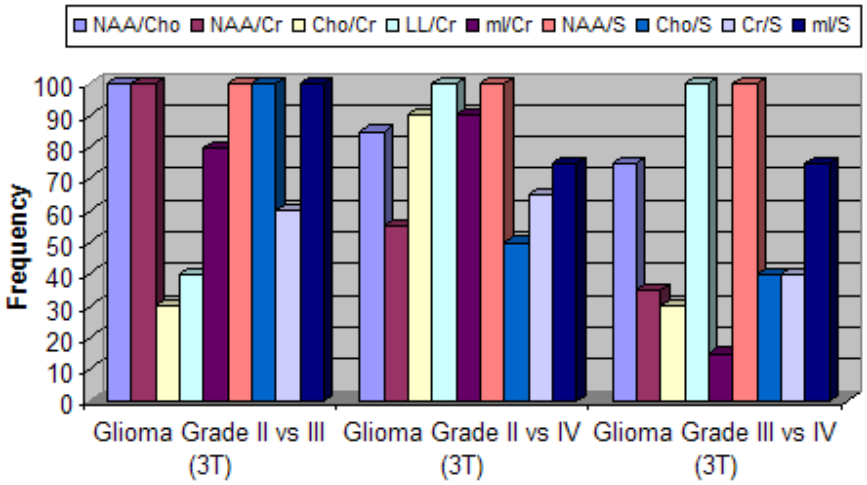


Figure 3.18 - Frequencies of features' participation in each binary classification scheme of the 3T dataset. The ratio markers, from left to right, correspond to the bars in the same order

Table 3.5 - Classification results for each system (1.5T and 3T)

	AUROC values		
Centers	Glioma grade II vs. III	Glioma grade II vs. IV	Glioma grade III vs. IV
1.5T (UMCN)	**0.83** (± 0.02)	**0.96** (± 0.04)	**0.95** (± 0.03)
3T (LUH)	0.77 (± 0.03)	0.85 (± 0.03)	0.75 (± 0.02)

Bold numbers indicate the highest AUROC values obtained. The numbers in the brackets represent the Confidence Intervals measured for each AUROC value

3.4.9 Discussion

The first important aspect of comparison relates to the classification results presented in Table 3.5. The classification values derived from the 1.5T dataset are higher than those obtained from the 3T. Even though the 3T system acquires metabolic spectra of better quality, it is clear that this is not enough to produce a substantial improvement in the classification of population-based data. The lack of a satisfactory amount of data renders the SVM classifier quite inefficient in the case of 3T, despite the quality improvement in spectral resolution. It is expected though that if the 3T dataset involves more voxels per patient then the classification accuracy would be drastically improved over that of 1.5T. This is an aspect that will be studied in the future since the 3T dataset is being constantly improved and enhanced with new cases.

The dependence of signal-to-noise ratio (SNR) on field strength is a complicated issue, both from a theoretical aspect and in terms of experimental comparisons. In cases of brain tumours examined at 3T, Kim J.H. et al observed that better spectral resolution was achieved at short TE (35ms); however, little improvement in spectral resolution was detected at intermediate TE (144ms). More specifically, at short TE, Cho/Cr and Cho/NAA ratios were significantly lower, and LL/Cr and mI/Cr were significantly higher, compared with those at intermediate TE, regardless of tumor grade *[Kim J.H. et al 2006(a)]*. In another study of Kim J.H. et al, at 1.5T and 3T, it was found that the SNR of major brain metabolites at 3T demonstrated 49-73% increase at short TE (35ms) and only 2-12% increase at an intermediate TE (144ms) compared with those of 1.5T *[Kim J.H. et al 2006(b)]*. Considering the observations of both studies of Kim J.H. et al, at intermediate TE (144ms) at 3T, we can say that the TE parameter plays a significant role in the discrimination of tumor grades, as shown in our study too.

With respect to the ratio features studied, the most significant and often encountered features are depicted in Figures 3.17 and 3.18, for the 1.5T and 3T systems, respectively. In general, we can conclude that both sets are capable of discriminating complex cases of gliomas. This is more obvious at the glioma grade II vs. III scheme, where in both datasets (1.5T, 3T) the grade II classes contained mixed low grade gliomas, such as astrocytomas and oligodendrogliomas. An important observation in this specific classification scheme is that the mI/S ratio feature has a strong influence, stronger than the mI/Cr, in the discrimination of the 3T dataset compared to the 1.5T dataset. As also mentioned in the introduction, mI is better resolved in 3T systems. mI is considered a vital feature in this case, since low grade gliomas posses much higher mI values than the more aggressive tumors [*Castillo M. et al 2000*]. Other significant features that have strong influence in the 3T dataset in this binary scheme are the NAA/Cho, NAA/Cr, NAA/S and Cho/S. This is again in accordance with recent bibliography *[Kounelakis M.G. et al 2008 - Kounelakis M.G. et al 2009]*. In the 1.5T dataset, however, the most significant features are the LL/Cr, Cr/S and the mI/Cr. It worth's noticing here that the LL feature, although expected to be significant at 3T due to its better spectral resolution, it does not have a positive impact in classification. This can be explained through

the design of the 3T dataset, where the grade II class also contains patients of grade I with no Lactate influence.

A general observation in the glioma grade II vs. IV binary scheme is the contribution of almost all ratio features in the 1.5T classification case. As shown in Figure 3.17, 8 out of 9 features participate in achieving an AUROC value of 0.96. In contrast, only LL/Cr and NAA/S appear to be most significant in the 3T dataset (as shown in Figure 3.18). These features are expected to be significant, since in glioblastomas the lipids and lactate get high values, as shown in Figure 3.16(D). Other ratio features that highly contribute in the 3T classification process are the Cho/Cr and mI/Cr. As mentioned above, mI is high in low grade and NAA (as a neuronal marker) is very low in grade IV; notice that most brain tumours are of non-neuronal origin *[Heerschap A. 2007]*.

In the most aggressive case of glioma grade III vs. IV, where the differentiation at metabolic level is both difficult and critical, almost all features of the 1.5T system contribute to the discrimination of the two classes as shown in Figure 3.17. In addition, the most important (frequent) features are the NAA/Cr, along with those normalized to the same baseline (S). These include the NAA/S, Cho/S, Cr/S and mI/S ratios. As in the previous case, the performance of features is not the same in 3T dataset, where LL/Cr and NAA/S have the strongest influence. This is expected since lipids and lactate peaks (LL) are easier to be detected and measured in the 3T case due to better spectral resolution. Before the ^{1}H-MRS procedure at 3T, high resolved FLAIR T2-weighted or T2-FSE images are used allowing the discrimination of regions with necrosis, haemorrhage, cyst or calcification within the tumour. However, as also stated in the material and methods section, all spectroscopic measurements are applied before contrast administration to avoid signal diminish. Therefore, it is possible that intratumoral necrotic regions might be masked on T2-weighted images, which are observed in anatomical images only after contrast injection. Moreover, it is clear that an elevation of lipids is a known progression indicator for gliomas. Thus, inside a voxel it is rather difficult to avoid lipid contamination leading to glioma-grade overestimation. NAA on the other hand as a neuronal marker is found only in neurons and, since most brain tumours are of non-neuronal origin, it is reduced or absent. In MRS of tumours, the presence of NAA within a spectrum generally indicates the presence of viable neurons within an infiltrative tumour.

Another observation that needs to be addressed at this point is the small contribution of the Cho and Cr ratios (Cho/Cr, Cho/S and Cr/S) in this classification process. A reason for that could be the presence of necrosis within some voxels of grade IV patients, although careful voxel positioning has been applied to avoid the inclusion of obvious necrotic and/or cystic regions. The presence of necrosis is one important distinction between anaplastic astrocytomas (grade III) and glioblastoma multiforme (grade IV). Presence of high lipid peaks may suggest macroscopic necrosis due to membrane breakdown. Therefore, lipids do correlate with necrosis in high-grade glioma and so may also be useful in differentiating glioma grades *[Fan G. 2006]*. Nevertheless, the dominant peaks of LL in some voxels could

mask the contribution of Cho and Cr metabolites, affecting the classification's outcome.

The results of this study imply that NAA/S, Cho/S, Cr/S and mI/S ratios may enhance the discriminating power in complex cases like high grade gliomas. It was also observed that the two ratios LL/S and Ala/S (Alanine), also tested, did not significantly improve the AUROC measures in either 1.5T or 3T. A main reason for that is the fact that the LL/Cr ratio possibly masks the contribution of the LL/S in the classification method. Furthermore, the Ala/S ratio (observed mostly in meningiomas) cannot be considered a reliable marker because of the broader peaks of LL which overlaps with the neighbour peak of Alanine (Ala at 1.48 ppm) especially in the intermediate and high grade gliomas.

Finally, this study indicates that in comparison between 1.5T and 3T MRS scanners, the ratio markers NAA/S, Cho/S, Cr/S and mI/S are more sensitive in revealing metabolic differences in these types of gliomas compared to NAA/Cho, Cho/Cr, mI/Cr, LL/Cr. This is also supported by Kim J.H. et al 2006(b), who observed that the difference between the mean values of NAA/Cho, Cho/Cr, mI/Cr and LL/Cr in both 1.5T and 3T systems were not statistically significant. At this point we should also mention that the use of metabolite ratios tested in this study has several advantages. Since the reference signal is acquired simultaneously with the metabolite of interest, many potential sources of systematic errors are suppressed. Such errors relate to the signal dependency on tissue internal factors, the number of observed spins inside the volume of interest, the magnetic field applied and timing parameters of the sequence *[Gillard J. et al 2005]*.

The pattern recognition method used based on SVMs is one of the most commonly applied methods for brain tumors classification, during the last decade. SVM and its simpler form called Least Square SVM (LS-SVM) are usually preferred for the classification of high dimensional data. Furthermore, it has been found in *[Devos A. et al 2004 – Di Costanzo A. et al 2006 - Caprinelli G. et al 1996 - Howe F.A. et al 2003(b) - Kounelakis M.G. et al 2008 - Kounelakis M.G. et al 2009 - Ott D. et al 1993 - Kumar A. et al 2003]* that they perform quite satisfactory in complex tasks such as the gliomas classifications, especially with RBF kernel; their performance is better compared to other methods such as Linear Discriminant Analysis (LDA) and Neural Networks (NN). This was also the reason for adopting SVM in this study. Concerning the evaluation method used to assess classifier performance, the 10-fold CV method was adopted. In contrast to bootstrapping methods, k-fold CV methods are usually adopted, since they perform better in brain tumors discrimination problems, even though they lack of stability when small datasets are studied, like in our 3T case *[Kohavi R. 1995]*. To overcome this problem, we used repeated stratified CV runs but also careful selection of the train and test sets.

The present study reflects several limitations that do not allow for immediate generalization of our results. The MRS procedure was performed with different types of radio frequency (RF) coils (Circularly Phased Array-head coil at 1.5T and 4-channel birdcage head coil at 3T) and shimming procedures. Both were high quality volume coils of about same size, but they differ in transmission of the RF signals. This different design might influence sensitivity and thereby

the SNR for the metabolites differently *[Marion D. et al 1989]*. The shimming protocols were manufacturer supplied, automatic procedures, which might not give the optimal shim for each acquisition. The voxels of the patients considered with 3T are very few compared to those of 1.5T, which forms an important limitation for this study. In other words, if the 3T dataset contained more voxels per patient then its classification results could be much higher. Moreover, the patients have heterogeneous lesions from different locations, which may also affect the study. The pulse sequences used and the timing parameters for the two MR scanners were not identical, so that differences in their efficiency need to be investigated. For instance, variable amounts of signal loss may occur as a function of the accuracy of the flip angle, the spacing and the duration. Careful calibration of the flip angle and adjustment of crusher gradient amplitudes on the unsuppressed water signal was performed on both scanners in this study. Nevertheless, although the SNR of major brain metabolites at 3T is increased and better spectral resolution is obtained at short TE, little spectral resolution improvement is detected at intermediate TE (=144msec) compared with 1.5T *[Kim J.H et al 2006(b)]*. The two MR pulse sequences share different repetition time values (TR), namely 2000msec for 1.5T MRS scanner and 1000msec for 3T MR scanner respectively. Differences in T1-values for the metabolites at 1.5T and 3T can also influence the SNR at a particular repetition time (TR), thus further influencing our results *[Sjobakk T.E. et al 2006]*.

3.4.10 Conclusions

Non-invasive proton-MRS may be used efficiently to determine the presence or type of a cerebral tumour and can be a useful tool to confirm and grade brain neoplasms together with biopsy. Several MRS scanner modalities are available in today's practice related to brain cancer diagnostics and discrimination. The obvious goal of all these systems is to improve the quality of the spectrum acquired in order to obtain better metabolic profile of brain tumors. Nevertheless, strengthening the magnets' field is not the only parameter that has to be taken under consideration. Other factors related to data acquisition, like the number of voxels per patient, the TE value and the noise reduction have to be carefully considered in order to reach accurate diagnosis and successful treatment of these complex tumors. In addition, experts must carefully decide the type and the number of metabolic features towards a more individualized diagnosis.

This study attempts to investigate some of these issues by comparing the outcomes of a classification method applied to two datasets extracted from two different MRS scanners, using various spectral features. The study derives specific ratio features for each modality that are capable of discriminating these complex pathologies. A future objective of this study is to incorporate peritumoral and contralateral regions in order to enrich the 3T dataset and extend the pattern recognition methods towards a more accurate discrimination of gliomas. Another plan is to incorporate quantitative data from other MR-based methodologies like Diffusion Weighted Imaging (DWI) and Perfusion Weighted Imaging (PWI) for developing a diagnosis decision support system assisting experts to discriminate more complicated metabolic spectral profiles.

References

Armstrong P. et al 1991, 'Magnetic Resonance Imaging-1: basic principles of image production', *British Medical Journal*, vol. 303, pp. 35-40.

Bottomley P. A. 1987, 'Spatial localization in NMR-spectroscopy in vivo', *Annals of the New York Academy of Sciences*, vol. 508, pp. 333-348.

Bulakbasi N. et al 2003, 'Combination of single-voxel proton MR spectroscopy and apparent diffusion coefficient calculation in the evaluation of common brain tumors', *American Journal of Neuroradiology*, vol. 24, pp. 225-233.

Callot V. et al 2008, '1H MR spectroscopy of human brain tumours: a practical approach', European *Journal of Radiology*, vol. 67, pp. 268-274.

Caprinelli G. et al 1996, 'Differentiation of glioblastoma multiforme from astrocytomas by in vitro 1H MRS analysis of human brain tumors', *Anticancer Research*, vol. 16, pp.1559–1564.

Castillo M. et al 2000, 'Correlation of myo-inositol levels and grading of cerebral astrocytomas', *American Journal of Neuroradiology*, vol. 21, pp. 1645-1649.

Cortes C. et al 1995, 'Support-Vector Networks', *Machine Learning*, vol. 20, pp. 273-297.

Cousins J.P. 1995, 'Clinical MR spectroscopy: fundamentals, current applications, and future potential', *American Journal of Roentgenology*, vol. 164, pp. 1337-1347.

De Seze J. et al 2010, 'Magnetic resonance spectroscopy evaluation in patients with neuromyelitis optica', *Journal of Neurology Neurosurgery and Psychiatry*, vol. 81, pp. 409-411.

Devos A. et al 2004, 'Classification of brain tumours using short echo time ^{1}H MRS spectra', Journal of Magnetic Resonance, vol. 170, pp. 164-175.

Di Costanzo A. et al 2006, 'Multiparametric 3T MR approach to the assessment of cerebral gliomas: tumor extent and malignancy', *Neuroradiology*, vol. 48, pp. 622-631.

Dimou I.N. et al 2009a, 'Classification of pathological human brain lesions using Magnetic Resonance Spectroscopy at 3T', *World Congress on Medical Physics and Biomedical Engineering*, Munich, Germany.

Dimou I.N. et al 2009b, 'Classification of 3T MRS spectra using Support Vector Machines', *9th International Conference on Information Technology and Applications in Biomedicine*, Larnaca, Cyprus.

Dimou I.N. et al 2011, 'Brain lession classification using 3T MRS spectra and paired SVM kernels', *Elsevier Biomedical Signal Processing and Control*, vol. 6, pp. 314-320.

Drost D.J. et al 2002, 'Proton magnetic resonance spectroscopy in the brain: report of AAPM MR Task Group #9', *Medical Physics*, vol. 29, pp. 2177-2197.

Dunn W.B. et al 2005, 'Metabolomics: current analytical platforms and methodologies', *Trends in Analytical Chemistry*, vol. 24, pp. 285-294.

Fan G. 2006, 'Magnetic resonance spectroscopy and gliomas', *Cancer Imaging*, vol. 6, pp. 113-115.

Frahm J. 1987, 'Localized proton spectroscopy using stimulated echos,' *Journal of Magnetic Resonance*, vol. 72, pp. 502-508.

Fountas K.N. et al 2004, 'Noninvasive histologic grading of solid astrocytomas using proton magnetic resonance spectroscopy', *Stereotactic and Functional Neurosurgery*, vol. 82, pp. 90-97.

Galanaud D. et al 2006, 'Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy', *Magnetic Resonance in Medicine*, vol. 55, pp.1236-1245.

Gates S.C. et al 1978, 'Quantitative metabolic profiling based on gas chromatography', Clinical Chemistry, vol. 24, pp. 1663-73.

Gillard J. et al 2005, 'Clinical MR Neuroimaging', *Cambridge: Cambridge University Press*, 2005.

Gonen O. et al 2001, 'Multivoxel 3D proton spectroscopy in the brain at 1.5 versus 3.0 T: signal-to-noise ratio and resolution comparison', *American Journal of Neuroradiology*, vol. 22, pp. 1727–1731.

Govindaraju V. et al 2000, 'Proton NMR chemical shifts and coupling constants for brain metabolites', *NMR in Biomedicine*, vol. 13, pp.129-153.

Heina P.A. et al 2004, 'Diffusion-Weighted Imaging in the Follow-up of Treated High-Grade Gliomas: Tumor Recurrence versus Radiation Injury', *American Journal of Neuroradiology*, vol. 25, pp. 201-209.

Heerschap A. 2007, 'In vivo magnetic resonance spectroscopy in clinical oncology. in Shields A.F. and Price P. (eds) Cancer Drug discovery and Development In Vivo Imaging of Cancer Therapy, *Humana Press Inc.,* Totowa, pp. 241-258.

Horning M.G. et al 1971, 'Variations in urinary steroid profiles after birth', *Clinica Chimica Acta*, vol. 34, pp. 261-268.

Howe F.A. et al 2003a, '1H MR spectroscopy of brain tumours and masses', *NMR in Biomedicine*, vol.16, pp. 123-131.

Howe F.A. et al 2003b, 'Metabolic Profiles of Human Brain Tumors Using Quantitative In Vivo 1H Magnetic Resonance Spectroscopy', *Magnetic Resonance in Medicine*, vol. 49, pp.223-232.

Kell D.B. 2004, 'Metabolomics and systems biology: making sense of the soup', *Current Opinion in Microbiology*, vol. 7, pp. 296–307.

Kim J.H. et al 2006a, '3T 1H-MR Spectroscopy in Grading of Cerebral Gliomas: Comparison of Short and Intermediate Echo Time Sequences', *American Journal of Neuroradiology*, vol. 27, pp.1412-1418.

Kim J.H. et al 2006b, 'Comparison of 1.5T and 3T 1H MR spectroscopy in human brain tumors', *Korean Journal of Radiology*, vol. 7, pp. 156-161.

Kohavi R. 1995, 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *14th International Joint Conference on Artificial Intelligence'*, San Mateo, US.

Kounelakis M.G. et al 2008, 'Identification of significant Metabolic Markers from MRSI data for Brain Cancer Classification', *8th IEEE International Conference on BioInformatics and BioEngineering*, Athens, Greece.

Kounelakis M.G. et al 2009, 'Revealing the metabolic profile of brain tumors for diagnosis purposes', *31st IEEE Annual International Conference of the Engineering in Medicine and Biology Society*, Minnesota, USA.

Kounelakis M.G. et al 2010, 'Strengths and Weaknesses of 1.5T and 3T MRS Data in Brain Glioma Classification, *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 647-654, 2011.

Kumar A. et al 2003, 'Role of in vivo proton MR spectroscopy in the evaluation of adult brain lesions: Our preliminary experience', *Neurology India*, vol. 51, pp. 474-478.

Lehnhardt F.G. et al 2005, 'Metabolic differences between primary and recurrent human brain tumors: a 1H NMR a 1H NMR spectroscopic investigation', *NMR in Biomedicine*, vol. 18, pp. 371-382.

Likavcanová K. et al 2005, 'In vitro study of astrocytic tumor metabolism by proton magnetic resonance spectroscopy', *General Physiology and Biophysics*, vol. 24, pp. 327-335.

Majos C. et al 2004, 'Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE', *American Journal of Neuroradiology*, vol. 25, pp. 1696–1704.

Marion D. et al 1989, 'Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data', *Journal of Magnetic Resonance*, vol. 84, pp. 425–430.

Meyerand M.E. et al 1999, 'Classification of biopsy-confirmed brain tumors using single-voxel MR spectroscopy', *American Journal of Neuroradiology*, vol. 20, pp. 117-123.

Moller-Hartmann W. et al 2002, 'Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions', *Neuroradiology*, vol. 44, pp. 371-381.

Nelson S.J. 2003, 'Multivoxel magnetic resonance spectroscopy of brain tumors', *Molecular Cancer Therapeutics*, vol. 2, pp. 497-507.

Ott D. et al 1993, 'Human brain tumor: assessment with in vivo proton MR spectroscopy', *Radiology*, vol. 186, pp. 745-752.

Simonetti A.W. et al 2003, 'A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy', *Analytical Chemistry*, vol. 75, pp. 5352-5361.

Sjobakk T.E. et al 2006, 'Clinical 1H magnetic resonance spectroscopy of brain metastases at 1.5T and 3T', *Acta Radiologica*, vol. 47, pp. 501-508.

Smith I.C.P. et al 2002, 'Magnetic resonance spectroscopy in medicine: clinical impact, Progress in Nuclear Magnetic Resonance Spectroscopy', vol. 40, pp. 1-34.

Tanenbaum L.A. et al 2005, '3T MRI in clinical practice', *Applied Radiology*, vol. 34, pp. 8-17.

Wirt M.D. 2003, 'Magnetic resonance spectroscopy: A basic guide to data acquisition and interpretation', *Applied Radiology*, vol. 32, pp. 25-30.

Williams R. J. 1940, 'Vitamin Study at The University of Texas', *Science*, vol. 92, pp. 579.

Internet sources

Barrow Neurological Institute, (Bohnert B.J, Current Applications of MR Spectroscopy of the brain, http://www.thebarrow.org/Education_And_Resources/Barrow_Quarterly/205120)

Computer Science Source, http://www.computersciencesource.com

Cybaea, http://www.cybaea.net

Encognitive, http://www.encognitive.com

Expasy Proteomics Server, http://expasy.org

E-Radiography, (Bradley W.G, Fundamentals of MRI, http://www.e-radiography.net/mrict/fund%20mr1/MRI%20fund1.htm)

Griff Wason, http://www.griffwason.com

Haodf, http://www.haodf.com

How MRI Works, http://health.howstuffworks.com/medicine/tests-treatment/mri.htm

Human Metabolome Database, http://www.hmdb.ca
Imaios, http:///www.imaios.com

Kyoto Encyclopedia of Genes and Genome, http://www.genome.jp/kegg

Magnetic Resonance –Technology Information Portal, http://www.mr-tip.com

Mark's Psychiatry, http:///markspsychiatry.com

Matlab, http://www.mathworks.com

Medscape, http://www.medscape.com

Michigan State University (Radiology), http://www.rad.msu.edu

NASA, http://rst.gsfc.nasa.gov

PF Science, http://www.pfscience.com

Queensland Diagnostic Imaging, http://www.qdi.com.au

Secca M.F, Basic Principles of MRI and F-MRI in neurosciences, http://www.panrs.org/doc/syllabus/ch1part1.pdf

Rochester Institute of Technology (Hornak J.P, The Basics of MRI, http://www.cis.rit.edu/htbooks/mri/)
SPSS, http:///www.spss.com
The Comprehensive Enzyme Information System, http://www.brenda-enzymes.info

University of California, Center for Functional MRI (Hesselink J.R, Basic principles of MR imaging, http://spinwarp.ucsd.edu/NeuroWeb/Text/br-100.htm)

US National Institute of Health, http://publications.nigms.nih.gov/biobeat/gallery/index.html

Wikipedia, http://www.wikipedia.org

Wikimedia, http:///www.wikimedia.org

Wikidoc (Chao C. 'Basic MRI physics, http://www.wikidoc.org)

Wikibooks, http://www.wikibooks.org

World of Teaching, (Banu H.B, 'Applications of NMR in medicine',
http://www.worldofteaching.com/powerpoints/biology/NMR%20diagnosis.ppt)

Chapter 4

IDENTIFICATION OF SIGNIFICANT METABOLIC MARKERS FOR BRAIN TUMORS CLASSIFICATION: an application on peak area ratios

Based on Author's published studies:

- ***M.G. Kounelakis***, M.E. Zervakis, G. C. Giakos, G. J. Postma, L. M. C. Buydens and Xenofon Kotsiakis, Embedding filtering criteria into a wrapper marker selection method for brain tumors classification: An application on metabolic peak area ratios*, IOP Measurement Science & Technology, 2011*

- ***M. G. Kounelakis***, M. E. Zervakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis, Revealing the metabolic profile of brain tumors for diagnosis purposes , *31st IEEE Annual International Conference of the Engineering in Medicine and Biology Society, Minnesota, pp. 35-38, 2009*

- ***M. G. Kounelakis***, M. E. Zervakis, M. E. Blazadonakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis, Feature Selection for Brain Tumour Classification using Ratios of Metabolites' Peak Areas from MRSI Data, *6th European Symposium on Biomedical Engineering, Chania, pp. 1-6, 2008*

- ***M. G. Kounelakis***, M.E. Zervakis, M.E. Blazadonakis, G.J. Postma, L.M.C. Buydens, A. Heerschap and X. Kotsiakis, Identification of significant Metabolic Markers from MRSI data for Brain Cancer Classification, *8th IEEE International Conference on BioInformatics and BioEngineering, Athens, pp. 1-6, 2008*

- Blazadonakis M., Zervakis M., ***Kounelakis M***., Beganzoli E. and N. Lama, Support Vector Machines and Neural Networks as Marker Selectors for Cancer Gene Analysis, *3rd IEEE International Conference on Intelligent Systems, London, 2006*

4.1 Abstract

The purpose of this study is to identify reliable sets of metabolic markers that provide accurate classification of complex brain tumors and facilitate the process of clinical diagnosis. Several peak area ratios of metabolites (MRSI data) are tested alone, or in combination with imaging markers (MRI data). A wrapper feature selection and classification methodology is studied, employing the Fisher's criterion for ranking the markers. The set of extracted markers that express statistical significance is further studied in terms of biological behaviour with respect to the brain tumors type and grade.

The outcome of this study indicates that the proposed method, exploiting the intrinsic properties of data, can actually reveal reliable and biologically relevant sets of metabolic markers, which form an important adjunct towards a more accurate type and grade discrimination of complex brain tumors *[Kounelakis M.G. et al 2011]*.

4.2 Introduction

Feature selection is a critical issue in biomedical data analysis. Its primary goal is to extract biologically significant features that support a clinical decision in crucial medical tasks, such as tumor type and aggressiveness (grade) discrimination. Brain tumors form one of these crucial clinical tasks, where the type and grade discrimination becomes a challenge due to the data complexity. In this area, the evaluation of biologically significant features in conjunction with other clinical data (morphology, location, size etc) of a patient can assist clinicians to decide upon critical matters regarding the therapeutic pathways to be followed.

When accessible, most brain tumors are surgically removed, but there is a balance between removing as much tumor tissue as possible whilst maintaining vital brain functions. Therefore, a non-invasive and accurate assessment of tumor type can reduce unnecessary surgical intervention.

Towards this direction, ^{1}H or proton Magnetic Resonance Spectroscopy (^{1}H-MRS) can be used to provide information on the metabolic profile of tissue, facilitate a better non-invasive differential diagnosis, define the tumor grade and aggressiveness, monitor the tumor response to nonsurgical treatments and finally determine an earlier presence of tumor recurrence. Proton Magnetic Resonance (MR) spectra can be obtained from a single volume element (voxel) or from multiple voxels with the so-called MR Spectroscopic Imaging (MRSI). This modality enables the identification of the heterogeneity of a tumorous region, since spatial variations of tissue characteristics can be assessed at metabolite level. For each voxel, the intensity of the relevant metabolites can be determined in the spectral domain as shown in Figure 4.1. Earlier studies have shown that MRS has significant clinical value, particularly for the evaluation of diseases that affect brain tissue *[Callot V. et al 2008 - Ross B. et al 1994 – Howe F.A. et al 2003(a) – Roser W. et al 1997]*.

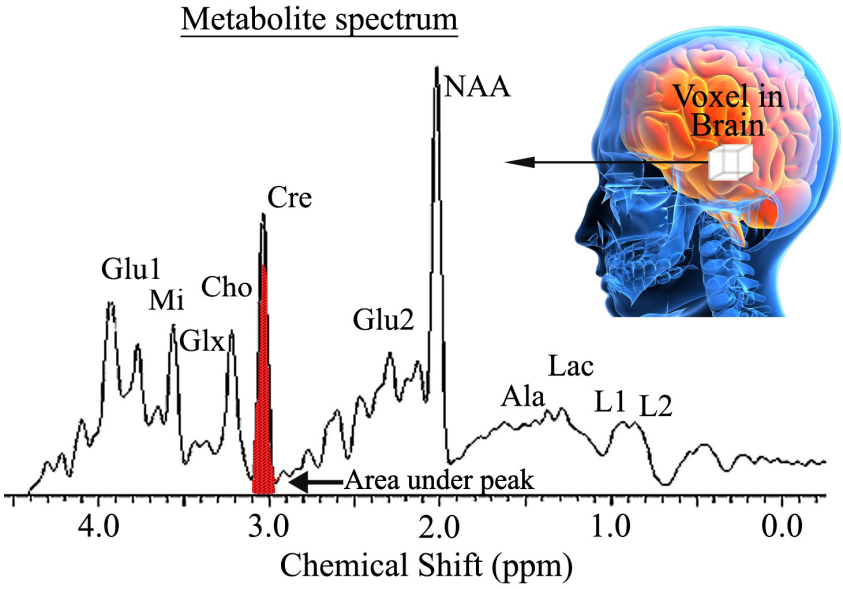


Figure 4.1 - The spectrum obtained from a voxel. Y axis: peak heights (proportional to metabolites concentration). X axis: frequency (position) in parts per million (ppm). NAA (N-

acetyl-aspartate), Cho (Choline), Cre (Creatine), Ala (Alanine), Lac (Lactate), mI (myo-Inositol), Glu1 (Glutamate), Glu2 (Glutamine), Glx (Glu1/Glu2), L1 and L2 (mobile lipids) are the metabolites considered – [Internet Sources: Hope Clinics - The UMKC Radiology Resident Resource] - (The figure was designed with Adobe Photoshop ver.CS5)

4.2.1 The peak area ratios case

The acquisition of the patient's metabolic spectra enables the determination of several types of features that can be used for brain tumor classification purposes. These may be the signal amplitudes, the area under the metabolite peaks (red area drawn in Figure 4.1) or the ratios of the peak areas of the metabolites. The amount of potentially useful features can be considerable. Feature selection methods can be applied to select the most informative subset of features providing accurate classification rates.

Ratio type features have been successfully used in several studies over the last decade for the discrimination of brain tumors *[Kumar A. et al 2003 – Howe F.A. et al 2003(b) – Lehnhardt F.G. et al 2005 – Galanaud D. et al 2006], but also other diseases using MRS data [Galanaud D. et al 2010]*.

Measuring ratios of metabolites has the advantage of cancelling out the effects of uncontrolled system and measurement variations on signal intensity. Quantification of in vivo cerebral metabolite concentration using MRS is quite complicated and time-consuming, to be routinely applied for clinical service. On the other hand, metabolite ratios can be derived directly from the outputs of a clinical MRS system, without the necessity of correcting for coil-loading and tissue characteristics. Therefore, many clinical MRS studies evaluate cerebral gliomas and related changes in terms of semi-quantitative (normalized) metabolite ratios rather than absolute concentrations *[Kumar A. et al 2003 – Howe F.A. et al 2003(b) – Lehnhardt F.G. et al 2005 – Galanaud D. et al 2006]*.

4.2.2 Main goals and achievements of the study

The main goal of this work is threefold. First, to reveal the potential of an embedded feature-selection method *[Blazadonakis M.E. et al 2006 – Blazadonakis M.E. et al 2008]* to identify the smallest and most compact sets of biologically significant features (markers) from brain spectra, which are capable to accurately classify different types of brain tumors such as gliomas and meningiomas. This method introduces a Recursive Feature Elimination (RFE) process that involves Fisher's filter criterion *[Fisher R.A. 1936]* in a wrapper selection scheme, within the operation of Support Vector Machines (SVMs) *[Dietterich T.G. 1998 – Vapnik N.V. 1999]*. Second, to exhibit the advantages of using peak area ratios extracted form short echo (Echo Time, TE = 20ms) Magnetic Resonance Spectroscopic Imaging (MRSI) data. To achieve this, we consider a dataset comprising of 18 ratio features, which stem from the biological significance of metabolites. This feature set (Table 4.2) includes metabolically significant ratio features that were not, in such extent, examined before. A third aim of this

work is to rigorously study the biological behavior of these markers with respect to the brain tumors type and grade.

Finally and in combination with the MRSI data, this study also examined the contribution of 4 MRI imaging features (at T1 = T1 weighted, T2 = weighted, PD = proton density and Gd = Gadolinium contrast enhancement agent) in the discrimination of these brain tumors.

The outcome of this study reveals that for each binary brain tumor discrimination scheme a dominant set of ratio markers exists, whose biological value is statistically verified (p-value). In six out of seven binary schemes (Healthy vs tumors, Healthy vs Gliomas, Healthy vs Meningiomas, Gliomas grade II vs Gliomas grade IV, Gliomas grade III vs Gliomas grade IV and Gliomas vs Meningiomas) the classification accuracy was greater than or equal to 0.97, in terms of Area under the Receiver Operating Characteristic Curve (AUROC). Only in the case of Gliomas grade II vs Gliomas grade III, the AUROC dropped to 0.84.

4.3 Materials and Methods

4.3.1 Dataset description

The dataset consists of MRI data and short echo MRSI data from 24 patients. From these patients, 21 have been diagnosed with World Health Organization (WHO) glial brain tumors of specific grade and 3 with meningiomas. More specifically the dataset contains 10 glioma subjects with low grade II (5 patients with diffuse Astrocytomas, 2 with Oligodendrogliomas and 3 patients with mixtures), 4 with intermediate grade III (1 patient with Anaplastic Astrocytoma, 2 patients with Oligodendrogliomas and 1 patient undefined), 7 with high grade IV (Glioblastoma multiforme) and 3 with meningiomas. Furthermore, in the same non-invasive way MRI and MRSI data were acquired from 4 healthy (control) volunteers and 4 patients. The initial core dataset, shown in Table 4.2, consisted of 569 pre-processed spectral sets containing data (MRI as well as MRSI) from 569 voxels of 24 brain tumor patients and 4 healthy persons.

The elaboration of the dataset was approved by the ethical committee of the University Medical Center of Nijmegen (UMCN), the Netherlands and written informed consent was obtained from all patients. Each patient-case passed strict quality control and validation procedures, including consensus histopathologic determination. Per tissue type, voxels were obtained from homogeneous regions.

The measurements were performed on a 1.5T Siemens Vision whole-body system using a circularly polarized (CP) head coil, shown in Figure 4.2. The system calibration, the acquisition techniques applied to obtain the four MRI images and the MRSI data, along with the MRSI data pre-processing methods, are discussed in earlier studies *[Simonetti A.W. et al 2003 – Simonetti A.W. et al 2002 – Pijnappel W.W.F et al 1992]*.

Figure 4.2 – Magnetom Aera 1.5 Tesla Magnetic Resonance Scanner – [Internet Source: Siemens]

Water normalization of the metabolites was performed by dividing each metabolite FID (free induction decay) with its calculated mean water signal. More specifically the mean water signal of all voxels of each subject was calculated first. Similarly, the calculated water signal of each voxel in a patient's dataset was estimated and divided by the mean water signal, to obtain inter-patient normalized water signal. This water normalization process, which is routinely used when pre-processing MRS data, was applied prior the metabolic ratio calculation.

The areas under the metabolites' peaks were obtained by peak integration, as shown in Figure 4.3(B). More specifically, these areas were estimated by integrating each metabolite's spectral intensities around its peak within a window of 0.13 ppm. The metabolites of interest are presented in Table 4.1.

Table 4.1 – Main ^{1}H-MRS metabolites examined, their function and chemical shift value

A/A	Name of Metabolite (symbol)	Metabolic function	Chemical shift
1	Glutamate (Glu1)	Neurotransmitter (neuron-glial interaction)	3.75 ppm
2	myo-Inositol (mI)	Glial cell marker	3.56 ppm
3	Glutamate/Glutamine (Glx)	Neurotransmitter (neuron-glial interaction)	3.44 ppm
4	Choline (Cho)	Cell membrane marker	3.20 ppm
5	Creatine (Cre)	Energy metabolism	3.02 ppm
6	Glutamine (Glu2)	Neurotransmitter (neuron-glial interaction)	2.20 ppm
7	N-acetyl-aspartate (NAA)	Marker of neuronal integrity and viability	2.02 ppm
8	Alanine (Ala)	In conjunction with Lactate increases in hypoxic regions	1.48 ppm
9	Lactate (Lac)	Product of anaerobic glycolysis	1.33 ppm
10	Lips (sum of the integrated peak areas of L1 and L2 mobile lipids)	Products of brain destruction	L1 at 1.30 ppm L2 at 0.90 ppm

The chemical shift value in ppm (parts per million) refers to the frequency value of metabolite's peak in the MRS spectrum

The 0.13 ppm window was selected as being a width covering most of the peaks of interest completely, without being contaminated with neighbouring peaks. In essence, this approach attempts to extract the most characteristic features and discard the redundancy produced by noise and artifacts.

In order to combine information from the spectroscopic data with information from the MR images they need to have the same resolution and must be spatially aligned. The MR images were aligned with respect to each other, by successively shifting the T1- weighted and Gd image with respect to the PD image until a maximum of spatial correlation was reached. The MRSI spectroscopic grid was aligned with the PD image, as they were acquired consecutively. After alignment of the MR images, the image pixels that did not fit within the boundary of the STEAM (STimulated Echo Acquisition Mode) box were discarded. Then, the resolution of the remaining part of the MR images was lowered to the resolution of the MRSI grid. This was performed by averaging the image pixels which were covered by each spectroscopic voxel. The values from each image were range-scaled in agreement to the range of the spectral data. After pre-processing, each voxel within the grid was then represented by a spectrum of 230 features (in the region between 0.5 to 4.0 ppm) and 4 image features (1 variable from each MR image). Among these 230 features, a set of ratios of quantified peak areas of metabolites (Glu1, mI, Glx, Cho, Cre, Glu2, NAA, Ala, Lac, and Lips) along with 1 variable from each MR image, were selected. This is schematically represented in Figure 4.3.

4.3.2 Design of binary classification schemes

The decision, regarding which classes of patients should be compared in order to examine specific characteristics of brain tumors, is always a crucial matter. Motivated from recent studies *[Callot V. et al 2008 - Ross B. et al 1994 – Howe F.A. et al 2003(a) – Roser W. et al 1997 - Kumar A. et al 2003 – Howe F.A. et al 2003(b) – Lehnhardt F.G. et al 2005 – Galanaud D. et al 2006 – Simonetti A.W. et al 2005 –Luts J. et al 2007 – Devos A. et al 2005 – Postma G.J. et al 2011 – Kounelakis M.G. et al 2008(a) – Kounelakis M.G. et al 2008(b) - Kounelakis M.G. et al 2009]* where ^{1}H-MRSI has been applied for brain-tumor discrimination purposes but also from today's clinical practices, seven binary classification schemes, shown in Table 4.2, were investigated. This study focuses on the most significant comparisons between healthy, gliomas and meningiomas tissues in an effort to reveal the discriminative potential of the ratio features selected in Table 4.3, for these types of brain tumors.

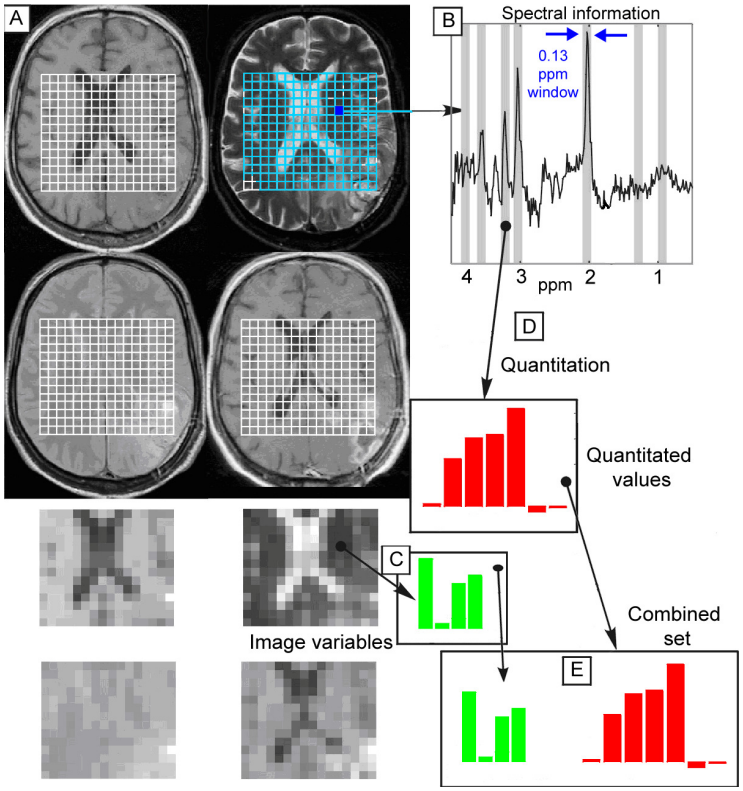


Figure 4.3 - Illustration of pre-processing and data reduction steps. A) The images with four contrasts are aligned (the lower right image has been shifted up) and the spectral grid is superimposed on the images; B) The spectrum after pre-processing obtained from the indicated voxel and the peak integration window (indicated with gray bands of 0.13 ppm width); C) The resolution of the images is reset to the MRSI resolution. Each voxel is represented by four image features; D) Data reduction is performed with quantification of important regions in the spectrum(in b); E) The image features (in green) from each voxel are combined with the MRSI features (in red) – [Internet Sources: The Radboud University]

Table 4.2 – Binary classification schemes

A/A	Binary schemes	Patients/class	Voxels/class
1	Healthy vs tumor	8 vs 24	218 vs 351
2	Healthy vs Glio	8 vs 21	218 vs 303
3	Healthy vs Mng	8 vs 3	218 vs 48
4	GRII vs GRIII	10 vs 4	176 vs 57
5	GRII vs GRIV	10 vs 7	176 vs 70
6	GRIII vs GRIV	4 vs 7	57 vs 70
7	Glio vs Mng	21 vs 3	303 vs 48

The binary classification schemes used in this study in terms of patient and voxel numbers. Glio (gliomas - integration of GRII, GRIII and GRIV), GRII (gliomas grade II), GRIII (gliomas grade III), GRIV (gliomas grade IV) and Mng (meningiomas)

4.3.3 Construction of feature sets

After a thorough investigation of recent literature, we noticed that different sets of ratio-type features have been examined in discriminating brain tumors. Most of these studies though employ ratios of only a few biologically-known metabolites (Creatine, Choline); while very few attempts combine spectroscopic data (MRS) with Magnetic Resonance Imaging (MRI) data, in the classification process.

The purpose of this study is to design a combined superset of metabolic-spectral ratio-type features (Table 4.3), not integrated in earlier studies in this way, and evaluate whether its predictive power could improve discrimination of brain tumors. For this purpose, two input feature ensembles were used, (a) the set of the 18 peak area ratios (Table 4.3) and (b) the set of 4 MR imaging intensities. As observed, the imaging intensities were also combined with the 18 peak area ratios set. Ratios of metabolites' peak areas were measured within each voxel independently, not in relation to healthy tissue. The S variable used in 5 out of the 18 ratios (NAA/S, Cho/S, Cre/S, mi/S and Ala/S) in the denominator is obtained as the sum of all metabolites within the same spectrum (voxel). This type of normalization (semi-quantitative evaluation) enables the evaluation of variations of a metabolite in the numerator *[Confort-Gouny S. et al 1993 - Galanaud D. et al 2006]*.

Many clinical MRS studies have evaluated cerebral gliomas and related changes in terms of semi-quantitative (normalized) metabolite ratios rather than absolute concentrations *[Galanaud D. et al 2010]*. Furthermore, it has been shown that the use of ratios has several advantages, with the most important being the elimination or smoothing of noise and abrupt variations observed in the initial data. Due to these reasons the normalization to S was also included in this study.

Notice here that in contrast to *[Galanaud D. et al 2006]* the S variable we use includes the mobile lipids L1 and L2 at 1.30 and 0.90 ppm respectively, because of their great importance in the evaluation of high grade gliomas spectra. These are often increased due to higher rates of lipogenesis and the existence of hypoxic-necrotic regions observed in these tumors that increase energy needs *[Auer D.P. et al 2001]*. Therefore lipids can play a significant role in gliomas discrimination.

Table 4.3 – The metabolic peak area ratios dataset

The 18 ratio features tested on the dataset					
NAA/Cre	NAA/Cho	NAA/S	Cho/Cre	Cho/S	mI/Cre
mI/Cho	mI/S	Lips/Cre	Lips/Cho	Lac/Cre	Lac/Cho
Ala/Cre	Ala/S	Cre/S	Glu1/Cre	Glu2/Cre	Glx/Cre

Lips denote the sum of L1 and L2 metabolite-peak areas and S denotes the sum of the 11 metabolite-peak areas

It was also noticed that some ratios (not including the S variable) computed on specific voxels of a patient could get extreme values This was the case in glioblastomas (GRIV), where specific ratios in 5 out of 70 voxels gave extremely high values due to near zero values of Cre's peak area in the denominator. These 5 voxels of two GRIV patients (2 from one patient and 3 from the other) were excluded from our analysis in order to avoid instabilities. Therefore, 65 out of 70 voxels were considered in GRIV. In order to evaluate the significance of metabolites from many aspects, we have used several other ratios based on their biological value, as verified in literature. It was also observed that in ratios with normalization to the S variable, the possibility of dividing by zero is further reduced.

4.3.4 Overview of applied methodology

Feature selection methods can be divided into three major categories, namely *filter, wrapper and embedded* methods *[Dietterich T.G. 1998 - Guyon I. et al 2003]*, as mentioned in Chapter 2 too. Filter approaches focus on the intrinsic properties of data in each feature direction, using various stochastic metrics such as Fisher's discriminant criterion, T-statistic (Student's t-test), Chi-square statistic (X^2 statistic), Information gain, Cross-entropy measure, Kruskall-Wallis test, Analysis of variance (ANOVA), Mann-Whitney U test and many others. Due to its operation, however, this type of feature selection methods ignores the impact of the learning algorithm.

Wrapper methods *[Kohavi R. et al 1997]* on the other hand work in a recursive way, where a classifier is used to assign a relevance weight to each feature and then the features with the lowest weights are eliminated, up to a point determined by the stopping criterion. The so called *Recursive Feature Elimination (RFE)* scheme is based on this approach. This scheme often operates on the basis of a support vector machines (SVM) classifier *[Vapnik N.V. 1999]* to evaluate features and remove those that have least value of classification, yielding the so-called RFE-SVM approach. The disadvantages of wrapper methods relate to the high computational cost of the search and their inability to take advantage of intrinsic data structures. An integration of the advantages of the above two feature selection schemes is adapted in the embedded methods *[Blum A. et al 1997]*. Such methods aim to immediately integrate the feature selection or weighting procedure into the learning algorithm of the classifier succeeding thus to retain the intrinsic characteristics of the data in the classification process.

In this study we employ an integrated feature selection method that embeds a Fisher's filter criterion within the RFE-SVM scheme. Based on the support vectors (which vary dynamically along the various steps of the feature elimination process), the Fisher's metric is calculated defining a new Fisher hyperplane at each iteration. This implies that the Fisher's criterion dynamically influences the decision based on the significance of features on the support vectors. A detailed explanation of the proposed method is presented in the following.

4.3.4.1 Feature selection and classification method

SVM was first applied in a linear, two-class (binary) classification form, where the learning scheme optimizes the separating hyperplane by maximizing the margin of class separation, as described in Chapter 2. The SVM solution is based primarily on those data samples that are at the margin of the decision boundary, called support vectors. In essence, SVM attempts to find the best separating hyperplane to distinguish between the two classes of interest, positive (+1) and negative (–1). This is done by maximizing the distance $2/\|w\|$ between the two parallel lines $(w \cdot x) + b = 1$ and $(w \cdot x) + b = -1$, which form the margin of separation of the two classes as shown in Figure 4.4. The vector w represents the direction of the hyperplane, x denotes the sample (or case) and b determines the offset of the hyperplane from the origin. The separating hyperplane passes through the middle of this margin with equation $(w \cdot x) + b = 0$.

At diagnosis stage, when a new sample x_{new} must be categorized to one of two classes (e.g. low grade tumor or high grade tumor), then the sign of the value returned by Eq. (1) indicates the predicted class associated with this new sample, while $|f(x_{new})|$ indicates the confidence level of the resulting decision.

$$f(x_{new}) = sign((w \cdot x_{new}) + b) \tag{1}$$

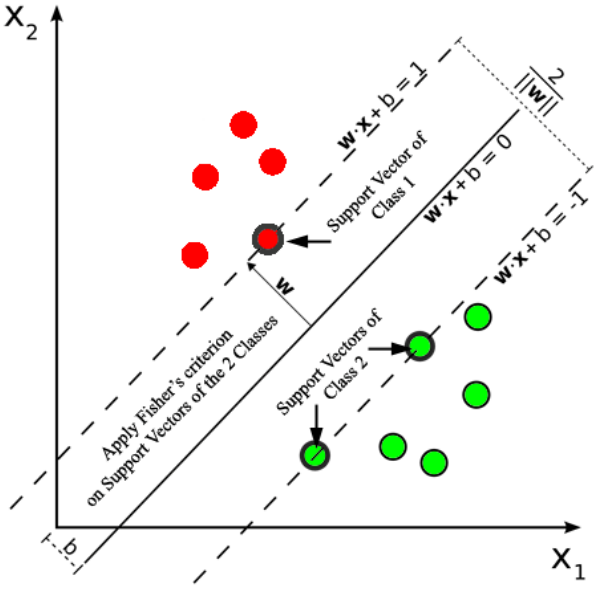


Figure 4.4 - Illustration of the binary classification topology, showing the margin of separation between the two classes; circled points represent the support vectors

According to the SVM theory, the direction of the separating hyperplane w is given as an expansion of the samples whose $\lambda_i's$ (Lagrange multipliers) are nonzero, i.e. the support vectors. If y_i denotes the label of the i^{th} sample x_i, then the direction vector w is specified as:

$$w = \sum_{i=1}^{\#SV} \lambda_i y_i x_i^{SV} \tag{2}$$

where SV indicate the support vectors on the margin of the separating hyperplane (Figure 4.4). Combining now Eq. (1) and Eq. (2) we get:

$$f(x_{new}) = sign\left(\sum_{i=1}^{\#SV} \lambda_i y_i x_i^{sv} \cdot x_{new} + b \right) \tag{3}$$

where $(\cdot)$ denotes the dot product of the two vectors

A significant property of SVM is that the learning process does not change as long as all the support vectors remain the same. Furthermore, a useful property of the support vectors is that they lie around the center (margin of separation) of the two classes which is the critical region to distinguish between the two classes of interest. Focusing on the support vectors, which actually determine the class borders, the SVM scheme elaborates on factors that could possibly misallocate the patient. Thus, the learning rule of SVMs is based on specific samples (SVs) that are supposed to be representative of the feature topology at the borders of classes. In order to further favour samples that reflect good discrimination power, we introduce a weight measure on the features, which attempts to incorporate the class-structure of the original data in the solution space. More specifically, we consider a slightly modified version of the Fisher metric *[Fisher R.A. 1936]* given as:

$$F(m_i) = \frac{|\mu_+(m_i) - \mu_-(m_i)|}{\sigma_+(m_i) + \sigma_-(m_i)} \tag{4}$$

where $\mu_{s+}, \mu_{s-}, \sigma_{s+}$ and σ_{s-} are the means and standard deviations of the distribution of the feature m_i in positive and negative classes, respectively. This metric reflects the discriminative power of the metabolic ratio m_i within the dataset of interest. In order to reflect this intrinsic characteristic right in the process of feature selection, we reformulate Eq. (2) based on this metric as follows:

$$w' = \sum_{i=1}^{\#SV} \lambda_i y_i x_i^{SV} \cdot \frac{|\mu_{s+}(m_i) - \mu_{s-}(m_i)|}{\sigma_{s+}(m_i) + \sigma_{s-}(m_i)} \tag{5}$$

Note that w' is computed based only on the support vectors, since λ_i is zero for non-support vectors, as explained in Chapter 2. Hence, the direction vector w' defined in Eq. (5) expresses a Fisher's hyperplane that passes through the origin and retains the same direction (sign) with that defined by the conventional SVMs approach. This new hyperplane can be used for defining the ranking criterion of surviving metabolic features (ratios). Rewriting Eq. (3) with respect to Eq. (5), we get in Eq. (6) a new decision function that incorporates the Fisher criterion as a weighting factor for support vectors.

$$f(x_{new}) = sign\left(\sum_{i=1}^{\#SV} \lambda_i y_i (x_i^{sv} \frac{|\mu_{s+}(m_i) - \mu_{s-}(m_i)|}{\sigma_{s+}(m_i) + \sigma_{s-}(m_i)} \cdot x_{new}) + b\right) \quad (6)$$

In the case where the two classes are not linearly separable, as in most discrimination problems related to brain tumours, then we can use different kinds of kernels, such as Polynomial, Gaussian Radial Basis Function (RBF) and Sigmoid. As stated in Chapter 2, a kernel actually creates a feature mapping from a lower dimension data space to a higher one, where separation of samples is easier (and becomes linearly separable). By using different kinds of kernels we are supplied with different sets of support vectors and, thus, different Fisher lines in our formulation. The kernel we have implemented is the widely used RBF, which is described by:

$$K(x_i^{SV}, x_{new}) = \exp(-\gamma \left\| x_i^{SV} - x_{new} \right\|^2) \quad (7)$$

where γ (gamma) is the RBF kernel's width. Finally, combining Eq. (6) with Eq. (7) we get the proposed decision function in Eq. (8):

$$f(x_{new}) = sign\left(\sum_{i=1}^{\#SV} \lambda_i y_i \exp(-\gamma \left\| x_i^{SV} \cdot \frac{|\mu_{s+}(m_i) - \mu_{s-}(m_i)|}{\sigma_{s+}(m_i) + \sigma_{s-}(m_i)} - x_{new} \right\|^2) + b\right) \quad (8)$$

Summarizing, we can say that the main goal of the proposed technique, which is also of biological interest, is the identification of the metabolic-feature topology that forces a specific patient to cross the border of separation from one class to the other. Focusing on and appropriately weighting the support vectors, this formulation reveals possible features responsible for misallocating a patient. Such features can then be eliminated from further consideration. Furthermore, by selecting different kinds of kernels (in addition to RBF used in this study) we can obtain a variety of support vectors, which can also be viewed as different sets of domain representatives.

4.3.4.2 Flowchart of the method

The functional operation of the feature selection and classification method is presented in Figure 4.5. The RFE process illustrates how the ratio features are selected as significant or not, based on their Fisher's value. More specifically, the criterion ranks the input features on the support vectors according to Eq. (4). It should be noticed here that the identification of support vectors is realized prior to the application of the Fisher's criterion. Thus, all data samples are considered in order to locate the support vectors and subsequently the Fisher's criterion is applied only on them.

The RFE process adopted removes one ratio feature per iteration. The feature with the smallest Fisher's value (least significant) is removed and the rest (most significant ones) are kept and tested for their discriminative potential, yielding a specific Area under the Receiver Operating Characteristic Curve (AUROC) value. The process stops when the highest AUROC accuracy is obtained. At this point, the smallest ratio set deriving the highest accuracy for this experiment is also obtained. This process is repeated 100 times (10folds X 10 times = 100 runs) within a 10-fold cross validation strategy. Each run derives a "highest" AUROC value and the average AUROC score is derived. Furthermore, the CV process derives 100 (smallest) sets of ratio features. From those 100 sets we calculate the frequency of appearance of each feature. Then, the derived final set of markers (Table 4.5) contains those ratio features with the highest frequency of appearance.

For the classification process that follows feature selection, both the SVM classifier and its least squares variant (LS-SVM) were tested. LS-SVM *[Suykens J.A.K et al 1999]* forms another type of SVM classifiers where instead of the inequality constraints (leading to quadratic programming), equality constraints are used leading to a system of linear equations**.** According to the theory of LS-SVM, every data point in the LS-SVM classifier is a support vector because, in general, none of the Lagrange multiplier (λ_i) equals zero. This implies that all cases are considered here, in contrast to the SVM case where only some cases at the borders are involved as SVs.

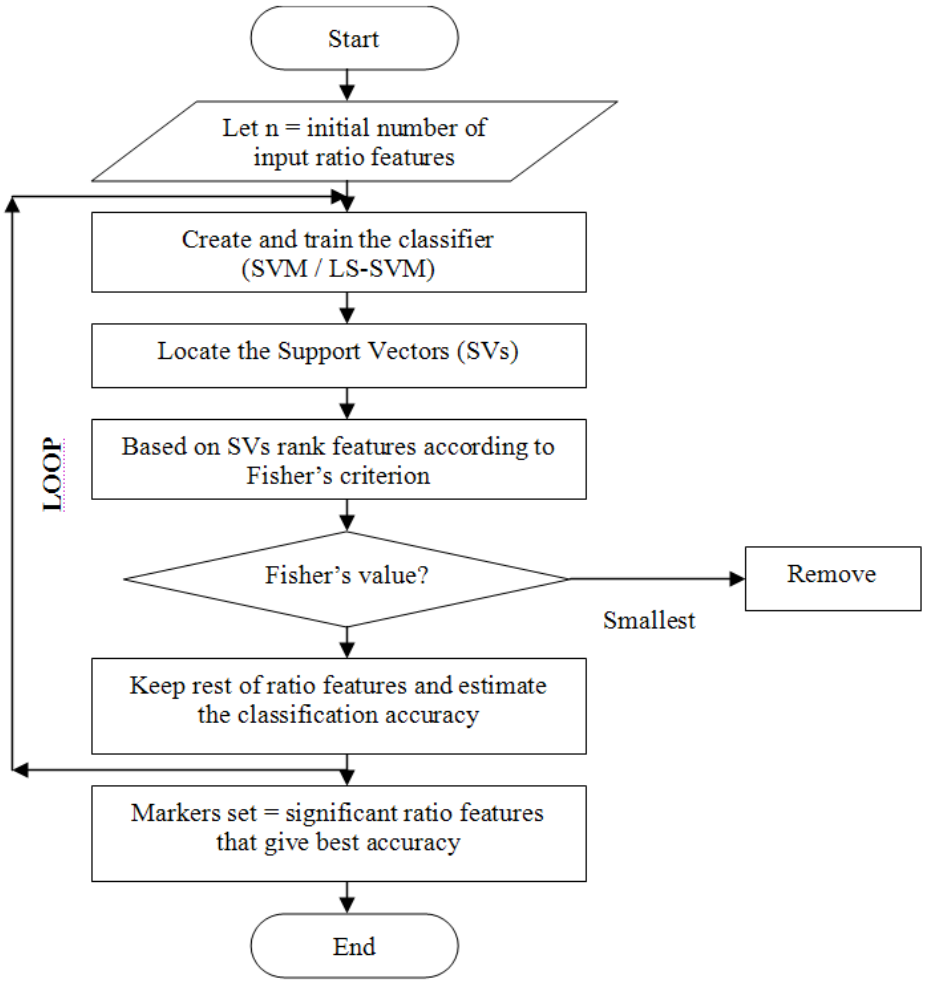


Figure 4.5 - The flowchart of the feature selection and classification method applied in this study

In both SVM and LS-SVM, a kernel is utilized to map the input data to a higher dimensional feature space, so that the problem becomes linearly separable. As mentioned before, the Gaussian RBF kernel in Eq. (7) was tested in this study, since it is considered a good choice when heterogeneous data are under scrutiny. In order to optimize the classifier's performance the two hyperplane parameters γ (gamma) and C (soft-margin) were adjusted. More specifically, the γ parameter controls the flexibility of the kernel applied. Large values of the γ parameter increase the curvature of the decision boundary, which inevitably leads to overfitting. In contrast, very small values make the kernel nearly liner. On the other hand, the C hyperplane parameter determines the margin-error penalty of the classifier, as mentioned in Chapter 2. Large values of this parameter decrease the margin of error, while small values increase it. The optimization of these parameters was achieved through a grid-search process. The two parameters were tuned in each one of the 100 runs of cross-validation, both at training and testing stage, in order to increase the classifier performance and avoid overfitting of data. It was noticed that the best pair of values was usually $\gamma = 10^{-4}$ and C=1000.

The software package used for the feature selection and classification process was implemented in Matlab ver. R2010 *[Internet Source: http://www.mathworks.com]*.

4.3.4.3 Classifier performance evaluation

The 10-fold Cross Validation (CV) and Leave-One-Patient-Out CV strategies were tested. Repeated stratified runs (10 folds X 10 times = 100 runs) were applied, which according to bibliography is suitable for cases where data samples are few, as in the glioma grade III and meningioma classes *[Kohavi R. 1995]*. Train sets contained 90% of the voxels of each of the two classes and the remaining 10% of these two classes were contained in the test sets. Instead, the Leave-One-Patient-Out CV scheme extracts all the voxels from each patient separately and keeps them out for testing. The remaining patients are used for the training set. Notice that the number of voxels of each patient in each class varies, as shown in Table 4.2. The test sets were constructed based on the principle that every patient from each class would be involved at least once in the test set, so that the entire dataset could be tested at least once.

4.3.5 Comparison with others studies

In comparison to recent studies *[Simonetti A.W. et al 2005 – Luts J et al 2007 – Devos A. et al 2005 – Postma G.J et al 2011]* using the same MRSI and MRI data for feature selection and classification of brain tumors, this study expands on several points.

Simonetti A.W. et al in 2005 applied four different feature reduction procedures: simple quantization, PCA, ICA and LCModel using as input the peak areas of 8 metabolites and 4 MRI features. In other words and in contrast to this study, Simonetti A.W. et al did not apply any wrapper evaluation scheme (with RFE) to rank these features according to their discriminative potential.

On the other hand Luts J. et al in 2007, applied 4 different feature selection methods (Fisher discriminant criterion, Kruskal-Wallis test, Relief-F, Automatic Relevance Determination), each one embedded in a LS-SVM classifier, using as input the peak areas of 10 metabolites and 4 MRI features. The major difference from our study is on the way these filter criteria are embedded in the classifier. In our study, the ranking of the ratio features based on Fisher's criterion involves only the support vectors whereas in Luts J. et al study, filtering is applied on whole data samples. Thus, the main goal of the proposed methodology, which is also of biological interest, is the identification of the feature topology that forces a specific patient to cross the border of separation from one class to the other. Focusing on the support vectors that determine the class borders, our approach attempts to identify the factors that could possibly misallocate a patient. Furthermore, the detailed characterization of the class-boundaries through the support vectors enhances the confidence of the classification results. In addition, in Luts J. et al study, 10 binary schemes were formulated for classification purposes, with the two classes of gliomas grade II and III being broken down to 6 subclasses.

Nevertheless, the exhaustive separation of patients into sub groups involves two important drawbacks. First, it does not provide good generalization potential on the classification results (statistical analysis) since the number of patients in each class is significantly reduced and second, significant information that could better explain the metabolic profile of these tumors might be masked within similar subclasses.

Devos A. et al in 2005 applied only binary classification procedures using two well known classifiers (LDA, LS-SVM) without focusing either on feature selection or on the discriminative value of the features involved.

Finally, Postma G.J. et al in 2011 compared 3 filter methods (Relief-F, Kruskal-Wallis test and Fisher discriminant criterion), 1 wrapper (forward selection with LDA) and 1 embedded (ARD with LS-SVM) based on metabolic features known from literature, in order to evaluate their relevance in differentiation of two brain tumors datasets. The direction of our study is quite different, focusing on the hypothesis that ratio-type features enhance the predictive power of the classifier. This was supported by the fact that the MRSI ratios alone provided high classification rates, whereas in Postma G.J. et al study the same was achieved using the MRI information.

4.4 Experimental Results

The evaluation of the classification performance was performed by means of the global metric representing the *Area under the Receiver Operating Characteristic Curve (AUROC)*. Confidence intervals (CI) around the best AUROCs were also estimated for statistical purposes. For each binary classification, shown in Table 4.2, the input dataset contained all the 18 peak area ratios. Using this dataset as input to the feature selection and classification algorithm described above, we estimate both the best AUROC in each binary classification scheme and the minimum number of ratio features needed (i.e. the ratio markers) to achieve that AUROC. The best AUROC values measured (average over CV runs) for the binary tests are reported in Table 4.4. The corresponding numbers of ratio features are also presented in Table 4.4. Bold figures indicate the highest AUROC scores achieved.

Comparing the CV approaches tested, it is observed that the 10-fold CV gives better AUROC results compared to the Leave-One-Patient-Out method and the SVM prediction model appears to be appropriate for such binary classifications. In other words, the SVM classifier in a 10-fold CV scheme provides the best classification rates in all binary classification schemes. This is expected when compared to the Leave-One-Patient-Out CV approach, which excludes all voxels from at least one patient from the training phase, thus reducing the learning capacity and rendering the testing process less efficient. This effect is more obvious for the GRII vs GRIII scheme were the number of GRIII patients is already small and does not allow loss of information from an entire patient. In this case, a larger pool of samples is needed to enable more accurate consideration of this validation scheme. Furthermore, the LS-SVM classifier, even though simpler in form, cannot adapt to the complex structure of the binary problems and generally yields lower accuracies compared to the SVM scheme. This probably relates to the fact that in LS-SVM every patient is considered as support vector, which makes the classifier less sensitive to class boundaries, leading to slightly smaller AUROC values and larger marker sets compared to the SVM classifier.

Table 4.4 – Best AUROC values measured (average over CV runs) for the binary tests and the numbers of most significant features.

Classifier evaluation method	***10-Fold CV***				***Leave-One-Patient-Out CV***	
Feature selection method	SVM^{R}	SVM^{R+I}	$LS\text{-}SVM^{R}$	$LS\text{-}SVM^{R+I}$	SVM^{R}	$LS\text{-}SVM^{R}$
Healthy vs tumor	**0.99** / **6** (0.004)	**0.99** / **8** (0.008)	0.98 / 9 (0.005)	0.98 / 10 (0.006)	0.97 / 5 (0.006)	0.98 / 8 (0.009)
Healthy vs Glio	**0.99** / **6** (0.005)	**0.99** / **7** (0.005)	0.99 / 9 (0.004)	0.98 / 10 (0.005)	0.98 / 8 (0.006)	0.98 / 6 (0.007)
Healthy vs Mng	**0.98** / **2** (0.009)	**0.99** / **6** (0.007)	0.98 / 4 (0.005)	0.98 / 7 (0.005)	0.90 / 9 (0.008)	0.97 / 13 (0.008)
GRII vs GRIII	**0.84** / **8** (0.028)	**0.85** / **12** (0.030)	0.84 / 9 (0.040)	0.81 / 11 (0.040)	0.78 / 7 (0.045)	0.63 / 12 (0.042)
GRII vs GRIV	**0.99** / **4** (0.004)	**0.99** / **6** (0.006)	0.99 / 5 (0.006)	0.98 / 7 (0.007)	0.99 / 4 (0.008)	0.97 / 4 (0.009)
GRIII vs GRIV	**0.98** / **2** (0.013)	**0.98** / **2** (0.014)	0.98 / 2 (0.015)	0.97 / 3 (0.015)	0.97 / 3 (0.016)	0.97 / 1 (0.016)
Glio vs Mng	**0.97** / **6** (0.010)	**0.97** / **10** (0.018)	0.92 / 8 (0.013)	0.92 / 10 (0.015)	0.89 / 10 (0.018)	0.84 / 10 (0.020)

Bold figures correspond to the highest AUROC scores achieved. The figures at the right side of the slash indicate the number of features achieved at best AUROC. The R superscript corresponds to the case of MRSI features only, while R+I correspond to the combination of MRSI and MRI features. The numbers in the parenthesis represent the confidence interval (CI) at best AUROCs

In healthy vs tumors, healthy vs gliomas and healthy vs meningiomas, the AUROC measures achieved by the SVM classifier in 10-fold CV are 0.99, 0.99 and 0.98 respectively. It is remarkable that only 2 ratio features can discriminate healthy from meningiomas with an AUROC value of 0.98. It is also observed that in healthy vs tumor test, the number of features needed to achieve 0.99 AUROC are 6; similar to the case of healthy vs gliomas, but with different ratio markers, as shown in Table 4.5. This is reasonable, since the tumor class is more abstract, containing both gliomas and meningiomas.

In the cases of gliomas, the AUROC measure at GRII vs GRIV and GRIII vs GRIV reach again high values of 0.99 and 0.98 respectively, with a small number of features. This is not the case though for GRII vs GRIII, where the AUROC score is 0.84. This is explained by the fact that these two classes present great heterogeneity (mixed cases of diffuse astrocytomas

and oligodendrogliomas in GRII and anaplastic astrocytomas and oligodendrogliomas in GRIII) as mentioned in subsection 4.3.1. Finally, in gliomas vs meningiomas case, the AUROC score is quite high (0.97).

Table 4.5 presents the smallest group of ratio features captured at the highest accuracy per binary scheme. We observe that the ratios of NAA, Cho, and Lips contribute most in the discrimination of healthy from the tumors (gliomas, meningiomas). The ratios of NAA, mI and Cho, in addition to those of Lac and Ala, assist the discrimination of healthy from gliomas. mI/S and NAA/S ratios are able to discriminate healthy from meningiomas. Ratios that involve Cre, Lips, Lac, along with Ala, NAA and mI can be used to discriminate low grade gliomas (GRII) from intermediate (GRIII) while Cre, Cho, mI and Ala can are capable to distinguish low grade form high grade (GRIII vs GRIV). The metabolic ratios of Cho and mI can be used in the GRIII vs GRIV discrimination Furthermore, ratios containing Ala, NAA and mI are important in gliomas vs meningiomas separation. It can also be observed that the ratios that involve the S variable as denominator are quite frequent in all binary schemes. The role of the brain metabolites mentioned above in brain cancer classification is also verified in recent literature, as presented in the Discussion section below.

Adding the MR image features into the feature selection and classification process, slightly improves the AUROC measures in 2 out of 7 classification schemes (Table 4.4). As shown, in Healthy vs Mng and GRII vs GRIII cases, there was an increase of the AUROC values. More specifically, the SVM classifier under a 10-fold CV scheme was applied in each binary scheme, but with input feature set now fixed to the ratio markers already derived (shown in Table 4.5) together with the 4 image features obtained form T1, T2, PD and Gd MR images. At this point we combine the derived set of MRSI ratio markers with the 4 MRI markers, since we want to explore if the inclusion of the MRI features influences (positively or negatively) the best AUROC values obtained from MRSI markers alone and therefore consider the incremental power of MRI features in class discrimination. The SVM classifier was once more selected since, compared to LS-SVM, provided the best AUROC values when only spectroscopic features were used. The 10-fold CV method was applied for similar reasons.

It was observed that all the spectroscopic markers (ratios) were selected as important for classification purposes when combined with the MRI features. In other words none of the ratio features was discarded by the process or substituted by an MRI feature, which verifies that the discriminative potential of the selected markers is significant. On the other hand, the number of features needed to obtain the same or slightly better classification rates was increased when MRI features included.

Among the four MRI features, T2 and Gd were selected in healthy vs tumor classification while only T2 in healthy vs gliomas. All MRI features (T1, T2, PD and Gd) were selected as important in healthy vs Mng, gliomas vs Mng and GRII vs GRIII. Also, features T2 and Gd found to be significant in GRII vs GRIV.

Using all MRI features in GRIII vs GRIV scheme the AUROC achieved was not differentiated. Overall it can be claimed that in most binary schemes, T2 and Gd have a

greater influence than T1 and PD. Finally it is worth noticing here that although the integration of MRI with MRSI features caused a slight improvement in the classification accuracies, which is also mentioned in recent studies *[Simonetti A.W. et al 2005 – Luts J et al 2007 – Devos A. et al 2005 – Postma G.J et al 2011]*, their discriminative value in our study was overshadowed by the strength of the MRSI features. The fact that MRSI features alone can give high classification rates without the support of MRI features has been also mentioned before *[Galanaud D. et al 2006 – Herminghaus S. et al 2003]*.

Table 4.5 – The metabolic behaviour of ratio markers in each binary classification scheme.

Binary schemes	***Ratio markers metabolic behavior and statistical significance***							
Healthy vs tumor	NAA/ Cho	NAA / S	Cho/ S	NAA/ Cre	Lips / Cho	Cho/ Cre		
	↓**	↓**	↑**	↓*	↑*	↑*		
Healthy vs Glio	NAA / Cho	mI / S	Cho / S	Cho / Cre	Lac / Cre	Ala / Cre		
	↓**	↓**	↑**	↑*	↑*	↑*		
Healthy vs Mng	mI / S	NAA / S						
	↓**	↓*						
GRII vs GRIII	Cre / S	Lips / Cre	Lac /Cre	Ala / S	Ala / Cre	NAA/ Cre	Lips / Cho	mI / S
	↓**	↑**	↑**	↑**	↑**	↓*	↑*	↓*
GRII vs GRIV	Cre / S	Cho / S	mI / S	Ala / S				
	↓**	↑**	↓**	↑**				
GRIII vs GRIV	Cho / S	mI / S						
	↑**	↓**						
Glio vs Mng	Ala / Cre	Ala / S	NAA / S	mI / S	Lips / Cre	Lac / Cre		
	↑**	↑**	↓**	↑**	↓*	↓*		

Double asterisks (**) correspond to highly-significant changes in the mean of the ratio marker, while single asterisks (*) indicate significant change. Upward direction of the arrow corresponds to an increase and downward direction to a decrease of the mean of the ratio marker in the second class

4.4.1 Statistical significance of the metabolic markers identified

Another important issue concerns the statistical significance of the features selected as most important by the RFE-SVM method, shown in Table 4.5. Towards this direction, the statistical significance of the difference-of-means, in terms of p-values, was computed using the Independent-Samples t-test (two-tailed). The independent-samples-t-test is actually used to re-validate the statistical significance (in terms of the difference of their means) of the features selected by the feature selection process. Although similar to Fisher's criterion, the t-test is applied independently on each feature as a filter criterion, whereas the Fisher metric is applied within the classifier as a wrapper criterion. The latter case considers the predictive power of group of features, whereas the former one expresses the class-discriminating power of each feature. The markers selected according to their Fisher's values are inputted into the SVM classifier that "decides" which of the markers contribute in discriminating glioma patients. These discriminating features might not necessarily be statistically significant. Thus, the application of the t-test on the RFE results further validates the efficiency of the derived markers. Table 4.5 illustrates the p-values computed. We considered two thresholds of significance. When $p < 0,05$ the difference of means is considered statistically significant and when $p < 0,001$ it is considered highly significant. Statistical analysis was applied using SPSS ver. 19.0 software tool [*http://www.spss.com*].

The examination of changes in the means of ratio markers between classes and the associated statistical significance, allows for a clear interpretation of the metabolic behavior of these markers, as indicated in Table 4.5. In this interpretation, the upward direction of the arrow indicates an increase in the mean of the ratio in the second class while the downward direction a decrease. An arrow with a subsequent double asterisk reflects a highly significant change in the mean of the ratio feature overhead, while a single asterisk a significant change.

4.5 Discussion

In contrast to recent studies that mainly focus on estimating the efficiency of a predefined feature set in the discrimination of brain tumors, we applied a wrapper feature selection and classification method aiming to extract the smallest possible sets of markers that enable accurate classification of a new brain tumor patient into one of the classes considered. In other words, starting from a large input feature set with 18 potentially useful ratio features, we concluded to 7 compact feature sets (one for each binary classification scheme), as shown in Table 4.5. These sets are derived from the methodology applied and contain the fewest and most significant ratio features, which yield the highest AUROC scores in the corresponding binary tests. In addition, the methodology applied is voxel-based; it formulates one feature vector for each voxel from the MRSI (and MRI) data, so that the most significant features are extracted from the dominant voxels emerging as support vectors that best separate the classes. The cross-validated results obtained indicate that this methodology can provide high classification rates, even when some classes contain only a few patients (as in meningioma).

In related attempts, Kumar A. et al in 2003 investigated the importance of Cho/Cre, NAA/Cre and NAA/Cho ratio features for the discrimination of gliomas and meningiomas. Howe F.A. et al in 2003(b) used mI/Cho Cho/Cre, to examine their potential in gliomas, meningiomas and metastasis discrimination. Lehnhardt F.G. et al in 2005 examined the power of Cho/Cre, Ala/Cre, Glycine (Gly)/Cre, Glu1/Glu2 ratio type features to discriminate primary from recurrent gliomas and meningiomas. Galanaud D. et al in 2006 investigated only three ratio type features, NAA/Cre, NAA/Cho and Cho/Cre, originating from multi voxel spectra (TE 135 ms) and a limited set of ratio features originating from single voxel spectroscopy. Furthermore, Galanaud D. et al introduced a new variable called S (sum of all metabolites' peak areas of the same spectrum, except lipids) used as denominator in the ratio features, to distinguish different human brain tumors.

Although the above cited studies also focus on revealing the classification significance of specific ratio-type features, they differ from our study in both the methodology and the limited number of investigated ratio features. In contrast to Galanaud D. et al who perform one-versus-all classification, our study examines seven binary classification schemes enabling the identification of both similarities and differences between stages of the pathology. Furthermore, in our study the ratio features are measured in an intratumoral way, not in respect to a normal (healthy) reference as in Galanaud D. et al study. This fact reveals the intrinsic characteristics of these tumors enabling the SVM-based classifiers to reach high classification rates. Compared to Galanaud D. et al, our study achieves higher classifications rates (0.84 vs 0.67); especially in the case of gliomas grade II vs grade III, using only MRSI features. Finally, in Galanaud D. et al study the 4 MRI-based features are assigned to each patient by manually scoring the images, similar to Asari S. et al in 1994. In contrast, the 4 MRI features used in our study are generated by averaging the image intensities within each voxel of the T1, T2, Pd and Gd images of each patient.

Most of the recent studies mainly use Creatine and Choline as denominators in their input ratio-feature sets. Apart from these two metabolites, we make use of the S and Lips features. It should be noticed here that the S variable in our work, in contrast to Galanaud D. et al work, contains the mobile lipids at 0.9 and 1.3 ppm. As mentioned before, the lipids are often increased in high grade tumors as a result of higher rates of lipogenesis due to their increased energy needs. Therefore, they can play a significant role in high grade gliomas discrimination *[Auer D.P. et al 2001]*. This fact facilitates the easier discrimination of GRII vs GRIII gliomas classes, where the lipids play an important role as observed in Table 4.5.

Most of the features selected by our proposed methodology for the healthy vs gliomas classification correspond to those mentioned in literature *[Howe F.A et al 2003(a) – Kumar A. et al 2003 -Majós C. et al 2002 – Castillo M. et al 2000 – Heerschap A. 2007]*, where the behavior of known metabolites in brain tumors is analyzed. Low grade gliomas (GRII) are discriminated from intermediate and high grade gliomas (GRIII and GRIV) using the ratios of Cre, Lips, Lac, Cho, mI and Ala metabolites' peak areas. Furthermore, the contribution of the ratios using the S in their denominator it was remarkable in the gliomas binary schemes as shown in Table 4.4 and also stated in relevant studies *[Galanaud D. et al 2006 – Tan W. et al 2008 - Likavcanová K. et al 2005]*. This is more obvious in the cases of GRII vs GRIV (Cre/S, Cho/S, mI/S and Ala/S) and GRIII vs GRIV (Cho/S, mI/S). Finally, gliomas can be discriminated from meningiomas using ratios of Ala, NAA, mI as well as Lips and Lac metabolites' peak areas. Overall, comparing the behavior of means of these markers in Table 4.5, it becomes obvious that NAA and Cre decrease in tumors, while Cho increases. In addition, Lipids and Lac show a significant increase in gliomas, whereas Ala and mI increase in meningiomas. These results are in agreement with the findings of the aforementioned references.

Although this study mainly focuses in revealing the smallest size and the most discriminant group of ratio features per binary scheme, it reveals additional metabolic ratio features, whose importance has to be addressed, as well. Thus, in healthy vs meningiomas test the Ala/Cre ratio is increased and this change was found to be statistically significant. In addition, the ratios Cho/Cre, Cho/NAA, Lips/Cre, Lips/Cho, Lac/Cre, Lac/Cho, all increased significantly in intermediate and high grade gliomas (GRIII and GRIV). The reason that these metabolic ratio features were not included in the markers' sets (Table 4.5) is mainly because they have not further improved the classification accuracy within the RFE-SVM process and they were less frequently selected by the classifier compared to these already presented in Table 4.5. In other words the features presented in this table are considered enough to achieve the highest accuracies.

Comparing the classification results of this study with those achieved by similar studies *[Simonetti A.W. et al 2005 – Luts J et al 2007 – Devos A. et al 2005 – Postma G.J et al 2011]* using the same dataset, high classification rates (above 0.97 AUROC) were also obtained in our study, in 6 out of 7 binary schemes. In GRII vs GRIII case, commonly described as the most heterogeneous binary scheme, the AUROC value obtained was 0.84, which is

considered quite satisfactory regarding the complexity of these classes. This fact reveals the reliability and efficacy of the methodology applied.

Comparing now the ratio markers derived in this study with those examined in these studies, we can conclude the following.

Similar to Kumar A. et al in 2003, this study also supports that Cho/Cre, NAA/Cre and NAA/CHO are significant in discriminating tumors from normal (healthy) tissues. Nevertheless, in this study exhibits the significance of other ratio type features like NAA/S, Cho/S and Lips/Cho which must also be considered in order to achieve high classification results. In Howe F.A et al 2003(b) study, mI/Cho found to play important role in tumor discrimination. Indeed, in our study too, mI was found to be very low in meningiomas and high in low grade gliomas, whereas Cho increased in gliomas as grade increased. Furthermore, Cho/Cre (or Cre/Cho), but also Lips and Lac all increased in tumors and were considered important in contrast to Howe F.A et al study. In relation to Galanaud D. et al study, the results of our study agree on the fact that the S variable plays a significant role in the classification process. In our study though the inclusion of mobile lipids in the S variable, not considered in Galanaud D. et al study, reveals also the important role of lipogenesis in brain gliomas, especially in those with high grade of aggressiveness.

Combining the spectroscopic features with the MRI features derived from T1, T2, PD and Gd imaging modalities slightly improved the classification results in 2 out of 7 classification cases. Among the 4 imaging features, T2 and Gd were found most significant in this study. Their role has been identified in recent studies *[De Edelenyi F.S. et al 2000 – Agnoli A.L. et al 1987]* too. Nevertheless, the integration of MRI features, although important, increased the number of features needed to achieve the same AUROC values. This fact emphasizes that the discriminative potential of the derived MRSI ratio markers is very significant. In essence, the ratio features have the strength to reveal intrinsic characteristics of these brain tumors and the combination with MRI features improves only slightly the classification accuracy.

4.6 Conclusions

The identification of biologically and clinically significant sets of markers is necessary as to assist today's medical practice in achieving better discrimination of complex brain tumours *[Kounelakis M.G. et al 2008(a,b) – Kounelakis M.G. et al 2009]*. This is mainly due to the fact that, even though the clinical profile between two patients might be similar, their metabolic profile may differ. Therefore, these metabolic differences must be taken under consideration by the clinicians in order to provide a more efficient patient-based treatment.

The outcome of this study indicates that embedded feature selection and classification methods, where the intrinsic properties of patient data are taken under consideration, can actually assist the derivation of reliable sets of metabolic markers towards a more accurate type and grade discrimination of complex brain tumours. Apart form the already known metabolic ratios, others such as NAA/S, mI/S, Cho/S, Cre/S and Ala/S introduced in this study, are highly significant for the classification of new brain gliomas or meningiomas patients. The importance of these markers has been revealed through both statistical and biological means. In addition to the evaluation of marker efficiency, this study derives the minimal sets of significant markers for specific binary classification tests.

References

Agnoli A.L. et al 1987, 'Magnetic resonance imaging of brain tumors: application of gadolinium-DTPA and comparison to computed tomography', *Neurosurgical Review*, vol. 10, pp. 25-29.

Asari S. et al 1994, 'Assessment of the pathological grade of astrocytic gliomas using an MRI score', *Neuroradiology*, vol. 36, pp. 308-310.

Auer D.P. et al 2001, 'Improved analysis of ^{1}H-MR spectra in the presence of mobile lipids', *Magnetic Resonance in Medicine*, vol. 46, pp. 615–618.

Blazadonakis M.E. et al 2006, 'Support Vector Machines and Neural Networks as Marker Selectors for Cancer Gene Analysis', *3rd IEEE International Conference on Intelligent Systems*, London.

Blazadonakis M.E. et al 2008, 'Wrapper filtering criteria via linear neuron and kernel approaches', *Computers in Biology and Medicine*, vol. 38, pp. 894-912.

Blum A. et al 1997, 'Selection of relevant features and examples in machine learning', *Artificial Intelligence*, vol. 97, pp. 245–271.

Callot V. et al 2008, '^{1}H MR spectroscopy of human brain tumours: a practical approach', *European Journal of Radiology*, vol. 67, pp. 268-74.

Castillo M. et al 2000, 'Correlation of myo-inositol levels and grading of cerebral astrocytomas', *American Journal of Neuroradiology*, vol. 21, pp. 1645-1649.

Confort-Gouny S. et al 1993, 'A multiparametric data analysis showing the potential of localized proton MR spectroscopy in the brain in the metabolic characterization of neurological diseases', *Journal of Neurological Sciences*, vol.118, pp.123–133.

De Edelenyi F.S. et al 2000, 'A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images', *Nature Medicine*, vol. 6, pp. 1287-1289.

Devos A. et al 2005, 'The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification', *Journal of Magnetic Resonance*, vol. 173, pp. 218-28.

Dietterich T.G. 1998, 'Machine learning research: Four current directions', *The Artificial Intelligence Magazine*, vol. 18, pp. 97–136.

Fisher R.A. 1936, 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, vol. 7, pp. 179–188.

Galanaud D. et al 2006, 'Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy', *Magnetic Resonance in Medicine*, vol. 55, pp. 1236-1245.

Galanaud D. et al 2010, 'Combined Diffusion Imaging and MR Spectroscopy in the Diagnosis of Human Prion Diseases', *American Journal of Neuroradiology*, vol. 31, pp. 1311-1318.

Guyon I. et al 2003, 'An introduction to variable and feature selection, *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182.

Heerschap A. 2007, 'In vivo magnetic resonance spectroscopy in clinical oncology. in Shields A.F. and Price P. (eds) Cancer Drug discovery and Development In Vivo Imaging of Cancer Therapy, *Humana Press Inc.,* Totowa, pp. 241-258.

Herminghaus S. et al 2003, 'Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data', *Journal of Neurosurgery*, vol. 98, pp. 74–81.

Howe F.A. et al 2003a, '^{1}H MR spectroscopy of brain tumours and masses', *NMR in Biomedicine*, vol. 16, pp. 123-131.

Howe F.A. et al 2003b, 'Metabolic Profiles of Human Brain Tumors Using Quantitative In Vivo ^{1}H Magnetic Resonance Spectroscopy', *Magnetic Resonance in Medicine*, vol. 49, pp. 223-232.

Kohavi R. 1995, 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *Proceedings of 14th International Joint Conference on Artificial Intelligence*, pp 1137-1143.

Kohavi R. et al 1997, 'Wrappers for feature subset selection', *Artificial Intelligence*, vol. 97, pp. 273–324.

Kumar A. et al 2003, 'Role of in vivo proton MR spectroscopy in the evaluation of adult brain lesions: Our preliminary experience', *Neurology India*, vol. 51, pp. 474-478.

Kounelakis M.G. et al 2011, 'Embedding filtering criteria into a wrapper marker selection method for brain tumors classification: An application on metabolic peak area ratios', *IOP Measurement Science & Technology*.

Kounelakis M.G. et al 2009, 'Revealing the metabolic profile of brain tumors for diagnosis purposes', *31st IEEE Annual International Conference of the Engineering in Medicine and Biology Society*, Minnesota, pp. 35-38.

Kounelakis M.G. et al 2008a, 'Identification of significant Metabolic Markers from MRSI data for Brain Cancer Classification', *8th IEEE International Conference on BioInformatics and BioEngineering*, Athens, pp. 1-6.

Kounelakis M.G. et al 2008b, 'Feature Selection for Brain Tumour Classification using Ratios of Metabolites' Peak Areas from MRSI Data', *6th European Symposium on Biomedical Engineering*, Chania, pp. 1-6.

Lehnhardt F.G. et al 2005, 'Metabolic differences between primary and recurrent human brain tumors: a ^{1}H NMR a ^{1}H NMR spectroscopic investigation', *NMR in Biomedicine*, vol. 18, pp. 371-382.

Likavcanová K. et al 2005, 'In vitro study of astrocytic tumor metabolism by proton magnetic resonance spectroscopy, *General Physiology and Biophysics*, vol. 24, pp. 327-335.

Luts J. et al 2007, 'A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection', *Artificial Intelligence in Medicine*, vol. 40, pp. 87-102.

Majós C. et al 2002, 'Adult primitive neuroectodermal tumor: proton MR spectroscopic findings with possible application for differential diagnosis', *Radiology*, vol. 225, pp. 556-566.

Pijnappel W.W.F. et al 1992, 'SVD-based quantification of magnetic resonance signals', Journal of Magnetic Resonance, vol. 97. pp. 122–134.

Postma G.J. et al 2011, 'On the relevance of automatically selected single-voxel MRS and multimodal MRI and MRSI features for brain tumour differentiation', *Computers in Biology and Medicine*, vol. 41, pp. 87-97.

Roser W. et al 1997, 'Assignment of glial brain tumors in humans by in vivo H-magnetic resonance spectroscopy and multidimensional metabolic classification', *MAGMA*, vol. 5, pp. 179-183.

Ross B. et al 1994, 'Clinical applications of magnetic resonance spectroscopy', *Magnetic Resonance Quarterly*, vol. 10, pp. 191-247.

Simonetti A.W. et al 2002, 'Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction', *Journal of Magnetic Resonance*, vol. 159, pp. 151–157.

Simonetti A.W. et al 2003, 'A Chemometric Approach for Brain Tumor Classification Using Magnetic Resonance Imaging and Spectroscopy', *Analytical Chemistry*, vol. 75, pp. 5352-5361.

Simonetti A.W. et al 2005, 'Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification', *NMR in Biomedicine*, vol. 18, pp. 34-43.

Suykens J.A.K. et al 1999, Least squares support vector machine classifiers', *Neural Processing Letters*, vol. 9, pp 293-300.

Tan W. et al 2008, 'Preliminary study in vitro brain tumor with high resolution magic angle spinning proton magnetic resonance spectroscopy', *Chinese-German Journal of Clinical Oncology*, vol. 7, pp. 103-106.

Vapnik N. V. 1999, 'The Nature of Statistical Learning Theory', *Springer –Verlag*, New York.

Internet sources

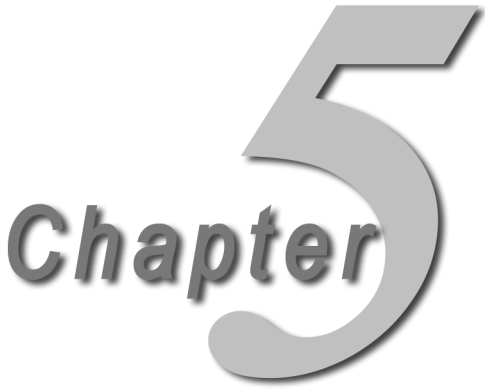
Hope Clinics, http://hope-clinics.com

Matlab, http://www.mathworks.com

SPSS, http:///www.spss.com

The Radboud University, http://www.ru.nl

The UMKC Radiology Resident Resource, http://www.umkcradres.org



REVEALING THE BIOLOGICAL SIGNIFICANCE OF GLYCOLYSIS AT GENOMIC AND METABOLIC LEVEL: a study on brain gliomas

Based on Author's published studies:

- ***M. G. Kounelakis***, M. E. Zervakis, G. C. Giakos, G. J. Postma, L. M. C. Buydens and X. Kotsiakis, On the relevance of glycolysis process on brain gliomas, *IEEE Transactions on Information Technology in BioMedicine, 2011 (Under Review)*

- ***M. G. Kounelakis***, M. Zervakis, G. C. Giakos and X. Kotsiakis, The role of the Glycolysis-related genes in Brain gliomas treatment design, *5th European Conference of IFMBE for Medical & Biomedical Engineering (EMBEC), Budapest, 2011*

- ***M. G. Kounelakis***, M. E. Zervakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis, Validation of MRS metabolic markers in the classification of brain gliomas and their correlation to energy metabolism, *12th Mediterranean Conference on Medical and Biological Engineering and Computing, Chalkidiki, vol. 29, pp. 33-36, 2010*

- ***M. G. Kounelakis***, M. E. Zervakis, G. C. Giakos, C. Narayan, S. Marotta, D. Natarajamani, G. J. Postma, L. M. C. Buydens and X. Kotsiakis, Targeting brain gliomas energy metabolism for classification purposes, *The 2010 IEEE International Conference on Imaging Systems and Techniques*, *Thessaloniki, pp. 36-40, 2010*

5.1 Introduction: The Energy Metabolism in a Eukaryotic Cell

Despite striking advances in proton Magnetic Resonance Spectroscopy (^{1}H-MRS) imaging of brain tumors during the last decades, recent research efforts focus on the evaluation of the role of the energy metabolism in cancer cells in order to identify possible additional potential markers at metabolic and genomic level that will provide a further in-depth analysis of tumors' behavior.

The energy requirements of the brain are amazingly high; indeed, while representing only 2% of the body mass, its oxygen and glucose utilization account for approximately 20% of those of the whole organism *[Magistretti P.J. 1999 – Magistretti P.J. 2006 – Pellerin L. et al 1994 - Waagepetersen H.S. et al 2009]*.

The energy metabolism of a cell, shown in Figure 5.1, also known as *cellular respiration*, is the most vital metabolic function in human body for life maintenance. Cellular respiration is a series of metabolic processes which all living cells use to produce energy in the form of adenosine triphosphate (ATP) molecules. In cellular respiration, the cell breaks down glucose (blood sugar) to produce large amounts of energy in the form of ATP. Cellular respiration can take two paths: *aerobic or anaerobic respiration*.

Aerobic respiration occurs when oxygen (O_2) is available, whereas anaerobic respiration occurs when oxygen is not available. The two paths of cellular respiration share the glycolysis ("lysis" of glucose or sugar splitting process) step, as shown in Figure 5.1. In presence of oxygen (aerobic process) the pyruvate produced enters the mitochondrion where 36 molecules of energy in the form of ATP are produced through the Krebs cycle (also known as TCA cycle or citric acid cycle) and the oxidative phosphorylation (OXPHOS). These 36 ATP molecules are added to the 2 ATP molecules produced prior to the synthesis of pyruvate resulting to 38 ATP molecules of energy within the cell. On the other hand, in the lack of oxygen (anaerobic process), pyruvate is converted to lactic acid or lactate through fermentation. Lactate then exits the cell to regenerate glucose through a process known as gluconeogenesis. This process forces glycolysis to run again.

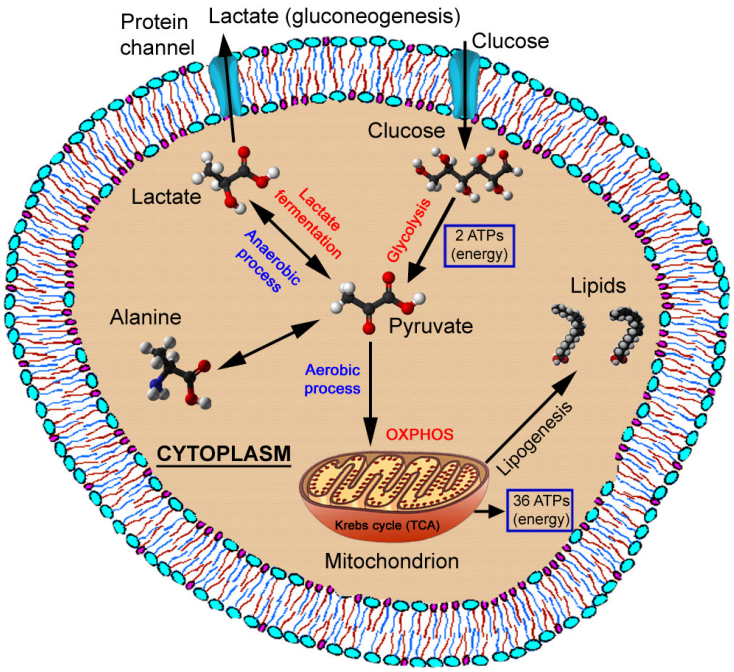


Figure 5.1 – The energy metabolism (cellular respiration process) in an eukaryotic cell. The two paths of glycolysis: OXPHOS (an aerobic process) and Lactate fermentation (an anaerobic process) – [Sources: Kounelakis M.G. et al 2010(a) – Elsom Research] – (The figure was designed with Adobe Photoshop ver. CS5)

Brain tumor cells, in order to meet the increased requirements of proliferation, often display fundamental changes in pathways of energy metabolism *[Garber K. 2006]*. This is due to the fact that many of the genetic mutations in tumor suppressors and oncogenes, explained next, which actually lead to the carcinogenesis, also drive the altered energy metabolism of tumor cells *[Vogelstein B. et al 2004]*.

However, increased energy demands in tumor's cells, set a great challenge in tumors. The energy resources can run out. This can easily generate a metabolic stress within the tumor. Thus, tumors face two distinct metabolic challenges: (1) how to modify their cellular metabolism to support enhanced cell growth and proliferation, and (2) how to engage strategies of metabolic adaptation to survive periods of metabolic stress and maintain viability as the cells accumulate *[Russel G. et al 2009]*.

5.2 The Glycolytic Profile of Brain Tumors

5.2.1 The brain tumors and the Warburg effect

In contrast to normal brain cells the malignant rapidly-growing glioma cells present very high glycolytic rates *[Warburg O. 1956 - Galaragga J. et al 1986]*. There are two common reasons for this fact. The classical explanation is that there is poor blood supply to tumors causing local depletion of oxygen. The other explanation stems from the well known hypothesis of Otto Warburg, who claimed that most cancer cells predominantly produce energy by glycolysis followed by lactate fermentation in the cytosol, rather than by oxidation of pyruvate in mitochondria like most normal cells *[Kim J.W et al 2006(a)]*. This occurs even if oxygen is plentiful. Warburg postulated that this change in metabolism is the fundamental cause of cancer, a claim now known as the '*Warburg effect'*, as shown in Figure 5.2. This effect may simply be a consequence of damage to the mitochondria in cancer cells, or an adaptation to low-oxygen environments within tumors, or a result of cancer genes shutting down the mitochondria because they are involved in the cell's apoptosis program, which would otherwise kill cancerous cells.

The Warburg effect may also be associated with cell proliferation. Since glycolysis provides most of the building blocks required for cell proliferation, it has been proposed that cancer cells may need to activate glycolysis, despite the presence of oxygen, in order to proliferate *[Lopez-Lazaro M. 2008]*. When oxygen is depleted, as for instance in hypoxic-necrotic tumorous tissues of gliomas, the dominant glycolytic product in many tissues is lactate and the process is known as anaerobic glycolysis.

Thus, lactate metabolite is a sensitive indicator of anaerobic glycolysis and reduced cellular oxygenation in living tissues. Furthermore, lipids and fatty acid syntheses are increased in gliomas *[Ledwozyw A. et al 1992]*. Therefore, apart from the already known metabolic markers (NAA, Cho, Cre, mI, Lac, Ala, Glx, Glut1, Glut2 and Lipids) studied in the first chapter for brain tumors evaluation, reliable estimates of the levels of glucose, pyruvate, lactate and lipid substrates are of special interest for the clinical management of brain gliomas patients.

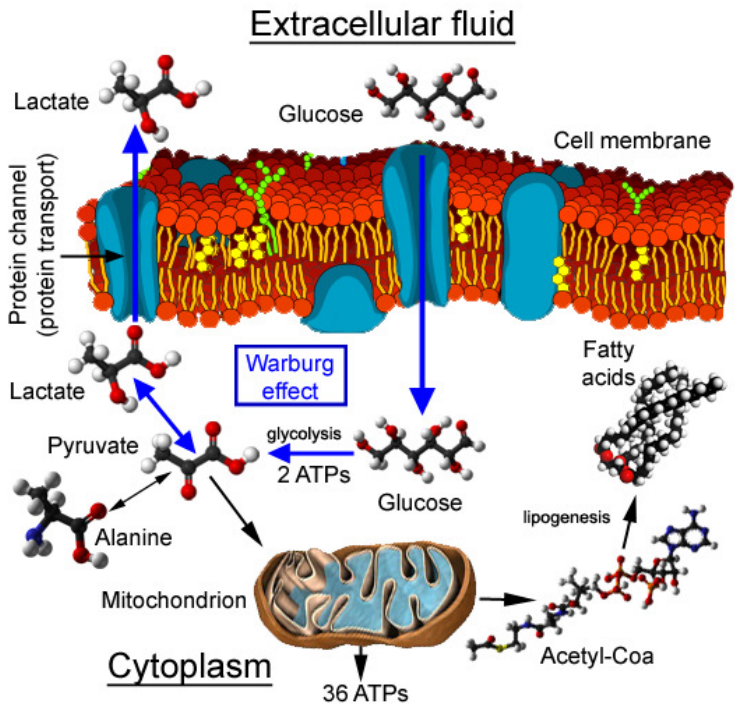


Figure 5.2 – The Warburg effect in a tumorous cell. The blue arrows show the anaerobic glycolysis pathway which results to lactate fermentation – [Sources: Kounelakis M.G. et al 2010(b) – The University of New South Wales] – (The figure was designed with Adobe Photoshop ver. CS5)

5.2.2 The neuron-astrocyte synergy in energy metabolism

Within the brain, the energy resources can be re-directed towards the regions of this organ that are undergoing increased signalling activity [*Kennedy C. et al., 1976*]. Depending on its energy needs the brain stimulates the energy metabolism for the production of the required energy in the form of ATP. Since the energy production process is basically regulated by the glycolysis pathway, as shown in Figures 5.1 and 5.2, special attention should be given to glycolysis when studying the metabolism of the brain. Furthermore and due to the fact that the glucose metabolite triggers the glycolytic activity, the evaluation of the glucose flux throughout the brain, is very important especially in pathological situations such as brain tumors where abnormal variations in the glycolytic process occur, as Otto Warburg also observed.

Another issue that is directly related to the glucose flux within the brain organ is the cellular structure of the brain. As described in the first chapter, the brain is made form two main types of cells, namely the neurons and the glial or supportive cells. Among the glial cells, astrocytes (star shaped glial cells) are the most important. The main role of the astrocytes, which form a supportive network between neurons, is to deliver energy to neurons whenever they need it. This is done through a bioenergetic synergy between neurons and astrocytes for glucose exchange purposes, as shown in Figure 5.3.

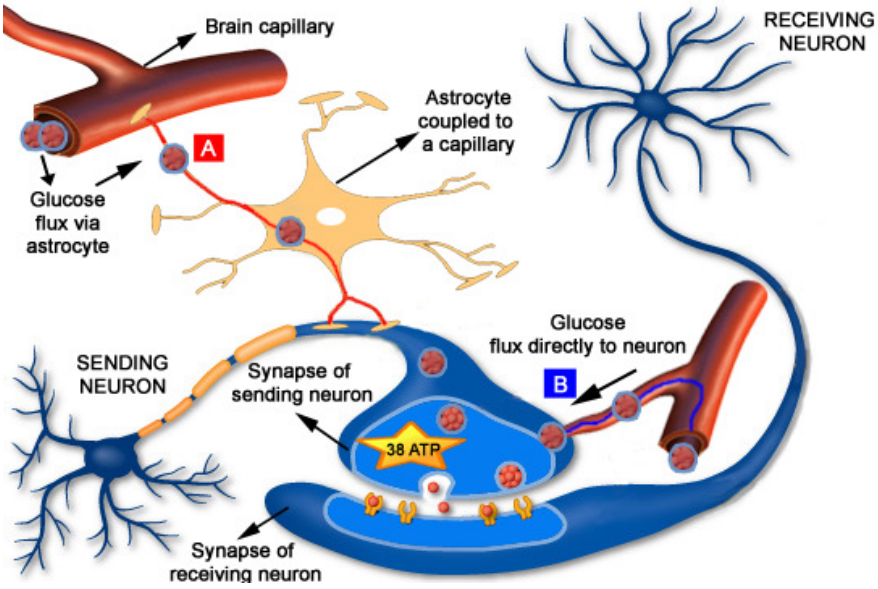


Figure 5.3 – The brain neuron-astrocyte metabolic coupling. The glucose's flux paths: In A (red path) the glucose fluxes to neuron via the astrocyte. In B (blue path) the glucose fluxes to neuron directly from capillary – [Internet Source: The University of Utah] - (The figure was designed with Adobe Photoshop ver.CS5)

As shown in Figure 5.3, the blood glucose molecules which 'travel' in the brain blood vessels (capillaries) follow two different paths in order to reach neurons where the energy production, will take place. The red colored path (A) starts from the brain capillary, passes through the astrocyte and finally reaches the neuron where the ATP energy molecules are generated. In a similar manner neurons also receive glucose directly from the brain capillaries nearby as described by the blue path (B). The overall procedure eventually generates 38 ATP energy molecules. Then through the synaptic activity, glucose is transferred from the sending neuron to the neighboring receiving neuron.

The most important glucose path is that through the astrocyte cells, i.e. the A path. This is due to the morphological associations between the vasculature, astrocytes and neurons. Perivascular astrocytes possess membrane extensions that contact the vasculature, other astrocytes, oligodendrocytes and cell bodies of neurons. This strategic positioning allows astrocytes to control the blood flow *[Gordon G.R.J. et al 2009]*. Furthermore, since a single astrocyte can make contacts with over 100,000 synapses, the demand on a single astrocyte for glucose delivery might be overwhelming *[Bushong E. A. et al 2002]*.

Another more realistic illustration of the glucose and oxygen exchange between the blood capillaries, the glial cells and the neurons is shown in Figure 5.4.

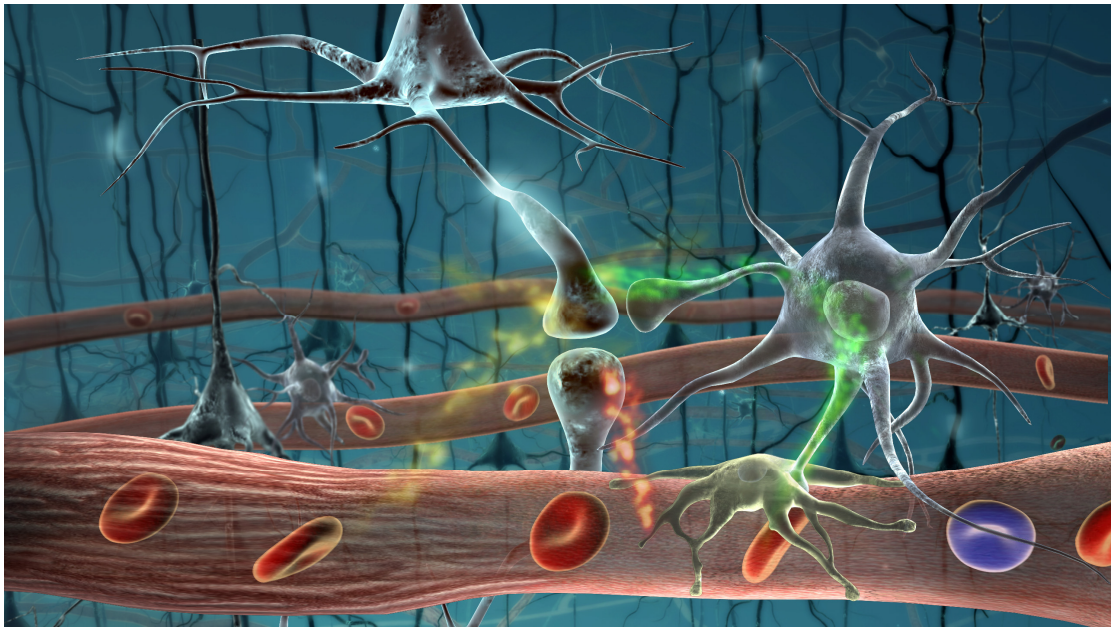


Figure 5.4 – A network of blood vessels, astrocytes and neurons in the brain which signal each other. The mists of colors (red, green, yellow) show the flow of important molecules like glucose and oxygen. This image is a snapshot from a 52-second simulation created by animation artist Kim Hager at the University of California, Los Angeles – [Internet Source: US National Institute of Health]

Under pathological conditions, such as in brain gliomas where the brain glia is gradually degenerated due to genetic and metabolic abnormalities and therefore the astrocytic bioenergetic network starts to collapse, the glucose flux is greatly differentiated.

All these observations and facts enhanced our will to further investigate the role of the glycolysis, both at genomic and metabolic level. Next, a study which focuses on the identification of novel genetic markers for low and high grade brain gliomas diagnostic and treatment purposes is presented.

Before that, an analysis of the currently studied genetic markers in brain gliomas evaluation is presented.

5.3 Genetic Markers Currently Studied in Gliomas

As stated in the first chapter of this thesis, cancer and so brain gliomas, are the result of genetic abnormalities (or abnormal mutations) of specific genetic indicators called *tumor suppressors and oncogenes*. Tumor suppressors are genes whose main role is to slow down the cell division (mitosis) process or cause cells to die at a programmed time (apoptosis). In contrast, oncogenes speed up the cell division process. Abnormal variations in these two types of genes can result in uncontrolled and excessive cellular growth which is the hallmark of malignant tumors.

5.3.1 The tumor suppressors and oncogenes in gliomas evaluation

Genes determine the form, function, and growth patterns of cells. Those that accelerate or suppress growth are often involved in cancer. For example, many cancers have an abnormality in a gene that is responsible for stimulating cellular growth and/or the gene that normally prevents cancer is not working properly.

Tumor genesis involves an interplay between at least two classes of genes: oncogenes and tumor suppressor genes. The most common scenario for inactivation of a tumor suppressor gene is mutation of one allelic copy, followed by loss of all or part of the chromosome bearing the second allele. Oncogene mutations, in contrast, generally involve a single allele because they are gain-of-function mutations. Gain-of-function mutations change the gene product such that it gains a new and abnormal function.

An allele (which comes form the Greek word “allelomorph”) is one of two or more versions of a gene. An individual inherits two alleles for each gene, one from each parent, as shown in Figure 5.5. If both alleles are the same then we say that the individual is homozygote. If the alleles are different then it is heterozygote *[Internet Source: National Human Genome Research Institute]*.

Unlike oncogenes, tumor suppressor genes generally follow the 'two-hit hypothesis', which implies that both alleles that code for a particular gene must be affected before a mutation is manifested. This is because if only one allele for the gene is damaged, the second can still produce the correct protein.

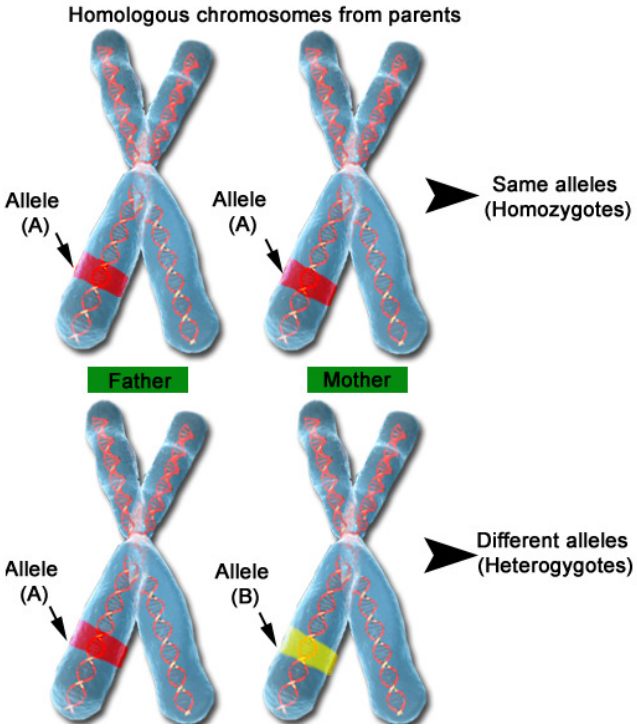


Figure 5.5 – Two homologous chromosomes from the two parents and their possible allelic copies - [Internet Source: Futurity] - (The figure was designed with Adobe Photoshop ver.CS5)

The most significant tumor suppressors and oncogenes examined nowadays for the evaluation of brain tumors are those presented in a well known genetic pathway, called the glioma pathway, provided by the Kyoto Encyclopedia of Genes and Genomes (KEGG), shown in Table 5.1 and Figure 5.6.

Table 5.1 – The tumor suppressors and oncogenes used for brain gliomas evaluation

Tumor suppressors	***Oncogenes***
• TP53 or p53 (Tumor Protein 53) • Rb (Retinoblastoma) • PTEN (Phosphatase and tensin homolog)	• EGFR (Epidermal growth factor receptor) • PDGF (Platelet-derived growth factor) • CDK4 (Cyclin-dependent kinase 4)

[Source: Kyoto Encyclopedia of Genes and Genome (KEGG)]

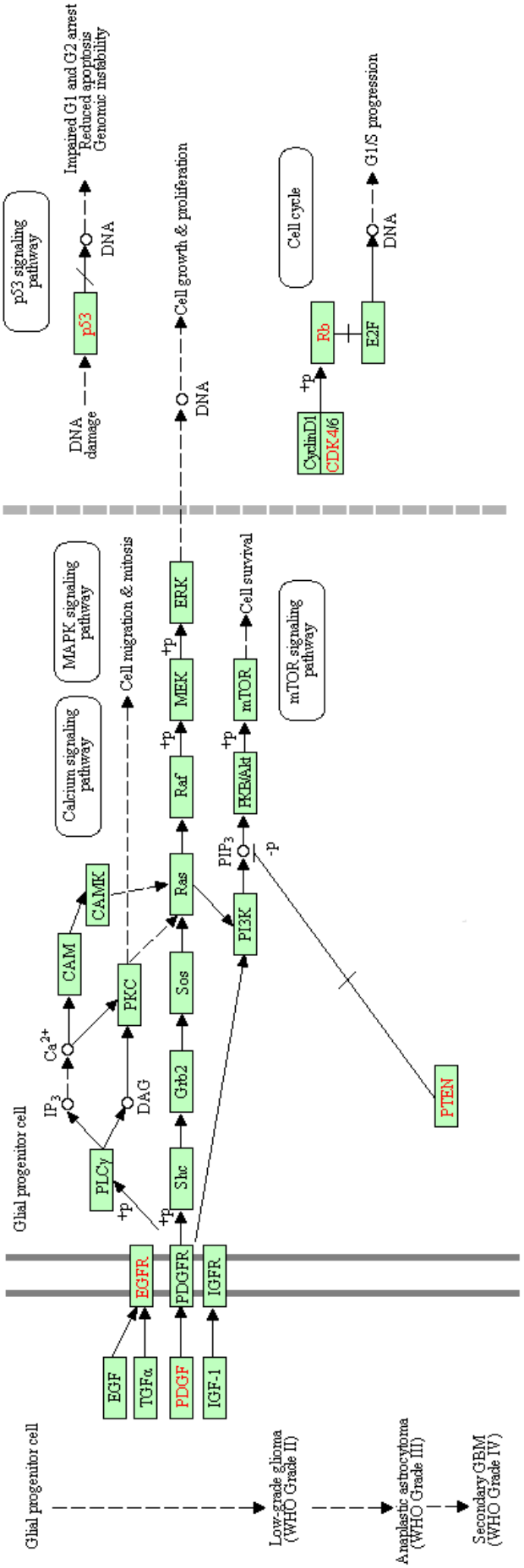


Figure 5.6 – Brain glioma pathway: The tumor suppressors and oncogenes expressed in glioma progression and the related signaling pathways. [Internet Source: KEGG Kanehisa Laboratories]

5.4 Identification of Significant Glycolysis-Related Genes for Brain Gliomas Management

5.4.1 Abstract

The proposed analysis considers aspects of both statistical and biological validation of glycolysis effect on brain gliomas, at both genomic and metabolic level. In particular, two independent datasets are analyzed in parallel, one engaging genomic (Microarray Expression) data and the other metabolomic (Magnetic Resonance Spectroscopy Imaging) data. The aim of this study is twofold. First to show that, apart from the already studied genes (markers), other genes such as those involved in the human cell glycolysis significantly contribute in gliomas' discrimination. Second, to demonstrate that the glycolysis process must be considered in the design of novel gliomas' treatment protocols.

The results of this analysis demonstrate that the combination of genes participating in the glycolytic process (ALDOA, ALDOC, ENO2, GAPDH, HK2, LDHA, LDHB, MDH1, PDHB, PFKM, PGI, PGK1, PGM1 and PKLR) with known from the bibliography tumor suppressors (PTEN, Rb, TP53), oncogenes (CDK4, EGFR, PDGF) and HIF-1, enhances the discrimination of low versus high-grade gliomas providing a prediction ability of 98%. Furthermore, the glycolytic metabolites, such as Glucose, Pyruvate, Lactate and Lipids, considered in combination with the genomic markers, achieve comparable classification accuracy.

Therefore, it is justified that the glycolytic profile of gliomas should be taken under consideration when diagnostic practices are applied. Finally, considering the glycolytic profile of glioma patients new therapeutic strategies must be generated to allow patient-specific treatment *[Kounelakis M.G. et al 2011(a) – Kounelakis M.G. et al 2011(b)]*.

5.4.2 The need to focus on glycolysis

Over the last decade a lot of research has focused on the identification of genetic alterations that play a vital role in brain glioma pathology. Most recent studies *[Shiraishi T. et al 2003 - Furnari F.B. et al 2007 - Ohgaki H. et al 2009]* conclude to a specific set of gene markers that has become a diagnostic standard to describe the type and grade of this complex and lethal cancer. Towards a better understanding of the interactions of these genes and the stage of their involvement in glioma grade progression, a genetic pathway in the form of a super network called glioma pathway, shown in Figure 5.5, has become available in KEGG. Within this network, specific pathways exemplify the role of oncogenes and tumor suppressor genes in crucial biological processes of the cell.

The mutations of oncogenes lead to uncontrolled cell division (mitosis) instead of the programmed cell death (apoptosis). On the other hand the tumor suppressor genes are considered the "guardians" of the cell. When these genes do not function correctly, the cells with DNA damage continue to divide contributing to the formation of cancerous cells *[Jones R.G. et al 2009]*.

The most important tumor suppressors and oncogenes involved in the glioma pathway (i.e. the gradual progress from low grade gliomas to intermediate and finally to high grade gliomas), include Phosphatase and tensin homolog (PTEN), Retinoblastoma (Rb), Tumor protein 53 (TP53) and Cyclin-dependent kinase 4 (CDK4), Epidermal growth factor receptor (EGFR) and Platelet-derived growth factor (PDGF) respectively *[Furnari F.B. et al 2007 - Ohgaki H. et al 2009 - Jones R.G. et al 2009]*. These genes are essential to be detected in glioma-discrimination procedures following either statistical or clinical diagnostic practices.

Nevertheless, additional molecular mechanisms should also be included in the prediction approach, especially when discriminating complex, overlapping classes and attempting to identify differences at gene level.

One such mechanism relates to the glycolysis process associated with the well known hypothesis of Otto Warburg. As mentioned, according to this hypothesis but others too *[Kim J.W. et al 2006(a)]*, most cancer cells predominantly produce energy by glycolysis followed by lactic acid (lactate) fermentation in the cytosol, rather than by oxidation of pyruvate in mitochondria like most normal cells. This occurs even if oxygen is plentiful.

Glycolysis pathway as also shown in Figure 5.7(A) schematically and (B) diagrammatically, involves a series of biochemical reactions (shown at the right hand side of Figure 5.7(B)) in which glucose is broken down to pyruvate in order to produce energy molecules. Pyruvate is then transformed to either lactate or enters the mitochondrion to activate the Citric Acid Cycle (TCA) depending on the cell's state *[Internet Source: Glycolysis-Gluconeogenesis Pathways (KEGG)]*.

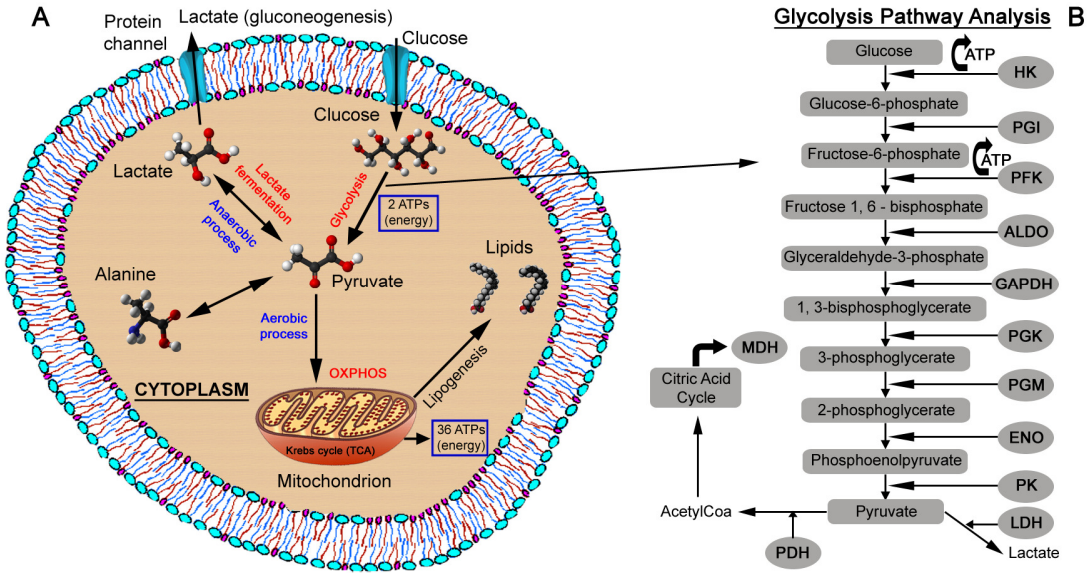


Figure 5.7 – A: Abstract presentation of glycolysis. - B. Analytical illustration of glycolysis pathway showing the protein/enzyme coding genes (in elliptic scheme) participating in the glucose break down to pyruvate - [Internet Sources: Elsom Research – Kyoto Encyclopedia of Genes and Genome – Human Metabolome Database] - (The figure was designed with Adobe Photoshop ver.CS5)

Under normal conditions, only about 13% of glycolytic pyruvate is converted to lactate, but larger proportions are expected in gliomas. Furthermore, lipids and fatty acid syntheses (FASN) produced through the lipogenesis process are increased in gliomas *[Lopez-Lazaro M. 2008]*.

5.4.3 Main goals and achievements of the study

The main goal of this study is to systematically study the involvement of glycolysis process in gliomas and reveal its role in brain gliomas discrimination, at both genomic and metabolic level. In other words, to identify the discrimination potential of the genetic and metabolic features (genes and metabolites) participating in the glycolysis. It is actually an attempt to form a biological "bridge" from genomics to metabolomics as to validate the fact that glycolysis must be taken under consideration in diagnosis and therefore treatment procedures.

Following a similar to Chapter 4 methodology, where the Fisher's filter criterion was embedded into a RFE wrapper selection process, the Relief-f filter method is now applied in a wrapper manner in order to select the most discriminant features. Furthermore, a comparison of these two methodologies is achieved in both classification and feature selection level.

The prediction ability of the selected features is estimated using an SVM classifier, evaluated through a 10-fold cross validation (CV) scheme. In addition to the individual testing of datasets, we formulate the combined feature vector composed of all significant genomic and metabolic markers and examine its predictive power through stochastic data perturbation techniques, again under a 10-fold CV scheme.

The outcome of this study exhibits the significance of the glycolysis-related genes in the diagnosis and new treatments design, suggesting that these genomic markers must be integrated in nowadays laboratory and clinical practices applied to face gliomas and improve patient's survival.

5.4.4 Materials and Methods

5.4.4.1 Datasets description

In order to achieve the goals of this study, two different datasets are investigated. The first dataset comes from the genomics while the second one from the metabolomics area. The reason for employing two different datasets is to obtain a twofold validation of the role of glycolysis both at genomic and metabolomic level.

The genomic dataset is available from a public functional genomics data repository of the National Center of Biotechnology Information *[Internet Source: Gene Expression Omnibus]*. This dataset was supplied by the Children's National Medical Center of Washington US and is an Affymetrix RNA array, first published in 2005 (last update in 2009) by Tobey MacDonald (contributor). This array consists of 12625 gene transcripts (rows) and 14 glioma patients

(columns) 8 of them characterized as low grade (GRII) and 6 as high grade (GRIV) or glioblastoma multiforme (GBM) and it was first used to identify significant gene expression profile differences between these two types of gliomas. GRII and GRIV abbreviations are according to World Health Organization (WHO), as explained in first chapter.

The second dataset consists of short echo ^{1}H-Magnetic Resonance Spectroscopy Imaging (MRSI) data from 17 glioma patients (10 of low grade and 7 of high grade). The two-dimensional MRSI data was collected by the Radboud University Medical Center and contains 246 pre-processed proton-MRSI (^{1}H-MRSI) volume elements (voxels) corresponding to 246 spectra. Each patient case passed strict quality control and validation procedures, including consensus histopathologic determination. Table 5.2 summarizes the data acquired from both genomics and metabolomics areas. Within each one of the 246 voxels obtained, specific metabolic features relating to glycolysis were measured. More specifically the areas under the peaks of Glucose (Gluc at 3.44 ppm), Pyruvate (Pyr at 2.37 ppm), Lactate (Lac at 1.33 ppm) and Lipids (Lips: the sum of mobile lipids levels at 0.90 and 1.30 ppm) were estimated by peak integration *[Simonetti A.W. et al 2003]*. The areas were estimated by integrating each metabolite's spectral intensities around its peak within a window of 0.13 ppm, as shown in Figure 5.8. Such a width covers most of the peaks of interest completely, without being contaminated with neighbouring peaks.

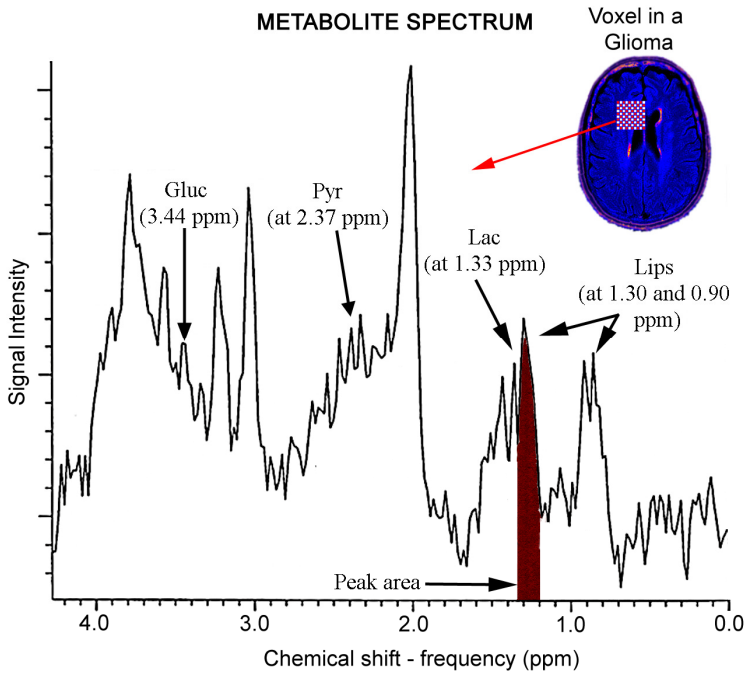


Figure 5-8: Spectrum of a GBM patient. Y axis: signal intensity (proportional to metabolites concentration). X axis: frequency (position) in parts per million. Gluc (Glucose at 3.44ppm), Pyr (Pyruvate at 2.37 ppm), Lac (Lactate at 1.33 ppm) and Lips (sum of Lipids at 1.3 and 0.90 ppm) are the metabolites observed. The shaded area is just for presentation purposes, not the real measured areas under the peaks – [Internet Sources: Radiology Spirit – Medscape] – (The figure was designed with Adobe Photoshop ver.CS5)

Table 5.2 – The datasets description

Tissue type	***Patients in Genomics Dataset***	***Patients in Metabolomics Dataset***
Low grade glioma (GRII)	8	10
High grade glioma (GRIV)	6	7
	Number of Genes	**Number of Voxels**
Number of features	12.625	246

Low grade gliomas correspond to grade II based on WHO - High grade gliomas correspond to grade IV based on WHO

5.4.4.2 Statistical analysis at genomic level

In order to identify the discriminative strength of the glycolytic genes, a subset of the original genomic dataset was manually derived based on the functional properties of genes. More specifically, this subset includes only those genes associated with two KEGG pathways (glycolysis and glioma pathways). In other words it contains the glycolysis genes shown in Figure 5.7(B) and their isoforms and the genes (tumor suppressors and oncogenes) shown in Figure 5.6. and their isoforms. A gene isoform (or allele) denotes any of the several different forms of this gene.

For gene annotation in both the dataset and the KEGG database, we used the gene ontology (GO) terminology *[Internet Source: The Gene Ontology]*. Overall, 62 unique gene identifiers out of 12625 initial gene transcripts were selected, as shown in Table 5.4. in section 5.7. The set of these 62 genes for each one of the 14 patients of the genomics dataset was used as input to the classifier. A Recursive Feature Elimination (RFE) process, based on the Relief-F filter method *[Kononenko I. 1994]*, was embedded into the well known support vector machines classifier (SVM). Following this combined feature selection and classification approach we can extract the most significant genes i.e. those providing the highest classification accuracy and examine whether the glycolytic genes are among them.

Following a similar strategy, the Fisher's filter criterion used in fourth chapter, was also applied in this study for comparison purposes.

5.4.4.3 The feature elimination process

The key goal of Relief-F is to rank the quality of features according to how well they distinguish between close-by instances. Relief-F attempts to find an estimate of the probability of the feature to differ significantly only across classes and assigns this probability as weight to each feature f . More specifically, the weight function has the form:

$$W[f] = P(different\ value\ of\ f\ in\ different\ class) - P(different\ value\ of\ f\ in\ same\ class) \quad (1)$$

In other words, given a randomly selected instance I, Relief-F searches for k nearest neighbors of I from the same class, called nearest Hits and also k nearest neighbors from the different class, called nearest Misses. The quality metric $W[f]$ for each feature f is updated depending on I, nearest Hits and nearest Misses. In the update formula, the contributions of all the $Hits$ and $Misses$ are averaged. An analytical explanation of the Relief-F algorithm is given in the following pseudocode:

Relief-F (C, n, m, k)

C : Training set,
n : Number of features;
m : Number of iterations,
k : Number of nearest neighbours

Initialize all weights $W[f]$, to zero

For $i = 1\ to\ m$ do begin

Randomly select instance I in C

Find its k nearest $Hits$ and k nearest $Misses$ from each class different from which I belongs to

For $f = 1\ to\ n$

$W[f] = W[f] - Avg(diff(f, I, nearestHit)) + Avg(diff(f, I, nearestMiss))$

End

End

Returns all weights $W[f]$

According to the Relief-F strategy, the feature (gene) with the least significant weight, according to the Relief-F strategy in Eq. (1), is eliminated and the rest (most significant ones) are kept for classification. In other words, at the end of each iteration i.e. after the elimination of the least significant gene, the prediction power of the remaining features is measured in terms of the Area under the Receiver Operating Characteristic curve (AUROC). Since 1 gene is eliminated per iteration and 62 genes were examined, at the end of the whole process 62 AUROC values are recorded. Furthermore, under a cross-validation scheme, this procedure was repeated 100 times and the highest classification accuracies were calculated. The total AUROC was obtained from the average value of the 100 runs. The genes most frequently

participating at the highest AUROC results were considered as the most informative markers for classification. This process is schematically presented in Figure 5.9. The software packages used to apply the Relief-F method was Matlab ver. R2010 and Weka ver. 3-6-4 *[Internet Sources: http://www.mathworks.com-http://www.cs.waikato.ac.nz]*.

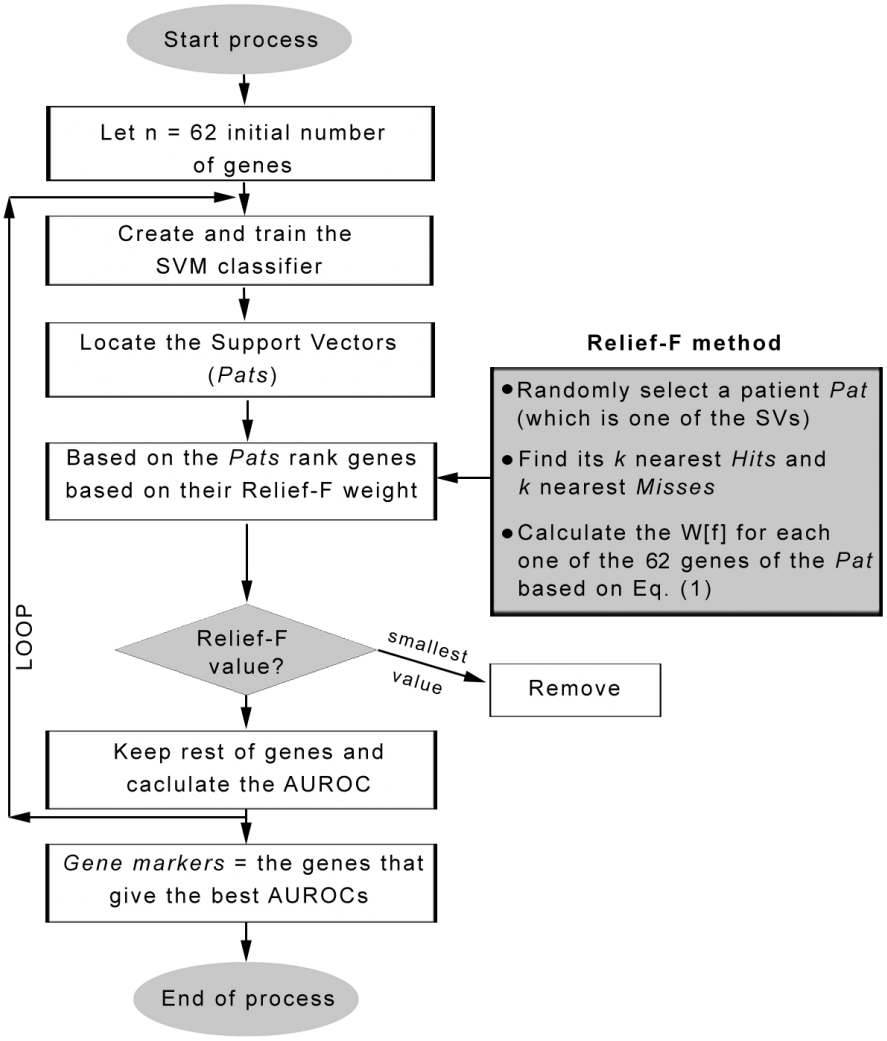


Figure 5-9: The flow chart of the proposed methodology for optimal feature selection and classification – (The figure was designed with Adobe Photoshop ver. CS5)

5.4.4.4 Classification process

The discriminative potential of the features in each dataset was evaluated using SVM classifier with a Radial Basis Function (RBF) kernel. The SVM solution is based only on those data samples (glioma patients here) that are at the margin of the decision boundary, called support vectors *[Abe S. 2005]*. The methodology applied focuses only on the support vectors, as in Chapter 4, adopting the basic property of the SVM classifier, as presented in Figure 5.9.

In essence, SVM attempts to find the best separating hyperplane to distinguish between the two classes of interest, positive (+1) and negative (–1). The separating hyperplane passes

through the middle of this margin with equation $(w \times x) + b = 0$ and the decision function is of the form:

$$f(x) = \mathrm{sgn}((w \cdot x) + b) \quad (2)$$

where w represents the direction vector of the hyperplane. The sign of the value returned by Eq. (2) indicates the predicted class associated with example x, while $|f(x)|$ indicates the confidence level of the resulting decision. By solving the corresponding dual structure, the problem can also be formulated in the kernel space *[Abe S. 2005]*, by defining a positive definite scalar kernel function $k(.,.)$ that measures the distance between two points (x_i, x_j) as in Eq. (3),

$$f(x_i) = sign\left[\sum_{j=1}^{N_{trn}} a_j y_j k(x_i, x_j) + b_i\right] \quad (3)$$

where a_j are the weight parameters (support values) of the training cases (j =1… N_{trn}) and y_j denote the corresponding hard labels. Kernel-based methods can perform well in processing high-dimensional and heterogeneous data. The parameters of RBF kernel (C, γ) used in our study were tuned in each binary classification, in such a way that the smallest possible number of support vectors would be retained for training purposes as to avoid overtraining.

5.4.4.5 Statistical analysis at metabolic level

At metabolic level, the input dataset was designed from the peak-area values of the four glycolytic metabolites (Glucose, Pyruvate, Lactate and Lipids) for each one of the 17 glioma patients of the metabolomics dataset. The recursive feature elimination was not applied on this dataset, since only four metabolic features are involved. Instead the SVM process focused only on the classification.

The statistical significance of the selected metabolites was also measured through Independent-Samples t-test, using the SPSS ver. 19.0 software *[Internet Source: http://www.spss.com]*. The difference of their mean values was estimated along the two classes and the p-values were computed, as shown in Table 5.3.

Table 5.3 – Independent samples t-test results

Metabolites	***GRII vs GRIV***
Glucose	HS (0,00093)
Pyruvate	S (0,041)
Lactate	HS (0,00085)
Lipids	HS (0,00084)

Estimation of the statistical significance of the four metabolic elements. The p-values are placed within the parentheses – The difference is statistically Highly Significant (HS) when p value < 0,001 and Significant (S) when p value < 0,05

5.4.4.6 Classifier performance evaluation

In order to evaluate the classifier's performance a 10-fold CV was applied to evaluate the classifier's performance. The classification accuracy was measured in terms of AUROC values. Data were stratified prior to being split into 10 folds. Stratification rearranged the data as to ensure that each fold is a good representative of the whole. In other words repeated stratified runs (10 folds X 10 times = 100 runs) were applied for each dataset (genomics and metabolomics). For each binary classification scheme the train sets involved almost 90% of the patients of each of the two classes and the remaining 10% were used in the test sets. This was done to ensure that a whole patient from each class is left for testing purposes.

Confidence intervals (CI) were also estimated for each dataset classification measure, using the SPSS ver. 19.0 software.

5.4.5 Experimental results

The application of the RFE process based on the Relief-F filter at the genomics dataset aims to derive the glioma-related genes with significant influence on classification. In particular, genes most frequently participating in signatures associated with maximum AUROC measures, along the iterations of the CV process, are identified. More specifically, using these 62 genetic signatures and after repeating the CV process 100 times (runs), the highest classification accuracies are recorded and the gene signatures are tabulated. The average of these 100 maxima AUROC measures, reaches a value of 0.94 (CI: $\pm$0.03), as shown in Table 5.4.

Overall, we derived 26 out of 62 genes, which are therefore considered as most informative markers for discrimination of grade II from grade IV gliomas. These genes are shown in Table 5.5 in bold, along with their frequency of appearance in 100 runs. The asterisk (*) symbol denotes tumor suppressors and oncogenes already known from the KEGG pathway for glioma while the (G) letter, genes participating in the glycolysis process within a cell. It can be observed that among the 26 significant genes, a large number (14) is directly related to the glycolysis process.

Repeating the classification process (100 times under CV), using this time only these specific 26 gene markers, we observed that the average AUROC value is slightly improved reaching the value of 0.98 (CI: $\pm$0.02), as presented in Table 5.4. This fact verifies the biological significance of these genes in the glioma grading process.

In a similar way, but without the feature elimination process since all four metabolites were found significant in terms of class-mean discrimination, the classification rates at metabolic

dataset were also recorded. The average AUROC value derived after 100 runs is 0.90 (CI: ±0.02), observed in Table 5.4.

Finally, following the dual testing of the hypothesis regarding the glycolysis involvement in gliomas and the selection of markers from individual datasets, we proceeded with a stochastic evaluation of the predictive power of the joint feature vector. Thus, we formulated in vector form an overall signature composed of the total markers (26 genomics and 4 metabolics) selected. Then, we formulated a test population of 60 subjects (30 in class 1 and 30 in class2) by randomly permuting the value of each marker within its corresponding range in the original datasets. We used this random permutation scheme to generate a larger population of patients with all necessary attributes that follows the stochastic distribution as the original data. The prediction ability of the joint signature was evaluated on this population using a 10-fold CV scheme. The average classification accuracy, measured in terms of AUROC, on the permuted dataset was estimated at 0.98 (CI: ±0.018), as presented in Table 5.4. Notice that the combination of both genomic and metabolic markers in a common dataset provided a similar AUROC value as the one obtained from the genomic markers.

The whole process was repeated using this time the SVM-Recursive Feature Elimination (SVM-RFE) feature selection and classification process used in Chapter 4, where the Fisher's criterion was embedded to score the most significant features. In Table 5.4 the classification accuracies achieved from both Relief-F and Fisher's process are shown, for comparison purposes. In Table 5.6 the common genomics features selected from both SVM-RFE methods are shown.

As it can be observed the SVM-RFE Relief-F method achieves better classification accuracies than SVM-RFE Fisher's method, in all cases. Furthermore in Table 5.6 it can be observed that the SVM-RFE Fisher's method reveals 30 genes, i.e. 4 more than the SVM-RFE Relief-F, as significant for glioma discrimination. Another important observation in this table relates to the number of the glycolysis-related genes selected based on the Fisher's criterion. As it can be seen this method selects 12 glycolysis-related genes and 18 tumor suppressors and oncogenes whereas the Relief-F criterion 14 glycolysis-related genes and 12 tumor suppressors and oncogenes.

Table 5.4 – Classification results obtained from both methodologies and confidence intervals

Feature sets	***Initial 62 genes (1)***	***Significant genes detected (2)***	***Only 4 metabolites (3)***	***Combined scheme (2+3)***
(Relief-F)	0.94(CI:±0.03)	0.98^{RF}(CI:±0.02)	0.90(CI:±0.02)	0.98(CI:±0.018)
(Fisher's)	0.91(CI:±0.02)	0.93^{FS}(CI:±0.02)	0.89(CI:±0.03)	0.95(CI:±0.012)

Classification accuracies are based on the AUROC values. The RF superscript corresponds to the result obtained from 26 genes identified as significant in the SVM-RFE Relief-f method. The FS superscript corresponds to the result obtained from 30 genes identified as significant in the SVM-RFE Fisher's method. Case (2+3) combines the 4 metabolites (Glucose, Pyruvate, lactate and Lipids) with the significant genes found from each methodology. CI stands for Confidence Intervals

Table 5.5 – Genetic markers selected from SVM-RFE Relief-f method in alphabetical order

Genes 1 - 16	***Genes 17 - 32***	***Genes 33 - 48***	***Genes 49 - 62***
ALDOA^{G} (90%)	GAPDHS^{G}	PDGFRL*	**RB1* (98%)**
ALDOB^{G}	GCK^{G}	PDHA1^{G}	RBBP4*
ALDOC^{G} (99%)	GRB2*	PDHA2^{G}	RBBP5*
BPGM^{G}	**HIF1A* (98%)**	**PDHB^{G} (97%)**	RBBP6*
CDK4* (100%)	**HK2^{G} (98%)**	PFKFB1^{G}	RBBP7*
CDKN2A*	**LDHA^{G} (95%)**	PFKL^{G}	RBBP8*
CDKN2B*	**LDHB^{G} (94%)**	**PFKM^{G} (87%)**	RBL1*
CDKN2C*	LDHC^{G}	PFKP^{G}	**RBL2* (98%)**
CDKN2D*	**MDH1^{G} (100%)**	**PGI^{G} (98%)**	**TP53* (98%)**
DLAT^{G}	MDH2^{G}	**PGK1^{G} (94%)**	TP53AP1*
ECD^{G}	OGDH^{G}	**PGM1^{G} (97%)**	**TP53BP1* (92%)**
EGFR* (95%)	PDAP1*	PGM2^{G}	TP53BP2*
ENO1^{G}	**PDGFA* (95%)**	**PKLR^{G} (97%)**	TP53I11*
ENO2^{G}(97%)	**PDGFB* (93%)**	PKM2^{G}	TP53I3*
ENO3^{G}	**PDGFRA* (97%)**	**PTEN* (99%)**	
GAPDH^{G} (82%)	**PDGFRB* (95%)**	PTENP1*	

Gene's names are based on The Entrez Gene Database. The asterisk (*) symbolizes the genes known from bibliography while the (G) letter the glycolysis-related genes. The genes in bold are the most significant genes based on the SVM-RFE Relief-F. The numbers in the parentheses correspond to the frequency of appearance of each gene at the highest AUROCs measured in a 100 runs process

Table 5.6 – Common and uncommon genomic markers found from each feature selection methodology

A/A	SVM-RFE (Relief-F)	SVM-RFE (Fisher's)	Common (■) Uncommon (□)
1	ALDOA[G]	ALDOA[G]	■
2	CDK4*	CDK4*	■
3	EGFR*	EGFR*	■
4	HIF1A*	HIF1A*	■
5	HK2[G]	HK2[G]	■
6	LDHA[G]	LDHA[G]	■
7	LDHB[G]	LDHB[G]	■
8	MDH1[G]	MDH1[G]	■
9	PDGFA*	PDGFA*	■
10	PDGFB*	PDGFB*	■
11	PDGFRA*	PDGFRA*	■
12	PFKM[G]	PFKM[G]	■
13	PGI[G]	PGI[G]	■
14	PGK1[G]	PGK1[G]	■
15	PTEN*	PTEN*	■
16	RB1*	RB1*	■
17	RBL2*	RBL2*	■
18	TP53*	TP53*	■
19	TP53BP1*	TP53BP1*	■
20	ALDOC[G]	ALDOB[G]	□
21	ENO2[G]	ENO3[G]	□
22	GAPDH[G]	CDKN2A*	□
23	PDGFRB*	CDKN2B*	□
24	PDHB[G]	ECD[G]	□
25	PGM1[G]	GRB2*	□
26	PKLR[G]	MDH2[G]	□
27		PDGFRL*	□
28		PTENP1*	□
29		RBL1*	□
30		TP53AP1*	□
Total	**26 genes**	**30 genes**	

The number of common genomic markers is 19 while the number of uncommon is 16 (6 come from SVM-RFE Relief-F and 10 from the SVM-RFE Fisher respectively)

5.4.6 Discussion

5.4.6.1 Statistical results analysis

The tumor suppressors and oncogenes involved in glioma grade progression, described in KEGG's glioma pathway, were established as a gold standard in today's diagnostic practices. Most of the related studies either evaluate the potential of already known markers from the published genetic networks or examine glycolytic genes for the classification of gliomas. Thus, they primarily focus on the analysis of gliomas at genetic level. In contrast, our study attempts to reveal the significance of the glycolytic genes in a different way. First, we focused on glycolysis at both genomic and metabolic level in order to examine whether its contribution is dominant in the discrimination of low from high grade brain gliomas. Second and based on its discrimination ability to exhibit the need to consider glycolysis effects in the design of novel treatments.

Both of these two goals were accomplished. As far as it concerns the classification potential of the glycolysis related genetic markers found, the results of the feature selection and classification process on the genomic dataset of 14 glioma patients clearly prove their significant contribution in gliomas discrimination. This fact was also verified through the estimation of their frequency of participation shown in Table 5.4. Furthermore the discrimination value of the four metabolites was also verified through the classification of 17 glioma patients. The significance of glucose, pyruvate, lactate and lipids in glioma discrimination and analysis is noticeable, reaching AUROC values of 0.90 (CI: $\pm$0.02) in the case of Relief-F method and 0.89 (CI: $\pm$0.03) in the case of Fisher's method respectively. The statistically significant difference of their means in Table 5.3 also implies that their metabolic activity differs in different stages or grade of gliomas *[Galarraga J. et al 1986 – Kounelakis M.G. et al 2010(a)]*. Nevertheless, the observation that these markers do not contribute additional efficiency in the classification process, when combined with the appropriately selected genomic markers, may be related to an assumption that metabolites convey more abstract information complementary to that of genetic markers.

As observed in this study the SVM-RFE methodology based on the Relief-F filter provided better results compared to those obtained from the Fisher's criterion application. More specifically the Relief-F method revealed 26 significant genes for glioma discrimination where 14 of them are glycolysis-related. In contrast the Fisher's criterion managed to select 30 significant genes where 12 of them are glycolysis-related, as shown in Table 5.6. Furthermore the Relief-F method managed to reach higher classification rates compared to the Fisher's classification results in all feature sets as shown in Table 5.4.

The differences in the results obtained (number of features selected and classification rates) from the two methodologies are justified by their different mechanisms in the selection of significant features. Fisher's criterion is a simple measure that scores each feature by calculating for that feature the (squared) distance between the class means and correcting

that distance with the within class variance. In contrast to most feature ranking methods, Relief-F does not assume conditional independence of the features. The significance of the features is obtained based on how good they discriminate between samples that are close. It is an iterative procedure in which random samples are picked from the dataset, and each time the k nearest neighbours of the same and the opposite class are calculated. Based on these neighbouring cases the weights of the features are updated. At the end the features are ranked based on the calculated weights.

Relief-F has been adopted in several studies on cancer discrimination, involving microarray data *[Robnik-Sikonja M. et al 2003]*. This approach has shown comparable performance in comparison with others already known methods such as Information Gain, Gain Ratio and χ^2 – statistic in various domains such as discrimination of leukemia (AML and MLL) and colon cancer *[Wang Y. et al 2004]* and brain tumors *[Postma G.J. et al 2011 - Luts J. et al 2007]*.

As far as it concerns the second goal, i.e. the role of glycolysis in the design of new treatments, the following sections clearly describes this issue.

5.4.6.2 The effect of the glycolytic gene-markers on glioma cells

Based on their biological function in the glycolysis process provided by the Gene Ontology database *[Internet Source: The Gene Ontology]*, the 14 genes (or protein/enzyme coding genes) detected (ALDOA, ALDOC, ENO2, GAPDH, HK2, LDHA, LDHB, MDH1, PDHB, PFKM, PGI, PGK1, PGM1 and PKLR) shown in Table 5.5, are categorized into six broad classes. These are the *Kinases, Isomerases, Aldolases, Mutases, Enolases and Dehydrogenases*.

Kinases: Between these 14 glycolytic genes revealed from this study, 4 of them belong to the class of kinases. These are the HK2, PFKM, PGK1 and PKLR which are isoforms of the genes, Hexokinase (HK), phosphofructokinase (PFK), phosphoglycerate kinase (PGK) and pyruvate kinase (PK) respectively. Kinases are the key regulators of cell function and constitute one of the largest and most functionally diverse gene families. They are particularly prominent in signal transduction and co-ordination of complex functions such as the cell cycle.

Regarding the 4 kinases found in our study, increased levels of HK have been identified in high grade gliomas (GRIV) in contrast to low levels found in low grade gliomas (GRII) *[Dominguez J.E et al 1987]*. In addition, several studies demonstrate that hexokinase, particularly the HK2 isoform, plays a critical role in initiating and maintaining the high glucose catabolic rates of rapidly growing tumors *[Mathupala S.P. et al 2006 - Pedersen P.L. et al 2007]*. Another major regulatory protein coding gene in glycolysis is the phosphofructokinase-muscle (PFKM or PFK1). This gene allows metabolic intermediates to be diverted into pathways other than glycolysis for example, the pentose phosphate pathway, as well as increases or decreases the rate of glycolysis depending on the energy status of the cell. Importantly, despite being a substrate for PFKM, ATP is a potent inhibitor of its activity which is probably the most important mechanism by which OXPHOS regulates glycolysis (Pasteur

Effect) *[Tennant D.A. et al 2009]*. Also, increased levels of PFKM have been noticed in high grade gliomas as glioblastoma multiforme (GRIV) *[Dominguez J.E et al 1987]*. PGK1 gene has been observed to be related with Hypoxia-inducible factor 1 (HIF-1) which is a gene mostly found in hypoxic tumorous areas. Over expression of HIF-1 caused elevation of PGK1 in recent study *[Zhao S. et al 2009]*. The final protein coding gene in glycolysis is pyruvate kinase (PK). This gene (and its isoform PKLR- pyruvate kinase-liver and RBC) allows the cell to sense the levels of anabolic precursors as well as the energy status of the cell. This regulatory mechanism is thought to allow tumor cells to survive in environments with varying oxygen and nutrients *[Tennant D.A. et al 2009 - Semenza G.L. et al 1996 - Pelicano H. et al 2006]*.

*Isomerases***:** The phosphoglucose isomerase (PGI) is the only gene of this category that has been found significant for glioma discrimination. Its function is to convert glucose (Glucose-6-phosphate) to fructose (Fructose-6-phosphate). Its role in both the glycolytic and gluconeogenesis pathways is important *[Yanagawa T. et al 2004]*.

Aldolases, Enolases and Mutases: Within the group of 14 gene markers involved in the glycolysis pathway, 4 of them belong to these three categories. These are the ALDOA, ALDOC, ENO2 and PGM1 genes, as shown in Figure 5.7(B). ALDO (with isoforms ALDOA, ALDOC) and ENO (with isoform ENO2) are directly related to the HIF-1 which is a key element of the mitochondria activity *[Semenza G.L. et al 1996]*. Phosphoglycerate mutase-1 (PGM1) is a glycolytic gene that catalyzes the conversion of 3-phosphoglycerate to 2-phosphoglycerate. The TP53 gene suppresses the expression of PGM and therefore, loss of TP53 in cancer cells will lead to increased PGM activity and enhanced glycolysis *[Levine A.J et al 2010]*.

Dehydrogenases: Among the 14 glycolytic genes markers found to be significant, 5 of them *(GAPDH, LDHA, LDHB, PDHB and MDH1) belong to the category of dehydrogenases.* Interestingly, Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is implicated in apoptosis when it is translocated into the nucleus, although the molecular mechanism responsible for its nuclear translocation and its role in cancer remain to be defined *[Pelicano H. et al 2006]*. The role of HIF-1 is not restricted to upregulation of the genes stimulating glucose utilization. Recent findings demonstrate that in addition, HIF-1 suppresses mitochondrial function in tumor cells, suggesting that it modulates the reciprocal relationship between glycolysis and OXPHOS. The switch between glycolysis and OXPHOS is controlled by the relative activities of two genes, pyruvate dehydrogenase (with isoform PDHB) and lactate dehydrogenase (with isoforms LDHA and LDHB) *[Simon M.C. et al 2006 - Kim J.W. et al 2006(b)]*. Furthermore, malate dehydrogenase (with the isoform MDH1) is an enzyme in the citric acid cycle, as shown in Figure 1(B), which catalyzes the conversion of malate into oxaloacetate. Malate dehydrogenase is also involved in gluconeogenesis, that results in the generation of glucose from non-carbohydrate carbon substrates such as lactate, glycerol, and glucogenic amino acids and it is used to keep blood glucose levels from dropping too low (hypoglycemia) *[Lee S.M. et al 2009]*.

Overall, our study has identified 14 glycolytic gene-markers and their corresponding metabolic elements (Glucose, Pyruvate Lactate and Lipids) controlling the energy production process in the tumors cells. It becomes evident so that the glycolytic profile of gliomas must be taken under consideration at the diagnosis and treatment stages.

5.4.6.3 The effect of tumor suppressors and oncogenes on glioma cells

Apart from the glycolytic genes, we considered 3 tumor suppressors and 3 oncogenes already known from bibliography *[Furnari F.B. et al 2007 - Ohgaki H. et al 2009 - Jones R.G. et al 2009]*. Their role was statistically verified in our study. More specifically and in agreement with current literature, our study verified the significance of the tumor suppressors (PTEN, Rb, TP53) and the oncogenes (CDK4, EGFR, PDGF), which play a crucial role in the control and regulation of vital functions of the cell such as cellular proliferation, cellular apoptosis, invasion and angiogenesis.

In addition to its well-known functions, such as maintenance of genomic integrity, cell cycle arrest, and apoptosis, recent evidence suggests that TP53 regulates mitochondrial respiration and glycolysis *[Matoba S. et al 2006 - Yin Y. et al 1992]*. TP53 influences the metabolic balance in cells between glycolysis and OXPHOS. Mutation of TP53 in tumors causes downregulation of mitochondrial respiration and a shift of cellular energy metabolism towards glycolysis. In terms of metabolism, loss of TP53 may provide a significant growth advantage to cancer cells since TP53-deficient tumors engaged in aerobic glycolysis are able to more readily access energetic and biosynthetic pathways that support anabolic synthesis. TP53 mutations are significantly more frequent in secondary glioblastomas than in primary glioblastomas (65% vs 28%) *[Ohgaki H. et al 2004]*.

The phosphatase and tensin homology (PTEN) tumor suppressor gene is encountered in high grade gliomas, in a percentage of 30%–44% *[Wang S.I. et al 1997]*. The high mutation rate of this gene in human cancers, points to the importance of cooperating genetic lesions in unleashing the full potential of PTEN loss of function in oncogenesis, progression, or tumor maintenance. In this regard, PTEN mutations are found in anaplastic astrocytomas (GRIII) and glioblastoma multiforme (GRIV), both primary and secondary *[Ohgaki H. et al 2004]*. Recent studies have shown that loss of function of PTEN function alters the relationship between glucose concentration and cell proliferation and increases glycolysis *[Ohgaki H. et al 2009 - Blouin M.J. et al 2010- Knobbe C.B. et al 2002]*.

Platelet-derived growth factor (PDGF) and epidermal growth factor receptor (EGFR) have important roles in gliogenesis. Amplification of EGFR gene occurs in about 40% of glioblastoma multiforme. Over expression of this gene has been found to be associated with poor prognosis. On the other hand, PDGF has a prominent over expression in low grade gliomas and is considered a critical regulator of gliogenesis *[Maher E.A. et al 2001]*. The retinoblastoma (Rb) is a tumor suppressor gene and its main function is the cell cycle progression. Due to this fact its mutations are encountered in high grade tumors such as glioblastoma rather than in low grade ones *[Maher E.A. et al 2001]*, Rb mutations are also

more often in secondary glioblastoma multiforme rather than in primary glioblastoma *[Ohgaki H. et al 2007]*. Furthermore, the function of the cyclin-dependent kinase 4 (CDK4) gene is vital since it is also a cell cycle regulation gene which promotes cell division *[Maher E.A. et al 2001 - Ohgaki H. et al 2007 - Rollbrocker B. et al 1996]*.

An important finding of this study is the fact that specific isoforms of these genes were also found significant. These were the PDGFA, PDGFB, PDGFRA, PDGFRB, Rb-1, RBL2, and TP53 binding protein-1 (TP53BP1). Furthermore the HIF-1 isoform Alpha was also detected. A common characteristic of rapidly growing tumors is that they become easily hypoxic owing to the inability of the local vasculature to supply adequate amount of oxygen. However, tumor cells can successfully escape hypoxia-mediated death by induction of the HIF-1 gene. Owing to the inability of mitochondria to provide enough ATP for cell survival under hypoxic conditions, tumor cells must upregulate the glycolytic pathway *[Simon M.C. et al 2006]*. Thus, hypoxia inducible factor-1alpha (HIF-1A) is a main responder to intracellular hypoxia and is over expressed in many human cancers, including gliomas *[Méndez O. et al 2010]*.

The isoforms Alpha and Beta of the PDGF and PDGFR genes that were found significant in our study have been reported to contribute to the gliomas growth. These genes are also expressed in the developing vasculature in both normal and pathological conditions, including tumor angiogenesis *[Hermanson M. et al 1992]*. Furthermore TP53BP1 mutations have an impact on gliomagenesis and are present in malignant gliomas, as illustrated in the Human Protein Atlas *[Internet Source: The Human Protein Atlas]*. Finally, it has been found that Rb-1 and RBL-2 retinoblastoma isoforms correlate with the degree of malignancy in gliomas *[Lapenna S. et al 2009]*. Therefore, the findings of our study support the conclusions of previous studies regarding these genes, but also enhance the list of glioma-related markers with glycolytic genes. The robust performance of the integrated marker list including both genomic and metabolic markers further enhances the importance of the glycolysis signature in glioma discrimination in support of Warburg hypothesis.

5.4.6.4 Targeting mitochondria for glioma treatment

According to Otto Warburg's effect the increased rate of glycolysis observed in gliomas is a result of mitochondrial defect which occurs due to low oxygenation in tumorous area. Limited oxygen in tumor causes inactivation of citric acid cycle (TCA or Krebs cycle) and oxidative phosphorylation (OXPHOS) in mitochondria and also inhibits the apoptotic function. The fact that apoptosis is disturbed leads to an anarchic tumor cell proliferation. In turn, the massive cell proliferation increases the need for oxygen within tumor areas.

It is well understood so that the impaired mitochondrial respiration is a key issue which should be given additional attention when new therapies and drugs are designed. The fact that the inactivation of mitochondrial activity which promotes unbalanced cell proliferation is a hallmark in glioma tumorigenesis makes this organelle attractive for further investigation aiming to design new therapeutic agents with maximum efficacy.

Recent experimental attempts in pharmacology focus on the identification of novel glycolysis inhibition agents that will stimulate the mitochondrial activity by downregulation of the glycolysis and lactate fermentation pathways. The main goal of these attempts is to reactivate the mitochondrial respiration (i.e. the citric acid cycle and OXPHOS) in glioma cells hoping that this will force tumor cells to choose the mitochondrial energy pathway instead of their usual energy supply via the glycolysis-lactate fermentation pathway. Waking up mitochondria in tumour cells will reactivate the apoptosis which will stop cell proliferation and so tumor growth.

Among the glycolysis inhibitors (agents) that are under investigation for their ability to stimulate the mitochondrial oxidation of glucose and suppress the glycolytic pathway in glioma cells, the most significant are shown in Table 5.7. In red color is a known drug (Gleevec) which is currently used in clinical practices.

Table 5.7 – Glycolysis inhibitors for anticancer treatment

Inhibitors	***Drug development stage***
2-Deoxyglucose (2-DG)	Clinical trials
Lonidamine (LND)	Clinical trials
3-Bromopyruvate (3-BrPA)	Pre-clinical
Imatinib (Gleevec)	Approved for clinical use
Oxythiamine (OT)	Pre-clinical

[Source: Pelicano H. et al 2006]

5.4.6.5 Controlling hypoxia for glioma treatment

The fact though that there is not enough oxygen to serve the rapid tumor cell growth, promotes hypoxia. Hypoxic tumors will eventually form necrotic areas due to insufficient oxygen supply. Although tumor tries to face the problem of low oxygenation by creating a microvascular environment to “feed” its cells, this will inevitably collapse due to structure and functionality defects, leading to even more severe hypoxia/necrosis. This is a common characteristic in approximately 80% of high grade gliomas such as glioblastoma multiforme (GBM) *[Kleihues P. et al 1997]*.

Under hypoxic conditions, oxidative phosphorylation in the mitochondria will not proceed normally because of insufficient oxygen supply, even if the mitochondria in cancer cells do not have structural defects. Increased glycolysis will result in elevated production of lactate through fermentation process, which leads to acidification of tumor tissue and provides a microenvironment that promotes and selects cells with malignant behaviors. Therefore, increased glycolysis may be viewed as cellular adaptation to hypoxia *[Gatenby R.A. et al 2004]*.

The cellular response to hypoxia is controlled by HIF-1 and especially the HIF-1a isoform which activates the expression of target genes involved in angiogenesis, glucose uptake, glycolysis, growth factor signalling, apoptosis, invasion, and metastasis *[Brahimi-Horn M.C. et al 2005]*. Importantly, hypoxia has been associated with chemotherapy resistance and reduced sensitivity to radiation therapy due to upregulation of HIF-1. Therapeutic resistance associated with hypoxia is a significant problem in clinical treatment of cancer, and inhibition of glycolysis may provide a novel approach to overcoming such resistance. Therefore efforts to prevent the formation of hypoxic environments within tumors' cells will enable chemotherapies to act more effectively. Finally this will significantly reduce the time period that the chemotherapy is applied and the toxicity that these drugs may cause.

5.4.7 Conclusions

Nowadays great efforts are made to identify reliable sets of markers for more effective diagnosis and treatment of aggressive brain tumors such as gliomas. Following this trend this study exhibits the role of the energy metabolism and especially glycolysis pathway, whose importance has been already reported since 1956 by Otto Warburg. Our results demonstrated that 14 glycolysis-related genes derived from the metabolic interplay of four known glycolytic metabolites such as glucose, pyruvate, lactate and lipids, have dominant role in the discrimination of low from high great gliomas. This fact points out the need to further evaluate their potential for diagnostic purposes especially at early stages of the disease.

Furthermore and in combination with the already known cancer inhibitors their contribution in the design of new therapeutic agents must be addressed. As shown in Table 5.7 such efforts have already started verifying the interest of the scientific community towards this direction. Many scientists now believe that the cancer's therapy has its roots in cell's metabolism. Recent findings demonstrate that stimulation of mitochondrial activity and restoration of the mechanisms of ATP generation, characteristics of non-malignant cells, might be an efficient tool in anticancer strategy. In particular, shifting cellular metabolism towards mitochondrial ATP production might overcome the effects of HIF-1-mediated upregulation of the glycolytic pathway. Several attempts have been made to prevent the formation of lactate and to redirect pyruvate metabolism towards oxidation in the mitochondria. Hence, combined strategies involving manipulation of both the glycolytic and the mitochondrial pathways might be useful tools in the elimination of cancer cells that would otherwise survive thanks to mitochondrial ATP production. This study confirms this necessity too.

References

Abe S. 2005, 'Support Vector Machines for Pattern Classification', Springer–Verlag, London.

Blouin M.J. et al 2010, 'Loss of function of PTEN alters the relationship between glucose concentration and cell proliferation, increases glycolysis, and sensitizes cells to 2-deoxyglucose', *Cancer Letters*, vol. 289, pp. 246-253.

Brahimi-Horn M.C. et al 2005, 'The hypoxia-inducible factor and tumor progression along the angiogenic pathway', *International Review of Cytology*, vol. 242, pp. 157–213.

Bushong E. A. et al 2002, 'Protoplasmic astrocytes in CA1 stratum radiatum occupy separate anatomical domains', *Journal of Neuroscience*, vol. 22, pp. 183-192.

Dominguez J.E et al 1987, 'Enzymes of glucose metabolism in cultured human gliomas: neoplasia is accompanied by altered hexokinase, phosphofructokinase, and glucose-6-phosphate dehydrogenase levels', *Metabolic Brain Disease*, vol. 2, pp. 17-30.

Furnari F.B. et al 2007, 'Malignant astrocytic glioma: genetics, biology, and paths to treatment', *Genes and Development*, vol. 21, pp. 2683-2710.

Galarraga J. et al 1986, 'Glucose metabolism in human gliomas: correspondence of in situ and in vitro metabolic rates and altered energy metabolism', *Metabolic Brain Disease*, vol. 1, pp. 279-91.

Garber K. 2006, 'Energy deregulation: Licensing tumors to grow', *Science*, vol. 312, pp. 1158–1159.

Gatenby R.A. et al 2004, 'Why do cancers have high aerobic glycolysis?', Nature Reviews Cancer, vol. 4, pp. 891-899.

Gordon, G. R. J. et al 2009, 'Astrocyte control of blood flow', n Parpura V., Haydon P.G. (eds) Astrocytes in (patho)physiology of the nervous system', *Springer-Verlag*, New York, pp. 461-486.

Greenblatt M.S. et al 1994, 'Mutations In The P53 Tumor Suppressor Gene: Clues To Cancer Etiology And Molecular Pathogenesis', *Cancer Research*, vol. 54, pp. 4855-4878.

Hermanson M. et al 1992, 'Platelet-derived growth factor and its receptors in human glioma tissue: expression of messenger RNA and protein suggests the presence of autocrine and paracrine loops', *Cancer Research*, vol. 52, pp. 3213-3219.

Jones R.G. et al 2009, 'Tumor suppressors and cell metabolism: a recipe for cancer growth', *Genes and Development*, vol. 23, pp. 537-548.

Kennedy C. et al 1976, 'Metabolic mapping of the primary visual system of the monkey by means of the autoradiographic [14C]deoxyglucose technique', *Proceedings of the National Academy of Sciences of USA* , pp. 4230-4234.

Kim J.W. et al 2006a, 'Cancer's molecular sweet tooth and the Warburg effect', *Cancer Research*, vol. 66, pp. 8927–8930.

Kim J.W. et al 2006b, 'HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia', *Cell Metabolism*, vol. 3, pp. 177–185.

Kleihues P. et al 1997, 'Pathology and Genetics of Tumours of the Nervous System', *International Agency for Research on Cancer*, Lyon.

Knobbe C.B. et al 2002, 'PTEN signalling in gliomas,' *Neuro-Oncology*, vol., pp. 4, pp. 196-211.

Kononenko I. 1994, 'Estimating attributes: analysis and extensions of relief', *Proceedings of ECML'94*, New York, 1994, pp. 171-182.

Kounelakis M.G. et al 2011a, 'On the relevance of glycolysis process on brain gliomas', *IEEE Transactions on Information Technology in BioMedicine*

Kounelakis M.G. et al 2011b, 'The role of the Glycolysis-related genes in Brain gliomas treatment design', *5th European Conference of IFMBE for Medical & Biomedical Engineering (MBEC)*, Budapest.

Kounelakis M.G et al 2010a, 'Targeting brain gliomas energy metabolism for classification purposes', *The 2010 IEEE International Conference on Imaging Systems and Techniques*, Thessaloniki.

Kounelakis M.G. et al 2010b, 'Validation of MRS metabolic markers in the classification of brain gliomas and their correlation to energy metabolism', *12th Mediterranean Conference on Medical and Biological Engineering and Computing*, Chalkidiki.

Lapenna S. et al 2009, 'Cell cycle kinases as therapeutic targets for cancer', *Nature Reviews Drug Discovery*, vol. 8, pp. 547-566.

Lopez-Lazaro M. 2008, 'The Warburg effect: why and how do cancer cells activate glycolysis in the presence of oxygen?', *Anticancer Agents in Medicinal Chemistry*., vol. 8, pp. 305–312.

Ledwozyw A. et al 1992, 'Phospholipids and fatty acids in human brain tumors', *Acta Physiologica Hungarica*, vol. 79, pp. 381-387.

Lee S.M. et al 2009, 'A nucleocytoplasmic malate dehydrogenase regulates p53 transcriptional activity in response to metabolic stress', *Cell Death and Differentiation*, vol. 16, pp. 738-748.

Levine A.J. et al 2010, 'The Control of the Metabolic Switch in Cancers by Oncogenes and Tumor Suppressor Genes', *Science*, vol. 330, pp. 1340-1344.

Luts J. et al 2007, 'A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection', *Artificial Intelligence in Medicine*, vol. 40, pp. 87-102.

Magistretti P.J. 1999, 'Brain energy metabolism', in Zigmond M., Bloom F.E., Landis S., Roberts J. and Squire L. (eds.) Fundamental Neuroscience, *Academic Press*, San Diego, pp. 389-413.

Magistretti P.J. 2006, 'Neuron–glia metabolic coupling and plasticity', *The Journal of Experimental Biology*, vol. 209, pp. 2304-2311.

Maher E.A. et al 2001, 'Malignant glioma: genetics and biology of a grave matter', *Genes and Development*, vol. 15, pp. 1311-1333.

Mathupala S.P. et al 2006, 'Hexokinase II: cancer's double-edged sword acting as both facilitator and gatekeeper of malignancy when bound to mitochondria', *Oncogene*, vol. 25, pp. 4777–4786.

Matoba S. et al 2006, 'p53 regulates mitochondrial respiration', Science, vol. 312, pp. 1650–1653.

Méndez O. et al 2010, 'Knock down of HIF-1α in glioma cells reduces migration in vitro and invasion in vivo and impairs their ability to form tumor spheres', *Molecular Cancer 2010*, vol. 9, pp. 133-143.

Ohgaki H. et al 2004, 'Genetic pathways to glioblastoma: a population-based study', *Cancer Research*, vol. 64, pp. 6892–6899.

Ohgaki H. et al 2007, 'Genetic Pathways to Primary and Secondary Glioblastoma', *The American Journal of Pathology*, vol. 170, pp. 1445-1453.

Ohgaki H. et al 2009, 'Genetic alterations and signalling pathways in the evolution of gliomas', *Cancer Science*, vol. 100, pp. 2235-2241.

Pedersen P.L. et al 2007, 'Warburg, me and Hexokinase 2: Multiple discoveries of key molecular events underlying one of cancers' most common phenotypes, the "Warburg Effect", i.e., elevated glycolysis in the presence of oxygen', *Journal of Bioenergetics and Biomembranes*, vol. 39, pp. 211–222.

Pellerin L. et al 1994, 'Glutamate uptake into astrocytes stimulates aerobic glycolysis: a mechanism coupling neuronal activity to glucose utilization', *Proceedings of the National Academy of Sciences of USA*, pp. 10625-10629.

Pelicano H. et al 2006, 'Glycolysis inhibition for anticancer treatment', *Oncogene*, vol. 25, pp. 4633–4646.

Postma G.J. et al 2011, 'On the relevance of automatically selected single-voxel MRS and multimodal MRI and MRSI features for brain tumour differentiation', *Computers in Biology and Medicine*, vol. 41, pp. 87-97.

Robnik-Sikonja M. et al 2003, 'Theoretical and empirical analysis of ReliefF and RReliefF', *Machine Learning*, vol. 53, pp. 23-69.

Rollbrocker B. et al 1996, 'Amplification of the cyclin-dependent kinase 4 (CDK4) gene is associated with high cdk4 protein levels in glioblastoma multiforme', *Acta Neuropathologica*, vol. 92, pp. 70-74.

Russel G. et al 2009, 'Tumor suppressors and cell metabolism: a recipe for cancer growth', *Genes & Development*, vol. 23, pp. 537-548.

Semenza G.L. et al 1996, 'Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1', *Journal of Biological Chemistry*, vol. 20, pp. 32529-32537.

Shiraishi T. et al 2003, 'Genetic alterations of human brain tumors as molecular prognostic factors', *Neuropathology*, vol. 23, pp. 95-108.

Simon M.C. et al 2006, 'Coming up for air: HIF-1 and mitochondrial oxygen consumption', *Cell Metabolism*, vol. 3, pp. 150-151.

Simonetti A.W. et al 2003, 'A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy', *Analytical Chemistry*, vol. 75, pp. 5352-5361.

Tennant D.A. et al 2009, 'Metabolic transformation in cancer', *Carcinogenesis*, vol. 30, pp. 1269–1280.

Vogelstein B. et al 2004, 'Cancer genes and the pathways they control', *Nature Medicine*, vol. 10, pp 789–799.

Waagepetersen H. S. et al 2009, 'Energy and amino acid neurotransmitter metabolism in astrocytes', in Parpura V., Haydon P.G. (eds) Astrocytes in (patho)physiology of the nervous system, *Springer-Verlag*, New York, pp. 177-200.

Wang S.I. et al 1997, "Somatic mutations of PTEN in glioblastoma multiforme," *Cancer Research*, vol. 57, pp. 4183–4186.

Wang Y. et al 2004, 'Application of Relief-F Feature Filtering Algorithm to Selecting Informative Genes for Cancer Classification Using Microarray Data', *Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference*, Stanford, 2004, pp. 497-498.

Warburg O. 1956 'On the origin of cancer cells', *Science*, vol.123, pp. 309–314.

Yanagawa T. et al 2004, 'Novel roles of the autocrine motility factor/phosphoglucose isomerase in tumor malignancy', *Endocrine Related Cancer*, vol. 11, pp. 749–759.

Yin Y. et al 1992, 'Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles,' *Cell*, vol. 70, pp. 937-948.

Zhao S. et al 2009, 'Glioma-Derived Mutations in IDH1 Dominantly Inhibit IDH1 Catalytic Activity and Induce HIF-1a', *Science*, vol. 324, pp. 261-265.

Internet sources

Elsom Research, http://www.elsomresearch.com

Futurity, http://www.futurity.org

Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo

KEGG Katehisa Laboratories, http://www.genome.jp/kegg/pathway/hsa/hsa05214.html

Glycolysis-Gluconeogenesis Pathways (KEGG), http://www.genome.jp/kegg/pathway/map/map00010.html

Human Metabolome Database, http://www.hmdb.ca

Radiology Spirit, http://radiologyspirit.blogspot.com

Kyoto Encyclopedia of Genes and genome (KEGG), http://www.jenome.jp/kegg

Matlab, http://www.mathworks.com

Medscape, http://www.medscape.com

National Human Genome Research Institute, http://www.genome.gov

SPSS, http:///www.spss.com

The Entrez Gene Database, http://www.ncbi.nlm.nih.gov/Entrez

The Gene Ontology, http://www.geneontology.org

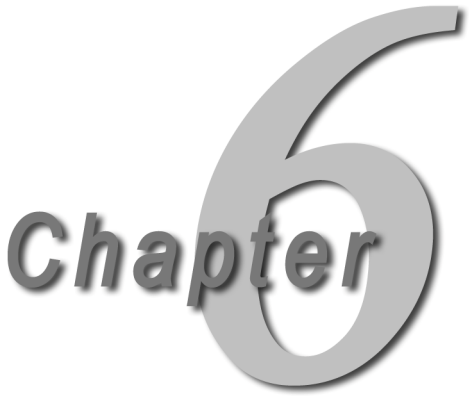
The Human Protein Atlas, http://www.proteinatlas.org/ENSG00000067369

The University of New South Wales, http://cellbiology.med.unsw.edu.au

The University of Utah, http://learn.genetics.utah.edu

US National Institute of Health, http://publications.nigms.nih.gov/biobeat/gallery/index.html

Weka, http://www.cs.waikato.ac.nz



GENERAL CONCLUSIONS & FUTURE PERSPECTIVES

Based on Author's published studies:

- ***M. G. Kounelakis***, M. Zervakis, G. C. Giakos and X. Kotsiakis, The role of the Glycolysis-related genes in Brain gliomas treatment design, *5th European Conference of IFMBE for Medical & Biomedical Engineering (EMBEC), Budapest, 2011*

6.1 Main Achievements of the Thesis

Fighting cancer and especially brain tumors like gliomas, is a tedious and continuous process. Experts world-widely strive hard to discover unique characteristics of tumors' pathological behavior in order to develop more effective methods for diagnosis, prognosis and treatment purposes. Although remarkable achievements in medicine and biomedicine have been recorded, particularly during the last decade, the fact that the majority of patients will not eventually survive more than 2 years from diagnosis, even if the most effective clinical practices are applied, remains an open challenge. The grand vision even today is to identify new ways to reliably integrate the knowledge extracted from all these efforts into the clinical practice in order to achieve a better management of this lethal disease and therefore prolong patients' life expectancy.

Sharing the same anxiety, our research aimed to add a small piece of knowledge and technical know-how into this puzzle. Motivated also by Socrates' (Ancient Greek Philosopher) say that *"the knowledge is in the whole, not in the part"* we decided to study the brain tumor as a pathological system. This decision came from the fact that the formation of a malignant tumor within the brain is the result of a series of abnormal biological processes rather than just an abnormal snapshot. By analyzing these biological processes we can derive valuable knowledge that can be used to better understand this type of cancer and identify methods to manage its unpredicted behavior.

These initial considerations directed our thoughts to the use of the "omics" science and technologies. More specifically we decided to study brain tumors at genomic and metabolic level (the two ends of the "omics" chain) and then exploit the extracted knowledge in order to derive important observations regarding brain tumors' behavior. Based on this strategy two main goals were determined. First to identify reliable genetic and metabolic features of brain tumor that will enable a more accurate discrimination of their type and grade but also open new ways towards the design of new treatment practices. Second, to propose a novel medical Decision Support System (DSS) which will integrate the knowledge extracted from this multimodal analysis (genomics and metabolomics) in order to develop an effective protocol for a non-invasive, patient specific, diagnostic and treatment decision making.

We believe that these two main goals were accomplished and justified by the results and their analyses obtained at both biological and statistical level, presented in this thesis. Our achievements are presented below.

6.1.1 Achievements at clinical level

- Reveal the benefits of using non-invasive imaging techniques such as Magnetic Resonance Spectroscopy (MRS) for the diagnosis of brain tumor and emphasize the need to develop even better ones in order to minimize the life risk behind the application of surgical operations (biopsy and tumor removal).
- Prove that the metabolic profile of brain tumors must be taken under consideration in the design of novel non-invasive treatments for brain tumors.
- Identify optimal sets of the most significant metabolic markers which can be used at clinical level to provide accurate discrimination of brain gliomas and meningiomas (Table 6.1).
- Understand the significant role of glycolysis and oxidative phosphorylation (OXPHOS) in tumor genesis and proliferation and provide a set of glycolysis-related genetic markers that can improve the diagnostic accuracy at clinical level (Table 6.2). These markers can also be combined with the already known tumor suppressors and oncogenes to further improve this accuracy.
- Exhibit the need to design new gene therapies based on glycolysis inhibitors and mitochondrial reactivation agents which will force cancer cells to death (i.e. reactivate apoptosis) and therefore inhibit proliferation of cancer cells.
- Propose a novel medical Decision Support System (DSS) which provides a protocol for non-invasive diagnosis and treatment based on the genetic and metabolic characteristics of brain tumors (Figure 6.1). According to this medical DSS the genetic and metabolic markers found can both support the conventional diagnostic and treatment practices and establish the foundations for the development of new non-invasive methodologies *(green arrows in Figure 6.1)*.

6.1.2 Achievements at theoretical level

- Exhibit the remarkable potential of maximal margin classifiers such as Support Vector Machines (SVM) and Least Squares - SVM in cancer discrimination problems, where data is complex and heterogeneous, due to their ability to non-linearly map the input data (kernel trick) to a higher feature space where data is linearly separable.
- Prove that embedding filter methods such as, Fisher's criterion and Relief-F ranking, which reveal the intrinsic characteristics of the data, into the SVM classifier improves feature selection and classification process.
- Show that the selection and preprocessing of the initial brain tumor features, inputted into the classifier, significantly influence the identification of the final cancer markers' set and so the classification of a new patient.

- Create a biological bridge from genomics to metabolomics, based on glycolysis effect on brain tumors, which confirms Otto Warburg's hypothesis that altered glycolytic rates are present in brain tumors.
- Show the strengths and weaknesses of increasing the MRS system magnet's power and its impact on data mining for brain tumor diagnostic purposes.

Table 6.1 – Optimal sets of metabolic markers (peak area's ratios) for the discrimination of brain gliomas and meningiomas – Optimal set is the set that includes the smallest possible number of markers providing the maximum discrimination accuracy.

Binary schemes	***Metabolic ratio markers identified***							
Healthy vs tumor	NAA / Cho	NAA / S	Cho/ S	NAA/ Cre	Lips/ Cho	Cho/ Cre		
Healthy vs Glio	NAA / Cho	mI / S	Cho / S	Cho/ Cre	Lac / Cre	Ala / Cre		
Healthy vs Mng	mI / S	NAA / S						
GRII vs GRIII	Cre / S	Lips/ Cre	Lac / Cre	Ala / S	Ala/Cre	NAA/ Cre	Lips/ Cho	mI / S
GRII vs GRIV	Cre / S	Cho / S	mI / S	Ala / S				
GRIII vs GRIV	Cho / S	mI / S						
Glio vs Mng	Ala / Cre	Ala / S	NAA / S	mI / S	Lips / Cre	Lac / Cre		

Table 6.2 – The optimal sets of genetic markers identified that can be used for the discrimination of low grade from high grade gliomas – Left column presents the 14 glycolysis-related markers while right column the 12 tumor suppressors and oncogenes

Glycolysis genetic markers		***Tumor suppressors and oncogenes***	
ALDOA	MDH1	RB1	CDK4
ALDOC	PDHB	RBL2	EGFR
ENO2	PFKM	TP53	PDGFA
GAPDH	PGI	TP53BP1	PDGFB
HK2	PGK1	PTEN	PDGFRA
LDHA	PGM1	HIF-1A	PDGFRB
LDHB	PKLR		

Genes' names are from the Entrez Gene Database

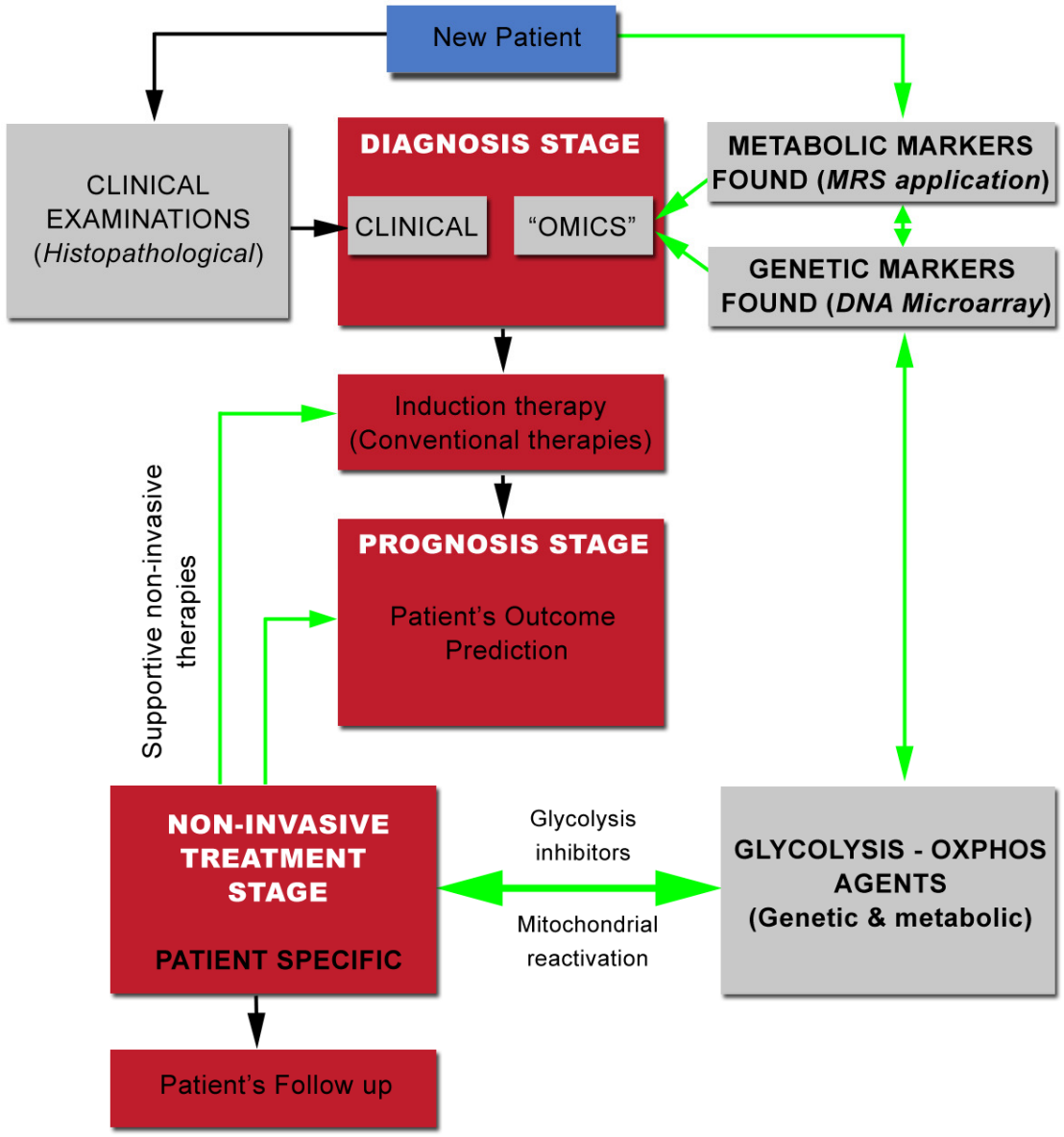


Figure 6.1 – The proposed medical Decision Support System – The "omics" technologies (green arrows) can both support the conventional diagnosis and treatment pathways but also provide an alternative non-invasive way to diagnose and treat brain tumors - (The figure was designed with Adobe Photoshop ver. CS5)

6.2 Future Perspectives

6.2.1 Improving the non-invasive diagnosis of brain tumors

Diagnosis of a disease is the first and most basic stage of a clinical procedure. Determining the type and stage of a disease has a direct impact on the selection of the therapeutic pathway. Especially in brain tumors this becomes a more urgent process since this type of cancer is highly aggressive and most of the times the prognosis is poor.

Our research has shown that the adoption of non-invasive diagnostic imaging techniques such as MRS and DNA Microarrays into clinical practices provides significant metabolic and genomic characteristics (markers) of brain tumors that have great diagnostic potential even in complex tumors such as gliomas. Based on the proposed medical DSS shown in Figure 6.1, these markers can be combined with those derived from the conventional clinical tests to increase the diagnostic outcome. *But is this actually the end of the story? The answer is no.*

If additional "omics" data from brain tumor patients are available then more significant characteristics can be possibly discovered, facilitating the design of more reliable diagnostic protocols. Such type of data can be obtained from several other non-invasive metabolic imaging techniques such as Diffusion Weighted Imaging (DWI), Perfusion Weighted Imaging (PWI) but also proteome analysis techniques such as Mass Spectrometry (MS) *[Riyadh N. et al 2006 - Lam W.W. et al 2002 - Provenzale J. et al 2006 - Calli C. et al 2006].* Furthermore, other very promising optical measurement systems such as polarimetric systems for cancer diagnosis can be integrated too *[Giakos G.C., Kounelakis M.G. et al 2010 - Giakos G.C., Kounelakis M.G. et al 2011(a) - Giakos G.C., Kounelakis M.G. et al 2011(b)]*.

All these methods allow a multimodal data analysis where fusion techniques can be applied to create visual diagnostic and treatment aids and the evaluation of an interactive graphical user interface to communicate and adapt the decision support system's statistical estimates to the clinician's feedback in real time.

While additional performance gains may be achievable by optimizing the medical DSS, the practical use of such decision support tools in a clinical environment relies heavily on the ability to provide coherent performance across multiple datasets and pathological classes. Unfortunately, finding new patients data and especially "omics" data for research use is not always a simple task. Several ethical, economic and political factors either prohibit or restrict the publication of such datasets for further research.

Further diagnostic improvement can also be achieved through the development of new pattern recognition methods for the selection of optimal cancer markers and classification of complex tumors. The most recent research tendency relates to the development of mathematical models that monitor cancer's behaviour at metabolic, genomic and proteomic level in order to facilitate a more accurate prediction of its metastatic and proliferating activity. The generation of mathematical growth models is an important tool for both clinical and research communities in oncology. They give us the opportunity to interpret and integrate experimental results made in diverse fields of cancer research by providing a common

mathematical ground to combine them in *[Hogea C. et al 2007 - Basanta D. et al 2008 - Hatzikirou H. et al 2005 - Clatz O. et al 2005 - Gu S. et al 2011].*

Towards this direction a future plan to optimise the decision making is by applying classifier's fusion techniques which will enable the identification of new biomarkers that can be used to define cancer growth models.

6.2.2 The need for a patient specific treatment protocol

Among the challenges that oncologists have to face today is the selection of the most suitable and effective therapy in order to obtain the best possible outcome for the patient. Based on the type and grade of the brain tumor a combination of conventional therapies (chemotherapy, radiotherapy and/or surgery) is applied. Then, based on the clinical image of the patient and the prognosis estimation of the disease, either a second therapeutic scheme or a maintenance scheme is followed.

The problem however is that most of the times the therapeutic protocols followed are more or less the same for two tumors presenting similar molecular profiles. For example, we might find that the histopathological characteristics in a subtype of gliomas are the same as those in a subtype of oligodendrogliomas. In that case, the treatment that works in the glioma may be appropriate for the oligodendroglioma. *But is this the right way to face these two different tumors? And if this is correct why then their response to the same therapy is often different? Again the answer is no.*

It is now known that each tumor must be managed as a unique case. No two tumours are the same, even within the same type of cancer. They may look the same under the microscope, but their molecular aberrations vary greatly. For these reasons modern medicine has moved from *diseased-based to patient-based treatment protocols [Van't Veer et al 2008].* The patient-specific therapies being developed are generally called *targeted therapies*.

Nearly all of today's advances in cancer treatment are based on the development of targeted therapies. Yet, although targeted therapies represent major advances in how cancer is treated, their main deficiency is that they are not leading to cures. They offer better outcomes, longer survival, better quality of life, but not the eradication of cancer (so far). The reason is that each drug only targets a few of the genes that drive a cancer. And there are dozens that must be targeted.

Our research has followed this tendency providing optimal sets of cancer biomarkers at both genomic and metabolomic level, that are capable to reveal the unique characteristics of brain tumors, facilitating the design of targeted therapies. Furthermore, based on the glycolysis pathway that has been proved to be very significant for the determination of the tumor behaviour *[Warburg O. 1956]*, new anti-cancer therapeutic agents can be developed, as our medical DSS suggests in Figure 6.1. For example generating glycolysis inhibitors which will force cancer cells to switch from glycolysis to OXPHOS for their energy demands, can be an effective way to reactivate the apoptosis and therefore stop their proliferation *[Pelicano H. et al 2006 – Kounelakis M.G. et al 2011]*. Towards this direction today's research

on pharmacogenetics has generated specific glycolysis agents that for the time being are under a trial period. These are shown in Table 6.3.

Closing this thesis we would like to express the wish and hope that the day where humanity will defeat cancer is not far away.

Table 6.3 – Prospective targeted therapies and their effects

Target gene	***Drugs***	***Effect***
HK2	2-DG	Glycolysis inhibitor leading to apoptosis
	3-BrPA	Decrease in intracellular ATP; Apoptosis
	Lonidamine	Glycolysis inhibitor
GAPDH	KA	Apoptosis
HIF-1a	FK228	HIF-1a downregulation
	NO	HIF-1a downregulation
Glut1	mAb	Reduced tumor proliferation

References

Basanta D. et al 2008, 'Studying the emergence of invasiveness in tumours using game theory, European Physical Journal B, vol. 63, pp. 393-397.

Calli C. et al 2006, 'Perfusion and diffusion MR imaging in enhancing malignant cerebral tumors', *European Journal of Radiology,* vol. 58, pp. 394–403.

Clatz O. et al 2005, 'Realistic simulation of the 3d growth of brain tumors in MR images coupling diffusion with biomechanical deformation', *IEEE Transactions on Medical Imaging,* vol. 24, pp. 1334-1346.

Giakos G.C., Kounelakis M.G. et al 2011a, 'Polarimetric phenomenology of photons with lung cancer tissue', *IOP Measurement Science & Technology,* vol. 22, doi:10.1088/0957-0233/22/11/114018.

Giakos G.C., Kounelakis M.G. et al 2011b, 'Near Infrared Light Interaction with Lung Cancer Cells', *IEEE International Instrumentation & Measurement Technology Conference, Binjiang-Hangzhou.*

Giakos G.C., Kounelakis M.G. et al 2010, 'Backscattered polarimetric detection from biological tissue', *IEEE International Conference on Imaging Systems and Techniques, Thessaloniki.*

Gu S. et al 2011, 'Applying a patient-specific bio-mathematical model of glioma growth to develop virtual [18F]-FMISO-PET images', Mathematical Medicine and Biology, in press.

Hatzikirou H. et al 2005, 'Mathematical modelling of gliobastoma tumour development: a review', *Mathematical Models and Methods in Applied Sciences,* vol. 151, pp. 1779-1794.

Hogea C. et al 2007, 'A robust framework for soft tissue simulations with application to modelling brain tumor mass effect in 3D MR images, *Physics in Medicine and Biology,* vol. 52, pp. 6893-6908.

Kounelakis M.G. et al 2011, 'The role of the Glycolysis-related genes in Brain gliomas treatment design', *5th European Conference of IFMBE for Medical & Biomedical Engineering (EMBEC), Budapest.*

Lam W. W. et al 2002, 'Diffusion MR imaging in glioma: does it have any role in the pre-operation determination of grading of glioma?, *Clinical Radiology,* vol. 57, pp. 219–225.

Pelicano H. et al 2006, 'Glycolysis inhibition for anticancer treatment', *Oncogene*, vol. 25, pp. 4633–4646.

Provenzale J. et al 2006, 'Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response', *Radiology,* vol. 239, pp. 632–49.

Riyadh N. et al 2006, 'Advanced MR Imaging Techniques in the Diagnosis of Intraaxial Brain Tumors in Adults', *RadioGraphics,* vol. 26, pp. 173-189.

Van't Veer L.J. et al 2008, 'Enabling personalized cancer medicine through analysis of gene-expression patterns', *Nature,* vol. 452, pp. 564–570.

Warburg O. 1956 'On the origin of cancer cells', *Science*, vol.123, pp. 309–314.

Internet sources

The Entrez Database, http://www.ncbi.nlm.nih.gov/Entrez

Appendices

ACRONYMS AND ABBREVIATIONS

ALDOA: *Aldolase-A (gene)*
ALDOC: *Aldolase-C (gene)*
AML: *Acute Myeloid Leukemia*
ATP: *Adenosine Triphosphate*
AUROC: *Area under ROC curve*
Bl_bm: *Number of blasts in bone marrow (refers to Acute Myeloid Leukemia)*
CDK4: *Cyclin-dependent kinase 4 (gene)*
Cho: *Choline (metabolite)*
CBTRUS *Central Brain Tumor Registry of the US*
CI: *Confidence Interval*
CNS: *Central Nervous System*
Cre: *Creatine (metabolite)*
CP: *Circularly Polarised*
CR: *Complete Remission (refers to Acute Myeloid Leukemia)*
CSF: *Cerebrospinal Fluid*
CSI: *Chemical Shift Imaging*
CT: *Computed (or Computerised) Tomography*
CTA: *CT Angiography*
CSI: *Chemical Shift Imaging*
CV: *Cross Validation*
DSS: *Decision Support System*
DTPA: *Diethylenetriaminepenta-acetic acid*
DWI: *Diffusion Weighted Imaging (or Diffusion MRI)*
EGFR: *Epidermal Growth Factor Receptor (gene)*
ENO2: *Enolase-2 (gene)*
Exm: *Extramedullary Infiltration (refers to Acute Myeloid Leukemia)*
FAB: *French-American-British*
FDA: *Fisher discriminant analysis*
FFT: *Fast Fourier Transform*
FID: *Free Induction Decay*
FLAIR: *Fluid Attenuated Inversion recovery*
FOV: *Field of view*
FN: *False Negative*
FNR: *False Negative Rate*
FP: *False Positive*
FPR: *False Positive Rate*
GAPDH: *Glyceraldehyde-3-phosphate dehydrogenase (gene)*
GC: *Gliomatosis Cerebri*

GBM: *Glioblastoma multiforme (or GRIV)*
Gd: *Gadolinium enhanced*
GE: *Gradient Echo*
GIMEMA: *Gruppo Italiano Malattie Ematologiche dell'Adulto*
Gluc: *Glucose (metabolite)*
Glu1: *Glutamate (metabolite)*
Glu2: *Glutamine (metabolite)*
Glx: *Glutamine and glutamate*
GRI: *Glioma grade I*
GRII: *Glioma grade II*
GRIII: *Glioma grade III*
GRIV: *Glioma grade IV (or GBM)*
Hb: *Hemoglobin (refers to Acute Myeloid Leukemia)*
HIF-1: *Hypoxia Inducible Factor (gene)*
HIF-1A *Hypoxia Inducible Factor-1A (gene)*
HK2: *Hexokinase-2 (gene)*
HLA: *Human Leukocyte Antigens*
HLSVD: *Hankel Lanczos Singular Value Decomposition*
HMP: *Human Metabolome Project*
^{1}H-MRS: *Proton Magnetic Resonance Spectroscopy*
ID: *Induction Death (refers to Acute Myeloid Leukemia)*
IR: *Inversion Recovery*
KEGG: *Kyoto Encyclopedia of Genes and Genome*
Lac: *Lactate (metabolite)*
LDA: *Linear Discriminant Analysis*
LDHA: *Lactate Dehydrogenase-A (gene)*
LDHB: *Lactate Dehydrogenase-B (gene)*
LOOCV: *Leave one out Cross Validation*
LS-SVM: *Least Squares Support Vector Machine*
Lips: *Lipids*
LL: *Lactate-Lipids*
LUH: *Larissa University Hospital*
MDH1: *Malate Dehydrogenase-1 (gene)*
mI: *Myo-inositol (metabolite)*
ML: *Myeloid Leukemia*
MR: *Magnetic Resonance*
MRA: *Magnetic Resonance Angiography*
MRI: *Magnetic Resonance Imaging*
MRS: *Magnetic Resonance Spectroscopy*
MRSI: *Magnetic Resonance Spectroscopy Imaging*

MS: *Mass Spectroscopy*
NAA: *N-acetyl-aspartate (metabolite)*
NEX: *Number of Excitations or Acquisitions*
NMR: *Nuclear Magnetic Resonance*
NMRI: *Nuclear Magnetic Resonance Imaging*
NN: *Neural Network*
OXPHOS: *Oxidative Phosphorylation*
PD: *Proton Density*
PDGF: *Platelet-derived growth factor (gene)*
PDHB: *Pyruvate Dehydrogenase-B (gene)*
PET: *Positron Emission Tomography*
PFKM: *Phosphofructokinase-M (gene)*
PGI: *Phosphoglucose Isomerise (gene)*
PGK1: *Phosphoglycerate Kinase-1 (gene)*
PGM1 *Phosphoglycerate mutase-1 (gene)*
PKLR: *Pyruvate Kinase-LR (gene)*
Plts: *Platelets in blood (refers to Acute Myeloid Leukemia)*
PNET *Primitive Neuroectodermal Tumors*
ppm: *Parts per million*
PR: *Partial Remission (refers to Acute Myeloid Leukemia)*
PRESS: *Point Resolved Spectroscopy*
PS: *Performance Status (refers to Acute Myeloid Leukemia)*
PTEN: *Phosphatase and tensin homology (gene)*
Pyr: *Pyruvate (metabolite)*
PWI: *Perfusion Weighted Imaging (or Perfusion MRI)*
QP: *Quadratic Programming*
Rb: *Retinoblastoma (gene)*
RBF: *Radial Basis Function*
Res: *Resistance (refers to Acute Myeloid Leukemia)*
RF: *Radio Frequency*
RFE: *Recursive Feature Elimination*
RNA: *Ribonucleic Acid*
ROC: *Receiver Operating Characteristic*
SE: *Spin Echo*
SNR: *Signal–to–Noise Ratio*
STEAM: *Stimulated-Echo Acquisition Mode*
SVD: *Singular Value Decomposition*
SVM: *Support Vector Machine*
SVS: *Single Voxel Spectroscopy*
T: *Tesla*

TCA: *The Kreb's or Citric Acid cycle in the mitochondrion*

TE: *Echo time*

TN: *True Negative*

TNR: *True Negative Rate*

TP: *True Positive*

TP53: *Tumor Protein-53 (gene)*

TP53BP1: *Tumor Protein-53-BP1 (gene)*

TPR: *True Positive Rate*

TR: *Repetition time*

T1: *Longitudinal relaxation time*

T2: *Transverse relaxation time*

UMCN: *University Medical Center of Nijmegen*

VOI: *Volume of interest*

WBC: *White Blood Cells*

WHO: *World Health Organisation*

MEDICAL, BIOMEDICAL AND MAGNETIC RESONANCE TERMINOLOGIES

Acquisition time: *the period of time required to collect the image data.*

Aerobic respiration: *the cellular respiration that requires oxygen to produce energy (ATP).*

Allele: *is one of two or more forms of a gene.*

Allogenic Bone Marrow Transplantation: *is a procedure in which a person receives stem cells from a genetically similar, but not identical, donor.*

Anabolism: *is the set of metabolic pathways that construct molecules from smaller units.*

Anaerobic respiration: *the cellular respiration without oxygen's presence.*

Anaplastic: *cancerous cells that divide rapidly and bear little or no resemblance to normal cells.*

Angiography: *x-ray examination of blood vessels.*

Angular Momentum: *a property of a mass or system of masses turning about some fixed point; it is conserved in the absence of the action of external forces.*

Autologous Bone Marrow Transplantation: *is a procedure in which a person receives stem cells from himself.*

Apoptosis: *is the programmed cell death in multicellular organisms.*

Astrocytes: *is a type of glial cells that support and nourish neurons. When the brain is injured, astrocytes form scar tissue that helps repair the damage.*

Astrocytoma: *a type of brain tumor that begins in the brain or spinal chord in small, star-shaped cells called astrocytes.*

Bases: *or nucleotide bases are a group of nitrogen-based molecules that are required to form nucleotides (adenine A - cytosine C - thymine T - guanine G).*

Benign: *a tumor that is not cancerous.*

Biopsy: *removal of a small piece of living tissue from a part of the body for microscopic examination.*

Blood-brain barrier: *a barrier between the blood and the tissues of the CNS (brain and spinal cord) that prevents the entry of many drugs, including chemotherapy drugs.*

Bone marrow: *is the flexible tissue found in the interior of bones. In humans, bone marrow in large bones produces new blood cells.*

Brain Stem: *is located at the bottom of the brain and controls many vitally important functions including motor and sensory pathways, cardiac and respiratory functions, and reflexes.*

Cancer: *a term for diseases in which abnormal cells divide without control.*

Calcification: *is the process in which calcium salts built up in soft tissue, causing it to harden.*

Catabolism: *is the set of metabolic pathways that break down molecules into smaller units and release energy.*

Cellular respiration: *is the set of the metabolic reactions and processes that take place in the cells of organisms to convert biochemical energy from nutrients into adenosine triphosphate (ATP), and then release waste products.*

Cerebrum: *is the largest part of the brain and is associated with conscious thought, movement and sensation. It consists of two halves, each controlling the opposite side of the body. Four lobes make up the cerebrum: the frontal, temporal, parietal, and occipital lobes.*

Cerebellum: *is located at the lower back of the head and is connected to the brain stem. It is the second largest structure of the brain and is made up of two hemispheres. The cerebellum controls complex motor functions such as walking, balance, posture, and general motor coordination.*

Cerebrospinal Fluid (CSF): *is a clear substance that circulates through the brain and spinal cord. It provides nutrients and serves to cushion the brain and therefore protect it from injury.*

Chemical Shift: *is the change in the nuclear magnetic resonance frequency of a nucleus depending on its electronic environment; used in Nuclear Magnetic Resonance spectroscopy to determine the structure of molecules.*

Chemotherapy: *the prevention or treatment of a tumor by the use of chemical substances.*

Chromosome: *is an organized structure of DNA and protein found in cells.*

CNS Lymphomas: a type of non-neuroepithelial brain tumor

Complete Remission (CR): *refers to the situation where the disease (here Acute Myeloid Leukemia) disappears completely with the treatment.*

Computed Tomography (CT) scan: *a computer aided x-ray used to provide a picture of the inside of the body.*

Coronal plane: *a tomographic imaging plane that is perpendicular to the ground.*

Craniopharyngiomas: *a type of brain tumor.*

Craniotomy: *surgical opening of a portion of the skull performed to expose a lesion of the brain.*

Cytosol: *is the liquid inside the cells.*

DNA: *is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms.*

Echo time (TE): *represents the time in milliseconds between the application of the 90° pulse and the peak of the echo signal in Spin Echo and Inversion Recovery pulse sequences*

Edema: *an accumulation of an excessive amount of watery fluid in cells, tissues or serous cavities.*

Eddy current: *an induced spurious electrical current produced by time-varying magnetic fields. Eddy currents can cause artifacts in images and may seriously degrade overall magnet performance.*

Ependymal cells: *These cells line the ventricles within the central part of the brain and spinal cord that provide the pathway through which cerebrospinal fluid travels.*

Ependymomas: *a type of neuroepithelial brain tumor arising in the ependymal cells.*

Ex vivo: *in biology refers to the experiment that takes place outside an organism.*

Excitation: *delivering (inducing, transferring) energy into the "spinning" nuclei via radio-frequency pulse(s), which puts the nuclei into a higher energy state. By producing a net transverse magnetization an MRI system can observe a response from the excited system.*

Flow Cytometry: *is a technique for counting and examining microscopic particles, such as cells and chromosomes, by suspending them in a stream of fluid and passing them by an electronic detection apparatus.*

Frontal Lobe: *is one of the four lobes of the cerebral hemisphere.*

Free Induction Decay (FID): *is the signal generated when the excited nuclei relax. Its amplitude becomes smaller (decays) over time as net magnetisation returns to equilibrium.*

Gene: *is a molecular unit of heredity of a living organism.*

Genome: *in modern molecular biology and genetics, the genome is the entirety of an organism's hereditary information (stored in genes).*

Genotype: *is the genetic makeup of a cell, an organism, or an individual*.

Germ cell tumor: *a type of a non-neuroepithelial brain tumor.*

Glial cells: *There are 3 types of glial cells: astrocytes, oligodendrocytes, and ependymal cells. Most brain and spinal cord tumors develop from glial cells.*

Glial tissue: *the special connective tissue of the central nervous system.*

Glioblastoma: *a general term that refers to malignant astrocytoma, a type of brain tumor.*

Glioblastoma multiforme (GBM): *a type of astrocytic brain tumor that forms from glial (supportive) tissue of the brain. It grows very quickly and has cells that look very different from normal cells. Also called grade IV astrocytoma.*

Glioma: *a cancer of the brain that begins in glial cells (cells that surround and support nerve cells).*

Gluconeogenesis: *is a metabolic pathway that results in the generation of glucose from non-carbohydrate carbon substrates such as lactate, glycerol, and glucogenic amino acids.*

Glycolysis: *is the 'lysis' of glucose (or break down of blood sugar) for the production of energy necessary for the cell.*

High grade tumor: *a tumor that grows quickly over a period of a few months.*

Histology: *the study of tissues and cells under a microscope.*

Histopathology: *the study of the microscopic structure, composition and function of diseased tissue.*

In vivo: *in biology refers to the experiment done on a living organism.*

In vitro: *in biology refers to an experiment done in an artificial environment outside the living organism*.

Induction Death (ID): *the patient's death after the induction treatment (refers to Acute Myeloid Leukemia).*

Karyotype: *is the number and appearance of chromosomes in the nucleus of a cell.*

Lactate fermentation: *the biological process by which glucose is converted into cellular energy and the metabolic byproduct lactate.*

Larmor Frequency: *the frequency at which magnetic resonance in a nucleus can be excited and detected. The frequency varies directly with magnetic field strength, and is normally in the radio frequency (RF) range.*

Lesion: area of tissue with impaired function as a result of damage by disease or wounding.

Low grade tumor: *a tumor that develops slowly over a number of years.*

Lymphocyte: *a small white blood cell (leukocyte) that plays a large role in defending the body against disease.*

Magnetic Resonance: *the absorption or emission of energy by atomic nuclei in an external magnetic field after the application of RF excitation pulses using frequencies which satisfy the conditions of the Larmor equation.*

Magnetic Resonance Imaging (MRI): *a procedure which shows a picture, in any plane, of the inside of the body. This procedure is non-invasive as it uses magnetic waves.*

Magnetic Resonance Spectroscopy (MRS): *a procedure which reveals the metabolic spectrum of the inside of the body. This procedure is non-invasive as it uses magnetic waves.*

Malignant: *cancerous*

Malignant tumor: *a tumor that invades and destroys the tissue where it originates and which can spread to other sites in the body.*

Medulloblastomas: *a type of neuroepithelial brain tumor.*

Membrane: *a thin layer of tissue surrounding an organ, lining a cavity or separating adjacent structures or cavities.*

Meninges: *the three connective tissue membranes that line the skull and vertebral canal and enclose the brain and spinal cord.*

Meningiomas: *a type of a non-neuroepithelial brain tumor arising in meninges*

Metabolism: *is the set of chemical reactions that happen in living organisms to sustain life. These processes allow organisms to grow and reproduce, maintain their structures, and respond to their environments. Metabolism is usually divided into two categories, catabolism and anabolism.*

Metabolites: *are the intermediates and products of metabolism. The term metabolite is usually restricted to small molecules. A primary metabolite is directly involved in normal growth, development, and reproduction.*

Metabolic pathway: *is series of chemical reactions occurring within a cell. The collection of pathways is called the metabolic network.*

Metabolome: *refers to the complete set of metabolites to be found within a biological sample, such as a single organism.*

Metabolomics: *is the field in biology which studies the metabolome of an organism. It is the last part in the "omics" chain (genomics-transcriptomics-proteomics-metabolomics).*

Metastasis: *the spread of cancer from one part of the body to another.*

Mitochondrion: *is a membrane-enclosed organelle found in cells.*

Mitosis: *the process where a single cell divides resulting in generally two identical cells, each containing the same number of chromosomes and genetic content as that of the original cell.*

Mutations: *are changes in the DNA sequence of a cell.*

Necrosis: *pathologic death of cells, tissue, or an organ from irreversible damage*.

Neoplasm: *is an abnormal mass of tissue such as a tumor.*

Neurons: *These are the most important cells within the brain. They carry signals through long, wire-like extensions called axons.*

Nuclear Magnetic Moment: *a measure of the net magnetic properties of a particle.*

Nuclear Magnetic Resonance (NMR): *is the physical phenomenon of the magnetic property of the nuclei.*

Nuclear spin: *also known as inherent spin, this defines the intrinsic property of certain nuclei (those with odd numbers of protons and/or neutrons in their nucleus) to exhibit angular momentum and a magnetic moment. Nuclei that do not exhibit this characteristic will not produce an NMR signal.*

Nucleotides: *are molecules that, when joined together, make up the structural units of DNA.*

Occipital lobe: *is one of the four lobes of the cerebral hemisphere. It is located in the back of the head and controls vision.*

Oligodendrocytes: *these cells make myelin. Myelin forms a layer that surrounds and insulates axons of the brain and spinal cord. In this way, oligodendrocytes help neurons transmit electric signals through axons.*

Oligodendroglioma: *a type of neuroepithelial brain tumor arising in oligodendrocytes.*

Oncogenes: *is a gene that has the potential to cause cancer. In tumor cells, they are often mutated or expressed at high levels.*

Oxidative Phosporylation: *is a metabolic pathway that uses energy released by the oxidation of nutrients to produce adenosine triphosphate (ATP).*

Parietal lobe: *is one of the four lobes of the cerebral hemisphere. It controls tactile sensation, response to internal stimuli, sensory comprehension, some language, reading, and some visual functions.*

Partial Remission (PR): *refers to the situation where the disease (Acute Myeloid Leukemia here) shrinks but not completely disappears completely with the treatment.*

Phenotype: *are an organism's observable characteristics such as its morphology, biochemical or physiological properties and behaviour.*

Pituitary Gland: *is a small, bean-sized organ that is located at the base of the brain.*

Primary tumor: *tumor that only occurs in the organ in which it started.*

Prognosis: *the assessment of the future course and outcome of a patient's condition.*

Proteins: *are biochemical compounds, essential parts of organisms and participate in virtually every process within cells. Many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism.*

Proteome: *the complete set of the proteins of a living organism.*

Proteomics: *the field of biology which studies the proteome.*

Proton: *a subatomic particle found in the atom's nucleus.*

Pulse sequence: *a preselected set of defined RF and gradient pulses, usually repeated many times during a scan.*

Radiotherapy: *a therapeutic procedure in which conditions are treated with energy in the form of waves or particles.*

Relaxation time: *after excitation the spins will tend to return to their equilibrium distribution in which there is no transverse magnetization and the longitudinal magnetization is at its maximum value and oriented in the direction of the static magnetic field. After excitation the transverse magnetization decays toward zero with a characteristic time constant T2, and the longitudinal magnetization returns toward equilibrium with a characteristic time constant T1.*

Repetition time (TR): *the amount of time that exists between successive pulse sequences applied to the same slice.*

Resistance (Res): *the situation where the disease (here Acute Myeloid Leukemia) remains even after treatment.*

Sagittal plane: *a plane that cuts the skull from front to back and continues down in the body in the same direction, dividing it into two parts.*

Schwann cells: *These cells make myelin outside the brain that surrounds and insulates axons in cranial nerves and in the peripheral nerves that connect the CNS to the rest of the body.*

Schwannomas: *a type of non-neuroepithelial brain tumor.*

Secondary tumor: *the spread of cancer from one part of the body to another.*

Spin: *the property exhibited by atomic nuclei that contain either an odd number of protons or neutrons, or both.*

Stem cells: *cells from which all blood cells develop.*

Stereotactic biopsy: *a surgical procedure guided by scans and a frame.*

Synapse: *is a structure that permits a neuron to pass an electrical or chemical signal to another neuron.*

Temporal lobe: *is one of the four lobes of the cerebral hemisphere of the cerebral hemisphere. It controls auditory and visual memories, language, some hearing and speech, language, plus some behaviour.*

Transcription: *is the process of creating a complementary RNA copy of a sequence of DNA.*

Transcriptome: *is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA produced in one or a population of cells.*

Transcriptomics: *is the field of biology which studies the genetic transcriptions (RNA).*

Tumor: *any abnormal swelling in or on any part of the body. This term is usually used when the swelling is as a result of an overgrowth of cells.*

Tumor suppressor genes: *are the genes that protect a cell from one step on the path to cancer.*

Voxel: *a volume element (usually of the brain).*

Internet Sources

Wikipedia, http://www.wikipedia.org

Magnetic Resonance Technology Information Portal, http://www.mr-tip.com

Medical Dictionary, http://www.medterms.com

MRI Glossary, http://www.fonar.com/glossary.htm

The Free Dictionary, http://www.thefreedictionary.com

JOURNALS

1. I. Ardoino, F. Ambrogi, S. Iacobelli, P. Boracchi, G. Manikis, **M. G. Kounelakis**, M. Zervakis, P.J.G. Lisboa, P. Fazi, M. Vignetti, E.M. Biganzoli and A. Starita "Evaluation of short- and long-term response to treatment in GIMEMA protocol for Acute Myeloid Leukaemia", *International Journal of Biomedical Engineering & Technology, vol. 3, pp. 329-348, 2010* ***(Published)***

2. **M. G. Kounelakis**, M.E. Zervakis, G. C. Giakos, G. J. Postma, L. M. C. Buydens and X. Kotsiakis, "Embedding filtering criteria into a wrapper marker selection method for brain tumors classification: An application on metabolic peak area ratios", *IOP Measurement Science & Technology, vol. 22, doi: 10.1088/0957-0233/22/114019, (pp.12), 2011* ***(Published)***

3. I.N. Dimou, I. Tsougos, E. Tsolaki, E. Kousi, E. Kapsalaki, K. Theodorou, **M. G. Kounelakis** and M. Zervakis, "Brain lession classification using 3T MRS spectra and paired SVM kernels", *Elsevier Biomedical Signal Processing and Control, vol. 6, pp. 314-320, 2011* ***(Published)***

4. **M. G. Kounelakis**, I. N. Dimou, M. E. Zervakis, I. Tsougos, E. Tsolaki, E. Kousi, E. Kapsalaki and K. Theodorou, "Strengths and Weaknesses of 1.5T and 3T MRS Data in Brain Glioma Classification", *IEEE Transactions on Information Technology in BioMedicine, vol. 15, pp. 647-654, 2011* ***(Published)***

5. G. C. Giakos, S. Marotta, C. Narayan, J. Petermann, S. Shrestha, J. Baluch, D. Pingili, D. B. Sheffer, L. Zhang, M. Zervakis, G. Livanos and **M. G. Kounelakis**, "Polarimetric phenomenology of photons with lung cancer tissue", *IOP Measurement Science & Technology, vol. 22, doi:10.1088/0957-0233/22/11/114018, 2011* ***(Published)***

6. **M. G. Kounelakis**, M. E. Zervakis, G. C. Giakos, G. J. Postma, L. M. C. Buydens and X. Kotsiakis, "On the relevance of glycolysis process on brain gliomas", *IEEE Transactions on Information Technology in BioMedicine, 2011* ***(Under review)***

BOOK CHAPTERS

1. **M. G. Kounelakis**, M. Zervakis and X. Kotsiakis, "The impact of Microarray Technology in Brain Cancer", *Outcome Prediction in Cancer, A. F. G. Taktak & A. C. Fisher (eds): A Multi-Perspective View on Outcome Prediction in Cancer, Elsevier Publishing Company, pp. 339-388, 2006* ***(Published)***
2. G. C. Manikis, **M. G. Kounelakis** and M. E. Zervakis, "Pattern Recognition Methods for AML-Induction- Treatment Analysis", *in Computational Intelligence and Bioengineering: Essays in Memory of Antonina Starita, Frontiers in Artificial Intelligence and Applications, F. Masulli, A. Micheli & A. Sperduti (eds) – IOS Press, vol. 196, pp. 69-80, 2009* ***(Published)***

CONFERENCES

1. **M. G. Kounelakis**, M. Blazantonakis, M. Zervakis and X. Kotsiakis, "The Impact Of Bioinformatics In Clinical Practices On Human Cancer", *European Conference on Emergent Aspects in Clinical Data Analysis, Pisa, 2005* ***(Published)***
2. M. Blazadonakis, M. Zervakis, **M. G. Kounelakis**, E. Beganzoli and N. Lama, "Support Vector Machines and Neural Networks as Marker Selectors for Cancer Gene Analysis", *3rd IEEE International Conference on Intelligent Systems , London, 2006* ***(Published)***
3. **M. G. Kounelakis**, M. E. Zervakis, M. E. Blazadonakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis, "Feature Selection for Brain Tumour Classification using Ratios of Metabolites' Peak Areas from MRSI Data", *6th European Symposium on Biomedical Engineering, Chania, 2008* ***(Published)***
4. **M. G. Kounelakis**, M.E. Zervakis, M.E. Blazadonakis, G.J. Postma, L.M.C. Buydens, A. Heerschap and X. Kotsiakis, "Identification of significant Metabolic Markers from MRSI data for Brain Cancer Classification", *8th IEEE International Conference on BioInformatics and BioEngineering, Athens, 2008* ***(Published)***
5. G. C. Manikis, **M. G. Kounelakis** and M. E. Zervakis, "Pattern Recognition Methods for Acute Myeloid Leukemia (AML) Induction-Treatment Analysis", *International Workshop on Computational Intelligence and Bioengineering, Pisa, 2009* ***(Published)***
6. **M. G. Kounelakis**, M. E. Zervakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis, "Revealing the metabolic profile of brain tumors for diagnosis purposes", *31st IEEE Annual International Conference of the Engineering in Medicine and Biology Society, Minnesota, 2009* ***(Published)***

7. **M. G. Kounelakis**, M. E. Zervakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis, "Validation of MRS metabolic markers in the classification of brain gliomas and their correlation to energy metabolism", *12th Mediterranean Conference on Medical and Biological Engineering and Computing, Chalkidiki, 2010* ***(Published)***

8. **M. G. Kounelakis**, M. E. Zervakis, G. C. Giakos, C. Narayan, S. Marotta, D. Natarajamani, G. J. Postma, L. M. C. Buydens and X. Kotsiakis, "Targeting brain gliomas energy metabolism for classification purposes", *IEEE International Conference on Imaging Systems and Techniques, Thessaloniki, 2010* ***(Published)***

9. G. C. Giakos, S. Marotta, K. Ambadipudi, K. Valluru, J. Petermann, C. Narayan, D. Natarajamani, D. Pingili, **M. G. Kounelakis**, G. Livanos and M.E. Zervakis, "Backscattered polarimetric detection from biological tissue", *IEEE International Conference on Imaging Systems and Techniques, Thessaloniki, 2010* ***(Published)***

10. G. C. Giakos, S. Marotta, C. Narayan, J. Petermann, D. Pingili, W. Xu, S. Shrestha, A. Tsokaktsidis, M. Zervakis, G. Livanos and **M. G. Kounelakis**, "Near Infrared Light Interaction with Lung Cancer Cells", *IEEE International Instrumentation & Measurement Technology Conference, Binjiang-Hangzhou, 2011* ***(Published)***

11. **M. G. Kounelakis**, M. Zervakis, G. C. Giakos and X. Kotsiakis, "The role of the Glycolysis-related genes in Brain gliomas treatment design", *5th European Conference of IFMBE for Medical & Biomedical Engineering (EMBEC), Budapest, 2011* ***(Published)***