

RECONHECIMENTO DE PADRÕES DE ESPECTROS DE RESSONÂNCIA MAGNÉTICA DE TUMORES CEREBRAIS EM HUMANOS

B. H. Vieira*, A. C. Santos**, C. G. Carlotti** e C. E. G. Salmon*

*FFCLRP-USP, Ribeirão Preto, Brasil

**FMRP, Ribeirão Preto, Brasil

e-mail: bruno.hebling.vieira@usp.br

Resumo: A facilitação do uso da espectroscopia por ressonância magnética no diagnóstico de tumores cerebrais pode ser atingida através da utilização de algoritmos de reconhecimento de padrões, em especial no que diz a discriminação de tipos de tumores e a classificação de tumores de tipo desconhecido, usando a quantificação dos metabólitos obtida na espectroscopia como atributo na implementação dos algoritmos. Utilizando alguns dos algoritmos mais comumente encontrados na literatura, os resultados obtidos em diferentes abordagens ao problema são explorados. Os espectros (72 no total, um valor razoavelmente alto comparado com os encontrados na literatura) foram divididos em quatro classes, segundo o tipo de tumor diagnosticado através de biópsia: meningiomas, tumores gliais de baixo grau, tumores agressivos (tumores gliais de alto grau e metástases) e outros tipos de tumor. Foram realizados quatro estudos sobre a classificação entre todas as classes, usando a LDA após a PCA (com 63,9 por cento de acertos), todas as concentrações de metabólitos como atributos na LDA (56,9 por cento de acertos), concentrações de metabólitos utilizadas na literatura (58,3 por cento de acertos) e apenas o PCA (61,1 por cento de acertos), e mais dois estudos sobre a classificação entre as classes de tumores gliais de baixo grau e tumores agressivos, utilizando atributos da literatura (77,4 por cento de acertos) e também se extraindo os atributos através da LDA (77,4 por cento de acertos).

Palavras-chave: Padrões, espectroscopia, ressonância magnética, tumores cerebrais.

Abstract: *The facilitation of the use of magnetic resonance spectroscopy for the diagnosis of brain tumors may be achieved through the employment of pattern recognition algorithms, in particular in the discrimination of tumor types and classification of tumors of unknown type, by using the quantification of the metabolites obtained in spectroscopy as an attribute in the implementation of the algorithms. Using some of the algorithms most commonly found in the literature, the results obtained in different approaches to the problem are explored. The spectra (72 in total, a fairly high number compared to those found in the literature) were divided into four classes according to the type of tumor diagnosed through biopsy: meningiomas, low-grade glial tumors, aggressive tumors (high-grade glial tumors and metastases) and other types of tumors. Four*

studies on the classification of all classes were performed using the LDA after the PCA (63.9 percent correct), all metabolite concentrations as attributes in the LDA (56.9 percent correct), metabolite concentrations used in the literature (58.3 percent accuracy) and only the PCA (61.1 percent correct), and two studies on the classification between the low grade glial tumors and aggressive tumors classes, using attributes from the literature (77.4 percent correct) and also through LDA feature extraction (77.4 percent correct).

Keywords: *Patterns, spectroscopy, magnetic resonance, brain tumors.*

Introdução

Entre as várias ferramentas utilizadas no diagnóstico de tumores cerebrais se encontra a espectroscopia por ressonância magnética, considerada uma das principais técnicas para o estudo bioquímico de tecido vivo [1]. Apesar de ferramenta freqüente no meio acadêmico, o uso da espectroscopia por ressonância magnética não é tão difundida como o da imagem por ressonância magnética, por exemplo, no uso clínico, muitas vezes pela dificuldade na interpretação dos espectros e pela necessidade excessiva da intervenção humana, que requer conhecimento específico [2-3]. De maneira geral, processo pode ser auxiliado através do uso de sistemas de apoio à decisão clínica (CDSS). Tais sistemas podem se basear, por exemplo, em princípios estatísticos a fim de facilitar a interpretação de dados.

A disciplina de reconhecimentos de padrões se mostra apta à tarefa de discriminação e classificação de espectros de ressonância magnética. Aqui, alguns conceitos básicos são identificados: denomina-se observação o conjunto lógico ou numérico que define uma figura análoga a um vetor, correspondendo ao dado de entrada ou input, sendo atributo uma característica medida em todas as observações e ainda classe ou categoria como sendo um conjunto de observações tendo alguma característica, includentes ou excludentes, em comum, de modo que o conjunto de todas as observações preenche o espaço amostral, enquanto que suas dimensões são dadas pelos atributos.

Para a tarefa de discriminação, ou seja, a descrição de uma série de medidas visando maximizar a separação de diferentes categorias, dois algoritmos de ampla utilização são a *Principal Component Analysis* (PCA),

capaz de determinar as direções ortogonais no espaço amostral que mantém a maior variância das observações projetadas sobre elas [3], e a *Linear Discriminant Analysis* (LDA), análise que determina as direções ortogonais no espaço amostral que maximizam a variância entre categorias e minimizam a variância entre observações de mesma categoria, assumindo-se uma distribuição normal das classes [3]. Já na classificação, isto é, a tarefa de auferir regras para a atribuição de categorias pré-definidas para novas observações, o algoritmo mais robusto é o *K-nearest neighbors* (KNN), onde a categoria de uma nova observação é dada pelos K vizinhos mais próximos desta no espaço amostral, em esquema de votação [3]. Outro ponto importante é a validação dos resultados, ou seja, o estudo de como os modelos obtidos se adequam a dados independentes. Em geral, utilizam-se dados não incluídos em sua construção. Pode-se, no entanto, utilizar o esquema de validação cruzada, onde a precisão é testada usando-se partições do espaço amostral, auferindo-se modelos em uma partição e testando-os em outra. Uma redução desse método é o chamado *leave-one-out*, ou seja, os modelos são construídos sobre todo o espaço a menos de uma observação, que será classificada baseada neste [3]. O processo se repete até que todas as observações tenham sido submetidas, podendo-se então determinar sua precisão.

Pode-se assim, desenvolver uma metodologia para o reconhecimento de padrões de espectros de tumores usando-se as concentrações dos metabólitos obtidas na espectroscopia por ressonância magnética como atributos em algoritmos de reconhecimentos de padrões. Ainda, é possível se estimar a adequação de cada algoritmo à base de dados estudada e aos tipos de tumor incluídos. Dada a grande variabilidade apresentada na literatura sobre escolha dos atributos, se torna oportuno o estudo da influência da escolha dos atributos sobre os resultados da classificação de espectros.

Materiais e métodos

Para realização do estudo, foram necessárias as seguintes etapas:

Obtenção dos espectros– Os espectros foram obtidos retrospectivamente no servidor do Centro de Ciências da Imagem e Física Médica (CCIFM) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HC-FMRP) após autorização e consentimento do Prof. Dr. Antonio Carlos dos Santos. Conjuntamente foi adquirida a confirmação histopatológica, obtida através de biópsia.

Seleção– Foram incluídos apenas espectros obtidos sob a modalidade *single voxel* (SV), com tempo ao eco (TE) médio (144 ms), o que resultou na seleção de setenta e dois espectros de tumores. Todos os espectros foram obtidos em uma máquina de ressonância magnética sob intensidade de campo de 3.0 T. Através da confirmação histopatológica, os tumores foram separados em quatro classes, conforme utilizado no projeto INTERPRET [4], mostradas na Tabela 1.

Tabela 1: Conjunto de dados incluído no estudo.

Classificação	Tipo de tumor	Tamanho da amostra	Sigla
Meningioma	Meningioma	6	MEN
Tumor glial de baixo grau	Astrocitoma, oligoastrocitoma, oligodendroglioma	24	LGG
Tumor agressivo de alto grau	Glioblastoma, gliossarcoma, metástase Linfoma, PNET, hemangioblastoma, ganglioglioma,	29	AGG
Outro	meduloblastoma, germinoma, adenocarcinoma	13	OTH

Processamento– O processamento dos espectros foi feito no software LCMoel [5], incluindo o dimensionamento do conteúdo de água (*water-scaling*), obtendo-se 29 concentrações, incluindo metabólitos e combinações de metabólitos reais e virtuais utilizadas pelo software na quantificação dos espectros.

Análise de dados– A análise de dados foi realizada no software R 3.0.1, utilizando-se vários algoritmos implementados neste.

Algoritmos estudados– Foram estudados quatro métodos de classificação sobre todas as classes e dois métodos de classificação sobre as classes LGG e AGG.

Método 1– Foi realizada a LDA após a PCA sobre todas as classes (onde as componentes principais foram mantidas a partir do critério de Kaiser, onde componentes com variância inferior a um são descartadas).

Método 2– Foram utilizadas todas as concentrações de metabólitos como atributos na LDA sobre todas as classes.

Método 3– Foram utilizadas concentrações selecionadas através da literatura como atributos na LDA sobre todas as classes (especificamente, Alanina, Lactato, o conjunto Colina e Fosfocolina, o conjunto N-acetilaspártato e N-acetil-aspartil-glutamato (NAA e NAAG), o conjunto Creatina e Fosfocreatina e o conjunto dos lipídeos no entorno de 1.3 ppm).

Método 4– Também a classificação obtida usando-se o K-NN, sendo K determinado através de um algoritmo iterativo [6], após o PCA sobre todas as classes.

Método 5– Foi estudada a diferenciação entre as classes LGG e AGG utilizando os mesmos atributos do Método 3 através da literatura na LDA

Método 6– Também foram utilizados os atributos mais significativos (obtidos através do uso da LDA sobre todas as características normalizadas) para diferenciar as classes LGG e AGG.

Validação– Em todos os casos, o método de validação utilizado foi o método de validação cruzada *leave-one-out*.

Resultados

Os resultados para os métodos enumerados constam nas seguintes tabelas.

Tabela 2: Resultados referentes ao Método 1.

Tipo de tumor	MEN	LGG	AGG	OTH	% Acertos
Meningioma	4	1	0	1	66,7
Tumor glial de baixo grau	0	22	2	0	91,7
Tumor agressivo de alto grau	0	9	19	1	65,5
Outro	0	10	2	1	07,7

Tabela 3: Resultados referentes ao Método 2.

Tipo de tumor	MEN	LGG	AGG	OTH	% Acertos
Meningioma	3	0	1	2	50,0
Tumor glial de baixo grau	0	17	3	4	70,8
Tumor agressivo de alto grau	0	6	18	5	62,1
Outro	0	7	3	3	23,1

Tabela 4: Resultados referentes ao Método 3.

Tipo de tumor	MEN	LGG	AGG	OTH	% Acertos
Meningioma	1	2	1	2	16,7
Tumor glial de baixo grau	0	22	1	1	91,7
Tumor agressivo de alto grau	2	8	17	2	58,6

Outro	0	6	5	2	15,4
-------	---	---	---	---	------

Tabela 5: Resultados referentes ao Método 4.

Tipo de tumor	MEN	LGG	AGG	OTH	% Acertos
Meningioma	3	2	1	0	50,0
Tumor glial de baixo grau	0	21	3	0	87,5
Tumor agressivo de alto grau	0	8	20	1	69,0
Outro	0	6	7	0	00,0

Tabela 6: Resultados referentes ao Método 5.

Tipo de tumor	LGG	AGG	% Acertos
Tumor glial de baixo grau	21	3	87,5
Tumor agressivo de alto grau	9	20	69,0

Para o Método 6 foi preciso estimar quais metabólitos mais contribuem para a diferenciação entre tumores gliais de baixo grau e tumores agressivos, conforme detalhado na metodologia.

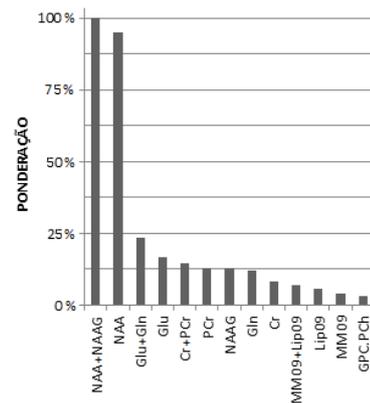


Figura 1: Ponderação relativa da concentração de alguns metabólitos para a diferenciação de tumores gliais de baixo grau e tumores agressivos.

A Figura 1 mostra a ponderação da concentração de diversos metabólitos (vários de menor ponderação foram suprimidos da figura). Descartando-se concentrações que são sabidas altamente correlacionadas (isso se dá entre as concentrações de

conjuntos e seus metabólitos constituintes), foram selecionadas as concentrações dos conjuntos N-Acetilaspártato e N-Acetilaspártilglutamato (NAA+NAAG), Glutamato e Glutamina (Glu+Gln) e Creatina e Fosfocreatina (Cr+PCr) como atributos para a classificação.

Tabela 7: Resultados referentes ao Método 6.

Tipo de tumor	LGG	AGG	% Acertos
Tumor glial de baixo grau	17	7	70,8
Tumor agressivo de alto grau	5	24	82,6

A Tabela 8 reúne os índices de acerto, indiferente quanto às classes, de todo o estudo.

Tabela 8: Índices de acerto sobre todos os métodos.

Método	% Acertos
Método 1	63,9
Método 2	56,9
Método 3	58,3
Método 4	61,1
Método 5	77,4
Método 6	77,4

Discussão

Talvez devido a sua grande presença no espaço amostral, a classe LGG atingiu bons índices de acerto em todos os métodos.

De maneira análoga, a classe MEN obteve resultados, de maneira geral, razoáveis. Contudo, esta é a menor classe, e a possibilidade de sobreajuste não pode ser descartada.

A classe OTH obteve os piores índices de acerto. Isso muito provavelmente se dá graças à heterogeneidade de seus componentes, incluindo diversos tipos de tumor sem distinção.

Incluindo-se apenas as classes LGG e AGG, o método da seleção de atributos pela LDA atingiu resultados praticamente idênticos ao método da seleção de atributos pela literatura quando se diferencia apenas as classes LGG e AGG, o que demonstra ainda a possibilidade de se usar a LDA como extrator de atributos. Segundo vemos na Figura 1, o conjunto N-acetilaspártato (NAA) e N-acetil-aspartil-glutamato (NAAG) obtêm alta ponderação na separação das classes LGG e AGG no LDA. Isso muito provavelmente se deve pelo fato de que, já que esses metabólitos são sintetizados pelos neurônios [3], a agressividade do tumor está correlacionada à perda de funcionalidade e a morte neuronal.

A seleção dos atributos através da literatura demonstrou-se mais eficaz quando o número de observações por classe é maior, ou seja, quando se diferencia apenas as classes LGG e AGG.

Conclusão

Observa-se que, quando se inclui todas as classes na diferenciação, os métodos A, B, C e D obtiveram boa diferenciação das classes LGG e AGG. A classe OTH obteve, invariavelmente, baixos índices médios de acertos, o que se dá pela heterogeneidade dessa classe. Já a classe MEN obteve resultados mistos, por ser uma amostra reduzida e pelo seu constante agrupamento ao redor da classe LGG.

A melhor diferenciação para a classe MEN se dá no Método 1 com 66,7% de acertos, para a classe LGG se dá nos Métodos 1 e 3, ambos com 91,7% de acertos, para a classe AGG no Método 4 com 69% de acertos e da classe OTH no Método 3 com 23,1% de acertos.

Apesar de possíveis erros ainda observa-se que os classificadores obtêm resultados muito superiores à suposição aleatória, que resultaria em um índice de acertos de 25% [3], e são análogos a resultados observados na literatura [2-3], em análises similares, porém não idênticas às apresentadas nesse projeto.

Agradecimentos

Agradecimentos à Pró-Reitoria de Pesquisa (PRP) da USP que possibilitou a bolsa-auxílio RUSP (Institucional).

Referências

- [1] El-Deredy, W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. *NMR in Biomedicine*, v. 10, n. 3, p. 99-124, 1997.
- [2] Tate, AR. Pattern recognition analysis of in vivo magnetic resonance spectra. 1996. Tese de Doutorado. University of Sussex.
- [3] Ladroue, CLC. et al. Pattern recognition techniques for the study of magnetic resonance spectra of brain tumours. 2004. Tese de Doutorado. University of London.
- [4] Tate, AR. et al. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. *Magnetic Resonance in Medicine*, v. 49, n. 1, p. 29-36, 2003.
- [5] Provencher, SW. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic Resonance in Medicine*, v. 30, n. 6, p. 672-679, 1993.
- [6] Kuhn, M. et al. caret: Classification and Regression Training. R package version 5.07-001, 2011. Disponível em: <<http://CRAN.R-project.org/package=caret>>, acesso em 31 jul.2014.