

Classification of Astrocytomas and Oligodendrogliomas from Mass Spectrometry Data Using Sparse Kernel Machines

Jacob Huang, Behnood Gholami, Nathalie Y. R. Agar, Isaiah Norton, Wassim M. Haddad,
and Allen R. Tannenbaum

Abstract—Glioma histologies are the primary factor in prognostic estimates and are used in determining the proper course of treatment. Furthermore, due to the sensitivity of cranial environments, real-time tumor-cell classification and boundary detection can aid in the precision and completeness of tumor resection. A recent improvement to mass spectrometry known as desorption electrospray ionization operates in an ambient environment without the application of a preparation compound. This allows for a real-time acquisition of mass spectra during surgeries and other live operations. In this paper, we present a framework using sparse kernel machines to determine a glioma sample's histopathological subtype by analyzing its chemical composition acquired by desorption electrospray ionization mass spectrometry.

I. INTRODUCTION

Brain cancer is the second leading cause of death in children and young adults [1]. As of 2007, 126,000 cases of brain tumor have been identified in the U.S. [2]. The most common type of malignant brain cancers are gliomas, which encompass about 80% of the 88,000 incidents in the U.S. [3]. Patients with glioblastoma multiforme (GBM), the most common glioma form, have a median survival time of only 10.6 months [4].

Part of a glioma's diagnosis is its histopathological type, or subtype, qualitatively based on the type of cell from which it originated or most closely resembles [4]. The World Health Organization (WHO) releases the International Classification of Diseases for Oncology (ICD-O), which classifies gliomas into three subtypes, namely, ependymomas, astrocytomas, and oligodendrogliomas, each originating from ependymal cells, astrocytes, and oligodendrocytes, respectively [5]. The National Cancer Institute breaks down the current cases of gliomas into 76% astrocytomas, 6.5% oligodendrogliomas, and 5.9% ependymomas [6]. Tumors are further categorized by a histopathological grade, based on how differentiated the

tumor cells are from the original cells [4]. The American Joint Commission on Cancer grades most tumors on a scale from G1 through G4, where G1 tumors are composed of cells similar to the original tissue, and G4 cells are the most differentiated [4]. These histopathological categories are the primary factors used in determining the prognosis of the tumor and planning treatment [4], [7].

Due to the sensitive nature of the cranial environment, the identification of cancerous tissue is of utmost importance [8]. Histopathological classification is currently performed by sending a resected sample for evaluation by an expert [9]. This process is inefficient during a neural surgery where accurate tumor classification and boundary detection are crucial.

Numerous techniques exist for identifying and analyzing cancerous cells. A well-established technique known as mass spectrometry (MS) is often used for analysis of complex proteins and lipids [10], [11]. MS is appropriate for this application due to brain matter being primarily composed of lipids and proteins [12]. It has been shown that certain disorders and conditions cause a measurable change in the composition of these molecules [11]. Specifically, gliomas have been shown to exhibit differences in lipid composition compared to healthy glial tissue [12], [13]. MS is a powerful tool to measure this change, where it is capable of identifying the molecular composition of a given sample. Specifically, the sample molecules are ionized into gas and identified using a property of the ions known as the *mass-to-charge ratio* (m/z) [10]. MS is specifically advantageous in a clinical setting due to multiple reasons. First, it has a high sensitivity and provides a large amount of chemical information. In addition, the administration of contrast agents can be avoided [8].

Traditional MS methods, including matrix-assisted laser desorption/ionization (MALDI) [14], [15], require sample preparation, which introduces a significant delay between the sample resection and the mass spectrum acquisition for clinical applications. A recent improvement to MS known as desorption electrospray ionization mass spectrometry (DESI-MS) involves ionization of molecules in an ambient environment without the application of a preparation compound [16]. This allows for a real-time acquisition of mass spectra during surgeries and other live operations.

Prior work has shown that mass spectrometry can be successfully used as a technique for cancer detection [17]. Furthermore, mass data from DESI-MS has been used to successfully diagnose bladder carcinomas [11]. Glioma subtype classification has been explored using gene expression data [18]. In [9], the authors present a preliminary study for classification of glioma subtypes using mass spectrometry.

In this paper, we aim to bridge the advantage of *in vivo* DESI-MS and real-time classification of glioma subtypes.

This research was supported in part by a grant from NIH (NAC P41 RR-13218) through Brigham and Women's Hospital, and by grants from the Air Force Office of Scientific Research, Army Research Office, the National Science Foundation, the Brain Science Foundation, Daniel E. Ponton Fund for the Neurosciences, and the NIH Director's New Innovator Award DP2 OD007383. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149.

J. Huang, B. Gholami, and A. R. Tannenbaum are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 jhuang@gatech.edu, behnood@gatech.edu, tannenba@ece.gatech.edu

N. Y. R. Agar and I. Norton are with the Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115 nagar@rics.bwh.harvard.edu, inorton@partners.org

W. M. Haddad is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 wm.haddad@aerospace.gatech.edu

Specifically, we show that machine learning techniques can be used to provide neurosurgeons with real-time information, which can be of critical importance for precise tumor resection. There exists a plethora of techniques for data classification. However, sparse kernel machines were chosen for their superior performance in dealing with high-dimensional data [19]. Here, we consider two sparse kernel methods for glioma subtype classification, namely, support vector machines and relevance vector machines. We show that sparse kernel machines can be used for glioma subtype classification with a high accuracy.

II. SPARSE KERNEL MACHINES

In this section, we discuss two sparse kernel-based algorithms, namely, the support vector machine and the relevance vector machine. For a more comprehensive discussion on these methods, see [20].

A. Support Vector Machine Algorithm

The support vector machine (SVM) algorithm [20] is a sparse kernel algorithm used in classification and regression problems. Here we will briefly discuss the SVM framework for the two-class classification problem. Let the *training set* be given by x_1, x_2, \dots, x_N , with *target values* given by z_1, z_2, \dots, z_N , respectively, where $x_n \in \mathbb{R}^D$ and $z_n \in \{-1, 1\}$, $n = 1, 2, \dots, N$. Moreover, assume that this training set is linearly separable in a feature space \mathbb{R}^M defined by the transformation $\phi: \mathbb{R}^D \rightarrow \mathbb{R}^M$; that is, there exists a linear decision boundary in the feature space separating the two classes.

To classify a new data point $x \in \mathbb{R}^D$ by predicting its target value z define $y(x) \triangleq w^T \phi(x) + b$, where $w \in \mathbb{R}^M$ is a weight vector and $b \in \mathbb{R}$ is a bias parameter. This representation can be rewritten in terms of a kernel function as $y(x) = \sum_{n=1}^N a_n z_n k(x, x_n) + b$, where a_n , $n = 1, 2, \dots, N$, and b are parameters determined by the training set x_n and z_n , $n = 1, 2, \dots, N$, and $k(\cdot, \cdot)$ is the kernel function. The sign of the function $y(x)$ determines the class of x . More specifically, for a new data point x , the target value is given by $z = \text{sgn}(y(x))$, where $\text{sgn } y \triangleq \frac{y}{|y|}$, $y \neq 0$, and $\text{sgn}(0) \triangleq 0$. In the SVM approach the parameters w and b are chosen such that the *margin*, that is, the minimum distance between the decision boundary and the data points, is maximized. Hence, only a subset of the training data (i.e., support vectors) is used to determine the decision boundary. It can be shown that the solution to the SVM problem results in a convex optimization problem [20], and hence, a global optimum is guaranteed.

In the case where there is an overlap between the two data classes, the SVM algorithm can be modified by allowing misclassification of the data points. In this case, the margin is maximized while penalizing misclassified points. Such a trade-off is controlled by a positive complexity parameter C , which is determined using a hold-out method such as cross-validation [20].

B. Relevance Vector Machine Algorithm

The relevance vector machine (RVM) algorithm [21] is a Bayesian sparse kernel algorithm, which can be regarded as the Bayesian extension of the SVM algorithm.

Next, we briefly review the method for the classification problem involving two data classes, namely \mathcal{C}_1 and \mathcal{C}_2 . Let

the training set be given by x_1, x_2, \dots, x_N , with target values given by z_1, z_2, \dots, z_N , where $x_n \in \mathbb{R}^D$ and $z_n \in \{0, 1\}$, $n = 1, 2, \dots, N$, $x_n \in \mathcal{C}_1$ if $z_n = 1$, and $x_n \in \mathcal{C}_2$ if $z_n = 0$. For a new data point $x \in \mathbb{R}^D$, we predict the associated class membership posterior probability distribution, namely, $p(\mathcal{C}_k|x)$, $k = 1, 2$, where $p(\mathcal{C}_k|x)$ is the conditional probability of the data class \mathcal{C}_k given the data point x . The class membership posterior probability for a given data point x is given by

$$p(\mathcal{C}_k|x) = \sigma(w^T \phi(x)), \quad (1)$$

where $\phi: \mathbb{R}^D \rightarrow \mathbb{R}^M$ is a fixed feature-space transformation, with components $\phi(x) = [\phi_1(x), \phi_2(x), \dots, \phi_M(x)]^T \in \mathbb{R}^M$, $w = [w_1, w_2, \dots, w_M]^T \in \mathbb{R}^M$ is the weight vector, and $\sigma(\cdot)$ is the logistic sigmoidal function defined by $\sigma(s) = \frac{1}{1+e^{-s}}$. Note that the RVM algorithm is a special case of the above model. Specifically, in the RVM algorithm $w^T \phi(x)$ in (1) has a special form (similar to the SVM algorithm) given by $\sum_{n=1}^N w_n k(x, x_n) + b$, where $k(\cdot, \cdot)$ is the kernel function. Hence, the class membership posterior probability for a given data point x is given by

$$p(\mathcal{C}_k|x) = \sigma\left(\sum_{n=1}^N w_n k(x, x_n) + b\right). \quad (2)$$

In the sequel, we consider the general formulation (1). Each weight parameter w_i , $i = 1, \dots, M$, in (1) is assumed to have a zero-mean Gaussian distribution, and hence, the weight prior distribution is given by

$$p(w|\alpha) = \prod_{i=1}^M \mathcal{N}(w_i|0, \alpha_i^{-1}), \quad (3)$$

where α_i , $i = 1, 2, \dots, M$, is the precision corresponding to the weight component w_i , $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]^T \in \mathbb{R}^M$, and $\mathcal{N}(x|\mu, \sigma^2)$ represents the normal distribution with mean μ and variance σ^2 . The parameters α_i , $i = 1, 2, \dots, M$, in the prior distribution (3) are called the *hyperparameters*.

The hyperparameters α_i , $i = 1, 2, \dots, M$, can be determined by maximizing the marginal likelihood distribution $p(w|z, \alpha)$, where $z = [z_1, z_2, \dots, z_N]^T \in \mathbb{R}^N$. As a result of the maximization of the marginal likelihood distribution, a number of the hyperparameters α_i approach infinity. Thus, the corresponding weight parameter w_i will be centered at zero, and hence, the corresponding component of the feature vector $\phi_i(x)$ plays no role in the prediction, resulting in a sparse predictive model. For further details of this approach, see [20], [21].

III. GLIOMA TYPE CLASSIFICATION

In this section, we use the sparse kernel machines described in Section II to classify the subtype of a glioma sample. The data were collected from research subjects at the Brigham and Women's Hospital, Boston, MA. In this study, 28 glioma samples were acquired from multiple research subjects, where the samples were either astrocytomas (A) or oligodendrogliomas (O). For the purposes of maximizing spatial resolution, the scanning pattern shown in Figure 1 was used to analyze each sample [22]. In order to account for the chemical variation within a sample, numerous spectra were extracted from each sample (one spectrum from each scanned row). The number of spectra varied depending on the amount of usable tissue in the sample. In our data set,

each sample contained between 7 to 45 spectra. A total of 19 astrocytoma samples and 9 oligodendroglioma samples were scanned, where the corresponding number of spectra was 426 and 205, respectively. To account for large-scale intensity differences between samples, each spectrum was normalized to zero mean and unit standard deviation.

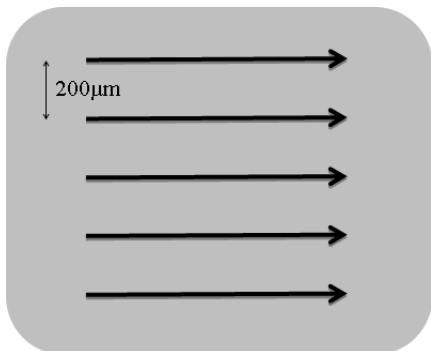


Fig. 1. Scan pattern

The mass spectrum is regarded as a mapping $MS : \mathcal{M} \rightarrow \mathbb{R}_+$, where $\mathcal{M} \triangleq \{m_1, \dots, m_D\}$ denotes the m/z values ranging from 150 to 1,000, with a resolution of 0.0833. Moreover, for a given m/z value $m_i, i \in \{1, \dots, D\}$, $MS(m_i)$ denotes the number of corresponding ions detected. These intensities collectively can be represented by a vector $x = [s_1, \dots, s_D]^T \in \mathbb{R}^D$, where $s_i = MS(m_i), i = 1, \dots, D$, and s_i is referred to as a *feature point*. For a given set of mass spectra $\{x_1, \dots, x_N\}, x_i \in \mathbb{R}^D, i = 1, \dots, N$, the target value $z_i \in \{0, 1\}, i = 1, \dots, N$, is provided by the pathologist, where $z_i = 0$ (resp., $z_i = 1$) indicates that x_i was from an astrocytoma (resp., oligodendroglioma) sample.

The SVM and RVM frameworks were chosen for the classification problem, both with a linear kernel $k(u, v) \triangleq u^T v, u, v \in \mathbb{R}^D$. We applied the leave-one-out method for validation [20]. Specifically, one spectrum x_j and its corresponding target value $z_j, j \in \{1, \dots, N\}$, are excluded, and the remaining spectra and their corresponding target values are used as the training set. A class, y_j , is assigned to x_j using the trained model. This process is repeated for each spectrum, and the outputs are compared to the target values z_j to determine the accuracy of the classifier.

The goal of cross-validation is to measure how well a classifier generalizes to a data set independent from the set used for training. As can be expected, the spectra extracted from the same sample exhibit a large degree of dependence. To compensate for this, spectra from the same sample as x_j are also removed from the training set. Another issue was the difference in the number of spectra from each sample as discussed earlier in this section. This was remedied by trimming the training set to use the same number of spectra from each sample. This trimming, however, was carried out randomly, introducing some variation into the performance of cross-validation. Thus, cross-validation was repeated 10 times, and the accuracies were averaged in the results of Table I. The MATLAB implementation of SVM was used with the penalty $C = 1$. The MATLAB RVM library [23] was used for RVM classification.

Table I shows the performance of the SVM and RVM for the problem of classifying mass spectra corresponding to astrocytoma and oligodendroglioma. Each row lists the sample number, the subtype determined by the pathologist (A for

TABLE I
CLASSIFICATION RESULTS PER SAMPLE

Sample	Subtype	SVM	RVM
1	A	100%	100%
2	A	0%	0%
3	A	0%	0%
4	A	100%	100%
5	A	100%	100%
6	A	100%	100%
7	A	100%	100%
8	A	100%	100%
9	A	100%	100%
10	A	76%	100%
11	A	100%	84%
12	A	100%	100%
13	A	100%	100%
14	A	100%	100%
15	A	97%	100%
16	A	100%	100%
17	A	100%	100%
18	A	75%	100%
19	A	0%	100%
Astrocytoma		81%	89%
20	O	70%	78%
21	O	92%	17%
22	O	100%	100%
23	O	88%	0%
24	O	38%	9%
25	O	0%	0%
26	O	100%	100%
27	O	100%	7%
28	O	100%	100%
Oligodendroglioma		77%	76%
Overall		79%	82%

astrocytoma and O for oligodendroglioma), the percentage of the spectra in this sample that were properly classified using SVM and RVM. Comprehensive classification accuracies are listed at the bottom of each table, representing the percentage of all astrocytoma and oligodendroglioma spectra that were properly classified as well as the overall classification accuracy. These accuracies may be analyzed to identify samples that are given incorrect subtypes by the pathologists, or to identify samples that are chemically dissimilar to other samples in the same subtype.

Finally, a feature selection technique was applied to the spectra. With thousands of points in the spectrum to describe two classes, redundancy in the information is expected [24]. Feature selection not only can improve the classification performance but also can aid in identifying the appropriate biomarkers for each glioma type. We implemented an *iterative method* of reducing the number of features [24]. The feature points were split into 10 subsets of size $P \in \mathbb{Z}_+$, $\{x_1, \dots, x_P\}, \{x_{P+1}, \dots, x_{2P}\}, \dots, \{x_{9P+1}, \dots, x_M\}$. The purpose was to identify the subset with the least effect on classification performance; these features were assumed to be either redundant or noisy, and were removed. The classification performance was measured by cross-validation, using an SVM classifier. This process was repeated iteratively, removing the features that contribute least to the classification performance in each iteration. A summary of these results are shown in Table II. As shown in Table II, the classification performance increases to 98% by selecting the most significant features.

TABLE II
ITERATIVE FEATURE SELECTION RESULTS

Iteration	# Features	Accuracy
0	10,200	79.4%
5	6,023	85.0%
10	3,556	85.4%
15	2,100	88.1%
20	1,240	91.9%
25	733	94.1%
30	433	98.3%
35	255	97.8%
40	151	96.6%
45	90	97.6%
50	54	97.1%
55	32	97.0%
60	19	98.3%
65	9	97.0%
69	2	82.3%

IV. CONCLUSION

In this paper, we proposed an approach for glioma subtype classification using mass spectrometry data. We considered two sparse kernel methods for glioma subtype classification, namely, SVM and RVM, and showed that sparse kernel machines can be used for glioma subtype classification with a high accuracy. Future work includes exploring alternative preprocessing, feature selection, and classification methods, extending this framework to the multi-class glioma classification problem as well as biomarker detection. Identifying the most significant molecules using a feature selection method can provide a deeper insight into the pathology of the condition and can improve treatment methods.

ACKNOWLEDGMENTS

J. Huang and B. Gholami acknowledge several fruitful discussions with Dr. Vandana Mohan.

REFERENCES

- [1] A. Jemal, R. Siegel, J. Xu, and E. Ward, "Cancer Statistics, 2010." *Cancer J. Clin.*, Jul. 2010, <http://caonline.amcancersoc.org/>.
- [2] S. F. Altekruse, C. L. Kosary, M. Krapcho, N. Neyman, R. Aminou, W. Waldron, J. Ruhl, N. Howlander, Z. Tatalovich, H. Cho, A. Mariotto, M. P. Eisner, D. R. Lewis, K. Cronin, H. S. Chen, E. J. Feuer, D. G. Stinchcomb, and B. K. Edwards, "SEER Cancer Statistics Review 1975-2007," Bethesda, MD, http://www.seer.cancer.gov/csr/1975_2007/.
- [3] American Cancer Society, "Brain and Spinal Cord Tumors in Adults," Atlanta, GA, 2009, <http://www.cancer.org/cancer/brainstumorsinadults/>.
- [4] S. B. Edge and D. R. Byrd, *AJCC Cancer Staging Manual*, 7th ed. New York, NY: Springer-Verlag, 2009.
- [5] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathologica*, vol. 114, no. 2, pp. 97-109, Aug. 2007.
- [6] Central Brain Tumor Registry of the United States, "Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2007," Hinsdale, IL, 2011, <http://www.cbtrus.org/reports/reports.html>.
- [7] S. Cha, "Update on brain tumor imaging: from anatomy to physiology," *Am. J. Neurorad.*, vol. 27, no. 3, pp. 475-87, Mar. 2006.
- [8] L. S. Eberlin, A. L. Dill, A. J. Golby, K. L. Ligon, J. M. Wiseman, R. G. Cooks, and N. Y. R. Agar, "Discrimination of human astrocytoma subtypes by lipid analysis using desorption electrospray ionization imaging mass spectrometry," *Angew. Chem. Int. Ed. Engl.*, vol. 49, no. 34, pp. 5953-6, Aug. 2010.

- [9] V. Mohan, N. Agar, F. Jolesz, and A. Tannenbaum, "Automatic Classification of Glioma Subtypes and Biomarker Identification Using DESI Mass Spectrometry Imaging," in *MICCAI 2010 Workshop Comp. Imag. Biomark. Tumors*, 2010.
- [10] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics." *Nature*, vol. 422, no. 6928, pp. 198-207, Mar. 2003.
- [11] A. L. Dill, D. R. Ifa, N. E. Manicke, A. B. Costa, J. A. Ramos-Vara, D. W. Knapp, and R. G. Cooks, "Lipid profiles of canine invasive transitional cell carcinoma of the urinary bladder and adjacent normal tissue by desorption electrospray ionization imaging mass spectrometry." *Anal. Chem.*, vol. 81, no. 21, pp. 8758-64, Nov. 2009.
- [12] M. Köhler, S. Machill, R. Salzer, and C. Krafft, "Characterization of lipid extracts from brain tissue and tumors using Raman spectroscopy and mass spectrometry." *Anal. Bioanal. Chem.*, vol. 393, no. 5, pp. 1513-20, Mar. 2009.
- [13] A. Beljebbar, S. Dukic, N. Amharref, S. Bellefqih, and M. Manfait, "Monitoring of biochemical changes through the c6 gliomas progression and invasion by fourier transform infrared (FTIR) imaging." *Anal. Chem.*, vol. 81, no. 22, pp. 9247-56, Nov. 2009.
- [14] F. Hillenkamp, M. Karas, R. C. Beavis, and B. T. Chait, "Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers," *Anal. Chem.*, vol. 63, no. 24, pp. 1193A-1203A, Dec. 1991.
- [15] N. Y. R. Agar, J. G. Malcolm, V. Mohan, H. W. Yang, M. D. Johnson, A. Tannenbaum, J. N. Agar, and P. M. Black, "Imaging of meningioma progression by matrix-assisted laser desorption ionization time-of-flight mass spectrometry." *Anal. Chem.*, vol. 82, no. 7, pp. 2621-5, Apr. 2010.
- [16] R. G. Cooks, Z. Ouyang, Z. Takats, and J. M. Wiseman, "Detection Technologies. Ambient mass spectrometry." *Science*, vol. 311, no. 5767, pp. 1566-70, Mar. 2006.
- [17] O. J. Semmes, Z. Feng, B.-L. Adam, L. L. Banez, W. L. Bigbee, D. Campos, L. H. Cazares, D. W. Chan, W. E. Grizzle, E. Izbicka, J. Kagan, G. Malik, D. McLerran, J. W. Moul, A. Partin, P. Prasanna, J. Rosenzweig, L. J. Sokoll, S. Srivastava, S. Srivastava, I. Thompson, M. J. Welsh, N. White, M. Winget, Y. Yasui, Z. Zhang, and L. Zhu, "Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility." *Clin. Chem.*, vol. 51, no. 1, pp. 102-12, Jan. 2005.
- [18] A. Li, J. Walling, S. Ahn, Y. Kotliarov, Q. Su, M. Quezado, J. C. Oberholtzer, J. Park, J. C. Zenklusen, and H. A. Fine, "Unsupervised analysis of transcriptomic profiles reveals six glioma subtypes." *Cancer Res.*, vol. 69, no. 5, pp. 2091-9, Mar. 2009.
- [19] M. Wagner, D. Naik, and A. Pothén, "Protocols for disease classification from mass spectrometry data." *Proteomics*, vol. 3, no. 9, pp. 1692-8, Sep. 2003.
- [20] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer-Verlag, 2006.
- [21] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, no. 3, pp. 211-244, Aug. 2001.
- [22] Z. Takáts, J. M. Wiseman, and R. G. Cooks, "Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology." *J. Mass. Spec.*, vol. 40, no. 10, pp. 1261-75, Oct. 2005.
- [23] M. Tipping and A. Faul, "Fast Marginal Likelihood Maximisation for Sparse Bayesian Models," in *Proc. Ninth Int'l Workshop Artif. Intel. Stat.*, Key West, FL, 2003, <http://research.microsoft.com/en-us/um/cambridge/events/aistats2003/proceedings/papers.htm>.
- [24] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview." *NeuroImage*, vol. 45, no. 1 Suppl, pp. S199-209, Mar. 2009.