# Classification Algorithms for SIFT-MS Medical Diagnosis

K. Moorhead, D. Lee, J Chase, A. Moot, K. Ledingham, R. Allardyce, S. Senthilmohan, and Z. Endre

*Abstract*— **Selected Ion Flow Tube - Mass Spectrometry (SIFT-MS) is an analytical technique for the real-time quantification of trace gases in air or breath samples. The SIFT-MS system can potentially offer unique capability in the early and rapid detection of a wide variety of diseases, infectious bacteria and patient conditions, by using a classifier to differentiate between control and test groups. By identifying which masses and Volatile Organic Compounds (VOCs) contribute most strongly towards a successful classification, biomarkers for a particular disease state may be discovered.**

**A classification method is presented and validated in a simple study in which saturated nitrogen in tedlar bags was differentiated from dry nitrogen in tedlar bags. Several biomarkers were identified, with the most reliable being $N_2H^+.H_2O$, and isotopes and water clusters of $H_3O^+$, as expected. The classifier was then applied in a clinical setting to differentiate between patient breath samples after one and four hours of dialysis treatment. Biomarkers for classification were ammonia, acetaldehyde, ethanol, isoprene and acetone. The model classifies significantly better than random, with an ROC area of 0.89.**

## I. Introduction

SIFT-MS is a relatively new quantitative mass spectrometric technique for the real-time quantification of VOCs [1, 2]. It exploits the chemical ionization of positively charged precursor ions that react with the VOCs in an air or breath sample. $H_3O^+$, $NO^+$ and $O_2^+$ precursor ions are typically used, since they do not react with the main compounds ($N_2$, $O_2$, $CO_2$ and Ar) found in air or breath. The process steps are summarized below.

1. Precursor ions are generated by passing water through a microwave discharge
2. A quadrupole mass filter is used to select the required precursor ion based on mass/charge ratio
3. The precursor ion is injected into a fast-flowing inert carrier gas (helium), which carries the precursor ion, and test sample, along the flow tube
4. The precursor ion reacts with the VOCs from the sample to form product ions
5. A representative proportion of the product ions then pass into a differentially pumped quadrupole mass spectrometer that filters ions according to mass
6. The selected product ions pass to the channeltron particle multiplier/detector where they are counted

Two cases are presented. The first is a simple validation study to differentiate 'dry' nitrogen from 'wet' nitrogen. The second case study uses the classification model in a clinical setting to determine the differences between dialysis patients after 1 and 4 hours of dialysis treatment. The classification model is also able to determine which masses are most useful in this classification and therefore those compounds that act as biomarkers for kidney function.

## II. Methodology

### A. Experimental Design

A simple test study was conducted to validate the statistical classification model developed. Gas collected remotely from the SIFT-MS device is collected in Tedlar bags composed of polyvinyl fluoride, which the manufacturers claim to be chemically inert [3]. Before using the tedlar bags, the manufacturers recommend flushing the bags with purified air or nitrogen. Others have shown using Solid Phase Micro-extraction (SPME) that the Tedlar bags emit 15 different VOCs into samples stored in the bags for 24 hours. Phenol, acetone and acetic acid are persistently present even after purging the bags with purified nitrogen [4]. Hence, one outcome of this initial study is determining the ability to detect and account for these effects in this common form of breath sample collection.

In this study, multiple samples of $N_2$ in Tedlar bags were tested by performing mass scans over a range of 10-150 amu using the precursors $H_3O^+$, $NO^+$ and $O_2^+$. The bags were new and all flushed at least three times prior to testing directly from the bag.

$N_2$ in Tedlar bags was also vented to sterile glass bottles filled with water using a sterile stainless steel needle, and a polytetrafluoroethylene (PTFE) permeable septum. The gas in the bottles was then tested by performing mass scans. 25 mass scans were obtained using each method. The specific objectives of this first study are as follows:

1. Determine which VOCs are added when venting to a Tedlar bag containing purified nitrogen
2. Determine which VOCs, or their water clusters, increase or diminish when passed through sterile glass bottles filled with water
3. Validate the classifier using this 'clean' system

Manuscript received March 20, 2007. This research was conducted on a commercially available SIFT-MS instrument provided by Syft Technologies Ltd., and was supported by a New Zealand Tertiary Education Commission Top Achiever Doctoral Scholarship (UOCX5011). This study was approved was by the Regional Health and Disability Ethics Committee (URB/05/12/178).

K. T. Moorhead is with the Department of Mechanical Engineering at the University of Canterbury, Private Bag 4800, Christchurch, New Zealand. phone: +64 3 3642987; fax: +64 3 3642078; e-mail: ktm19@student.canterbury.ac.nz).

In the second part of the study, 1 patient with impaired kidney function underwent dialysis treatment on 7 separate occasions. Breath mass scans were performed 1 hour into the treatment and after 4 hours. These mass scans were taken over a range of 10-150 amu using the precursors $H_3O^+$, $NO^+$ and $O_2^+$. The aims of this study are:

1. To develop probability density profiles for the pre- and post-dialysis groups
2. To determine if it is possible to reliably differentiate between the 2 sample groups
3. To identify possible new biomarkers for kidney function in this type of cohort

The study is based on the fact that dialysis takes a patient from kidney failure to functional status (artificially). Hence, it provides a comparison for analyzing kidney function.

### *B. Classification*

Kernel density estimates are used for the classification of mass scan data. Test datasets of known classification are used to develop probability density profiles for each of two datasets: Groups *j* and *k*. An unknown sample is then tested against the datasets, with the result being a classification into either Group *j* or Group *k*. In a diagnostic analysis, *j* and *k* are the non-disease and diseased states.

For each mass, a mixed distribution made up of a kernel density and a dirac delta function is used to develop a density profile from each group using each sample's concentration value at that mass. Stronger, generally smoother density profiles are obtained with greater numbers of different mass scan samples.

When a mass scan from an unknown sample is obtained, at each mass, the probability densities for groups *j* and *k* at the concentration of the unknown sample are compared. Let $x_0$ denote the concentration of the unknown sample at the given mass, and $\hat{f}_j(x_0)$ and $\hat{f}_k(x_0)$ denote the probability densities of groups *j* and *k* respectively at the concentration of the unknown sample. This situation is illustrated in Fig. 1, where the plot shown would be useful for classification given the minimal overlap of distributions.

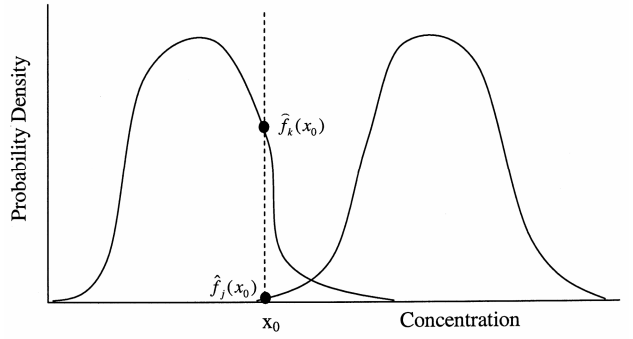


**Fig. 1. Probability densities for a given mass**

For each mass, (1) gives the probability of the given sample being from Group *j*, given the concentration value obtained at that mass is $x_0$.

$$\hat{\Pr}(j \mid x_0) = \frac{\hat{\pi}_j \hat{f}_j(x_0)}{\hat{\pi}_k \hat{f}_k(x_0) + \hat{\pi}_j \hat{f}_j(x_0)} \tag{1}$$

where $\hat{\pi}$ is the prior probability (prior) of the sample being in that group. When no information is known about the data, the priors are set to 0.5. If the ratio in (1) is greater than a specified threshold, *q*, the sample is classified in Group *j*, otherwise it is classified in Group *k*.

The log-odds ratio is defined as the natural log of the probability of a group *j* classification divided a group *k* classification.

$$LogoddsRatio = \ln\left(\frac{\hat{\Pr}(j \mid x_0)}{\hat{\Pr}(k \mid x_0)}\right) \tag{2}$$

Equation (2) can be broken down such that the log-odds ratios are effectively summed over all masses to obtain a final log-odds ratio over all masses:

$$\ln\frac{\hat{\Pr}(j \mid x_0)}{\hat{\Pr}(k \mid x_0)} = \ln\frac{\hat{\pi}_j}{\hat{\pi}_k} + \sum_{m=1}^{M} \ln\frac{\hat{f}_j(x_{0\cdot}, m)}{\hat{f}_k(x_{0\cdot}, m)} \tag{3}$$

A final log-odds ratio greater than $\ln[q/(1\text{-}q)]$ indicates the sample is in the numerator group (Group *j*), otherwise, it is in the denominator group (Group *k*).

Once density profiles have been created for the two datasets, the stratified bootstrap method is used to estimate the prediction error of the classification model. Bootstrap samples are created by choosing with replacement from the original sample until a bootstrap sample is created that is the same size as the original sample [5]. This process is repeated *B* times, producing *B* bootstrap datasets, where *B* is sufficiently large to ensure that all patients are left out of at least 1 bootstrap dataset. The bootstrap estimate of the classification error is defined:

$$\hat{E}^{(1)} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|C^{-i}|}\sum_{b \in C^{-i}} I(\hat{y}^b(x_i) \neq y_i) \tag{4}$$

where $N$ is the total number of samples; $|C^{-i}|$ is the number of bootstrap samples that do not contain sample *i;* $\hat{y}^b$ is the classifier trained on bootstrap sample *b*; $I(\hat{y}^b(x_i) \neq y_i)$ equals 1 if sample *i* is classified incorrectly and 0 otherwise; and $y_i$ is the group that $x_i$ belongs to.

The kernel classifier is trained with the bootstrap datasets that do not contain sample *i* as the training set, and then use sample *i* as the test set. This process is repeated for all bootstrap samples that do not contain a given sample *i*. The total number of incorrect classifications is summed and divided by the number of bootstrap samples that did not contain sample *i*. This overall process is repeated for each sample. It concludes by averaging the number of incorrect classifications over all samples, giving $\hat{E}^{(1)}$.

The bootstrap estimate is biased upward as an estimate of the true classification error, and this limitation is alleviated using the 0.632 estimator [5], defined:

$$\hat{E}^{(0.632)} = 0.368 \times \bar{e} + 0.632 \times \hat{E}^{(1)} \quad (5)$$

where $\bar{e}$ is the biased error, calculated using all the data as the training set, and testing each sample against this set. This result is biased downward of the true error because the test data is also found in the training set. The overall prediction error should be compared with random classification, for which the overall error rate, $P^*$, is given by:

$$P^* = p_j(1-\pi_j) + (1-p_j)\pi_j \quad (6)$$

With a prior probability set at 0.5, the overall error rate for a random classifier is 50%, regardless of group proportions ($p_j$ and $p_k$) in the sample.

Density profiles of the log-odds ratio obtained from the bootstrap method can be plotted for each group, and a reliability curve can be generated, as shown in (7) and Fig. 2.

$$\hat{\Pr}(S=j \mid \lambda) = \frac{\hat{f}(\lambda \mid S=j)\hat{\Pr}(S=j)}{\hat{f}(\lambda \mid S=j)\hat{\Pr}(S=j) + \hat{f}(\lambda \mid S=k)\hat{\Pr}(S=k)} \quad (7)$$

where $\hat{\Pr}(S=j \mid \lambda)$ is the probability of the sample, $S$, being in Group $j$ given the log-odds ratio, $\lambda$; $\hat{f}(\lambda \mid S=j)$ is the density of the log-odds ratio for Group $j$; $\hat{\Pr}(S=j)$ is the probability that the sample is in group $j$. This situation is illustrated in Fig. 2. A good dataset will have minimal overlap between the two density profiles because where a significant difference is detected between sample groups, a large log-odds ratio should always be obtained. However, the log-odds ratio obtained where group $j$ and $k$ profiles overlap, has a reliability of 50%, indicating an equally likely probability of a correct or incorrect classification.

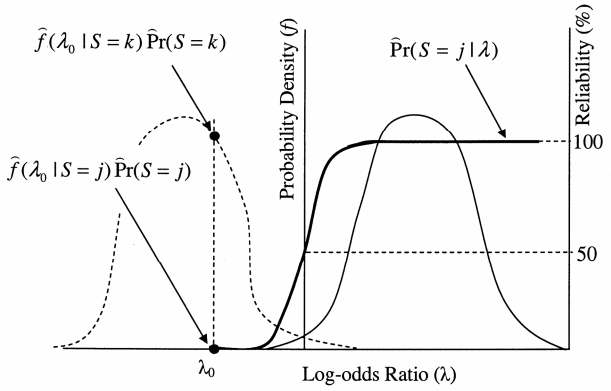


**Fig. 2. Reliability Curve**

The null hypothesis is that the unknown sample is in Group $j$. An ROC curve can be obtained by plotting sensitivity against non-specificity, where non-specificity is the probability of incorrectly rejecting the null hypothesis (classifying the sample as Group $k$ when in fact it is Group $j$), and sensitivity is the probability of correctly rejecting the null hypothesis (classifying the sample as Group $k$ when in fact it is Group $k$). An area under the ROC curve of 1 indicates perfect classification, and an area of 0.5 indicates a purely random classification.

Biomarkers are found by determining which masses have log-odds density profiles with minimal overlap. Density profiles are created as described previously for each mass, *individually*. The area of overlap is shaded in the schematic of Fig. 3 and is calculated using the Trapezoid Rule.

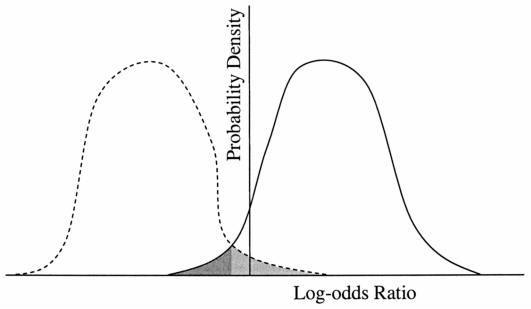


**Fig. 3. Biomarker Determination**

## III. Results and Discussion

### A. Validation Study - Nitrogen in Tedlar Bags

Using the 0.632 bootstrap estimator method to estimate prediction error, with $B$ = 500 bootstrap samples, there was 0% classification error over all precursors, indicating excellent differentiation between sample groups. A density profile is shown in Fig. 4 for the best biomarker. It can be seen that there is excellent separation between density profiles for the two groups. Squares indicate the concentrations of the raw data, and the curves are the probability density profiles fitted to that raw data.

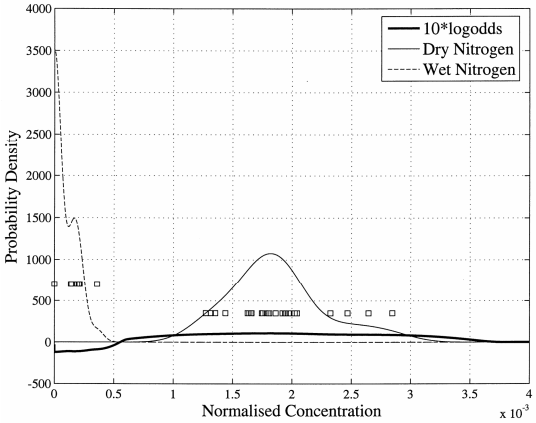


**Fig. 4. Probability density profile of mass 47 using $H_3O^+$ precursor**

The classification model found that for the $H_3O^+$ precursor, the masses indicated in Table I were most useful in aiding towards a correct classification, as ranked.

**Table I. Biomarkers for $H_3O^+$ classification in validation study**

| Product ions | Mass | Rank |
|---|---|---|
| $H_3O^+$ and its water clusters | 19, 55, 73 | 3, 4, 5 |
| Isotope of $H_3O^+$ & its $H_2O$ clusters | 57 | 2 |
| $C_4H_9NO.H^+$ | 88 | 7 |
| $C_6H_6O.H^+$ (product of phenol) | 95 | 6 |
| $N_2H^+.H_2O$ | 47 | 1 |

Density profiles were created from the final log-odds values obtained from each bootstrap sample entry for each of the test groups. It is observed in Fig. 5 that there is excellent separation between the two profiles, with consistently large log-odds values obtained. The area under the ROC curve was found to be 0.9998. These ideal results are to be expected for this validation study. Lesser values would indicate an invalid classifier.

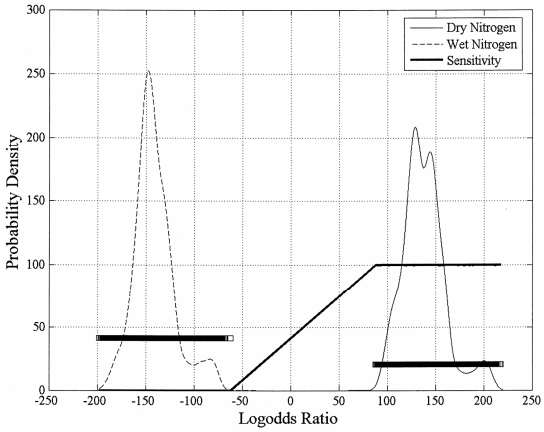


**Fig. 5. Validation study classification and reliability curves**

*B. Case Study - Dialysis*

7 repeat mass scans were taken of 1 dialysis patient over the course of 6 months. These mass scans were taken at time $t$ = 1 hour and $t$ = 4 hours into treatment. Mass scans were analyzed for key biomarkers that aid in a classification between each data set. A sample size of 7 was much smaller than desired, but sufficient to test the classifier concept for identifying known, different states.

With $B$ = 1000, the overall prediction error was estimated at 11.2%, 26.0%, and 18.2% using the $H_3O^+$, $NO^+$, and $O_2^+$ precursor, respectively. A density profile for the best biomarker, mass 36 corresponding to ammonia, is shown in Fig. 6. It can be seen that the density profile is relatively strong for the 4-hour dialysis group (Group *j*), because after dialysis treatment, levels of ammonia and other VOCs fall to normal levels, as expected. Depending on factors such as diet and the length of time since the last dialysis treatment, VOC levels can vary dramatically prior to treatment, as seen by the spread of data in the 1-hour dialysis group (Group *k*).

The classification model found that the masses indicated in Table II were most useful in aiding towards a correct classification, as ranked. These masses correspond primarily to ammonia and are expected results in dialysis, further validating the classifier.

**Table II. Biomarkers for $H_3O^+$ classification in dialysis case study**

| Precursor | Mass | Rank |
|---|---|---|
| $H_3O^+$ | 18, 36, 54 | 2,1,5 |
| | 17, 35 | 4, 3 |
| | 89, 47 | 6, 7 |
| $NO^+$ | 18, 36 | 2,3 |
| | 47 | 1 |
| $O_2^+$ | 17, 18, 35, 36, 53, 54 | 4,3,6,5,1,2 |
| | 77, 58 | 7,8 |

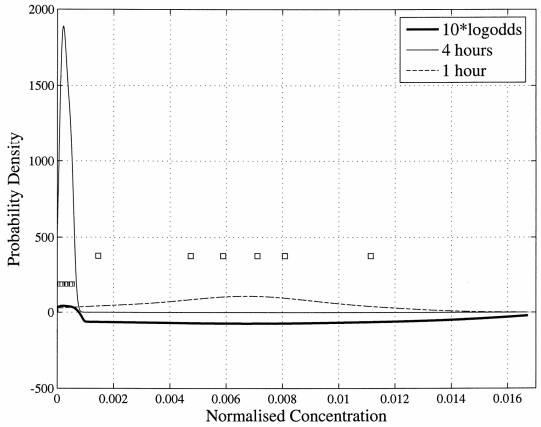


**Fig. 6. Probability density profile of mass 36 using $H_3O^+$ precursor**

Density profiles were created from the log-odds values obtained from each bootstrap sample entry for each of the test groups. Fig. 7 shows that there is much greater overlap in the two density profiles compared with the Nitrogen validation study, and an absolute log-odds value of approximately 50 must be obtained to classify with 90% certainty in Fig. 7. The model classifies significantly better than random, with an ROC area of 0.89. Hence, the model is further validated in an accepted clinical case with clinically noisy data.

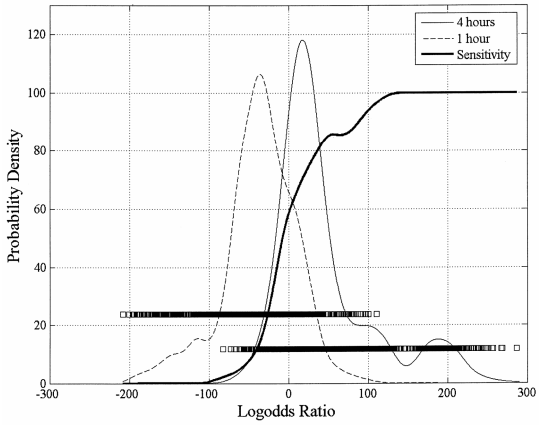


**Fig. 7. Dialysis case study classification and reliability curves**

## REFERENCES


[1] D. Smith, and P. Spanel "Application of ion chemistry and the SIFT technique to the quantitative analysis of trace gases in air and on breath". International Reviews in Physical Chemistry, 1996, 15(1), 231-271.

[2] C. G. Freeman, and M. J. McEwan. "Rapid Analysis of Trace Gases in Complex Mixtures Using Selected Ion Flow Tube - Mass Spectrometry". Australian Journal of Chemistry, 2002, 55, 491-494.

[3] SKC inc. Gas Sample Bags. http://www.skcinc.com/prod/tedlar.asp#tedlar

[4] D. B. Parker, M. B. Rhoades, J. Koziel, and J. Spinhirne. "Background odours in Tedlar Bags used for CAFO odour sampling". Conference proceedings, ASAE Annual International Meeting, Las Vegas, July 2003. Paper No. 034144.

[5] T. Hastie, R. Tibshirani, and J. H. Friedman. "Elements of Statistical Leaning: data mining, inference and prediction". New York: Springer, 2001.