Feature-based detection using Bayesian data fusion

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This article was published in the journal, International Journal of Image and Data Fusion [© Taylor and Francis] and the definitive version is available at: http://dx.doi.org/10.1080/19479832.2013.824029

Metadata Record: https://dspace.lboro.ac.uk/2134/14605

Version: Accepted for publication

Publisher: © Taylor and Francis

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
FEATURE BASED DETECTION USING BAYESIAN DATA FUSION

Ayodeji Akiwowo; Mahroo Eftekhari

ABSTRACT

Current cocaine detection techniques used at borders have their challenges which include cost of training specialised operators, the susceptibility of the operators to errors and the dangers involved in exposure of both operators and container contents to radioactive material. This paper describes a technique which utilises the benefits of data fusion to develop a non-invasive system which both relies less on the expertise of the operator whilst improving false positive rates. To improve the capabilities of cocaine detecting fibre optic sensor, the raw data was pre-processed and features were identified and extracted. The output of each feature is a decision on the classification and the conditional probability that it belongs to the chosen class based on the observed data which serve as input into a Bayesian data fusion module and outputs the probability that a sample belongs to a class based on the observed features and makes a decision based on the class with the higher probability. The results show that the Bayesian fusion module greatly improves the detection rates of individual feature.

Keywords: bayesian data fusion, feature selection, fibre optic sensor, classification.
Definitions

TPR (True Positive Rate): This is the percentage of all cocaine data which are correctly identified as cocaine data.

FPR (False Positive Rate): This is the percentage of non-cocaine data incorrectly classified as cocaine data. It is also known as false alarm.

TNR (True Negative Rate): This is the amount of correctly classified non-cocaine data.

FNR (False Negative Rate): This is percentage of incorrectly classified cocaine data. It is also known as missed detection.

1.0 Introduction

The concept of data fusion is one which is gaining grounds. Although it was initially developed for military applications for target tracking, automatic target recognition and smart weapons (Erhard & McGalliard, 2011; Hall & Llinas, 2009), in recent years with improved technology and with sensors becoming more affordable, it has found application in non-military areas including Robotics Zou, (Ho, Chua and Zhou 2000) medical diagnoses (Luo, Chang, and Lai 2011) and environmental monitoring. In addition to these applications, data fusion has also attracted interest in the area of border security. (Derrick, 2010) describes a data fusion technique for combining ultra-fast infrared and near-infrared cameras with a Laser Doppler Vibrometer (LDV) in a bid to capture psychophysiological and behavioural cues for deception to aid border security agents detect whether an interviewee is telling the truth or not. In cargo screening in particular, attempts have been made to develop techniques for detecting for biological weapons or drugs concealed in containers and luggage (Tromp, 2006). In drug detection, the need for a reliable system to detect concealed drugs is now even more pronounced. Cocaine is one of the most frequently abused
drugs and as such is a very viable product for smugglers (Grabherr, 2008). In 2009, a car arriving from Calais to Dover was discovered to have heroin worth over £1 million hidden in its airbag (Anon., 2009). Other similar cases have been discovered at various ports and sea borders (Gysin & Mills, 2008; UKBA Business Plan, 2009). Current techniques used in the detection of cocaine involve the use of the analysis of scatter radiation (UKBA Business Plan, 2009). One major issue is that operators need to be properly trained to be able to distinctly detect specific substances. Another issue is the exposure of container contents to x-ray and gamma ray radiation thus exposing not only the contents but also the operators to dangers (UKBA Business Plan, 2009). Another method currently used in detection is trace detection machines. These machines work by collecting sample of particles and detecting residues of cocaine. They are also used in the detection of explosives and other common contraband drugs like heroin. Swabs are used in collecting residues by rubbing them over container contents and then inserted into chemical analyzers for detection. Usually, the analysers detect within a few seconds but it takes minutes to swab container contents. To maintain the effectiveness of this detector, it has to be thoroughly cleaned for first use and after every positive detection, failure to do this will lead to false positives (NIJ Report, 1999). Usually, a trace detector is ‘challenged’ with blank swabs until it gives a negative detection before being used on a fresh sample. This may require the detector to be left idle for several hours at a high temperature to ensure it is clean from chemical traces (NIJ Report, 1999). The high chances of a false alarm after a positive detection are some of the challenges of this system. False alarms usually lead to a distrust of the system by operators. There is thus a need for a non-invasive system which both relies less on the expertise of the operator and can also improve false positive rates. This paper describes a technique which utilises the benefits of data fusion to develop such a system. The system uses the data from a fibre optic sensor developed for detecting cocaine (Nguyen,
Hardwick, Sun, & Grattan, 2011) as input data into a data fusion process. The results from (Nguyen, Hardwick, Sun, & Grattan, 2011) show that the fibre optic sensor exhibits an “increase in fluorescence intensity in response to cocaine” and in addition, it shows good selectivity over other drugs. However, although the sensor shows good selectivity over other compounds, detection is possible only when the concentration of the sample being tested is known a priori. Therefore, to improve its capabilities in providing on the spot non-invasive means of detecting for cocaine at borders, it would need to overcome this limitation. This paper outlines efforts towards improving the capabilities of this sensor. Using spectra data sourced from experiments carried out and outlined in (Nguyen, Hardwick, Sun, & Grattan, 2011), the raw data was pre-processed and features were identified and extracted. These features served as inputs to a data fusion module. The feature selection process is outlined in another paper but is summarised in this paper. The performance of these features is although may look satisfactory is further improved by combining the decisions using the Bayesian algorithm. The output of each feature is a decision on the classification of a sample and its conditional probability that it belongs to the chosen class based on the observed data. These serve as input into a Bayesian data fusion module which then combines the feature probabilities and outputs the probability that a sample belongs to a class based on the observed features and makes a decision based on the class with the higher probability. The sensor system for the purpose of this project is delivered to a cargo container via a platform robot (called ferret robot). The ferret will move around the container and collect raw data samples using the fibre-optic sensor at time intervals $t_1...t_n$ and this data is subsequently analysed.

The next section will discuss data fusion models as extracted from literature. This will lead to introduction of the single sensor fusion and its appropriateness for the system implemented in this paper. Section 1.3 will present the fusion model and
highlight the algorithm used for single sensor fusion. In section 2, the methods covering data set, data pre-processing and feature selection, Bayesian fusion algorithm and Classification and Feature Performance will be discussed. The results and discussion sections will then follow accordingly.

1.1 Data Fusion Models

When it comes to fusion of data from a single sensor at various time intervals or from multiple sensors, it is imperative to decide at which level of a data fusion algorithm the sensor data will be combined. It is important because the choice of the design may affect the quality of the output (Hall & Llinas, 1997). Based on different systems and data-fusion problems, experts have come up with a number of architecture. In the mid-1980s, the Joint Directors of Laboratories (JDL) from the United States Department of Defence recommended one of the earliest data fusion models (Estebani, Starr, Willetts, Hannah, & Bryanston-Cross, 2004). This JDL model The JDL model (figure 1) has four levels (or five levels since level zero was added in 1998 (Steinberg, Bowman, & White, 1999)) – The first level (level 0) describes the estimation of signal states. This state implements identification of patterns inferred from sensor measurements; at the next level, (level 1) using the patterns identified at level 1, estimation and prediction of parametric and attributive states of the entity to be identified are made; it is at the third level (level 2) that an assessment of the situation is projected. This involves evaluating an estimate of the relationships among entities and the implication of these relationships for the state of the entities; at the fourth level (level 3), an impact assessment is conducted and it is at this stage that the system performs a self-check to estimate the costs such as signal, entity or situation states given the system’s alternative courses of action (Steinberg & Bowman, Revisions to the JDL Data Fusion Model, 2009); the final stage (level 4) conducts an assessment of all levels to rate their performance against
expectations. This original version of the JDL model now has several variants since some researchers have critiqued it because of its military style description (Carvalho, Heinzelman, Murphy, & Coelho, 2003) (Steinberg, Bowman, & White, 1998). The argument is that the military styled model would make it difficult to associate data fusion concepts and thus apply it to other applications. Thus researchers like Steinberg et al (1998) have revised the model by updating its taxonomy by replacing level 1 (object refinement) and level 3 (threat assessment) with more flexible event management and impact assessment levels.

![The Joint Directors for Laboratories Data Fusion Model](image)

Figure 1 The Joint Directors for Laboratories Data Fusion Model

One other suggested shortcoming of the JDL model is that it is described as capable of only allowing for a sequential ordering between levels and no flexibility and although Bedworth and Obrien (1999) have refuted this claim, the still admit that the JDL model may not fit all Data-Fusion application and is not rightly implemented due to its military terms and non-robust definitions. As a results of the above, many have come out with various models including the Thomopoulos model (Thomopoulos, Sensor Integration and Data Fusion, 1989), Omnibus model (Bedworth & O'Brien, 1999) and so on. The above models mainly serve as guidelines for data fusion process. The problem to be solved usually dictates what
model to follow or if a new model should be designed. In this paper, a ‘new’ model using existing models as guidelines was developed. However, despite the fact that the model was designed for this research, it is flexible enough to fit other applications into it.

1.2 Single Sensor Fusion

The fusion of data can be performed on data coming from multiple sensors or data from a single sensor at different time intervals. When data from multiple sensors are fused, it is known as multi-sensor data fusion. However, when a single sensor is used (for example, in the case of fusing video images from a single camera panning a wide area coverage over a period of time), this is known as single sensor tracking or filtering (Koks & Challa, 2005). This research is focused on single sensor fusion where data is collected over time from a fibre optic sensor with the measurements fused after processing.

In single sensor tracking, to identify or classify an object normally follows a process of first extraction of features from the pre-processed raw data collected by the sensor before the identity declaration can take place (Hall D., 1992). Technically, some of the models described in the previous section will work for single sensor fusion and the levels can proceed just as they would for multi-sensor fusion. One main issue to look at and which serves as advantage for single sensor fusion over fusion of different sensors is that for single sensor fusion, there is less difficulty when it comes to computations especially with regards to data alignment.

Data in a substance detecting system collected at different time intervals can be combined at three different levels some of which are described below:

1. Sensor level: The raw data acquired from the sensor can be pre-processed and fused to generate new data from which features can be extracted. For example, in substance detection, raw data collected at two time intervals (from different areas of
a cargo container for instance) are fused together to form a new set which would then be subjected to feature extraction and identity identification process.

2. Feature level: Feature sets may be extracted can be combined thus creating a new feature set which now represents the substance. The new feature set can then be fused to make a decision on the identity of the substance. For example, in substance detection, extracted features can be combined to form a joint feature vector which will then represent the substance. Additional step of feature selection process may then be performed to reduce the dimensionality of the features.

3. Decision level: At this level, individual extracted features make decisions on identity detection based. These features are features extracted from pre-processed raw data. The decisions from these features are then combined to output a single decision with accompanying uncertainty level. Common techniques used for decision fusion include Bayesian, majority voting and Dempster-Shafer techniques.

Data can be combined at any of these three levels depending on the application. Combining at raw data level however may lead to over-fitting of the data as dictated by the curse of dimensionality (Trunk, 1979) due to the vast amount of data available at this level. Thus extraction of features will serve to minimize the amount of data by extracting useful information from the raw data. The implication of this is that there is less computational work involved and thus an improvement in processing time. Combination of data at feature level has its limitations in real life applications including that there might be no known link between the features making it difficult to combine them and the fusion of two features may lead to a joint feature data with a high dimension resulting in a curse of dimensionality problem. Due to its simplicity and use of information at feature level, plus the disadvantages of implementing fusion at other levels highlighted, decision level fusion will be implemented in this paper.
In developing a model for the purpose of this research, two factors were taken into consideration – (1) that it is a single sensor system and thus (2) data will be fused over time. As mentioned earlier, this system is being designed for a cargo screening application. A robotic ferret has been designed to convey the sensor into the cargo container. The ferret is controlled by a border security official from a remote position. The overall aim is to collect data from different positions (and time) in the container as determined by the user and information collected combined to give a decision on probability of substance present or otherwise. A step by step algorithm was first outlined in this research as follows:

1. Sensor collects raw data \( (X_t) \) at time \( t \)
2. Data is pre-processed
3. Features \( (Y_t) \) are extracted from pre-processed data and decision is made based on individual feature data
4. Posterior probability at time \( t \) based on fusion of decisions made by each feature is calculated
5. Sensor collects raw data \( (X_{t+1}) \) at time \( t+1 \)
6. Data is pre-processed
7. Features \( (Y_{t+1}) \) are extracted from pre-processed data
8. Posterior probability at time \( t+1 \) based on fusion of decision made by each feature is calculated
9. Posterior probability at time \( t+1 \), \( P(X_{t+1}|Y_{t+1}) \) that the analyte observed is cocaine given all previous probabilities (steps 4 and 8) is calculated.
10. The prior probability used at time \( t+1 \) is the posterior probability from time \( t \).
At the start of the algorithm, the prior is assumed to be non-informative and evenly distributed as there is no prior information available. Subsequently, as posteriors are evaluated, they are assigned priors for the following evaluation of posteriors.

The fusion model used in this research is as shown in figure 4. $S_t$ and $S_{t+1}$ represent the sensor with data supplied by them at times $t$ and $t+1$..... When raw data is retrieved by the sensor at time $t$, it is pre-processed and features are extracted. Individual features then make a decision and a probability matrix containing the true positive rate (TPR), false positive rate (FPR), True negative and false negative rates is then outlined. The decisions of the individual features are then combined to give a fused decision. Using pre-determined threshold (see section 2.3 for threshold selection), the system will then decide whether to decide on detection or no-detection. If further analysis is needed, the system will extract another set of data and the repeat the process (Steps 5 to 8) above and the result fused with the posterior probability resulting from the first data sample.

2 Methods

2.1 Data Set

The initial set of data consisted of 115 spectra recorded based on experiments performed using a fibre optic sensor on various cocaine and non-cocaine samples (Nguyen, Hardwick, Sun, & Grattan, 2011). Individual spectrum is made up of pairs of intensity (counts) vs. wavelength (nm) data with wavelength ranging from 450nm to 750nm (see figure 2).

The spectra data collected were for six compounds used to test the selectivity of the sensor: codeine, ketamine, ecegonine methyl ester, amphetamine sulphate, buprenorphine HCl. However, since the focus of the research is to uniquely identify
cocaine data, the data after pre-processing were grouped into cocaine and non-cocaine data with the last 5 sets of data representing the non-cocaine data set thus providing 59 sets of cocaine spectra and 56 sets of non-cocaine spectra.

2.2 Feature selection and Data pre-processing

One way of selecting features used in this research was to see the performance of the proposed features in their classification of the samples into cocaine and non-cocaine. In this research, 12 features were selected for identity declaration. The features were selected using ‘activity points’ detected on the spectra. The activity points are obvious distinctions based on visual analysis of the spectra. These distinctions are points that show a change in gradient. The spectra where then segmented based on this change in gradient as against segmenting the spectra by equally dividing the wavelength.

![Figure 2. Intensity vs Wavelength (nm) for cocaine at 1000µM](image)

For example, as shown in figure 2, one would see that between wavelengths of 450nm and 470nm, the spectrum is seemingly smooth with constant gradient.
However, the gradient changes from this point into a positive gradient and becomes even steeper as it rises to around 500nm. The exact positions of these activity points were found and recorded based on analysis of the data using Microsoft Excel spreadsheet. Using these activity points, the spectra were segmented into six (6) sections and with these sections, four features were selected from the changes in intensity and two features from the gradient changes. The peak was also added as additional feature. In all, a total of 12 unique features were identified: Peak (PK), Steepest Slope (SS), Maximum Negative Slope (MNS), Most Positive change in intensity (MPI), Average intensity in regions (AIR) (there are six regions and therefore six values representing each spectrum), Most negative change in intensity, Average change in intensity.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak (P)</td>
<td>The peak intensity value is the maximum value for the spectrum after normalization.</td>
</tr>
<tr>
<td>Average change in intensity (AVI)</td>
<td>The AVI is the mean of the difference between intensity values corresponding to extreme wavelength values for each segment.</td>
</tr>
<tr>
<td>Steepest Slope (SS)</td>
<td>The slope of each region is the gradient at the mid-points of the region. When the slopes for all regions are evaluated, the steepest slope is the gradient with the highest value.</td>
</tr>
<tr>
<td>Most Negative Slope</td>
<td>Once the slopes have been calculated as in SS, the maximum negative slope is the gradient with the most negative value.</td>
</tr>
<tr>
<td>Most Positive change in Intensity (MPCI)</td>
<td>The change in intensity is the difference between intensity values range for each</td>
</tr>
<tr>
<td>Segment with positive and negative values. The MPCI is the highest positive change in intensity value.</td>
<td></td>
</tr>
<tr>
<td>Most Negative change in Intensity (MNCI)</td>
<td>As in the MPCI, after the change in intensity has been calculated for each partition, the most negative change in intensity is the intensity change with the most negative value.</td>
</tr>
<tr>
<td>Average Intensity in Regions (AIR)</td>
<td>For all the regions, the mean of the intensity values (corresponding to various wavelength values) in each region is also calculated and tabulated as AIR1, AIR2, AIR3, AIR4, AIR5 and AIR6.</td>
</tr>
</tbody>
</table>

### 2.3 *Classification and Feature Performance*

Classification of data is performed experimentally. Each of the features is used to discriminate the data into cocaine and non-cocaine sets using a pre-determined threshold. The threshold chosen is determined based on the tests carried out and on False positive and True positives. By experimentally varying the threshold, false positive and true positive values are determined and the ‘best’ threshold is that which gives the lowest false positive rate side by side a high true positive value. In cargo screening application, using the four decision variables i.e. True positives, false positives, true negative and false negative, it would be seen that whilst it is easy to statistically account for false positives, it is difficult to determine the amount of positive containers (containers with suspect substance) passed off as negative (that is containing no contraband substance), negative containers correctly passed off as negative and positive containers wrongly passed off as negative. Thus, in this system, priority is achieving a low false positive rate above other performance criteria. An
example from the actual data is used below to give a more detailed explanation of the threshold selection process.

Figure 3 showing distribution for sample cocaine and non-cocaine data. (threshold is the thick vertical line)

Figure 3 shows two distributions, one representing cocaine data and the other non-cocaine data. If the vertical line parallel to the y-axis represents the threshold (T), and if all data to the right of the threshold gives off an alarm for positive detection of cocaine, then for the cocaine data, all data to the right of the threshold will be correct detection of cocaine while all data to the left of T of the cocaine distribution will be false negative (FNR). Consequently for the non-cocaine data distribution, all data to the right of T will be false positives (FPR) and all data to the left of T will be correct non-detections (TNR).
Table 1 TPR and FAR for varying threshold

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TPR</th>
<th>FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>1</td>
<td>0.9756</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.7805</td>
</tr>
<tr>
<td>0.70</td>
<td>0.80</td>
<td>0.4390</td>
</tr>
<tr>
<td>0.65</td>
<td>0.40</td>
<td>0.00</td>
</tr>
</tbody>
</table>

An example using actual data for one of the features is summarised in table 1. The threshold value is varied from 0.65 and above. The corresponding TPR and FPR values are calculated as stated above. From table 1, it can be seen that there are extremes of 100% TPR value and on the other end, for a threshold of 0.65, there is a FPR of 0%. Whilst these might seem tempting, their respective FPR and TPR pairs are not acceptable. Therefore, the thresholds of 0.85, 0.75 and 0.65 are rejected. A more realistic threshold value for TPR and FPR values (0.8 and 0.44 of 0.70) is accepted.

For classifier validation, 15 test runs were performed which was used in the data were pseudo randomly divided into two in the ratio 70% to 30% to ensure that both groups contained the two sets of samples. After the grouping, 47% of the test data were cocaine data while the remaining 53% were non-cocaine data. The training data was used in designing the classifiers i.e. for evaluating a threshold value. Based on the selected threshold during training, the test data was used in evaluating the performance of the classifier.
2.4 Bayesian fusion algorithm

Sensors typically relay responses in binary 1 or 0 form representing a ‘detection’ or ‘no detection’ response with no level of uncertainty feedback leading to a loss in useful information (Fennelly et al. 1992). In the case of the fibre optic sensor used in this research, the response given is in the form of an intensity v. wavelength spectrum as explained earlier which has its own limitations as described in section I. Once the sensor has taken data, processed and analysed it as above, it gives a decision on the presence of the analyte or not with a level of uncertainty. As shown in table 1, all 12 features have varying degrees of uncertainties. In a view to improve on these performance criteria values, the sensor is used to collect data again at a time ‘t+1’ and the decision at this point with its uncertainty, is used to update the decision at time t.

The Bayesian theory is known as the canonical method for statistical inference problems (Wu, 2003). Its basic principle lies in treating all unknowns as random variables and thus they can be represented by a probability distribution. The Bayesian methodology states that the probability of an event is represented by the level of belief in the occurrence of that event. This level of belief is associated with a probability which can be updated based on additional information. In essence, based on initial probabilities (called prior probability), new information is used to update this prior probability to obtain a posterior probability. In classification, suppose one wants to classify a substance as belonging to class A or B, then given prior information as regards the class, this information is updated by new observations to provide a new posterior probability.

Mathematically, if \( p(C|X_{obs}) \) is the posterior probability of a sample being classified as cocaine (C) given observation \( X_{obs} \) and \( p(C) \) the prior probability, then
Conversely, if $p(\sim C|X_{obs})$ is the posterior probability of the sample declared as not cocaine, then

$$
p(\sim C|X_{obs}) = \frac{P(X_{obs}|\sim C)P(C)}{P(X)}
$$

where $p(X)$ is the normalizing factor which ensures that the sum of the posterior probabilities sum up to unity and is expressed mathematically as the sum of the products of likelihood probability $P(X_{obs}|C)$ and the prior probability $P(C)$. Equal prior probabilities has been assigned (assumption of no prior information) to avoid bias prediction results.

The aim in this research is to classify samples into one of two groups based on classifications made by individual features. Many Bayesian algorithms have been developed to solve this type of problem. Suppose $X$ is a target group belonging to one of $K$ classes (in this case $K = 2$). Then $X$ will take a value of one of $K=1,2$. If the selected independent features are $F_1, F_2, F_3, \ldots F_{12}$, then the conditional probability required is thus

$$
P(X|F_1,F_2,\ldots F_{12}) = \frac{1}{C} \prod_{i=1}^{12} P(F_i|X)
$$

where $C$ is the normalization constant. A decision is made for the probability estimate with the larger value.

2.5 Sensor fusion architecture

Sensor fusion in this research is performed at two levels. At the first level, the decisions reached by each of the 12 features are fused to get a single decision of presence or otherwise of cocaine. The second level fusion is performed as a means of
updating an initial decision made by the sensor. The architecture used for fusion in this research is shown in figure 4.

![Bayesian Fusion model](image)

**Figure 4 Bayesian Fusion model**

The two fusion stages consist of exactly the same pre-processing. The data is first normalised to ensure that all the values are kept between a range of 0 and 1. Once this process is completed, the feature selection process extracts all features from the spectra. The performances of the features are tested on a 13-1-1 neural network using k-fold algorithm. A detailed explanation of the extracted features is given in another paper but the results are outlined in the next section. Individual features then using
pre-set thresholds for each feature, then make an identity declaration i.e. cocaine or not cocaine with attached posterior probabilities. The Bayesian fusion fuses these decisions at the first fusion centre and updates it with the fusion decision at time t+1.

3 Results

A summary of the feature performance results is given below using the TP, TR, FP and FN as performance criteria.

**Data Set:** 59 instances of cocaine samples and 56 instances of non-cocaine for 12 features using 5-fold cross validation.

Table 2  Confusion Matrix Representation for K-fold cross validation ANN

<table>
<thead>
<tr>
<th>Test Mode: Five-fold cross validation</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>312 iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Fold cross validation</td>
</tr>
<tr>
<td>Confusion Matrix</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

(a) Confusion matrix for 1<sup>st</sup> fold cross validation

<table>
<thead>
<tr>
<th>332 iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Fold cross validation</td>
</tr>
<tr>
<td>Confusion Matrix</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

(b) Confusion matrix for 2<sup>nd</sup> fold cross validation
### 3rd Fold cross validation
**Confusion Matrix**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Predicted as</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>a = cocaine</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>b = not cocaine</td>
</tr>
</tbody>
</table>

(c) Confusion matrix for 3rd fold cross validation

### 4th Fold cross validation
**Confusion Matrix**

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>Predicted as</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0</td>
<td>a = cocaine</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>b = not cocaine</td>
</tr>
</tbody>
</table>

(d) Confusion matrix for 4th fold cross validation

### 5th Fold cross validation
**Confusion Matrix**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Predicted as</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>a = cocaine</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>b = not cocaine</td>
</tr>
</tbody>
</table>

(e) Confusion matrix for 5th fold cross validation

**208 iterations**

**Correctly classified instances**: 103, **Incorrectly classified instances**: 12

**Detailed Accuracy by Class**

<table>
<thead>
<tr>
<th></th>
<th>TP Rate (%)</th>
<th>FP Rate (%)</th>
<th>Precision</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cocaine</td>
<td>87.5</td>
<td>8.5</td>
<td>90.72</td>
<td>0.881</td>
</tr>
<tr>
<td>Not Cocaine</td>
<td>91.5</td>
<td>12.5</td>
<td>88.52</td>
<td>0.881</td>
</tr>
</tbody>
</table>
The confusion matrix in table 2 shows the result of all 5 stages of the 5-fold cross validation process. The major diagonals show the correctly classified samples of both cocaine and not cocaine samples while the minor diagonals indicate the confusion or incorrectly labelled samples. The false positive and true positive rates are computed as the ratio of incorrect cocaine prediction to total cocaine feature data in the data set and the ratio of correct cocaine prediction to total cocaine feature data in the data set.

In table 2(b), of the 23 sample feature data, 12 represent cocaine and 11 not cocaine data. The classifier correctly detects all 12 cocaine data as cocaine and all 11 not cocaine data as not cocaine representing 100% detection rate and 0 error.

Tables 2(c) and 2(d) show the result of the third and fourth stages of the cross validation process. In (c) there is 100% classification rate for the 7 and 16 cocaine and not cocaine feature samples which also implies a 0% error. The f fifth stage of the 5-fold cross validation process is shown in table 2(e). The classifier detects approximately 83% of the 12 cocaine samples and correctly classifies about 64% of the not cocaine samples. The error at this stage is thus 0.261. The total error estimate calculated as the average of all the error estimates is thus 0.104.
In summary, table 2(f) presents the performance of the overall system in terms of the True positive rates, the false positive rates, the precision and the area under the Receiver Operator Characteristics (ROC) curve.

The percentage of incorrectly classified data (cocaine and not cocaine samples) is 10.43% may indicates a satisfactory performance. The True Positive rate is 87.5%, True negative rate of 91.5%, the false positive rate is 8.5% and false negative rate is 12.5%. The aim is to use the Bayesian fusion algorithm to improve on this performance as presented in the next section.

Table 3 shows the classification output for each feature using both training and test data. The training data are those used for designing the classifier whilst the test data are those which were presented to the classifier to test its performance. From table 2, it will be seen that the features have their respective performances in terms of ability to classify. For the training data, for example, the average intensity (AVI) gives a false positive rate of 15% and a true positive rate of 80% whilst giving a 5% and 83% false and true positive rates using the test data. All the features have varying degrees of performance for both training and test data with the Average Intensity Value for region 4 giving the worst FPR for training data and test data whilst many of the features seemed to have similar results for the test data.

Table 3 FAR and TPR rates for feature training and test data

<table>
<thead>
<tr>
<th>Feature</th>
<th>AVI</th>
<th>SS</th>
<th>MNS</th>
<th>MPCI</th>
<th>MNCI</th>
<th>PK</th>
<th>AIR1</th>
<th>AIR2</th>
<th>AIR3</th>
<th>AIR4</th>
<th>AIR5</th>
<th>AIR6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training FAR</td>
<td>0.15</td>
<td>0.098</td>
<td>0.146</td>
<td>0.146</td>
<td>0.195</td>
<td>0.293</td>
<td>0.195</td>
<td>0.171</td>
<td>0.146</td>
<td>0.439</td>
<td>0.098</td>
<td>0.293</td>
</tr>
<tr>
<td>TPR</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.625</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Test FAR</td>
<td>0.05</td>
<td>0.095</td>
<td>0.05</td>
<td>0.095</td>
<td>0.095</td>
<td>0.05</td>
<td>0.095</td>
<td>0.095</td>
<td>0.095</td>
<td>0.24</td>
<td>0.05</td>
<td>0.095</td>
</tr>
<tr>
<td>TPR</td>
<td>0.83</td>
<td>0.83</td>
<td>0.61</td>
<td>0.83</td>
<td>0.83</td>
<td>0.61</td>
<td>0.83</td>
<td>0.83</td>
<td>0.61</td>
<td>0.83</td>
<td>0.67</td>
<td>0.83</td>
</tr>
</tbody>
</table>

3.1 Bayesian implementation
From the table in the previous section, it will be seen that features have varying performances. Although some of them have similar performance levels for example, the AIR1 and AIR2 have similar true positive rates of 80%, it should be noted that the samples they tested rightly or otherwise were noticeably different.

The conditional probability distributions for each individual feature \( p(F_i|X) \) are obtained using the training data. These are then used in the Bayesian fusion module to calculate the conditional probabilities \( p(X|F_i) \) and the results shown in Table 4. The results show a 100% classification rate when the threshold is set at 0.5

<table>
<thead>
<tr>
<th>Feature</th>
<th>AVI</th>
<th>SS</th>
<th>MNS</th>
<th>MPCI</th>
<th>MNCI</th>
<th>PK</th>
<th>AIR1</th>
<th>AIR2</th>
<th>AIR3</th>
<th>AIR4</th>
<th>AIR5</th>
<th>AIR6</th>
<th>FUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>% correctly classified</td>
<td>90</td>
<td>87</td>
<td>79</td>
<td>87</td>
<td>87</td>
<td>74</td>
<td>87</td>
<td>87</td>
<td>77</td>
<td>82</td>
<td>87</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>% wrongly classified</td>
<td>10</td>
<td>13</td>
<td>21</td>
<td>13</td>
<td>13</td>
<td>26</td>
<td>13</td>
<td>13</td>
<td>23</td>
<td>21</td>
<td>18</td>
<td>13</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5 Confusion matrix for fused feature data

<table>
<thead>
<tr>
<th>a</th>
<th>B</th>
<th>Predicted as</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>0</td>
<td>a = cocaine</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>b = not cocaine</td>
</tr>
</tbody>
</table>

Correct classification = 100%
Incorrect classification = 0%

4. Discussion

Figure 5 shows the performance of the fusion of all 12 features over 18 cocaine test data samples and the posterior probability of the fused feature level decisions (in black). The sensor is tested with 18 cocaine data and using the pre-
processing/processing methods described above, each feature decides on detection or no-detection. The correct classification rates for individual features are shown in table 3. Each feature decision is colour coded. The posterior probabilities are plotted against each sample representing time \( t, t+1 \ldots t+18 \). Immediate past posteriors at time \( t \) are used as priors to update data at time \( t+1 \) to calculate new posteriors at time \( t+1 \).

![Posterior probability graph](image)

**Figure 5** Posterior probability of feature classification being correct after feature has declared for cocaine \( P(T^\dagger|C) \).

Starting from the first sample point, all features rightly decide that the sample is cocaine although with varying degrees of uncertainty and the fusion of their decisions rightly reflects this. At decision points 3, 4 and 13, the features make the wrong decision on substance detected but interestingly, the fused decision gives a decision with uncertainty above the 0.5 threshold value. Although threshold value selected for this project is higher (from 0.8), a decision above 0.5 will at least warrant further investigation. Setting a threshold of a value lower than 0.6 gives a 100% decision rate, but with a threshold of a value higher than 0.6, a correct decision
rate of over 89% is achieved (a threshold of 0.8 gives 89% correct decision) which matches the best performing feature in terms of correct decisions.

![Figure 6 Posterior probability of feature making wrong decision based on positive declaration for cocaine P(T|C)](image)

The experiment above is repeated with 22 samples of non-cocaine data. The resulting probabilities are shown in figure 6. Again, the posterior for each step acts as prior for the immediate next step. It should be noted that ideally, each feature posterior (shown with various colour codes) should tend to zero for each data sample, however because of false positive detections, this ideal situation does not happen. Nevertheless, with the fused data (black), the ideal situation is mirrored leading to a zero number of false positives.
5. Conclusion

The results of fusion are shown in the tables above based on the test data. The confusion matrix in table 4 shows that the Bayesian fusion module performs at 100% correctly classifying all samples into cocaine and non-cocaine groups at threshold values above 0.5. Although in terms of correct decisions on cocaine data, the fused decision matches the best performing feature i.e. the average change in intensity (AVI), it should be noted that in terms of classification on non-cocaine data, the fused system outperforms all the features with no false positive. Therefore overall, the fused system gives an improved performance over all the features. It can thus be seen that the feature based Bayesian fusion algorithm performs better than the any of the individual features. Bayesian fusion in general shows promising results for feature level data fusion. This paper shows the steps involved in achieving this type of data fusion. First it builds the network model that would implement the data fusion explicitly highlighting the underlying stages. Rather than work with raw data from the sensor, unique independent features are experimentally selected and the features then act as inputs into the data fusion module. Decision on the class a sample belongs is evaluated by determining the probability of the sample belonging to the class given the feature decisions and a decision is made for the probability with the larger value. It should be noted that although the features used in this paper are obtained from raw data from the same sensor, the concept can be applied to multiple sensors with applications abound.
References


Tables and Figures

Table 1 TPR and FAR for varying threshold
Table 2 Confusion Matrix Representation for K-fold cross validation ANN
Table 3 FAR and TPR rates for feature training and test data
Table 4 Percentage correctly and wrongly classified for 12 features
Table 5 Confusion matrix for fused feature data

Figure 1 The Joint Directors for Laboratories Data Fusion Model
Figure 2 Intensity vs Wavelength (nm) for cocaine at 1000µM
Figure 3 showing distribution for sample cocaine and non-cocaine data (threshold is the thick vertical line)
Figure 4 Bayesian Fusion model
Figure 5 Posterior probability of feature classification being correct after feature has declared for cocaine P(T^+|C).
Figure 6 Posterior probability of feature making wrong decision based on positive declaration for cocaine P(T^-|C)