# Developing Integrated Data Fusion Algorithms for a Portable Cargo Screening Detection System

## Using Dempster-Shafer and Bayesian Theories

by

**Ayodeji Akiwowo**

**PhD Thesis**

**Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University**

**March 2012**

# Abstract

Towards having a one size fits all solution to cocaine detection at borders; this thesis proposes a systematic cocaine detection methodology that can use raw data output from a fibre optic sensor to produce a set of unique features whose decisions can be combined to lead to reliable output. This multidisciplinary research makes use of real data sourced from cocaine analyte detecting fibre optic sensor developed by one of the collaborators - City University, London.

This research advocates a two-step approach: For the first step, the raw sensor data are collected and stored. Level one fusion i.e. analyses, pre-processing and feature extraction is performed at this stage. In step two, using experimentally pre-determined thresholds, each feature decides on detection of cocaine or otherwise with a corresponding posterior probability. High level sensor fusion is then performed on this output locally to combine these decisions and their probabilities at time intervals. Output from every time interval is stored in the database and used as prior data for the next time interval. The final output is a decision on detection of cocaine.

The key contributions of this thesis includes investigating the use of data fusion techniques as a solution for overcoming challenges in the real time detection of cocaine using fibre optic sensor technology together with an innovative user interface design. A generalizable sensor fusion architecture is suggested and implemented using the Bayesian and Dempster-Shafer techniques. The results from implemented experiments show great promise with this architecture especially in overcoming sensor limitations. A 5-fold cross validation system using a 12 – 13 - 1 Neural Network was used in validating the feature selection process. This validation step yielded 89.5% and 10.5% true positive and false alarm rates with 0.8 correlation coefficient. Using the Bayesian Technique, it is possible to achieve 100% detection whilst the Dempster Shafer technique achieves a 95% detection using the same features as inputs to the DF system.

Key words: Data Fusion, Bayesian, Dempster Shafer, feature selection, spectral, Neural Network

# Acknowledgements

I am very grateful to my supervisor Dr Mahroo Eftekhari for her relentless help and support. I appreciate all her contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. Words are not enough to show my appreciation.

I also appreciate the support and help of all my friends at Loughborough University; in particular, all researchers at the Civil and Building Engineering research hub and all members of the Redeem Christian Church of God, you guys are the best!

I will like to thank my family including my mother, sisters and brother in-law for their support and understanding. Studying for a PhD can not only be intellectually tasking but also emotionally and psychologically demanding. My family was there all through this period of my life and I thank them.

Finally, to the Immortal and Invincible One, I give all the glory.

# Table of Contents

## Table of Figures

## Definitions and Acronyms

**Basic Probability Assignment** (bpa) – the basic probability assignment represented by *m* is a basic measure representing confidence in a hupothesis.

**Cargo** – A cargo can be defined as any kind of goods ready and packed for shipping.

**Cocaine** – Cocaine is as addictive drug C17H21NO4 which is derived from coca plant and can also be prepared artificially.

**Contraband** - Goods prohibited by law or treaty from being imported or exported

**DF** – Data Fusion

**FAR** – False Alarm Rate (see False Positives)

**False Negatives** – False negatives usually known as missed detections are simply cases where real threats are passed off as non-threats.

**False Positives** (1 – Specificity) – False positives also known as false alarms can be defined as a situation where for a system that detects threats for example, non-threats are inappropriately signalled as threats.

**Features** – A feature is a distinctive attribute of a substance which identifies it from other substances.

**Ferret** - A weasel like, usually albino mammal often trained to hunt rats or rabbits.

**Frame of discernment** (Ω) – this is a set containing all possible events.

**Mutually Exclusive events** - Two events are mutually exclusive (or disjoint) if it is impossible for them to occur together

**Plausibility ((*Pls(.)*)** – The plausibility is defined as the total mass of all states which do not contradict the target state.

**ROC** – Receiver Operator Characteristics curve is a graphical curve which shows the performance of a system as its discriminant threshold is varied. In this case, it is the plot of the true positives vs the false positive rate.

**Sensor Fusion** is the combining of sensory data or data derived from sensory data such that the resulting information is in some sense better than would be possible when these sources were used individually.

**Support (*Spt (.)*)** – The support for a target is defined as the total mass of all states implying the target state.

**True Negatives** (Specificity) – True negatives are cases where non-threats are correctly classified as non-threats.

**True Positives** (Sensitivity) – True positives or true detection are cases where real threats are correctly identified as threats

**Precision** - The ratio of the true positive rate to the sum of the true positive rate and false positive rate or represents the probability that substance A is detected given that the sample is actually substance A.

# CHAPTER

# 1

# INTRODUCTION AND MOTIVATION

Sensors are typically used for gathering information from the environment. This step of gathering information from the environment is the first step in building intelligent Human – Computer Interaction (HCI) (Wu, 2003). In building intelligent systems, it is important to retrieve as much information from the environment as possible. In many cases, the data received is usually in raw form and not useful unless further processing is performed on it to extract useful features. Depending on the aim of the intelligent system, multiple sensors may be required to obtain as much information as possible. These sensors may differ in terms of what physical quantities they measure or in what format they generate data. Using the combined information from these sensors thus outlines the importance of a data fusion process. In the case of a single sensor, asides from smart sensors which can extract

features automatically from raw data, many sensors deliver only raw data which need further processing. Extracting features from the data is necessary for further processing so as not to overburden the processing system. The decisions from each individual feature can also then be combined to give a more ‘intelligent’ result. Sensor fusion is thus in addition to fusion of data from multiple sensors, also a combination of information /decisions from features of the same sensor.

According to Erhard et al (2011), data fusion technology was initially developed for military applications mainly due to the high costs of sensors used. However, as sensors got cheaper and technology advanced, it began to have more civilian applications. Today, sensor data fusion is applied in robotics, and biomedical applications (Luo et al., 2011). For border security, data fusion has attracted varying research into how to improve current border security systems. For example, a combination of ultra-fast infrared and near-infrared cameras in combination with a Laser Doppler Vibrometer (LDV) was used in the study of how to capture psychophysiological and behavioural cues for deception to help border security agents determine when an interviewee is lying (Derrick et al., 2010). Tromp (2006) mentions fusion of information from existing detection systems at United States security borders to help detect, screen and intercept chemical, biological, radiological, nuclear and explosive (CBRNE) materials. In particular, the challenge at cargo screening is increasing as smugglers take advantage of low ratio of border agents to cargo entering via the borders on an annual basis. The fact that the

cargo usually has to undergo spot checking exposes a challenge of difficulty in checking all cargo and an increase in false positives.

In addition, smugglers have devised new means of ferrying drugs via borders. In a recent event, Heroin worth up to £1.5 million has been found by customs officers partly concealed in a car's airbag (Anon., 2009). The drug was also discovered in a false bulkhead behind the rear seats of the Kia Clarus car after it arrived at Dover on a ferry from Calais. In a similar case, a former England sportsperson was accused of trying to smuggle in cocaine with an estimated street value of £200,000 through Gatwick Airport (Gysin & Mills, 2008).

Also on the 10th of November 2008, 15.6 kilos of cocaine valued at between £750,000 and £1million were discovered in a passenger's luggage by UK border Agency officers working at Edinburgh airport. This was described as Scotland's largest class 'A' drugs seizure as an airport (UKBA, 2009).

Cocaine is one of the most frequently abused drugs and as such is a very viable product for smugglers (Grabherr et al., 2008). Accordingly, smugglers have devised several ways of smuggling cocaine and other drugs including body packing (Hergan et al., 2004) (Beck & Hale, 1993) which is a dangerous means since it involves hiding packs of concealed drugs in one's body (rectum, vagina and bowels via swallowing) and hiding in luggage or cargo (Vogel & Brogdon, 2003). In both cases, application of industrial x-rays, gamma ray, conventional and back scatter radiography and

sonography[1] all of which are imaging techniques have been proved to be successful in detection of hidden stowaways and other contrabands (Vogel & Brogdon, 2003); (Vogel et al., 2006); (Grabherr et al., 2008) (see Figure 1.1).

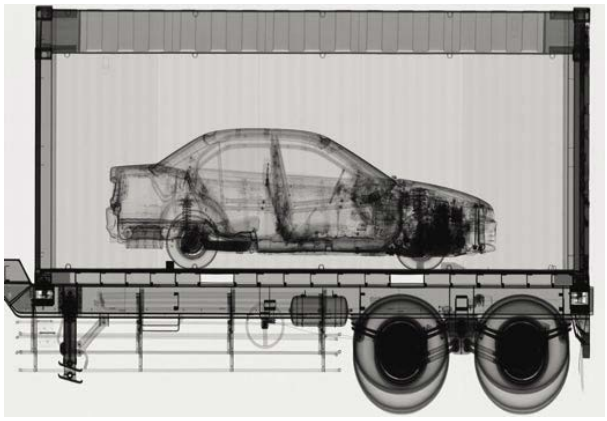


Figure. 1.1 A vehicle going through an X-ray Scanner

These imaging techniques have the capability of displaying high detailed images of both moving and stationary vehicles. Chemical compounds such as drugs are mainly detected via analysis of scatter radiation. This however requires highly trained operators to be able to distinctly detect specific substances. The dangers involved in the exposure of food and human to x-ray and gamma ray radiation are disadvantages of some of these techniques (Vogel & Brogdon, 2003). This exposure is not only a risk to the contents of the cargo but also a risk to the operators of the detecting devices (Vogel & Brogdon, 2003). Therefore the challenge of dangerous exposure to radiation and high dependence on trained operators to adequately interpret images is

[1] Sonography is the use of sound frequencies to produce images

some of the short coming of the above methods. There is thus a need for a non-invasive system that relies less on the operator but still able to deliver on accuracy.

Other existing methods available for the analysis of cocaine include gas chromatography mass spectrometry (GC/MS) which distinguishes and quantifies cocaine and its metabolites and other coca alkaloids. Its high sensitivity of 1-5 ng/ml makes it a more attractive option than High-pressure liquid chromatography which is less sensitive with a sensitivity of 20ng/ml) (Warner, 1993). However, its shortcoming lies in the fact that it requires a clean-up of the sample and derivatization of cocaine and its metabolites before use rendering it an expensive process (Devine et al., 1995). Other commercially available kits for screening for cocaine have lower sensitivities (>300 ng/ml) (Devine et al., 1995). Current cargo screening technology involves swabbing suspicious containers and testing with a trace-detection portal machine[2] (Figure 1.2) which will give off an audible/visual alarm if cocaine is detected. The challenge with this is that the operator has to physically enter the container to swab the contents. This might be dangerous especially if there are any harmful substances in the container. In addition, the operator has to know where/what to swipe to get an accurate detection.

[2] **Trace Detection portal machine** is a machine used to detect trace elements of a class A drug and/or explosives. The operator uses a swab which he rubs against suspect items. The swab is then inserted into the machine which gives an indication if the chemical substance is present.

To date, this is done randomly and may lead to missed detections. Having to move from container to trace-detector machine also means that sample may be contaminated mid-way leading to false positives. To meet this challenge will require a means of taking the sensor right inside the container.

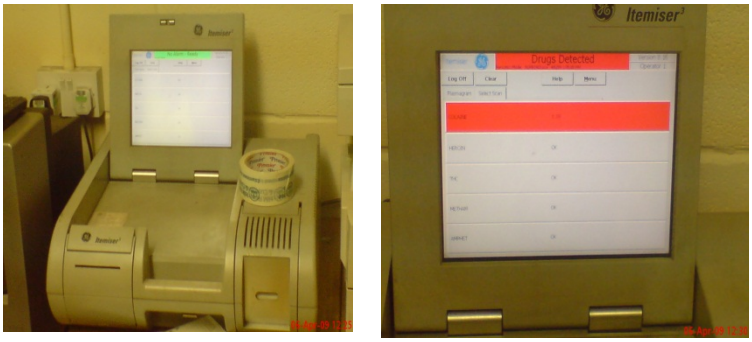



Figure.1.2. Trace-detector portal machine

## 1.1 Fibre Optic Sensors

The recent advances and reduction in cost have increased the interest in fibre optic sensors. In addition, they have the advantage of being small, lightweight and able to transmit light over long distances with minimal loss in energy. They have found applications in a diverse fields including sensing of positions, vibration, chemicals and other environmental factors (Fidanboylu & Efendioglu, 2009). In border security, fibre optic sensors have been successfully used in the detection of biological threats like *Staphylococcal enterotoxin*[3] type-B, *Francisella tularensis*[4]*, Bacillus anthracis,* and *Bacillus globigii spores* (Jung et al., 2003). In particular, for cocaine detection, fibre

[3] *Staphylococcal enterotoxin* is an enterotoxin produced by the bacterium *Staphylococcus aureus* and is a common cause of food poisoning,

[4] ***Francisella tularensis*** is a pathogenic species of gram-negative bacteria and the causative agent of rabbit fever. Due to its ease of spread by aerosol and its high virulence, *F. tularensis* is classified as a Class A agent by the U.S. government

optics technology has been used in detecting cocaine, its metabolites and other coca alkaloids using a monoclonal antibody (mAb) against a derivatized benzoylecgonine (Devine et al., 1995) and for detecting cocaine metabolites in urine (Nath et al., 1999). The impracticality of the above cases for cargo screening application includes the need to treat the cocaine with other compounds in both cases.

The cases presented above show that the current techniques available for detection of cocaine hidden in cargo containers at borders have their short comings. The imaging systems which make use of gamma rays, x-rays, sonography technology in spite of their abilities produce high quality images are reliant on the ability of the operator to visually identify contrabands. New generation x-ray machines for cargo screening not only ensures greater penetration allowing for deeper scrutiny of objects but also allows operators to view images in 3 dimensions and also in colour while being able to manipulate a containers contents' 3-D image on the computer screen. This also implies greater reliance on the operator to reliably determine if a hidden object is cocaine or if it just fits the parameters of a packed cocaine object. The implication of this is that any error of judgement by the operator would mean a missed detection or a false alarm. In addition, the exposure of the contents of container and operators to radiation in the case of gamma and x-rays is an increasing concern. The use of trace-detector portal machine is less dependent on operator's training compared with the techniques above. However, using a trace detector implies that container contents would be

unloaded and swiped with a pad which is then inserted into the machine. Although not all containers are opened and the border agency security relies on intelligence to determine which container to screen based on the most at-risk container, the time taken to open selected containers and swipe the contents is a cause for concern especially if it turns out to be a false alarm; this will greatly affect operator confidence. Also, the fact that operators need to enter containers to unload them of their contents before performing tests exposes the operators to unnecessary risks.

## 1.2 Data Fusion

The concept of receiving information can be broken down into two stages – sensor measuring and data processing. The sensor measuring aspect involves the collection of data or measurements of quantities dependent on the characteristics of the immediate environment. In the data processing stage, the measurements from the sensor measuring stage is taken through different stages including removal of noise, alignment and extraction of useful information. The final result is intelligent information which offers a more satisfactory meaning to the user than the raw data (Wald, 2000). In some cases, one set of data from a sensor may not be enough to totally characterise a system. In such a case, multiple data is collected at different time intervals. In other cases, one sensor may not be enough to characterize the system and as such, additional sensors are brought in to complement the sensor. The two cases mentioned are called Data Fusion. It is a multilevel, multifaceted process dealing with the automatic detection, association,

correlation, estimation and combination of data from single and multiple sources (U.S. Department of defense, 1991).

In general, data fusion techniques can combine data either from multiple sensors or as explained above from a single sensor over time (tracking) with information/data from databases with the aim of giving improved results than can be achieved from raw data alone.

Applications of data fusion can be found in diverse engineering and science disciplines including remote sensing (Haack & Bechdol, 1999), robotics (Zou et al., 2000), automated target recognition (Bethel & Paras, 1998) (Brown & Swonger, 1989) and medical applications (Hernandez et al., 1996) (Hernandez et al., 1999). Fusion of data can be achieved via various techniques including (Ma, 2001): Control theory (Sossai et al., 1999), signal processing (Soumekh, 1999), statistical estimation (Smith & Kelly, 1999) and artificial intelligence (Matia & Jimenez, 1998). The technique to use is dependent on various aspects including type of data, requirements of the application and the grade of reliability targeted (Carvalho et al., 2003).

## 1.3 Background to the collaborative research

This thesis (referred to as Loughborough University in this section) is a part of a collaborative research (called the Cargo Screening Ferret Project)) with the University of Sheffield and City University in conjunction with the British Home Office and Qinetiq funded by the Engineering and Physical Sciences Research Council (EPSRC).

The University of Sheffield will investigate different methods for robot designs with the aim of building a robot that will manoeuvre around the cargo container. The cargo container's design and fabrication material will guide the design selection for the robot. For this aspect of the collaborative research, this thesis will contribute to the design for the robot controls and communications while the British Home Office will aid in terms of possible design support. The role of the robot is to deliver the sensor into the container. This fibre optic sensor will be built and calibrated at City University in London. After fabrication and development of the sensor, City University and Loughborough University will perform specific experiments designed by Loughborough University with the aim of Loughborough University processing the raw data and extracting features which will serve as inputs into the data fusion model also developed at Loughborough University. In addition, the user interface to display the detection results and to provide motion control for the ferret will be developed by this thesis.

The overall aim of this collaboration is thus to build a ferret to detect illegal substances (in this case cocaine) being smuggled into the UK via its borders.

Figure1.4 below shows the interaction and contribution of this thesis and other project partners.

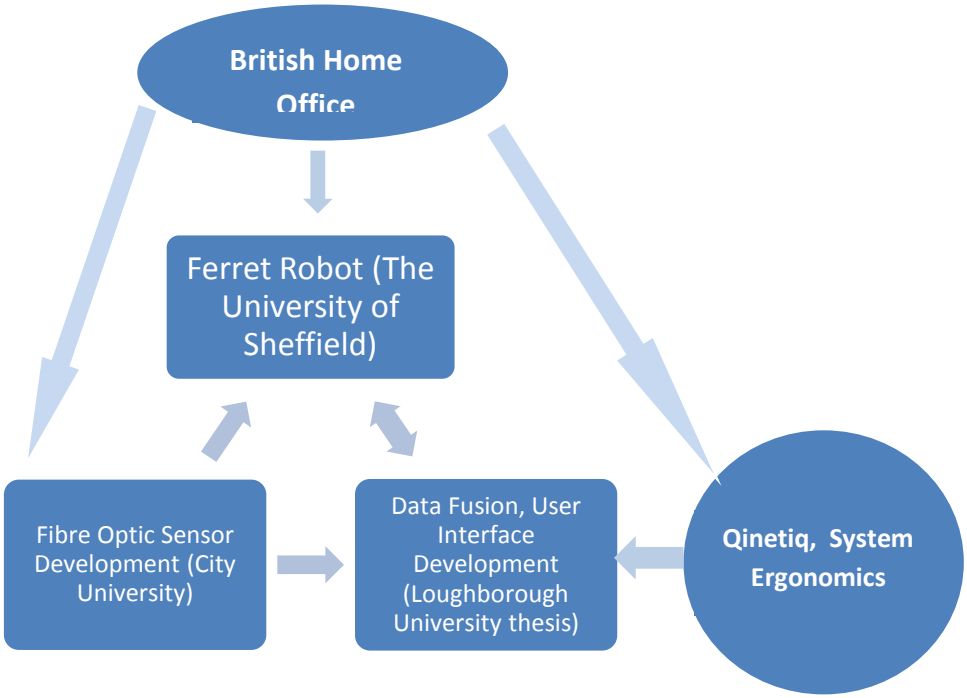



Figure1.4 Interaction between project partners

## 1.4 Aim and Objectives

This thesis is an important part of the collaborative research and aims to devise an automatic sensor fusion system for positive detection of cocaine using an optical fibre sensor. Specifically, using data fusion techniques, characteristic features from the sensor data will be extracted and combined in such a way as to provide improved results on the sensor output.

The objectives of this thesis are to –

- To explore existing data fusion techniques so as to identify suitable/robust ones.

- Characterise the characteristics of the cocaine detecting optical fibre sensor by analysing the raw data with close collaboration with City University
- Extract unique features from the sensor data
- Develop a model for the fusion of multiple decisions from sensor
- Capture User requirements for this system (in collaboration with Qinetiq, The University of Sheffield and the Home Office)
- Design a user interface to display appropriate information in a meaningful way and instilling confidence in the operator. The user interface will also serve as a controller for robot manipulation (with close collaboration with the University of Sheffield)
- Make recommendations for the implementation of an integrated cocaine detection system

This sensor developed for this thesis is a cocaine detecting sensor using optical fibre technology. The fibre optic sensor was developed by City University in London and the results detailed in a report (Nguyen, et al., 2010). The fibre-optic chemical sensor developed is based on molecularly imprinted polymer (MIP) which contains fluorescence in moiety as the signalling group and exhibits an increase in fluorescence intensity in response to cocaine in concentration range of 0 - 500μM in aqueous acetonitrile mixtures and has good reproducibility over 24 h (Nguyen, et al., 2010). Chapter 2 of this report will discuss more details on the optical fibre sensor.

## 1.5 Outline of the dissertation

The outline of this thesis is explained below.

**Chapter 2** of this thesis is a comprehensive literature review of sensors, data fusion and border security in general. Specifically, the literature review will look at the concept of data fusion as discussed in several researches and its applications. The way data is fused is dependent on application. The chapter will thus also look at suggested methods by which data can be fused. This chapter will also discuss some of the advantages of data fusion including applications of data fusion in both military and non-military environments. It will then narrow down to current techniques used in cargo screening at borders including detection of cocaine at borders and the challenges faced by these techniques. Finally, the chapter will discuss the principle behind fibre optic sensor detection with a follow up introduction into detection of cocaine using fibre optic sensors.

**Chapter 3** of the dissertation will discuss the different models used for data fusion and human computer interaction issues. It will begin by looking into the JDL model developed by the Data Fusion Group set up by the Joint Directors of Laboratories (JDL). All six levels of this JDL model will be discussed. Other models including the waterfall, omnibus amongst others will also be discussed. The various levels and support systems in this model will be highlighted in chapter 3. The HCI is an important part of many data fusion models thus this chapter will also dovetail into the issues surrounding the human computer interaction.

**Chapter 4** will give a brief background into the experimental set up and data collection. The chapter will then discuss data fusion algorithms used in this thesis – Bayesian and Dempster Shafer with the aim of giving an insight into the systematic process of developing a model for single sensor fusion for cocaine detection.

**Chapters 5** will outline the process of extraction of valid features from the raw data. The feature selection stage is preceded by a data pre-processing stage which involves cleaning the data, normalising and other processes. Cleaning the data removes noise which could be due to background errors, normalizing ensures all data from the sensor are in the same domain. Finally, the steps involved in validation using neural networks will be highlighted and the results will be discussed.

**Chapters 6** will outline the implementation of the data into the data fusion model using the Bayesian and Dempster Shafer techniques. The model developed in **Chapter 4** is the same model used for both implementations but because of the underlying differences in outputs for the two approaches, the output results may defer. The chapters will show the results given by each approach in terms of percentage true positive, false positive, true negative and false negative. A comparison of the output results from these techniques will also be outlined. The chapter will round off by discussing the Human Computer Interaction and highlight the features of the Interface designed to output results. It will also give a brief description of the ferret robot.

**Chapter 7** concludes the dissertation and will summarise the achievements and will also highlight some of the thesis contributions, its limitations and suggestions to industry.

# CHAPTER

# 2

# REVIEW OF DATA FUSION SYSTEMS AND CARGO SCREENING

Many practical problems arising in monitoring and detection can be modelled with the aid of parametric models in which the parameters are subject to abrupt changes at unknown time instants. These changes are normally associated with some form of disorder, which is highly undesirable and should be quickly detected with a few false alarms as possible. *Multiple sensors are used in these systems in order to reduce uncertainty and obtain more complete knowledge of the state*. Data fusion helps to combine the data from these sensors leading to a more efficient and thus reliable system.

This chapter will outline a review of sensors, sensor systems, data fusion and its applications and then x-ray cargo screening at borders. It will begin

by discussing sensors and sensor data with a look at different types of sensors, especially the fibre optic sensor which is the sensor used in this dissertation. The applications of the fibre optic sensor and how it is used in detection of certain compounds will also be discussed. Following this will be a highlight of data fusion with definitions, applications and its advantages presented. In addition, an overview of current applications of data fusion with respect to border security will be outlined. The application of data fusion is dependent on the specific issue to be resolved. Data can be fused over time from a single sensor or by combining results from multiple sensors. The applications of data fusion will also be examined in this chapter which will then delve into cargo screening and the use of data fusion in cargo screening.

Finally, the chapter will discuss the menace of cocaine smuggling via borders citing recent examples.

## 2.1 Data Fusion

The history of data fusion dates back to over forty years ago when it was used as a mathematical model for data manipulation. In 1986, the US Department of Defence founded the Data Fusion Group as a Data fusion subpanel of the Joint Directors of Laboratories (JDL) to unify research terminology and to promote technology transfer and cooperation between groups (Hall, et al., 2009).

In general, data fusion involves combinations of data or of sensory data or data derived from sensory data from disparate sources such that the

resulting information is in some sense, more accurate, more complete or more dependable than it would be possible when these sources were used individually. However, more particularly, several bodies and individuals have come up with different definitions for data fusion, the basics however, remain the same.

In terms of application, at the early stages, data fusion was at first used mainly in military intelligence. However, with time the application has been extended to weather prediction, remote sensing, air traffic control and navigation. Other applications include Intelligent Transportation System (ITS) and missile detection and tracking.

### 2.1.1 Definitions

One of the earliest definitions of Data Fusion came from the North American Joint Directors of Laboratories who defined Data Fusion as (U.S. Department of defense, 1987)

> "*A multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation and combination of data from single and/or multiple sources*".

However, Charniak *et* al (1987) further modified this definition by defining data fusion as follows:

> "*A multilevel, multifaceted process dealing with the detection, association, correlation, estimation and combination of data and information from multiple sources to achieve a refined state and*

*identity estimation, complete and timely assessments of situation and threat*".

These definitions focus on three (3) major aspects –

1. that data fusion is a process
2. that data fusion includes detection, association, correlation, estimation and combination of data.
3. that the results of data fusion include state and identity estimates at the lower levels and assessments of overall tactical situations at the higherlevels.

Also, Mitchell (2007) defines Data Fusion as

"*The theory, techniques and tools which are used for combining sensor data, or data derived from sensory data, into a common representational format*'' (Mitchell, 2007).

This definition further adds that the data fusion process uses both sensor data and data derived from sensory data. The importance of this is that it shows that the process of data fusion not only fuses data but also exploits all combinations of data that may be available thus making it an efficient process.

In summary, data fusion helps to improve confidence in the system by reducing uncertainty in the measurement outputs (Keller, 2008). The use of data fusion for any process should however be weighed against factors such as additional costs and complexity.

### 2.1.2 Structure of data fusion

Data is fused from either a single sensor taken at separate times or by combining data from multiple sensors at a specified time. Bearing this in mind, it can then be said that data can be fused across sensors, across attributes, across domains or across time (Boudjemma, et al., 2004).

Across Sensors – In this case, a number of sensors measuring the same property are fused together

Across Attributes – Sensors measuring different properties but for the same experimental situation are fused.

Across Domains – sensors measure the same attribute over a number of different ranges.

Fusion across time (temporal) – For this, several measurements are taken over time and current measurements are fused with historical information.

For un-identical data i.e. data from different types of sensors, the data is first analysed and converted to the same format whereas, if the sensors to be fused together are identical, this is not necessary as the data should already be in similar format.

Dasarathy classifies a multisensory data fusion system according to its input/output characteristics as shown in Table 2.1 (Dasarathy, 1994):

**Table 2.1 Dasarathy's Input/Output Data Fusion Model (Dasarathy, 1994)**

| Name | Description |
|---|---|
| Data Input/Data Output | Input is filtered |
| Data Input/Feature Output | Generation of features from input data |
| Feature Input/Feature Output | Input features are fused and new features are generated |
| Decision Input/Decision Output | Input decisions are fused together to give a final output decision |
| Feature Input/Decision Output | Input features are fused together to give output decision |

### 2.1.3 Advantages of Data Fusion

There are several advantages of fusing data from sensors, however, as stated above, data fusion is not always the best solution and the process should be weighed against other factors. Factors to be considered include:

- Improving the performance of a system by providing more information than a single sensor would normally provide. For instance, in a camera surveillance system, a single roving camera while scanning a particular area is bound to have a lower coverage within a time frame than a system of multiple cameras.
- Reducing false positives and false negatives. False positives are otherwise known as false alarms. They occur when a system detects a problem when there is none. However, false negatives are more dangerous because they do not detect when there is something to be detected. False positives can be managed but will in time reduce

operator confidence. By increasing the number of sensors/ sensor data, the system becomes less prone to false alarms and will be more accurate thus reducing or eliminating false negatives.

- Sometimes, data fusion involves fusing data from different types of sensors. This thesis for instance, will be fusing data from sensors that are individually set to achieve different purposes. One sensor will search for illegal drugs and another will search for contraband and human beings. Individually, the sensors will achieve only a specific task. Fusing them together gives a more complete and balanced purpose and a wider task is achieved.

In addition to the above, Thomopoulos (1989) gives further advantages of data fusion including:

- Higher signal-to-noise ratio;
- Information regarding independent features in the system can be obtained;
- Increased robustness and reliability in the evident of sensor failure;
- Increased dimensionality of the measurement;
- Improved resolution;
- Reduction in measurement time, and possibly costs - there is a trade-off to consider in this issue. Thus, an optimal number of sensors to extract the required information from a system should be ideally pursued.

- Increased hypothesis discrimination with the aid of more complete information arriving from multiple sensors;

### 2.1.4 Applications

In the past, data fusion has been mainly utilized in the area of defence. However, in recent times, the application of data fusion has spread over to non-military areas. Some examples include (Hall, et al., 2009): Monitoring of manufacturing processes, environmental monitoring, robotics and medical fields.

#### 2.1.4.1 Military Applications.

In 1986, the Joint Directors of Laboratories (JDL) came together and introduced terminologies related to data fusion. This was necessary because of the lack of a unifying terminology stood as a barrier to technology transfer in data fusion. Within the U.S. Department of Defence (DoD), sensor fusion falls in the same category as the overall definition of information fusion, as specified by the Joint Directors of Laboratories (JDL). The JDL's definition of information fusion includes four levels, ranging from identifying and tracking targets of interest, to determining whether these targets are threats.

Specifically, military applications include (Hall, et al., 2009):

**Table2.2 Data Fusion applications for Military Systems (Hall et al., 2009)**

| Application | Purpose |
|---|---|
| Ocean Surveillance | Detection, tracking, identification of targets. |
| Air-to-air and surface-to-air defense | Detection, tracking, Identification of aircraft |
| Intelligence/surveillance and target acquisition | Detection and identification of potential ground targets |
| Strategic warning and defense | Detection of impeding strategic actions, detection and tracking of ballistic missiles and warheads |

### 2.1.4.2 Non Military Applications

In addition to military systems, data fusion also has non-military applications. In recent times, the study of data fusion techniques has received high interests in this area. Sample applications are (Hall, et al., 2009) shown in table 2.3.

**Table 2.3 Application of Data Fusion in non military systems [4]**

| Application | Purpose |
|---|---|
| Environmental Monitoring | Used basically in identification of natural phenomena like weather, earthquakes, etc |
| Medical diagnoses | Identification of tumours, abnormalities and disease. |
| Robotics | Object detection and avoidance |
| Multimodal biometric Systems | Detection and identification of traits to uniquelydefine a human being |
| Surveillance | Detection and tracking |

## 2.2 Cargo Screening

The Oxford English Dictionary defines cargo as

car·go / ˈkärgō/ • n. (pl. -goes or -gos) goods carried on a ship, aircraft, or motor vehicle: *transportation of bulk cargo* | *a cargo of oil* (Oxford, 2011)

These goods are transported over the border enclosed in containers. Goods are transported for various uses. People moving from one country to another have to transport the bulk of their personal belongings in these large containers.

The sizes for a cargo container according to the International Standard Organization vary by shipping and air freight containers. For shipping containers, there are five common standard lengths: 20-ft (6.1 m), 40-ft (12.2 m), 45-ft (13.7 m), 48-ft (14.6 m), and 53-ft (16.2 m). Container capacity is often expressed in *twenty-foot equivalent units* (*teu*). An equivalent unit is a measure of containerized cargo capacity which does not consider the height of the container and is equal to one standard 20 ft (length) × 8 ft (width) container (Emase, 2007).

The maximum gross mass for a 6.1 m dry cargo container is 30,480 kg, and for a 12m including the 2.87 m high cube container, it is 34,000 kg. Allowing for the tare mass of the container, the maximum payload mass is therefore reduced to approximately 28,380 kg for 6.1 m, and 30,100 kg for 12 m containers (Emase, 2007).

A picture of a typical container is shown in Figure 2.1:

Figure 2.1 A typical standard sea container (Transit Logistics)

Cargos represent a security issue to most countries especially developed nations like the US and the UK especially after the September 11 and July 7th terrorist attacks on both nations respectively. The challenge therefore, is for border security agents to be able to scan these cargos for illegal goods, contrabands and illegal immigrants.

### 2.2.1 Cargo Screening in the UK

Cargo screening refers to non-destructive methods of inspecting and identifying goods in transportation systems. In the UK, it is the responsibility of the Home Office and the UK Border force to make sure that cargo is checked for illegal substances.

Hitherto, sensors have been used individually to detect for drugs, contraband, stowaways and explosive elements in cargo. *The novelty of this*

*research is in the use of data fusion for fusing data from a single sensor for the detection of cocaine to provide a more effective and a more reliable result.*

Key to many data fusion systems is the User interface. The user interface provides a means for which an operator interprets results from a data fusion system. This thesis also discusses a user interface that was developed for this project and will be outlined in subsequent chapters.

The Oxford English Dictionary defines contraband as “goods that have been imported or exported illegally” (Oxford Dictionary 2005). Goods which are illegal to possess, such as stolen materials, are also called contraband. However, this thesis is focused on contraband as goods which are imported illegally. Consequently, contraband goods will vary from country to country as what may be illegal for importation into the UK, for instance, may be legal in another country.

### 2.2.2 Existing Technologies in Cargo Screening

#### 2.2.2.1 Background

The issue of border security has always been a problem for most countries. As early as 1904, in the United States of America, Mounted watchmen of the U.S. Immigration Service patrolled the border in an effort to prevent illegal crossings. These were called Mounted Guards and they patrolled the borders at El Paso, Texas. However, in March 1915, the US Congress authorized a separate group called the Mounted Inspectors which was a

separate group from the Mounted Guards and some of whom rode on horseback while others rode cars and some patrolled in boats.

Over the years, with the advances in technology, a number of technologies have been developed for screening illegal substances. As technologies are developed, the offenders became even more sophisticated and constantly seek new ways of beating new controls. This threat has increased in various borders world wide. In the UK, for instance, the UK Border control stopped over 21,700 people trying to cross the Channel illegally between April and December 2008 (UKBA 2009). It should be noted that over 809,000 frieght vehicles were searched to make the above discovery (UKBA 2009). Also, within the same time frame, officers seized in excess of 800 million cigarettes, representing a potential loss of £149 million in tax revenue, £260.6 million worth of prohibited drugs (UKBA, 2009). The police, border agencies, immigration and other security agencies involved in screening have more responsibility in constantly improving methods of detection. More importantly, screening technology must be improved to be more effective in terms of accuracy of detection and speed.

Current technology in use can be grouped in two (2) categories – Imaging and Non-Imaging technologies.

- *Imaging Technologies*

These include X-ray, Gamma ray and Neutron Technologies. The challenge for X-ray technology is in its ability to maintain a balance of being dense enough to penetrate the densest cargo while not being dense enough to

cause health issues. There is also a challenge of image quality. A sample of X-ray image is shown in Figure. 2.2.

Figure 2.2(a) X-Ray Ganter scanning a vehicle (side view)

Figure.2.2(b) X-Ray Ganter scanning a vehicle (rear view)

Gamma ray technology has a lower radiation field when compared to a similar X-ray technology, thus providing a smaller safety exclusion zone (Neumann, 2008).

The Neutron system creates gamma-ray signals when it interacts with the elemental ingredients of the inspected object. The gamma-ray energies are unique to the elements in the inspected object. If the gamma-ray signatures match those in a threat database, the system automatically alarms indicating the possible presence of the threat.

In all, imaging technology depends much on the quality of the image and the penetration of the rays used. Their disadvantages however include their size (see figure 2.2), and the time delay caused by analysing cargo using this technique. The use of highly dangerous radioactive materials also makes it

important that a safety exclusion zone must be created when the system is in use thus adding to the already large space needed.

- *Non-imaging Technologies*

Non-imaging technologies as the name depicts do not require a computer rendering of the internal contents of a container. In this case, the observed object is scanned and an alarm is given if the object's features match signature features stored on the database. Examples include dogs and vapour phase chemical detection systems.

The use of dogs as drug detectors can be termed as a non-imaging technique. Trained K-9 dogs are trained to detect certain drugs using their ability to detect even very faint scents (Marks, 2007). These dogs scent individuals and luggage as they pass through security at borders and let off an alarm by barking whenever it detects a scent that matches the scent he is trained to detect (see Figure 2.3).

The challenges here are obvious. Dogs and human detectors may become tired and thus less effective over time. Repetition of the same duty may lead also to boredom leading to the same consequence. This will lead to a significant number of false negatives, allowing for illegal substances going undetected.

Also for gamma resonance technology which is used for detecting explosives, the Gamma Resonance occurs when the energy of a gamma beam is precisely tuned to coincide with a nuclear excitation level in a

nucleus of an element of interest. Similar to the imaging technology counterpart, they are usually too large. Some of them are mobile and can be transported from point to point but locating them at the border is a major challenge. Also, the situation of exposing operators to likely dangerous materials is also an issue.

Figure.2.3 Customs Inspectors using specially trained dogs to sniff out drugs and other contraband (courtesy US Customs Service)

Other non-imaging techniques that exist include:

**Vapour Phase Chemical detection Systems**

Olfactory sensing is the means by which sensors are developed to mimic the human nose in the detection of substances usually in based on their signature 'scent'. It is a complex but very specific system which has attracted the interest of researchers over the years (Stubbs et al., 2005). The applications of this type of sensors are diverse including in areas like medicine, anti-terrorism, environment and biotechnology.

Generally, there are two main features that characterise a vapour sensing detection system – sensitivity and specificity (Stubbs et al., 2005). The specificity is defined by the nature of the 'chemically sensitive' film on the surface of the system while the sensitivity is defined by the sensor's modality and design.

Usually, the aim for many researchers in this field is the quest to develop commercially viable and portable systems which would take the place of trained dogs as is being used today (Figure. 2.3). As can be seen in Figure 2.2, dogs are trained and used at ports to sniff through baggage for contraband mostly explosives and illegal drugs. However, as mentioned in the previous section, there are issues which arise in the use of these dogs. The cost of training and caring for the dogs and some other issues has been previously discussed. For cocaine detection, Stubbs et al (2005) suggest that because it is still unclear what the dogs are actually detecting which may vary depending on what dog is used, the response of dogs to a cocaine sample may thus also defer. Hence whilst dogs have been shown to be effective in the detection of illegal compounds, their effectiveness is limited by the issues raised and by the inability of researchers to perfectly analyse the system behind their process of detection.

#### 2.2.2.2 Existing cocaine detection sensors

Much research has been done in the area of developing commercially viable portable sensors for real time, on-site detection of a range of compounds including cocaine due to the potential numerous applications of such sensors

including healthcare, environmental monitoring and so on (Yu & Yi, 2011; Daar, 2002; Fan et al., 2005). The key phrases that should be noted in the above are “portable” and “commercially viable”. Montagnana et al., 2009 and Yu & Yi, 2011, have postulated that the Personal Glucose Meter (PGM) (Figure 2.4 below) is arguably the most successful of such sensors. However, (Yu & Yi, 2011) stated that one of the obvious short comings of the PGM is that it has only one target. They then went on to develop a novel methodology which extends the ability of the PGM from a single target to the detection and quantification of a wide range of targets including biological cofactors, such as adenosine and of course, contraband drug, cocaine. The proposed extension work uses functional DNA-conjugated invertase to link the detection of glucose to that of other targets such as cocaine and at the same time, use the concentration of glucose to estimate that of the targets of interest.


Figure.2.4 A simple Personal Glucose Meter (PGM)

The success of this research was based on its ability to detect more targets in addition to glucose. However, despite its low cost, the “extended – PGM”

cannot be used in cargo screening due to the 'pre-processing' that needs to be done and the method by which the analyte is passed on to the meter[5].

Another attempt at developing a cocaine sensing device was made by (Stubbs et al., n.d.). Using an ST-Quartz resonator with centre frequency of about 250MHz, the researchers were able to detect a shift in the transient frequency alongside a baseline frequency shift. They claimed in their research that they were able to achieve real time molecular detection of cocaine molecules using their anti-benzoylegonine coated sensor. Again, while this may seem interesting, its application to cargo screening does not show any advantage over any of the current techniques discussed above.

*Challenges*

The preceding section has shown the various attempts at developing portable sensors that can detect a range of compounds and in particular cocaine and their limitations also highlighted. Most importantly, the issue of how effective in terms of detection rates and reliability of the detection system are major areas of concern.

Although there are different approaches to cocaine detection as discussed above, the commercially available systems use a test strip to test for the presence of cocaine based on the reaction of the compound on the strip. They are known as immunoassay tests (Crouch et al., 1998; Concheiro et al., 2007). The shortcoming of this test is that it may need preparation of a

[5] The 'extended PGM works by using a specific custom designed analyte reagent and then mix with a little amount of sugar which is converted into glucose. The PGM is then put into the vial and it goes on to measure a target compound.

sample of the compound before testing thus requiring that the staff be trained to be able to use and interpret the result. Another disadvantage of the staff having to use the test strips especially in the area of cargo screening is the time expended in swiping the material with the strip and then moving to test for the detection of cocaine.

To add to the cocaine detection short comings, traditional cargo screening methods also inhibit the flow of passengers, baggage and cargo mainly due to the size of the systems. This, in turn, adds a high price to operations in terms of added costs, overheads, delays and lost business. Thus, there is a pressing need to make the screening process more reliable, effective, efficient and less intrusive. There is also the challenge of having to offload containers for full inspection and inspectors having to go into containers thereby exposing themselves to contaminants which may be in such containers.

Unfortunately, due to the challenges raised above, only a very small percentage of suspect items are thus inspected leading to vulnerability of borders. Thus, a new, cost efficient, efficient, reliable and fast means of detecting illegal substances are required.

This thesis which introduces techniques in the detection of cocaine and interpretation of results aims at reducing the checking time. Using a sensor fusion approach, necessary features will be extracted from newly developed cocaine sensors and the features will be combined using data fusion

algorithms to lead to more reliable results than individual features. The expected advantages of this approach are

(1) By depending on more than one feature, the reliability of the system is improved upon, with lower false positives and higher detection rates and

(2) Updating the system with information from past data also adds to the reliability because the system can now learn from past experience.

Its application will be in the cargo screening at ports and borders but can be applied also in the areas of medicine, environment, etc. It is also expected to be more accurate in terms of detection rate.

For the overall project, by carrying the sensors to the containers rather than working remotely like current systems do, safety of border patrol personnel is guaranteed and sensitivity of the sensors need not be too high while still expecting better results in terms of reduced false negatives and false positives.

## 2.3 Cocaine detection at Borders

Illegal substances smuggling is increasingly becoming a challenge for border agency officials. Cocaine is one of the illegally smuggled drugs over the UK borders. Chapter one of this thesis has highlighted some high profile cases of cocaine smuggling around the world. One common way of smuggling in cocaine over the borders is by a method called body packing where the drug courier dangerously inserts well packed cocaine packets into their rectum or vagina or by swallowing them (DeMarco et al., 1999) (Hergan et al., 2004).

This thesis is focused on the smuggling of cocaine via the sea and land borders. Using sea and land borders, Cocaine is also smuggled via cargo containers (see fig2.1 above for the picture of a typical container). Using this means, smugglers sometimes hide the cocaine in wax (Jellema, 2011), beer or milk cans (Anon., 2011), or even baseball caps (Anon., 2003). To combat the threats of smugglers, some of the techniques used have been discussed in Chapter one. The various techniques used are as follows:

Figure.2.5 Cocaine hidden in baseball caps (courtesy the United States Department of Justice Drug Enforcement Administration)

Back- scatter technology uses computer algorithms to develop an outline of a container and displaying on a computer screen any item inside the container. A trained technician is needed to detect which of the items displayed on the screen is suspected to be illegal drugs. Due to the fact that this is mainly a visual system, it is almost impossible to decide whether the concealed item is cocaine or not, without further testing.

Figure 2.6 Back scatter technology revealing smuggled cigarettes (in red box). (Image courtesy American Science and Engineering, Inc.)

Detection of cocaine using this technology is in two stage. The first stage consists of the container passing through a ganter (see Figure 1.2 in Chapter one) and the x-ray image of the container displayed on a screen. Once the border agency official suspects packed substances may contain cocaine, he/she then recommends the container for further inspection. Further inspection involves unloading the container and passing the suspected item/items through further tests.

One can see that one major challenge with this is that detection is largely dependent on the border agency official. The operator needs to be able to make an informed decision based on what he/she can see. Any oversight will lead to false alarms and missed detections. Additional drawbacks include the size of the detectors, availability and costs (Meijer & Bots, 2003)

In addition to the above, there is also the health issue. In addition to the smuggling of cocaine, smugglers also try to smuggle in stowaways. Although the Transportation Security Administration (TSA) contends that the levels of ionizing radiation emitted by approved X-ray back-scatter technology is well below levels considered safe for human exposure (TSA, 2011), some

researchers have disputed this claiming that although the radiation dose received from the system would be safe if distributed throughout the body but because it is concentrated only on the skin and underlying tissue, the dose to the skin may be “dangerously high” (Sedat et al., 2010). Therefore, in terms of safety, this technology was proven to be unsafe for use at borders.

Another method in the detection of cocaine at borders in Cargo containers is in the use of high speed gas chromatograph with surface acoustic wave sensor (Staples & Viswanathan, 2008). This procedure uses a single, uncoated, high Q surface acoustic wave sensor, along with a high-speed chromatograph and column, a programmable gate array microprocessor, and a vapour pre-concentrator.

In some cases, the cocaine contraband is dissolved in liquids and extracted once it has reached its destination. In cases such as this, current scanning methods involve the use of an immunologic test using a drug-test panel. These panels use a random sample of the cargo contents as a control sample and this is opened and the test performed on it. Unfortunately, there is a possibility of missing the boxes containing the drugs because they are well hidden amongst legitimate boxes. As a result, there are high cases of false alarms and missed detections in cases like this as mentioned by border officials during the author’s visit to the site.

All the above show the drawbacks in the detection of contraband at borders especially in cargo containers. The drawbacks stem from over reliance on the ability of border officials to make the right judgement based on visuals from x-ray pictures and/or failure of existing systems to give reliable results in

terms of false alarms and true detections. Another drawback is the intrusive nature of existing systems. It is thus imperative that a non-intrusive system that will lead to reliable high true positives and low false positives while delivering real-time solutions is required. In addition, a well-designed user interface that increases operator confidence is needed. This thesis implements a two-step methodology to implement a data fusion algorithm for the detection of cocaine using a fibre optic sensor. The sensor developed and the methodology are described in Chapter three of this thesis.

## 2.4 Sensors

A sensor can be described as an instrument that detect or measure physical phenomena. Specifically,

“Sensors are devices that convert a physical parameter such as room temperature, humidity, smell or wind speed into a signal that can be measured electrically or sometimes, visually (e.g. visual output from a glass thermometer)” (Waltz & Llinas, 1990).

Many sensors are typically just data extracting systems. They can either be active (laser fluoro-sensor, radars, x-ray machines) or passive (cameras, x-ray detectors) with the latter simply observing emissions from a target and the former provides its own energy source and emits this energy to induce a detectable phenomenon from an observed target (Hall, 1992). Due to the fact that active sensors do not depend on an external source of energy, they have the advantage of being able to gather measurements at any time of the

day regardless of the energy from for example, the sun. The fibre optic sensor used in this thesis is an active sensor.

Due to the advances in sensor technology and the subsequent increase in funding for sensor systems there has been an increase in the volume of data in sensor systems. Research shows that this volume will continue to increase due to the increasing interest in sensor technology. According to researchers at ON World consultancy, sensor sales have been growing "well over 50 per cent a year" for three years, and will keep growing at least that fast for the next few years. In fact according to the San Diego-based consultancy, by 2012, ON World projects, sales of wireless sensor networking equipment in just four areas – industrial, commercial building, advanced metering infrastructure and residential applications – will total $14 billion (Smith, 2008).

Applications of sensors are now seen everywhere around us. Temperature sensors, light sensors, are all a part of our everyday lives. The human body is a combination of several sensors (sensory organs) each working independently and also in conjunction with other sensors (sensory organs) to detect or track objects.

The eyes for instance can detect the size, shape and colour of an object and can track the movement of objects too. The nose, another human sensor, can ‘smell’ the odour coming from different objects sending out a smell (signal) while the skin can detect how hot or cold (temperature) an object is.

Similar to every sensor, human sensors work based on the signal sent out by objects they are meant to sense. If the signals are not strong enough (e.g. wind direction taking the smell of food away from the position of the person) or the strength of the sensor has been weakened (e.g. a short sighted or long sighted person), then the accuracy of detection/tracking is reduced. In most cases, the human body adapts to this problem by using additional sensory parts to complement the effort of one. For instance, if the nose cannot detect what a substance is from its smell, the person can touch to feel or taste to determine what it is. Sometimes in drug (cocaine) detection, human detectors combine the use of the taste buds and touch to detect the drug. The coordination is performed in the brain.

Man-made sensors work using basically the same principles. They are manufactured to 'sense' (detect/measure) a physical phenomenon. Thermometers measure the temperature of a body using heat variations, potentiometers detect change in voltage and so on and they can measure either directly or indirectly. When sensors are not needed to make direct contact with the object, it is called "remote sensing" (Hall, 1992). Examples are in satellite imaging, thermal imaging, etc. Other sensors however, need physical or near physical contact with the detection phenomena (Waltz & Llinas, 1990). Examples include switches, thermometers, transducers, etc.

In plain terms therefore, it can be seen that for every sensor, there is a parameter that would be measured; there is an output for every input. This is explained as the system's transfer function which gives us the relationship

between the input and output (Fowler & Schmalzel, 2004). The input into the sensor is the parameter to be measured also called the measurand. Some measurands can be measured by more than one sensor and the sensor used depends on the property of the material and the system. For example, in a temperature measuring system, the bi-metallic strip, the property that metals expand and contract at changes in temperature is used. The two metals have different linear expansivity values and thus expand and contract in different direction when the temperature (measurand) heats or cools. In the clinical thermometer, mercury is the material used. It expands when temperature increases and contracts when temperature decreases. The readings for both are read off a standardized scale.

Properties of sensors to be considered before using one for a project includes the following (Waltz & Llinas, 1990) (Fowler & Schmalzel, 2004):

- Sensitivity – this is the smallest change in input that will yield an output. If a small change in the measurand can be measured by the sensor, then it is said to be very sensitive. If the reverse is the case, the system is said to be 'not sensitive'. The sensor's insensitivity to small measurand changes is also used to define its 'robustness'.
- Repeatability – this tells how many times a sensor can measure the same value and give an almost the same value at all times.
- Threshold – this is the maximum and minimum values outside of which the sensor will not give any value. For instance, in many

thermometers, the minimum temperature is $0^{o}$C (32F) and the maximum is $100^{o}$C (212F).

- Noise – this includes all additional 'unwanted' data which may be from the environment, the user or from the sensor itself.
- Accuracy – how much does the measured value deviate from the true value of the measurand? Does the sensor offer the accuracy required by the application?

In addition, the economics involved in terms of cost is also an important factor to consider. If there are more than one sensor offering the same qualities, then the cost can be a deciding factor in making a choice.

### 2.4.1 Sensor Data

The sensor data is an important aspect to be considered in multisensor data fusion since it serves in one form or the other, as the input to a data fusion system. In a multisensor data fusion system, there are three main sources of data inputs:

1. The sensors themselves
2. *a prio*ri data stored in the database
3. the inputs in form of commands by the users.

In an example of a typical active sensor, the sensor may emit energy such that a certain phenomenon is detected in a target. Examples of this are radars that send out short, high-intensity burst of high-frequency radio waves and receive the echo. It then uses this time to determine the distance of the object (example: ship or airplane). The radar can also measure the Doppler

shift very accurately and determine the speed of the airplane (Hall, 1992). Basically, in moving from input (sending out waves and receiving echo) to output (determining the distance and speed of plane), there are several steps in between. These are outlined in figure 2.5:

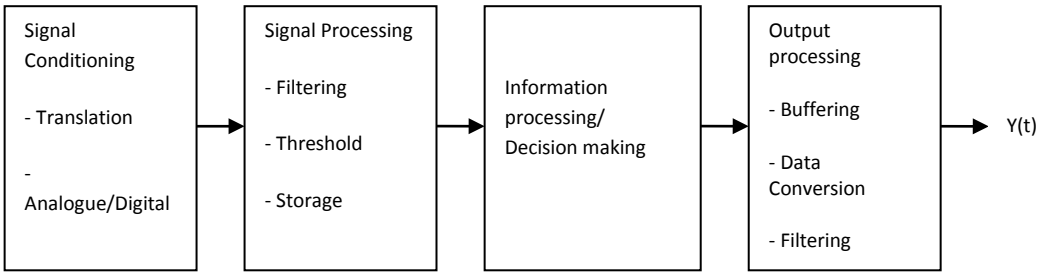



Figure 2.5 A sensor architecture (Hall, 1992)

Using the radar as example explained by Hall, when the radar sends out radio waves and receives the echo, it then goes on to produce an output showing the distance and (if required) speed of the airplane. To do this, as shown in the diagram above, the first step is a pre-processing stage which involves conditioning the signal so that it is referenced to a particular frequency. It is important to align signal at this stage so that all signals are at the same domain.

#### 2.4.1.1 Signal Processing

When a sensor receives energy, it receives not only the energy of interest, it also receives other forms of energy depending on the sensor. The signal processing stage isolates the energy of interest and may also include

transformation from time to frequency domain and background noise removal.

#### 2.4.1.2 Information Processing/ Decision making

Information processing involves the use of classification and pattern recognition techniques to identify patterns in the signals. It is also at this stage that a decision is made on the identity of the target. For instance, in the radar example, the sensor determines if the target is an airplane, bird, etc based on the shape and size detected. The decision is based on a comparison of the detected information with heuristic data stored in the database.

#### 2.4.1.3 Output Processing

Once the decision is made, the result is output. The processing at output level may include buffering, data unit conversion, transformation smoothing and filtering. For the radar example, the position and velocity of the plane may now be estimated. According to Hall[12], the form of the output of a sensor can be any of the following:

- A continous waveform (amplitude, frequency or phase versus time)
- A vector consisting of parametric positional data, target state data, or straight forward declaration of the target identity
- A 2-D image consisting of image coordinates and spectral data.

For the fibre active sensor used on this dissertation, the form of the output consists of intensity count data against the wavelength pair. The sensor also

simply using a light source, directs light towards the sample compound and based on the signature of the compound, a set of output data is generated.

Generally, there is no single perfect sensor. No single sensor can accurately detect, locate and identify targets under all conditions. Just as the sensors in the human body have their strengths (nose – smell, tongue – taste, eyes – sight, etc), man made sensors also have their strengths. Some sensors are accurate at detection, some are best suited for locating and tracking. Also, the environment which the sensor will work will also affect its performance. Therefore while some sensors perform very well in the atmosphere, some may not give accurate results. In addition, just as some human sensors require further processing of features to give an intelligent output, some sensors also need to extract features and fuse the information from these features to make more informed decisions, hence sensor data fusion.

## 2.5 Sensor Detection

A typical sensor is a device which makes observations by taking measurements of physical quantities such as temperature, angle, distance, using the mapping relationship which exists between the measured quantity and the state of nature to output the necessary information.

In this regard, the interpretation of sensor measurements and sensor environment is extremely important. However, physical descriptions of sensors (sensor models) are unavoidably only approximations owing to incomplete knowledge and understanding of the environment. This, coupled with the varying degrees of uncertainty inherent in a system itself and the

practical reality of occasional sensor failure, results in the lack of confidence in sensor measurements. Some researchers have claimed that despite advances in sensor technologies, no single sensor is capable of obtaining all the required information reliably, at all times, in often dynamic environments (Punska, 1999). One of the solutions to this is to use additional sensors to extract as much information as possible (Punska, 1999). Another solution would be to extract information over a period of time from the environment using the same sensor and then fusing the data so collected. In the case of the latter, multiple sensors of the same kind could be used to measure the same quantities. This is especially helpful in the case of sensor failure. In the cases mentioned above, the uncertainty is significantly reduced thus making the system more reliable.

### 2.5.1 Optical Fibre Sensors

Optical fibre sensors have become an important part of sensor technology and have found application in clinical, military and other areas. The advantage they have is the excellent ability to deliver light and ability to excite target molecules and also capture the emitted light from the target (Bosch et al., 2007). A fibre optic sensor is a silica glass or plastic optical fibre which in principle uses the principle of total internal reflection to transmit light. The fibre optic sensor produces a signal which is proportional to the concentration of a chemical to which the biochemical reacts (Bosch et al., 2007).

Fibre optic sensors have been used for the detection of several substances. In food technology for instance, it is used in the detection of bacteria in food samples. For example, it is used in the detection of foodborne pathogens (Morgan et al., 2006). Using a portable and automated fiber-optic biosensor called “RAPTOR”, Salmonella enteritidis bacteria was detected in food samples. The detection of these bacteria using this biosensor could detect $10^4$ cfu/ml of the bacteria in less than 10 minutes of the assay time. This time taken proves far better than conventional methods for pathogen detection which would normally take days (Morgan et al., 2006).

Also, in detecting bacteria in food samples, a fibre optic biosensor was used in detecting *Escherichia coli* O157:H7 in seeded ground beef samples (DeMarco et al., 1999). This biosensor worked on the principle of “a sandwich immunoassay using cyanine 5 dye-labelled polyclonal anti-*E.coli* O157:H7antibodies for generation of a specific fluorescent signal”. The biosensor developed detected *E.coli* O157:H7 to 3 to 30 CFU/ml in seeded ground beef samples (DeMarco et al., 1999).

#### 2.5.1.1 Fibre Optic Sensor for cocaine detection

This section looks at the Fibre Optic sensor in general, the operating principles and then delves into the fibre optic sensor developed and used in this research. Furthermore, this section will show how the sensor works for the detection of cocaine and the shortcomings giving a justification for this research.

An Optical fibre is a very thin (less than the diameter of a strand of human hair) glass made fibre which is typically flexible and transparent. Its main function is to transmit light from one end of a fibre to the other with minimal loss of signal using the total internal reflection (TIR) phenomenon. It consists of a core which is surrounded by a *cladding* layer made of dielectric materials. Optical fibres can be used in the fabrication of sensors (remote sensing) and also as light guides in medical applications. They are also used in signs, toys, alarms, and Christmas trees. Their main positives include low cost, small size and portability.

A typical Fibre-Optic Fibre sensor works by using absorbance measurements to determine any change in concentration of analytes that absorb a given wavelength of light (Bosch et al., 2007). There are three types of fibres used in developing a fibre optic sensor: (1) Step Index Multimode (2) Graded Index Multimode and (3) Single mode. For this thesis, a UV multimode fibre was used.

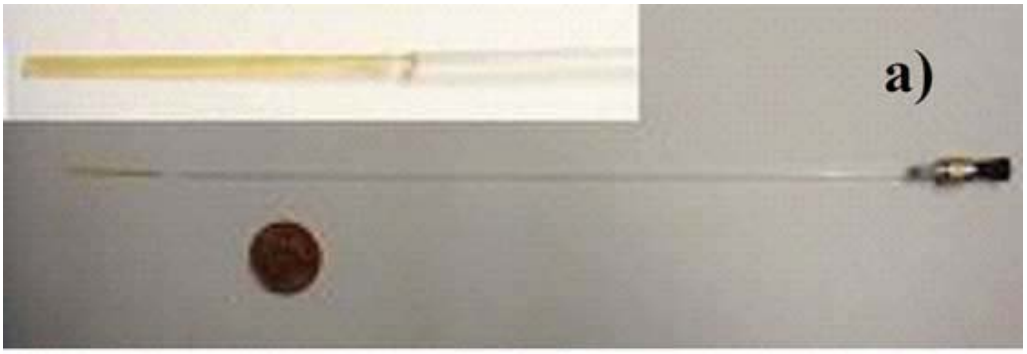



Figure 2.6. Cocaine probe prepared in this work showing the active distal end of the sensor

The aim was to develop an optical fibre based sensor for cocaine detection which would also allow a fluorescence enhancement response to cocaine. The focus was on cocaine in solution while targeting vapour detection.

The fibre optic sensor is based on a molecularly imprinted polymer (MIP[6]) containing a fluoresce in moiety as the signalling the presence and concentration of the analyte. The molecular imprinting is used as a method for generating chemically selective binding sites.

The resulting data is a pair of wavelength (nm) and number of counts and the graph is plotted. Though one can differentiate the presence of cocaine from other substances from the graph when the concentration of the substance is known *a priori*, it is not possible to do this without further analysis. For practical purposes, it is not always possible to know the concentration of the substances *a priori* and as such, there is need for a fusion algorithm which would use extracted features from the data as inputs into a model to determine the presence or otherwise of cocaine from a solution.

## 2.6. Need for research

As shown in section 2.2.2, existing cargo screening technologies have their short comings when it comes to cocaine detection. Some of them rely on the ability of the operator to be able to visually detect the substance. For others, the dangerous emission of rays when in contact with humans implies that their use is limited.

[6] MIP – Molecularly Imprinted Polymer

Also in section 2, existing current cocaine detection systems were discussed and their limitations highlighted. Some of the limitations include that many of these systems require that a swab be used to wipe the surface of container contents and passed through a cocaine detecting device (for example a trace detector). The time taken to wipe surfaces of all container contents implies that operators may randomly select contents to wipe with swab and thus there is a possibility of missed detection. In addition, there the false alarm rate in this case is high due to possible contamination of swab from external sources (the author was informed about this by operators during a site tour of one of the borders in the UK). This possible contamination will lead to false alarms and if the false alarm rate is high, it will in turn lead to a lack of operator confidence in the system. Perhaps one major downside of some of the current systems is that operators have to physically enter into containers to unload their contents thus exposing them to danger or possible contamination by harmful substances.

The optical fibre sensor provides one solution to cocaine detection. However, in the fibre sensor developed for this thesis, its shortcomings ensure that detection of cocaine cannot be achieved without prior knowledge of the concentration of the substance, a scenario which is not ideal in real life.

All the above show that there is therefore a need for a non-invasive system which will take the sensor to the container via a platform that will be remotely monitored by an operator. This system should provide high reliability with low false alarms and high true positives to ensure operator confidence. The drawback of the sensor also implies that a new method must be designed

which will not depend on prior knowledge of the concentration to determine presence or otherwise of cocaine. In this thesis, features will be extracted from sensor measurements. These features will be trained to individually test for cocaine and their results will be fused using models also developed in this thesis.

## 2.7. Chapter Summary

This chapter has presented a literature review of data fusion techniques. It should be noted that there are different ways data can be fused and the various ways were highlighted to include Data Input/Data Output, Decision Input/Decision Output and so on. In this dissertation, a decision input, decision output model is implemented. This chapter also shows the advantages of using data from sensors, applications with focus on applications in cargo screening. This section of the chapter concluded that data fusion helps to increase confidence by reducing uncertainty in measurement outputs. To serve as a form of background to the overall project, this chapter also looked into border security issues and current techniques used at borders to detect for contraband with a focus on several means used by smugglers to traffic cocaine via borders. The challenges are in two fold – the first is that the goods are stored in a container and the second is that usually, cocaine is usually discretely hidden within the goods. Current techniques are shown to be inadequate due to over reliance on the abilities of the operator which could to an increase in false alarms and missed detections. In addition, smugglers have continuously devised new

techniques in cocaine smuggling and some of them were discussed in the chapter. There is thus a need for a robust portable system aimed at increasing true positives and reducing false alarms. The subsequent chapters discuss the development of this system. In terms of sensors, this chapter also highlights a review of the principle behind sensor detection dovetailing into optical fibre sensors and then a short introduction to the optical fibre sensor for cocaine detection developed for this dissertation.

# CHAPTER

# 3

# REVIEW OF SENSOR FUSION ARCHITECTURES

This chapter discusses data fusion models and investigates possible fusion techniques based on the outcome of highlighted research in literature. Deciding on the appropriate method suitable for cocaine detecting fibre optic sensor with cargo screening application requires an understanding of the challenges identified in the previous chapter. For example, the D-S [7] technique has advantages over some other techniques when the application is trying to mimic human response and that is why it is used in applications like context awareness (Wu, 2003) and robotics (Zou et al., 2000) even though it may be computationally demanding (Koks & Challa, 2005).

[7] D-S or DST is the Dempster Shafer Technique is a generalization of the Bayesian technique which combines evidence from multiple sources and gives a belief function which takes into account all the available evidenced. It was developed by the works of Arthur P. Dempster (1968) and Glenn Shafer (1976) (Shafer, 1990).

However, before deciding on a technique, it is important to decide on where to combine the data as the design choice also affects the quality of the eventual output (Hall & Llinas, 1997). Researchers have come up with a number of data fusion architectures (Asheq, 2004) (Hall, 1992). Some of these architectures will be discussed in the following section.

## 3.1 Data Fusion Architectures

### 3.1.1 The JDL Architecture

A general architecture was prescribed by the Joint Directors of Laboratories (JDL) from the United States Department of Defence (figure 3.1 below) in 1986 (Estebani et al., 2004). The JDL model is shown to include four levels (or five levels since level zero was added in 1998 (Steinberg et al., 1999)) – The first level (level 0) involves the estimation of signal states. This implies identifying patterns inferred from sensor measurement; the second level (level 1) estimates and predicts the parametric and attributive states of the entity to be identified; the third level (level 2) is the situation assessment level and involves estimation of relationships among entities and the implication of these relationships for the state of the entities; the fourth level (level 3) is the impact assessment stage where the system performs a self-check to estimate the cost of signal, entity or situation states, given the system's alternative courses of action (Steinberg & Bowman, 2009); the fifth level (level 4) performs an assessment of all the remaining levels to rate their performance against expectations. Variants of the JDL model have been used in research.

Since the JDL model was introduced, various other architectures and algorithms have been proposed as data fusion methods to improve performances of sensor systems.

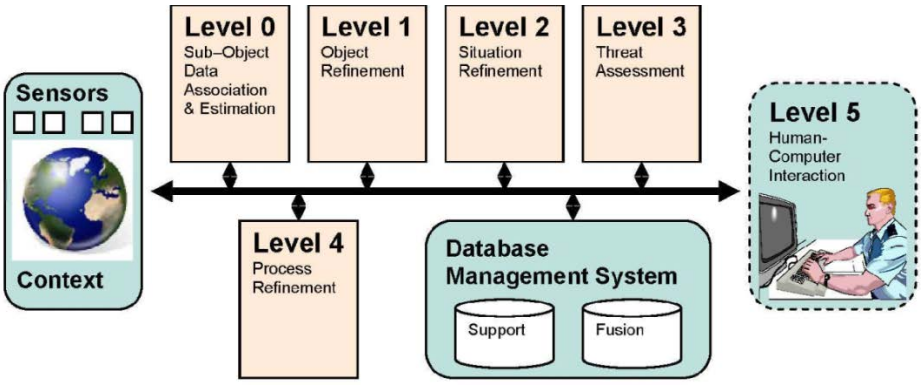



Figure.3.1(a) The JDL Data Fusion model

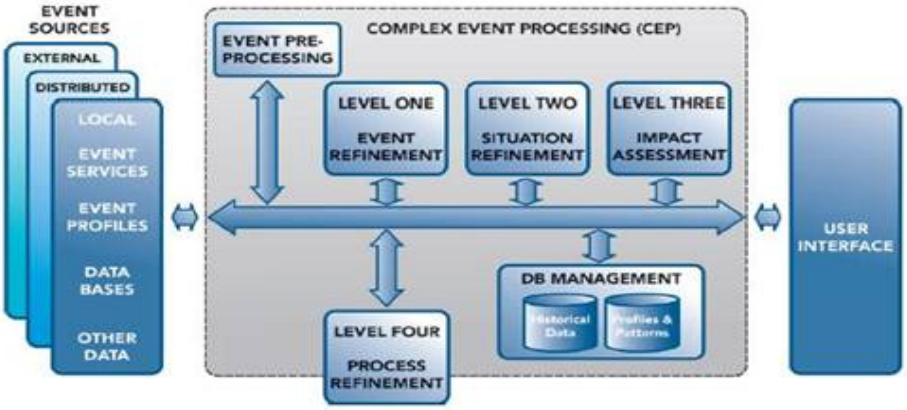



Figure.3.1(b) The revised JDL model (Estebani et al., 2004)

Figure. 3.1. JDL DF model (Steinberg et al., 1998)

Despite the attempt at making this model robust, there have been efforts in recent times to revise this model with the aim of expanding it to 'remedy some deficiencies' (Steinberg et al., 1998). It is described as a *paper* model which should not be used as a blue print for system design (Hall & Garga, 1999). Some presumed deficiencies of the JDL model is that it has a military focus (Steinberg et al., 1998). This explains why the JDL model's taxonomy

is filled with military jargons such as threat refinement, targets and so on. This made it difficult to associate data fusion concept with other applications. Steinberg *et al* tried to revise this by updating the taxonomy and replacing the levels with more robust and refined levels. The refined JDL model replaced the previous levels 1, 3 with the Event Management and Impact Assessment levels (Figure 3.1).

Besides this, another presumed shortcoming of the JDL model is that it only allows for a sequential ordering of the flow from level 0 to level 4. This strict adherence to processing flow does not allow for flexibility amongst levels. However, Bedworth and O'Brien claimed that this assumption is not correct (Bedworth & O'Brien, 1999). They claim that the JDL model was not intended to be strictly implemented sequentially from the first to the last level.

The revisions and development of further models show that the JDL model may not fit all data fusion applications. Even Bedworth & O'Brien (1999) admit that the JDL model is sometimes not appropriately implemented which may be due to its non-robust definition of levels and militarised taxonomy.

### 3.1.2 The Thomopoulos Architecture

The Thomopoulos architecture was proposed in 1989 (Thomopoulos, 1989) as a three-level architecture. These levels - signal, evidence and dynamics levels collate data measurements in such a way that the new set of data is integrated with prior data/information using a predetermined order (Velosos et al., 2009).

Signal level fusion – at this level, data measurements taken from a sample the measurements are processed and correlated with prior information from database using learning techniques.

Evidence level fusion – at this level, using a statistical model and decision making assessment specified by the user, the data from the signal level is combined.

Dynamics level fusion – it is at this level that a mathematical model is used for fusing the data at different levels of inference.

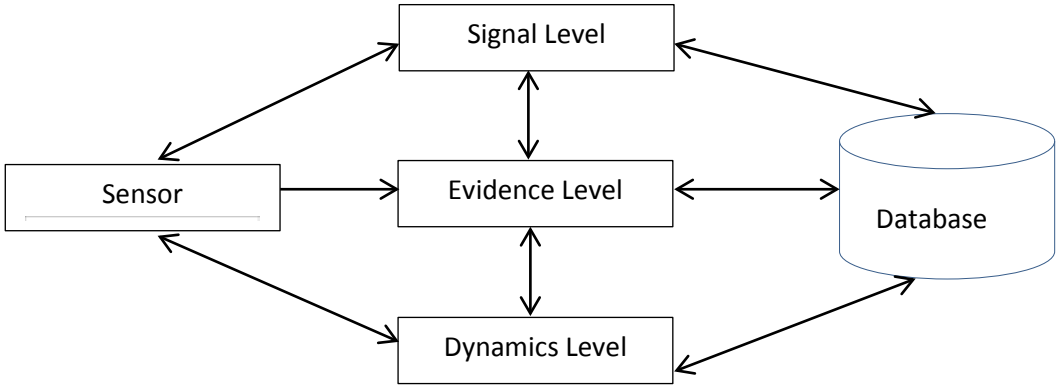



Figure 3.2. The Thomopoulos architecture (Thomopoulos, 1989)

The figure above shows the three levels as proposed by Thomopoulos and their interactions with each other and an important data fusion assessor – database. The three levels can work together either sequentially or interchangeably so for instance, in the case of cocaine detection, the signal level will represent new data measurements from sensor being correlated with information previously stored in the database while the evidence level will deal with the use of statistical models to make detection decisions based

on features. The dynamics level will involve the use of data fusion techniques to fuse the decisions made by the features.

### 3.1.3 The Omnibus Architecture

The Omnibus architecture (also called the omnibus model) was proposed by Bedworth and O'Brien in 1999 (Bedworth & O'Brien, 1999). This model similar to the Thomopoulos architecture, involves three levels – Observe, Orientate and Decide (OOD). The outline of these levels is as follows:

- Observe – this level like the name suggests involves measuring and gathering of data from the environment using sensors. It also involves processing of the data collated. The processed data is then passed on to the next level.
- Orientate – this level accepts processed data from the Observe level and fuses the data while extracting main features from the data using feature extraction and selection algorithms. The extraction of features helps to reduce the amount of data.
- Decide – at this level, the processed data from level 2 is presented to the human operator and then acts on the environment.

Unlike the Thomopoulos architecture, this model must be performed sequentially and forms a closed loop with a control module which is used for calibrating the sensors (Velosos et al., 2009).

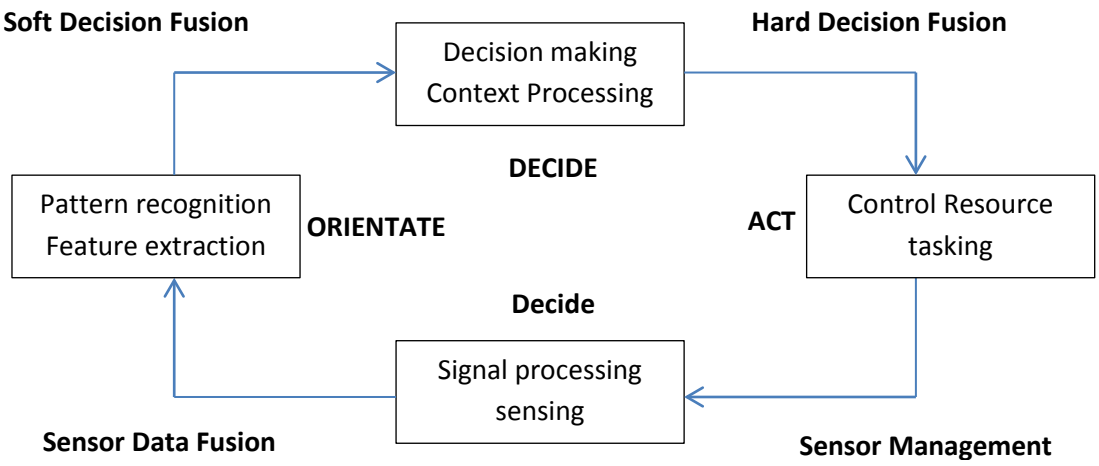



Figure.3.3. Omnibus model

### 3.1.4 The Waterfall Architecture

As proposed by Harris *et al* (Harris et al., 1998), the waterfall architecture is a stratified architecture involving three levels with each level encompassing some data processing steps.

Level one of the waterfall model is the signal level. This level involves the gathering of data from the environment using the sensors. The data is then pre-processed and the processed data and sensor information is passed on to the next level.

Level two involves using feature extraction and feature selection methods to extract and select features from the pre-processed data from level 1. These features are then fused using fusion techniques. This thus reduces the amount of data transmitted from the previous level.

Level three is called the interrogation level and this is where the situation assessment and decision making takes place. It uses processed information

from previous levels to create possible events and possible course of actions.

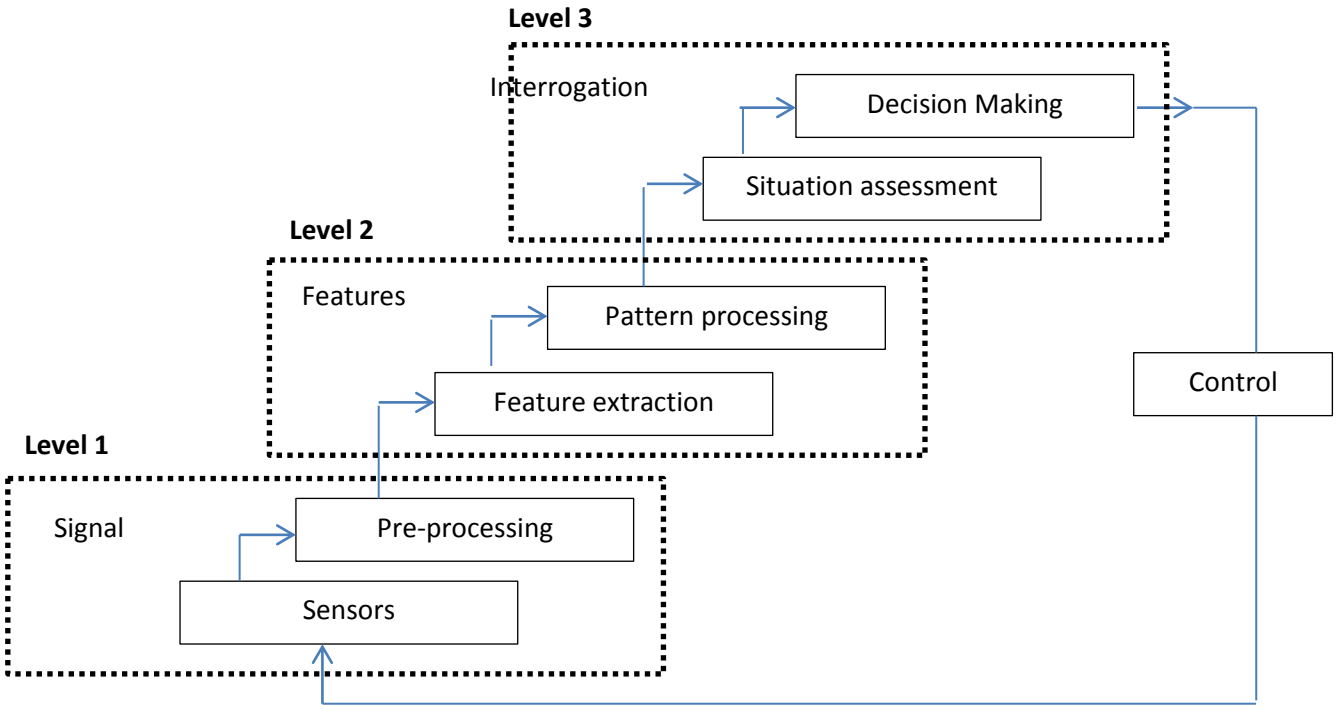



Figure.3.4 The Waterfall Model

Similar to the other models described, this model also has some similarities with the JDL model. The major drawback may be the omission of a feedback from this model (Bedworth & O'Brien, 1999) although this has been 'corrected' with the addition of a *control* stage which acts as feedback (fig.3.4). This does not however, make room for feedback in between levels. Usually, data fusion systems need iterative processes within levels to allow for updating of information before a final decision on identified target is made. Not providing a means for iteration within levels is a shortcoming of the Waterfall model.

In addition, the waterfall model suffers from the same criticism of the JDL model in that it is strict in the ordering of steps from level one to three. What is more, even within levels, the processes are strictly sequential so for example, the pattern processing must follow the feature extraction process.

### 3.1.5 Justification for a New Data Fusion Framework

The list of data fusion models highlighted above is certainly not exhaustive. They are however, attempts at providing a means to deal with multisensory data fusion issues. In most cases, they are constrained in the sense that their application is usually in situations where multiple sensors (similar or dissimilar) are involved. In situations where only a single sensor is used (as in this thesis), it is imperative that the models above be adapted to suit this need. To achieve, this, key features of the models and in general, of a data fusion model must be identified.

A cursory investigation of the models listed above will reveal that although they may differ in terms of implementation, three major levels are prominent in all investigated models. These are – Data collection, Feature extraction and Decision making levels. Within each of these levels, several processes may take place. In addition, acting as a form of complementary accessory is the database. The database stores information used in the fusion process which may include prior information as inputted by the operator, decision thresholds for features and so on.

It is therefore important that a data fusion process includes at least these three processes. These steps will form the basis of the new framework

developed in this thesis. Complementary steps within each process include data pre-processing which may take place after the data collection stage. This pre-processing involves cleaning of data to remove outliers and other forms of noise, normalization to align data in different forms into a common temporal and/or special space.

Depending on the application, the next stage should be the feature extraction level although in some applications if the data measure the same physical phenomena as in image fusion (Hall & Llinas, 1997). In the feature extraction and selection stage, the data is interrogated for unique features which will represent the raw data but have a lower dimension than the raw data. The features can then act as inputs into the data fusion system. In single sensor data fusion, once the features have been extracted, they go through a feature selection process to select an optimum number of features using established techniques like principal components analysis (PCA). The decisions of each selected feature can then be fused using statistical combination techniques.

## 3.2 Single Sensor Data Fusion

As mentioned in chapter two, data fusion can be performed on data either from single or multiple sensors. In the case of multiple sensors, it is known as multi-sensor data fusion. In the case of the former, it is sometimes called single sensor tracking or filtering (Koks & Challa, 2005).

In single sensor tracking, process of identity declaration follows a systematic process from the sensor output to feature extraction and then identity

declaration (Hall, 1992). Many of the architectures which apply to multi sensor data fusion can also apply to single sensor fusion. The main advantage in single sensor fusion relative to multi-sensor fusion is that in single sensor fusion, because the sensor measures the same entity and as such there is less computational challenge in terms of data alignment as in the case of multiple sensor fusion. Asides this however, single sensor fusion proceeds just as multi-sensor. The data output from the sensor can be used as input to a data fusion process or features extracted from this data output can act as the inputs. The levels at which data can be fused are described below:

**3.2.1 Raw Data level** (see Figure 3.5a): the raw data acquired from the sensor are fused directly if the data are of the same standards. If not, the data could be pre-processed before fusing. Data association is performed on the raw data to ensure that the data being fused measure the same substance. Once the data has been fused, features can be extracted from the new set of data generated. For example, in the case of detecting for cocaine presence, the raw data can be collected at two time intervals (from different areas of a cargo container) and the two sets of data are fused together to form a new set which would then be subjected to feature extraction and identity identification process.

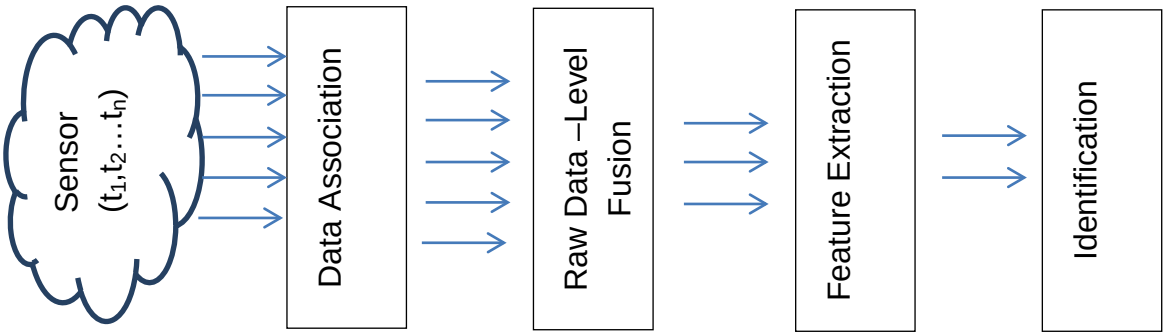



Figure.3.5(a) Raw Data-Level fusion

**3.2.2 Feature level fusion** (see Figure 3.5b)**:** a set of features may also be extracted from a data set and can then be fused to create a new set of features which can then be fused to make a decision on the identity of the substance. For example, in the identification of cocaine using a fibre optic sensor, features such as peaks, band size and so on can be extracted from the raw data and fused into a joint feature vector which will then represent the substance generated. Several feature extraction techniques are available and include neural network, cluster analysis, etc (Hall, 1992).

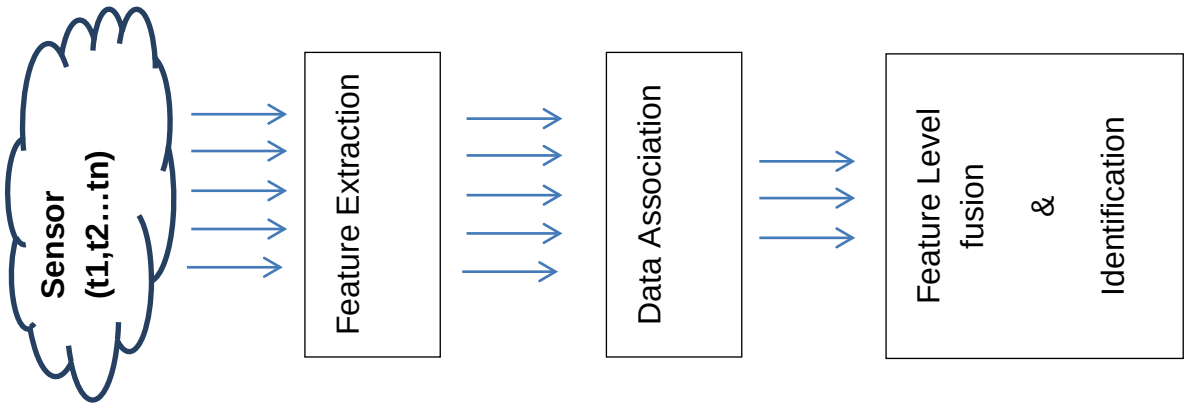



Figure.3.5(b) Feature level fusion

**3.2.3 Decision Level Fusion** (see Figure 3.5c)**:** At decision level, the sensor output data is analysed for features which then go on to form a

feature vector. The individual feature in this set is then used to make a decision on identity based on a pre-set threshold. This threshold is determined either by modelling the system or laboratory based experiments and then testing various thresholds based on sensor requirements. Once the decisions have been made, they are then passed through a fusion process. Common techniques used for fusion include Bayesian and Dempster-Shafer techniques.

In general, depending on the application, data fusion can be performed at any of these levels or a combination of these levels. Feature level fusion however has some advantages over others. In single sensor fusion where there is a vast amount of data in each set, fusing at the data level will cause problems due to the curse of dimensionality and may cause over fitting of the data. Extraction of features will help reduce the dimension of the data. It also implies less computational work and will improve processing time. These are the main reasons this was the selected method of fusion in this dissertation.

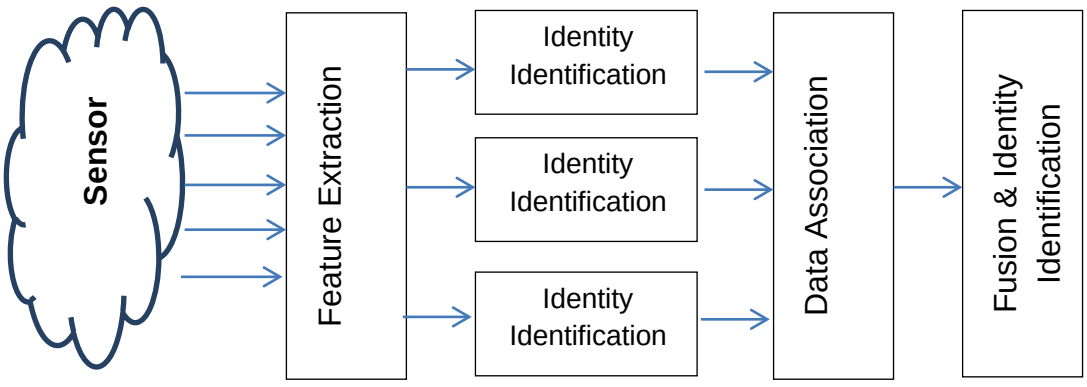



Figure.3.5(c) Decision Level Fusion

## 3.3 Human Computer Interaction

The design of a human computer interface affects effectiveness of a data fusion system (Hall, 1992; Waltz & Llinas, 1990). It provides the means by which the data fusion results are passed on to the operator whilst also giving the operator a means of inputting into the DF process. The human operator's input includes verification of data, inputting the *a priori* information and so on. In designing a user interface for a data fusion process, it is important for the interface to present to the user as much information as possible with care taken that this amount of information does not create a clutter on the interface. Usability criteria in the design of a user interface for DF systems are described in Chapter 7 of this dissertation.

### 3.3.1 Key Features

The user interface for the cargo screening ferret will be used for both robot control and manoeuvres and for displaying DF results. Key features of the user interface should include:

1. Visual aids: The operating environment of the robot will mostly be within cargo container. It is important for the operator to be able to view the environment real time from his remote position. The visual aid relaying feeds from a camera will make this possible.

2. User input: Prior information like details about the container under investigation should be inputted and stored by the operator via the interface.

3. Robot controls: A means of navigation for the robot should be made available on the interface

4. Cargo Summary: A summary of the container under investigation should be available on the interface. The summary should include information such as the name of shipper, origin of container, date, etc.

### 3.3.2 Key Considerations

This section will present high level considerations and guidelines in the development of a Human-Computer Interaction design. These guidelines form the background to the development of the User interface designed for this dissertation. As mentioned earlier, the user interface meets two uses – control of robot and presentation of DF results.

#### 3.3.2.1 Design

The design of as interface should take into the consideration the operator or user. The user's knowledge of the task is key and thus his strengths and weaknesses are important factors. The user in this case is the Border Force Officer (BFO) and his experience is based on the operation of different detection machines including x-ray ganders, trace detection machines and so on. The BFO is also experienced in the use of sniffer dogs as a contraband detection means. With the preceding factor in mind, and in terms

of design considerations, Farry (2009) recommends the following factors which should be considered:

1. User's perception of the user interface – it is important that the interface does not deviate too much from user interfaces the operator is used to. In terms of words, colours, icons, the operator must be familiar with these visual representations and where technical information is included; operator must be intimated with such via training or experience. This gives the operator a *feel* of the interface and makes it easy for his use.
2. Use of simple and easy to understand concepts – concepts and operations used in design must be simple and easy to understand and must not require expert knowledge.
3. The state of the system at any point in time based on the interface must be understood by the user.
4. The interface must provide clear feedback which should be in close proximity to the event that led to the feedback.
5. Robot location and effectiveness including current state of battery level must be provided on the interface. This can be done via video camera feeds and battery level icons on the interface.
6. Current information on successful detection must be provided in a clear and simple manner.
7. Control of the robot must be easy enough for anyone to manipulate.

8. Interface must be consistent both with other popular interfaces but also within itself.
9. Interface must be simple to operate. Not more than two keystrokes should be required for important operations.
10. Alarms indicating detection should be embedded on the interface. However, the scale should match the level of the problem. For instance, if there is a low probability of detection, it should still be reported but not with the same intensity as a high probability. This will give the operator the opportunity to make a decision on what to do (ignore/overlook the former).
11. The interface must provide error messages which in turn should be able clear and easy to understand. The error massages must also provide a simple explanation on how to avoid the error.

The steps above served as guidelines for the design of the Ferret Robot user interface. Additional guidelines including the interface developed are highlighted in chapter 7.

### 3.4 Chapter Summary

There are many different data fusion models and architectures. A few of these models are generic in the sense that other models can be extracted from these models. Some of these so called generic models were discussed in this chapter. There are also three levels common to majority of the models. The importance of these levels makes them important to data fusion architecture. In addition to these levels, there are also complementary

supports which aid the fusion process. The database is one of these supports and acts as a store for information such as prior probability and past data in time series sensor fusion. The chapter also includes a discussion on the levels at which data can be fused listing them as raw data level, feature level and decision level fusion. The furthest assessor on the right of the JDL model is the user interface. A discussion on the set of guidelines to be followed in the development of an interface for the ferret robot concludes the chapter.

# CHAPTER 4

# NEW FRAMEWORK FOR DATA FUSION

This chapter will outline the systematic development of a model for the detection of cocaine using a fibre optic sensor. The experimental set up and the process of collection of data and development of a generic model which can be adapted for different data fusion techniques will be discussed.

## 4.1 Data Collection

The process of fusion of data begins with the interaction of the sensor(s) with the environment and proceeds to the measurement and collection of data from sensor(s). Depending on the type of sensor, the measurements may need further processing before fusion can take place. In this dissertation, experiments were carried out in the laboratory at City University to collect measurements corresponding to the response of the fibre optic sensor to various analytes. An overview of sensor quality, experimental set up and results are discussed in the following sections.

### 4.1.1 Experimental Set Up

In developing this sensor, the following aims were targeted:

- Size and portability of the sensor
- Immunity to electromagnetic interference
- Resistance to chemicals
- Remote sensing capability
- Low cost and long shelf life

The fibre optic sensor fitted adequately with the above aims (Grattan & Meggitt, 1999) (Lee et al., 2001) (Haupt & Mosbach, 2000).

The preparation of the optical fibre only involved cleaning and polishing of the distal end multimode fibre.

The experiment was performed over a period of 3 months. Several tests were performed over this period to test for reproducibility. The tests were performed in a laboratory at City University, London.

The experimental set up is as shown in figure 4.1

The apparatus used are:

1. A light emitting diode (LED) emitting at a centre wavelength of 374nm used in exciting the material
2. Ocean Optics USB2000 spectrometer used for detection
3. A desktop computer
4. 2 by 1 Y fibre coupler connected using two multimode UV/Visible fibres
5. Sensor probe

The active region of the probe is at the distal end of the fibre. This region of the probe was inserted into the cocaine solution. Once this end comes in contact with the cocaine solution, the intensity (counts) for varying wavelengths is recorded and the graph of wavelength against intensity is displayed on the screen. The senor was tested on cocaine solutions with varying concentrations (from 0 to 1000 µM) in MeCN/$H_2O$*. All aqueous solutions were prepared using distilled water. The experiment is fully reported in (Nguyen et al., 2010).

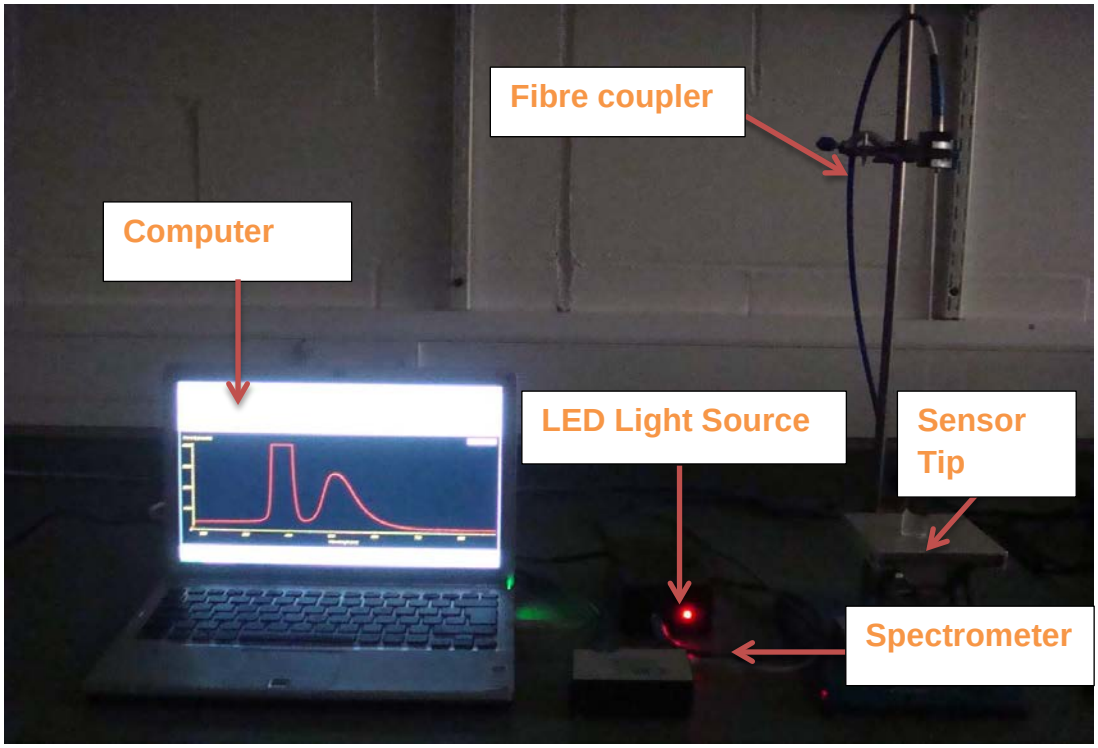



Figure. 4.1 Photo of experimental set up showing apparatus used. (Galbraith & Nguyen, 2009)

#### 4.1.1.1 Output

The output shows the response of the sensor probe to cocaine solution and is a graph of Intensity (counts) against the wavelength. The output graph is in two sections. The first section shows a high intensity value against

wavelength representing the light source and is not useful for the results. The second output is the cocaine fluorescence spectrum for the compound.

#### 4.1.1.2 Data Collection Results and discussion

The sensor was tested with different drugs asides from cocaine. The drugs codeine, ketamine, ecgonine methyl ester, amphetamine sulphate, buprenorphine.HCl were used to investigate selectivity of the sensor. The results showed that the sensor had lower fluorescence enhancing values for other drugs than cocaine (Nguyen et al., 2010). From the spectrum output, at similar concentrations, a study of the peaks shows that cocaine has the highest intensity followed by codeine, Amphetamine sulphate, Buprenorphine HCL, Ecgonine methyl ester and Ketamine in that order (Appendix B). The values of the intensity at those peaks vary depending on the concentration. However, the florescence enhancing values are calculated for drugs with similar concentrations thus, selectivity is dependent on prior knowledge of the concentration of the compound. When the concentrations of the drugs tested differ, detection of a particular drug using just the fluorescent spectrum cannot be achieved. To be able to detect for a specific drug (in this case cocaine) without prior knowledge of the concentration, this thesis will employ data fusion.

## 4.2 Data Fusion Algorithms

This section will review basic algorithms used in the literature for the fusion of data. It will focus on algorithms employed in this thesis – Neural Network,

Bayesian and Dempster Shafer. This review lays a foundation for the data fusion process undertaken in chapters 6.

Over the years, many algorithms have been developed and can be applied to data fusion. However, the common ones remain: Neural Network, Bayesian inference, Dempster-Shafer theory of evidence.

### 4.2.1 Classical inference

As a prelude to the Bayesian Inference discussed in the next section, this section will introduce the Classical Inference to serve as a background. The Classical inference method aims at validating a hypothesis at the expense on another based on empirical probabilities (Hall, 1992). It uses observed sample data to draw conclusions on an underlying distribution (Hall, 1992).

In simple terms, the empirical probability concept states that the relative frequency distribution of a long run trend of events is approximately equal to the probability. Mathematically,

$$P\{E_i\} = \lim_{\mathrm{k}\to\infty} \quad \left[\frac{K\{Ei\}}{K}\right] \qquad (4.1)$$

That is, for K trials of events, as K tends to infinity, the relative frequency of occurrence $E_i$ is equal to the probability of events $E_i$. For a sensor detecting the presence or otherwise of a contraband drug in a container, the use of classical inference will compare two hypotheses i.e. Null Hypothesis $H_0$ which supports that the observed sample data are caused by the presence of the drug and the alternative hypothesis $H_1$ which supports that the observed sample data are NOT caused by the presence of the drug.

It then proceeds by assuming that $H_0$ is true and goes on to calculate the probability of the observed data given that $H_0$ is true. An hypothesis test is performed and if the probability of observing the data is high based on $H_0$ being true, then the declaration is made that the data does not contradict $H_0$ and if otherwise, the declaration is made that the data contradicts $H_0$. Possible means of hypothesis testing include (Hall, 1992):

1. *Maximum likelihood* which accepts the null hypothesis $H_0$ as true if $[P(H_0|y)]*[P(y|H_0)]$ is greater than $P(H_1)*P(y|H_1)$.
2. *Maximum a posteriori* which accepts hypothesis $H_0$ as true if $[P(H_0|y)]$ is greater than $P(H_1|y)$.
3. *Neyman-Pearson* which accepts $H_0$ as true if the ratio $H_0:H_1$ is less than or equal to a threshold c.

Given that

$$[P(H_0|y)]*[P(y|H_0)] > P(H_1)*P(y|H_1).$$

As can be seen, this probability can only be applied to repeatable events and thus lies one of its disadvantages in data fusion application. Other limitations include (Klein, 1999): (1) only two hypotheses can be assessed at the same time, (2) there are complexities which arise when multivariate data is encountered, (3) its non-ability to make use of *a priori* likelihood probabilities.

### 4.2.2 Bayesian inference method

Some of the limitations of the classical inference are resolved using the Bayesian inference (Berger, 1980). This technique, given new observations and a previous likelihood estimate, updates the likelihood of a hypothesis.

Other advantages of the Bayesian inference is that it also uses prior estimates of the likelihood of a hypothesis being true and in the absence of the empirical data, Bayesian inference allows the use of subjective probability which does not require probability density functions (Wu, 2003).

For mutually exclusive hypotheses, H1 .... Hi for an event E occurring (say an illegal substance being discovered), Bayesian inference suggests that if (Klein, 1999)

$$\sum_i P(H_i) = 1 \qquad (4.2)$$

then,

$$P(H_i|E) = \frac{P(E|H_i)*P(H_i)}{\sum_i P(E|H_i)*P(H_i)} \qquad (4.3)$$

where

$P(H_i)$ is the *a priori* probability that an illegal substance has been detected

$P(H_i|E)$ is the *a posterior* probability or the likelihood of the illegal substance i.e. $H_i$ being detected and

$P(E|H_i)$ is the probability of observing evidence *E* given that an illegal substance has been detected.

Therefore, for multiple detections, i.e suppose a sensor can detect more than one substance (say 3 substances – *S1, S2 and S3*), and observation *O1* from the sensor then the likelihood of each will be

$$P(Substance1|O1) = \frac{P(O1|S1)P(S1)}{\sum_i P(O1|S_i)P(S_i)} \qquad (4.4)$$

$$P(Substance2|O1) = \frac{P(O1|S2)P(S2)}{\sum_i P(O1|S_i)P(S_i)} \qquad (4.5)$$

$$P(Substance3|O1) = \frac{P(O1|S3)P(S3)}{\sum_i P(O1|S_i)P(S_i)} \qquad (4.6)$$

Using the *maximum a posterior* (MAP) criterion, the detected substance is most probably the substance whose joint probability (from Eq. 4.4, 4.5 and 4.6) is a maximum.

In spite of the strong points demonstrated by the Bayesian inference over the classical inference, it however has its own drawbacks. As shown above, unlike the classical inference, the Bayesian inference can be used when more than two hypotheses are compared.

However, (Klein, 1999) lists the limitations of the Bayesian inference method as:

(a) The prior probability incorporated in determining the posterior probability is difficult to define in some cases. For this reason, non-informative priors are used in some cases to provide unbiased prior information (this is explained further in section 2 of Chapter 6).

(b) Complexities when there are multiple potential hypotheses and multiple conditionally dependent events. However, for this thesis, only two hypotheses are involved - detection of cocaine or non-detection of cocaine.

(c) Difficulty in maintaining mutual exclusivity of hypotheses as required for the Bayesian inference. Again, this does not arise in this thesis as the events are clearly mutually exclusive.

(d) Difficulty in being able to assign general uncertainty for example, when the sensor is not able to determine if the substance detected is cocaine or not cocaine.

The Dempster-Shafer theory of evidence methods seeks to improve on the Bayesian evidence by addressing (c) and (d) above.

### 4.2.3 Dempster-Shafer Theory of Evidence method

The Dempster-Shafer (D-S) method is a generalization of the Bayesian theory which corrects the Bayesian inference limitation (d) mentioned above (Casti, 1990). That is, the D-S method allows for a general level of uncertainty. Based on multiple evidences, the D-S method determines the likelihood of hypotheses using probability and uncertainty intervals.

In particular, the D-S method is closer to the way humans think which involves the assignment of measures of belief to unions of hypotheses rather than assigning these evidence/belief to a set of mutually exclusive and exhaustive hypotheses. By doing this, the D-S gives room to allow for uncertainty in the likelihood function and allowance for cases where there is an 'unknown' event.

According to David, et al., (2008), when the situation under consideration contains hypotheses that are mutually exclusive and there are no general levels of uncertainty i.e. the set of hypotheses is exhaustive, then the D-S and Bayesian will yield identical results.

General propositions in D-S are obtained by using Boolean operator 'OR' to combine elementary propositions therefore, a situation of overlapping or conflicting hypotheses may arise.

**Basic definitions** (Campos, et al., 2005)

### *A. Frame of Discernment ($\Theta$)*

In D-S, the Frame of Discernment is the set of hypotheses which contains all possible hypotheses defining a system. It must satisfy two major rules:

1. It must be complete and exhaustive containing all possible hypotheses and
2. The subset hypotheses must be mutually exclusive elements.

### *B. Mass Function (m)*

The mass function or basic probability assignment (bpa) assigns belief, m(A) to the hypotheses in the Frame of Discernment. The bpa assigns a number between [0,1] with 0 implying no belief in the hypothesis and 1 implying total belief in the hypothesis. For a given frame of discernment, the sum of all m(A*i*), where *i* is the number of hypotheses in the Frame of Discernment, is equal to 1.

### *C. Belief Function (Bel)*

The belief function *bel(.),* is defined as the sum of all the masses of subsets of the set of interest. It is mathematically given as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \qquad (4.7)$$

### *D. Plausibility (Pl)*

The plausibility function Pl(.), is a measure of the extent to which evidence supporting the proposition leaves room for belief in the proposition. Mathematically for a proposition A,

$$Pl(A) = 1 - Bel(\sim A) \quad (4.8)$$

As an example, suppose a sensor's data was analysed and a feature from the sensor gives the belief of 0.6 and a plausibility of 0.7 for the proposition, A = 'presence of cocaine'. This can be interpreted as availability of evidence that supports the proposition is true with a confidence of 0.6 and a confidence of 0.3 (1- 0.7) which supports the proposition B = 'no cocaine'. The difference 0.6 – 0.3 = 0.3 is the possibility of either cocaine or not cocaine detection or a state of 'unknown' as illustrated in the table below.

**Table 4.1 D-S process for a single feature**

| Proposition | Mass | Belief | Plausibility |
|---|---|---|---|
| Null (~A and ~B) | 0 | 0 | 0 |
| A | 0.6 | 0.6 | 0.7 |
| B | 0.3 | 0.3 | 0.4 |
| Either (A or B) | 0.1 | 1.0 | 1.0 |

From the table above, the null proposition does not exist as it negates both the cocaine and not cocaine detection. The probability masses for A and B i.e. m(a) and m(B) are 0.6 and 0.3 respectively while the mass for the either proposition m(AUB) ensures that the previous two masses sums to unity by taking up the balance.

### *E. Confidence Interval*

The confidence interval is simply the interval covering the Belief function as an upper limit and the Plausibility function as the lower limit i.e. [*Bel(A)*, *Pl(B)*]

In Dempster-Shafer technique, if $\theta = \{A_1, A_2, A_3, \dots, A^n\}$ is a set of n elementary (mutually exclusive and exhaustive) propositions, then $\theta$ is called the *frame of discernment*. In illegal substance detection using sensor $S_1$, elements $A_1$ ... $A_n$ represents all possible illegal substances that can be detected by the sensor (say for example 6) and $2^\theta$ (or $2^6$ = 64) is the power set of $\theta$ which contains all possible general propositions. Thus, $A_1$ ... $A_6$ represent – Cocaine, Ketamine, Codeine, Amphetamine sulphate, Ecgonine methyl ester and Buprenorphine HCL. **...** The aim is to determine, at unknown concentration, when cocaine is detected.

Thus,

$$\Theta = \{C, Ke, Co, Am, Ec, Bu, \{C, Ke\}, \dots \{C, Ke, Co, Am, Ec, Bu\}, \{\varphi\}\}. \qquad (4.9)$$

*where C* = Cocaine*

*Ke* = Ketamine*

*Co* = Codeine*

*Am* = Amphetamine sulphate*

*Ec* = Ecgonine methyl ester*

*Bu* = Buprenorphine HCL.*

$\varphi$ *= Unknown*

**EQ.4.9** show that there are $2^6$ (64) possible scenarios representing possible detections. This corresponds to a need for a large computational ability and represents one of the challenges of D-S method. For example, the first six scenarios indicate detection of ONLY cocaine or any of the other five substances was detected. The following five scenarios from **EQ4.9** indicate the presence of one of the combination: *cocaine* OR *ketamine*, *cocaine* OR

*codeine*, *cocaine* OR *Amphetamine sulphate*, *cocaine* OR *Ecgonine methyl ester*, *cocaine* OR *Buprenorphine HCL*. Other scenarios include combinations of any two of the substances with no combination repeating itself i.e. *{cocaine, codeine}* is the same as *{codeine, cocaine}* and so only one is represented in the *frame of discernment,* $\Theta$. The next set of combinations in $\Theta$, take any combination of three of the substances indicating detection of one of the three for example, *{...{cocaine,codeine,ketamine}...}* indicates detection of one of *cocaine*, *codeine*, *ketamine*. This format continues and the next set takes a combination of any four and then five of the substances. The last but one 'subset' *{...{*$C, Ke, Co, Am, Ec, Bu$*}...}* indicates detection of any one of the six substances and is an indication of ignorance and the final scenario, $'\varphi'$ is an indication of exception (Wu, 2003).

In Bayesian inference, probability is assigned to all hypotheses. However, the D-S approach assigns evidence to all propositions including single and general propositions once all elements of the *frame of discernment,* $\Theta$ are defined. At this point, the D-S assigns a probability mass, m($\theta$) representing each evidence supporting the belief. The total belief equates to unity just as total probability equals to 1 in Bayesian and classical inference. Thus,

$$\sum \boldsymbol{m}(\boldsymbol{\theta}) = \mathbf{1} \qquad (4.10)$$

The probability of a proposition, also known as the *Support (Spt)* or *Belief*, is thus the sum of all probability mass, $\boldsymbol{m}(\boldsymbol{\theta})$ for the elements in the frame of discernment, $\Theta$, i.e.

$$Spt(A_i) = \sum_{H_i \epsilon \theta_1} m(\theta) \qquad (4.11)$$

which translates that the belief in $H_i$ is the sum total of all probability masses m($\theta$) that supports $H_i$ only. Note that as **EQ4.9** has shown, evidence is assigned not only to mutually exclusive propositions (i.e. *Ai* in $\Theta$), but also to general propositions that involve overlapping and nonexclusive general propositions. From **EQ4.9**, it is also shown that a probability mass can also be assigned to the general proposition $\{...\{C, Ke, Co, Am, Ec, Bu\}...\}$ which implies that the sensor cannot determine which of the substances it has detected.

However, in the case where there is evidence against the support of $A_i$'s exclusivity, this is called the *plausibility* of Ai and is simply

$$Pls(A_i) = 1 - Spt(\sim A_i) \quad (4.12)$$

The interval between the Belief (or Support) and the Plausibility in D-S is known as the *Confidence Interval* which measures the support and respective plausibility for $A_i$ and is shown below:

**[***Belief*($A_i$)*, Plausibility*($A_i$)**]** or (4.13a)

**[***Belief*($A_1$)*, Plausibility*($A_1$)**]** (4.13b)

**[***Belief*($A_2$)*, Plausibility*($A_2$)**]**

⋮

**[***Belief*($A_n$)*, Plausibility*($A_n$)**]**

Thus for a D-S, the inputs are probability masses, m($A_i$), while the outputs are the confidence intervals (EQ4.13b).

All the above represent scenarios where only one sensor is used. However, when more than one sensor is used, with all sensor measurements independent of each other, the D-S combination rule, similar to the Bayes

formula for combining probabilities, provides a means for combining their probability masses. For example, if there are two sensors detecting for illegal substances and for *i* number of propositions, then the D-S combination rule states that the joint probability mass of propositions $m_{1,2}(A_i)$ is given as:

$$m_{1,2}(A_i) = \sum_{A=A_k \cap A_j} \frac{m_1(A_k) m_2(A_j)}{1 - \sum_{A_k \cap A_j = \emptyset} m_1(A_k) m_2(A_j)} \qquad (4.14)$$

The denominator in EQ4.14 is the normalization factor which takes into consideration, all propositions that conflict (i.e. should have an empty set) but has been assigned with non-zero values. The numerator gives the probability mass function of the products of the observed evidence of the two sensors which gives proposition *A*. The order of the combinations has no effect on the joint probability masses implying that they are commutative and associative (Lowrance & Garvey, 1982).

Two main shortcomings of the Bayesian Inference highlighted in the previous section are addressed in this section.

a. Difficulty in maintaining mutual exclusivity of hypotheses as required for the Bayesian inference and which the D-S addresses by providing a means of assigning evidence to hypotheses which overlap.
b. Difficulty in being able to assign general uncertainty which the D-S addresses by allowing assignment of evidence to a union of all possible hypotheses which is the sensor's way of declaring its inability to determine what substance is present (Klir, 1999) (Rocha, n.d.).

The D-S method has done well in addressing the short comings faced by the Bayesian method. However, D-S itself has its own shortcomings some of which are (Liu et al., 2002):

a. Impractical assumption that evidence are independent
b. The D-S will only be applicable when sets of hypotheses are mutually exclusive and exhaustive
c. Compared with Bayesian, the D-S has a higher level of complexity and thus requires more computational resources

In explaining (c) above, **EQ4.9** reveals that the number of elements (general propositions) in the frame of discernment Θ, increases exponentially ($2^n$) with the increasing number of elemental propositions, n. One of the requirements of D-S is that all mutually exclusive and exhaustive hypotheses including the case of the 'unknown' must be represented in Θ. This ensures that in many cases, there is a sufficiently large number of elemental propositions. With increasing n and thus increasing size of Θ, more computational resources will be needed and thus lays one limitation of the D-S method.

However, in this dissertation, while the sensor can detect for cocaine and five other substances, its application relies on its being able to detect the presence or not of cocaine. Thus, there are only three propositions i.e. cocaine, not cocaine and unknown with the 'unknown' case capturing situations where the model may not be able to decide whether the detected substance is cocaine or not cocaine.

### 4.2.4 Neural network technique

As stated above, the D-S method compared with the Bayesian method needs more computational power, especially as the size of the elemental propositions increases. When the problem is not well defined and requires a high computational power, one method that can be used in this situation is the Neural Network method. Neural networks are robust and versatile when it comes to characterizing input-output behaviour of unknown systems (David, et al., 2008), (Bishop, 1995).

A neural network consists of separate layers of interconnected nodes. The first layer on the leftmost side of the network of a neural network is known as the *input* layer. It consists of one or more several nodes representing the input to a multisensory data fusion system. At the far right side of the network is the *output* layer. Artificial Neural Networks (ANNs) mimic the biological neurons of the nervous system.

As the data in entered into the network via the input nodes, the network performs a nonlinear transformation achieved by weights attached to each node in the layers, eventually giving an output as shown in figure 4.2. In between the input and output layers is the *hidden* layer. In a network, it is possible to have more than one hidden layer depending on the complexity of the system. The number of nodes in each hidden layer is also not fixed.

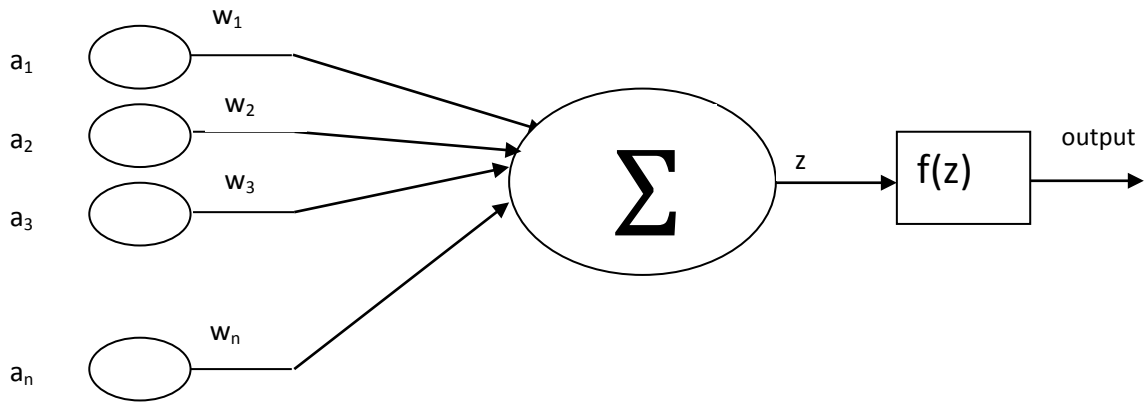



Figure. 4.2 showing inputs and output for a 2 layer NN

When data is entered into the network via N number of input nodes, the inputs a = ($a_1$, $a_2$, $a_3$, ... $a_n$) is non-linearly transformed to give an output z. If 'b' is a bias or *threshold,* this transformation is a computation of weighted values of 'a', i.e.

$$z = f[\sum_{i=0}^{n} w_i a_i + b] \qquad (4.15)$$

The function f(.) can be any of a number of non-linear transformation, including a step function, sigmoid function.

The output value, z, is then compared to the expected output value say z'. The mean square difference between the expected and real output values (the error) for each node, starting from the output node, is then used to iteratively adjust the weights to reduce the errors until the desired output is achieved (Rumelhart, et al., 1986). This process is called *training* the network and this method of iteratively adjusting weights from the output backwards is known as *back-propagation algorithm*.

The input data can be the set of data collected by a sensor which can detect any of six substances if given a liquid containing one of the substances. If the

content of the liquid is previously known and the data collected is mapped to the known output, for instance say a user wants to detect for cocaine and the remaining substances are not important for our purpose, then data may be collected for when the substance is cocaine and used as the input while the output may be given as unity (‘1’) and data is also collected for other substances and output is tagged as ‘0’. The network may then be trained using a specified number of layers such that when a new ‘unknown’ set of data is inputted into the system, it will (with little error) classify whether the set belongs to the ‘1’ or ‘0’ category. This is discussed more in Chapter 3 of this thesis.

Other applications of neural network abounds and includes Robotics: Joris, *et al* (1996) used neural network in converting sensor data for an autonomous mobile robot, Hu (2010), used fuzzy logic and neural network to solve the motion planning problem of a mobile robot; Remote sensing: Neural network is used in the attitude control of remote sensing satellites (Wei-feng, et al., 2002);

In spite of the strong positives presented by the use of ANNs, there are limitations too. A common limitation when back propagation is used is the local minima challenge – since back propagation uses a gradient descent technique and this exists on a non-linear surface, the result may end up with local minima rather than a global solution (David, et al., 2008). Another common limitation is that the training of a network is generally slow.

## 4.3 Model Development

The methodology implemented in this thesis was based on the target results. We aim to collect data from the environment with the sensor as the ferret robot moves around the container. Data collected is analysed at time intervals $t_1$, $t_2$ … $t_n$. the Time interval is determined by how long it takes for the sensor to collect data and output decisions. The final output is a decision on whether cocaine is detected or not along with a probability of detection.

The architecture shown in Figure 4.3 assumes that sensors A and B acquire two different measurements which characterise three different contrabands. The combination from the two probabilities of the sensors decides on which substance is detected (Akiwowo & Eftekhari, 2010). The trackers are used for updating probabilities after each time interval. This architecture is expanded into the model used in this thesis.

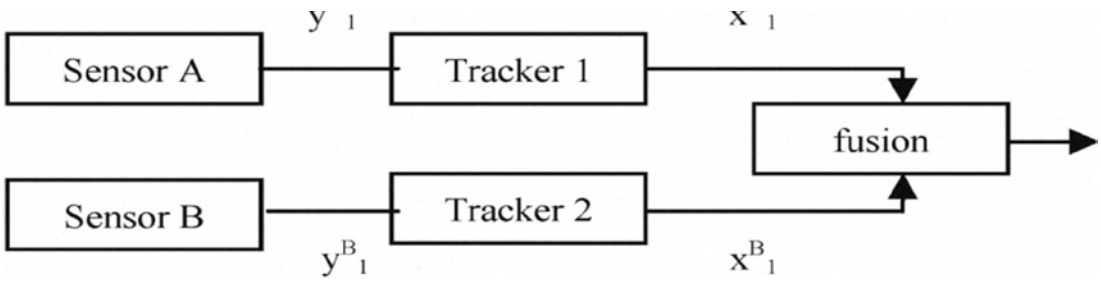



Figure. 4.3 Data Fusion model with updating tracks (Koks & Challa, 2005) (Akiwowo & Eftekhari, 2010)

The model adopted in this dissertation has two stages. In the first stage, the sensor collects data from the surrounding atmosphere and stores the intensity values against corresponding wavelength. Once data has been collected, stage two involves pre-processing and feature extraction processes. The features based on a certain threshold, $T_1$ determined

empirically (see section 5.7 in chapter 5), will make decisions on cocaine detection or otherwise with a corresponding probability of detection ($D_1 \ldots D_{12}$ with corresponding $P(D_n|H_n)$. This threshold is determined by comparing cocaine and non-cocaine data measurements for each feature and determining what threshold value gives the optimum performance criteria (low false alarm, high true positives). The steps involved are highlighted in Appendix C. At this stage, a fusion of all decisions is performed using the posterior probabilities from above. The result from this based on a pre-set threshold, $T_2$ (this threshold is determined by the operator and is dependent on *a priori* knowledge of the origin of the shipment). If the posterior probability exceeds this threshold, a positive detection of cocaine is concluded. However, if the result is not conclusive or negative, the ferret robot moves to a new location, collects data and is analysed as above.

Figure. 4.4 gives a diagrammatic model of the description.

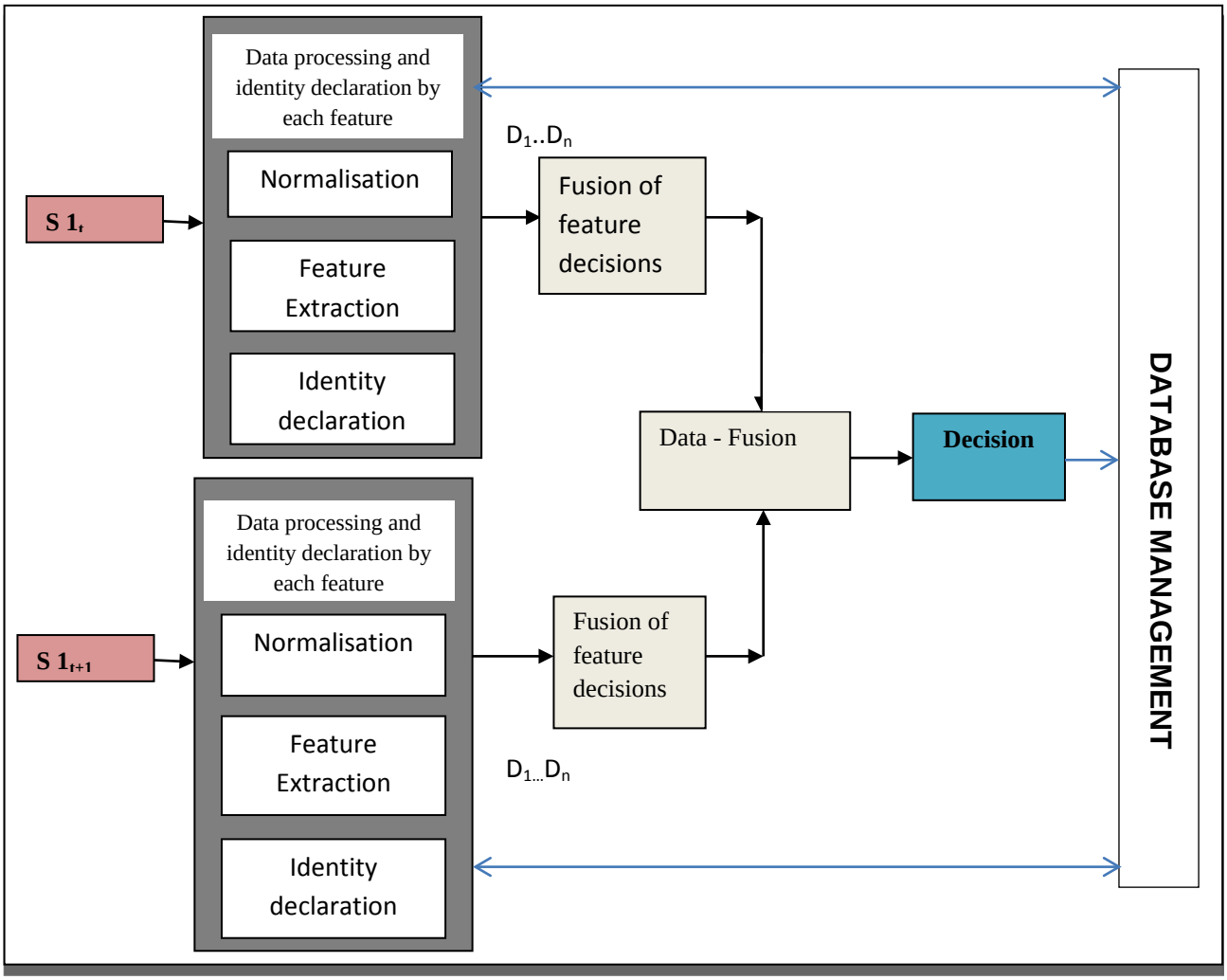



Figure 4.4 Data Fusion Model

## 4.5 Conclusions

The model employed for any data fusion process is critical to the success of the data fusion itself. Whilst there are many algorithms for combining data, this chapter has highlighted a few and discussed their advantages and short coming. The chapter began by describing the data collection and experimental set up used in this thesis. The results from the data collected were also highlighted and discussed and the need for data fusion was given. Smuggling of cocaine into the UK via the sea borders in containers has

posed a big threat to UK border officials and the economy at large. The challenges are in two fold – the first is that the goods are stored in a container and the second is that usually, cocaine is hidden within the goods. Current technology used by border officials is not sufficient enough to detect the contraband. This thesis proposes a two-stage methodology to be used to meet the challenge of detecting cocaine. In the first stage, a state of the art fibre optic sensor developed for this purpose is used to identify cocaine signature from the environment. This signature also matches that of many other similar compounds at same concentration. This leads to the second stage of the methodology. At this stage, using data fusion techniques, the output of the sensor is broken down into 12 separate features in a feature extraction process. Based on the training data, each feature then makes a decision on whether cocaine is detected or not and these separate decisions then act as inputs into a data fusion system. The resulting output is the probability of detection of cocaine.

The methodology takes into consideration that the process of data collection is separate from data analysis and the fusion process. The first step takes care of collection of data while the second step takes care of data analysis. The analysis of the data including pre-processing of the data, feature extraction stages was also analysed in this chapter. Implementation of the model after pre-processing and feature extraction using neural network technique, the Dempster-Shafer technique and Bayesian Technique is then outlined in subsequent chapters with their corresponding results shown.

# CHAPTER 5

# FEATURE SELECTION AND VALIDATION

The inputs into a data fusion model are selected from the output of the fibre optic sensor. The output of the fibre optic sensor is the intensity (counts) versus the wavelength (nm). One set of data of intensity versus wavelength contains over a thousand pairs of raw data. It is important that features are extracted from these pair of dataset. In addition it must be ensured that the number of selected features is an efficient representation of the underlying structure.

This chapter will be looking at the method of extracting features from the raw data. In many data fusion applications, the feature extraction stage helps to identify important features from the raw data. It is these features that are then used as inputs into the data fusion system. Technically, using the raw data as inputs into the system is not advised and one of the reasons is to avoid the *curse of dimensionality,* this is discussed further in following

sections. However, before feature extraction can be performed on the data, the data needs to be pre-processed (cleaned[8], data alignment and data association).

## 5.1 Data Pre-processing

Data preprocessing is an important part of data fusion. It helps to identify data that do not fit into the overall pattern of the data. Issues such as outliers and missing values are identified at this stage and dealt with. Not dealing with these errors at this stage could lead to unreliable results. It also helps to align previously unaligned data either in the spacial or temporal domain or both.

The data used in this thesis were sourced from another arm of the project (Nguyen, Sun, Grattan, & Hardwick, 2010). One important issue to consider in the implementation of this sensor is its reproducibility in use (Nguyen, Sun, Grattan, & Hardwick, 2010). To test the reproducibility of the sensor, it was calibrated with different cocaine concentrations from 0 to 500 μM and recalibrated after 24 h. The results showed that the data were mostly the same with no significant differences, thus proving its repeatability and reproducibility. To maintain data integrity, the data used in the analysis stage of this thesis were collected on the same day within the same environmental conditions.

[8] Sensor data are usually corrupted with background noise and outliers thereby affecting the integrity of the data. Cleaning of data involves separation of noise and outliers from the data.

### 5.1.1 Normalization

Normalization is an important step in data pre-processing. It aligns all sets of data to fit into the same domain. In Neural Networks for example, it is imperative that output values are selected within the range of the activation function used (for example 0, 1 for sigmoid function). Therefore, to meet this requirement, all input and output values are transformed to within the required range; this transformation is called *normalization* (Haykin, 1998). Normalization also helps to align data which are not in the same spatial and/or temporal state.

In neural network data preprocessing, normalization is used to prepare raw data for training. It also helps in speeding up the training process for the Neural network.

There are several types of data normalization. They are used in scaling the raw data so that it is in the same range of values to reduce the bias between input features within the Neural Network. They can also be used to speed up the training time by starting the training process for each feature within the same scale (Jayalakshmi & Santhakumaran, 2011). Its usefullness can also be found when the input features are different scales. Normalization techniques include Min-Max, Median, Statistical Column, Sigmoid and Z-score normalization.

**Normalization Techniques**

1. **Min-Max Normalization**: Min-Max normalization involves the rescaling of raw data from one range of values to another range of values.

Common among the range of values used for min-max normalization is the 0 to 1 ([0 1]) range or the -1 to 1 range ( [-1, 1] ). The transformation is performed using the following formula

$$x^{norm} = (x_{max} - x_{min}) * \frac{(x_i - x_{min})}{(x_{max} - x_{min})} + x_{min} \quad (5.1)$$

$x^{norm}$ is the normalised value of the raw data

$x_{max}$ is the maximum of the normalization range

$x_{min}$ is the minimum value of the normalization range and

$x_i$ is the input raw data

The new normalized value, $x^{norm}$ of each raw data will lie between the target range. One major advantage of the min-max normalization technique is that it preserves the relationship between data values (Jayalakshmi & Santhakumaran, 2011).

2. **Median Normalization**: Median normalization is mainly used in situations where the raw data has extreme deviations. It is computed by finding the ratio between each raw datum and the median of all the raw data such that:

$$x^{norm} = \frac{x_i}{median\ (b_i)} \quad (5.2)$$

$x^{norm}$ is the normalised value of the raw data and

$x_i$ is the input raw data

$median\ (x_i)$ is the median of the set of data $x_i$

3. **Statistical Column Normalization:** Statistical Column Normalization follows two basic steps. Given an N x M matrix, the first step calculates the normalization of each column $M_i$ (i = 1,..M) by normalizing the columns to a unit length. In step 2, each raw data is now divided by the normalized column attribute and multiplied by a bias. The equation is shown below:

$$x^{norm} = \left(\frac{x_i}{n(c_a)} - 1\right) * b \qquad (5.3)$$

$x^{norm}$ is the normalised value of the raw data and

$x_i$ is the input raw data

$n(c_a)$ is the normalized attribute value and

*b* is the bias

4. **Sigmoid Normalization**: Sigmoid Nomalization like the min-max normalization, is used to scale the raw data to values between the range 0 and 1 or -1 and 1. It replaces each raw data with the non-linear sigmoid function. There are many available sigmoid functions including the logistic function (EQ. 5.4a) and the hyperbolic tangent function (EQ. 5.4b).

$$x^{norm} = \frac{1}{1+ e^{-x_i}} \qquad (5.4a)$$

$$x^{norm} = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}} \qquad (5.4b)$$

$x^{norm}$ is the normalised value of the raw data and

$x_i$ is the input raw data

5. **Z-score Normalization**: in the z-score normalization technique, every input raw data is normalized by using the mean and standard deviation as shown in EQ.5.5 below. The normalized data produced by this method has zero mean and standard deviation of 1. This technique has the advantage of being able to reduce the effect of outliers (Jayalakshmi & Santhakumaran, 2011). The key issue to be remembered when using the z-score normalization technique is that all data must be trained with the normalized data and the means and standard deviation calculated from the training data retained to be used later in the design process.

$$x^{norm} = (x_i - \mu_i)/\sigma_i \qquad (5.5)$$

$x^{norm}$ is the normalised value of the raw data

$x_i$ is the input raw data

$\mu_i$ is the mean and

$\sigma_i$ is the standard deviation

For this thesis, the min max normalization technique was adopted. This technique has been selected due to its ability to retain the relationships between the data values without introducing any bias to the sets of data. In addition, the min-max technique easily transforms the raw data to fit into the [0 1] range (using EQ. 5.1) required for back propagation algorithm used for

the network architecture. All normalization transformations where done using Microsoft Excel package and MatLAB® software. As can be seen in figure5.1, there is a dependence of intensity values on the concentration of the analyte. Normlization is performed to remove the effect of the concentration dependent intensity changes.

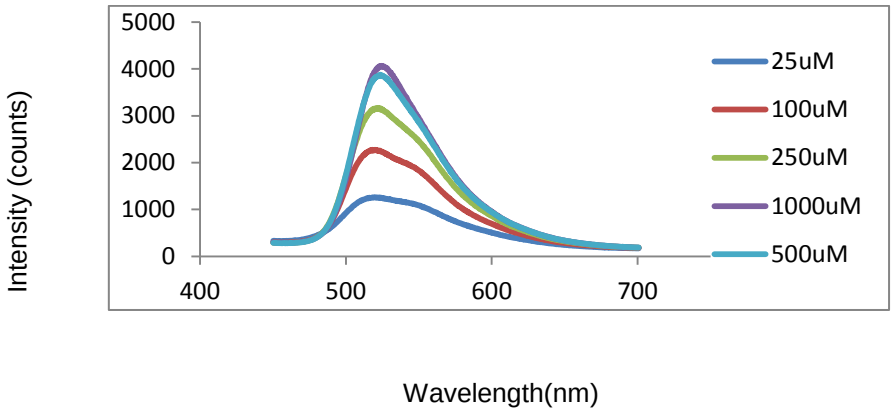



Figure. 5.1 Raw data plot of intensity vs. wavelength for cocaine at different concentrations before normalization

In calculating the features from the normalised plots for the compounds, each spectra was divided into six *activity points*. The activity points represent sections within the plots where a change in gradient occurs by visual inspection.

## Cocaine

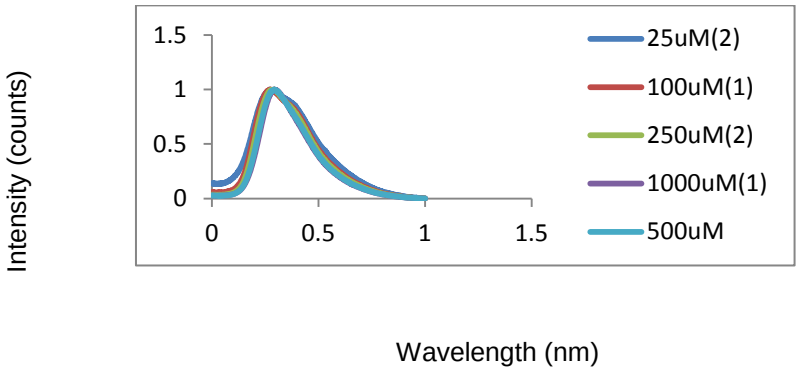



Figure.5.2 Normalised spectra for Cocaine at different concentrations

## Ketamine

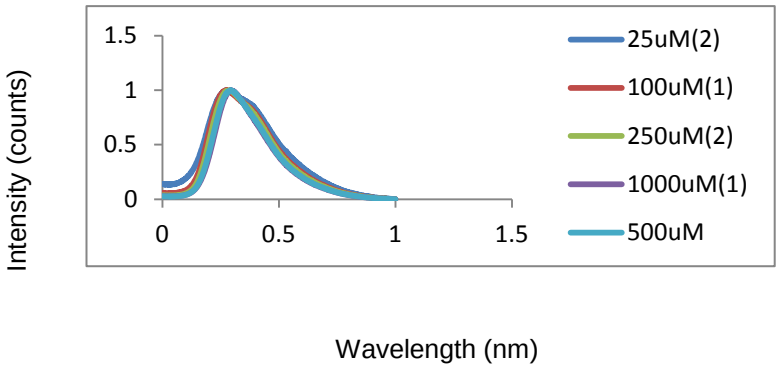



Figure.5.3 Normalised spectra for Ketamine at different concentrations

## Codeine

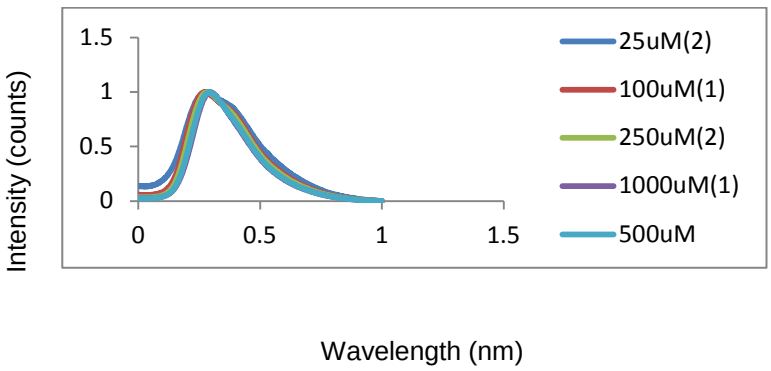



Figure.5.4 Normalised spectra for Codeine at different concentrations

## Amphetamine sulfate

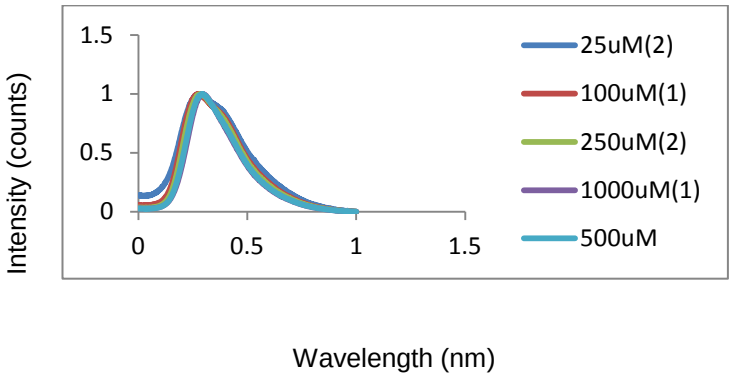



Figure.5.5 Normalised spectra for Amphetamine sulfate at different concentrations

## Ecgonine methyl ester

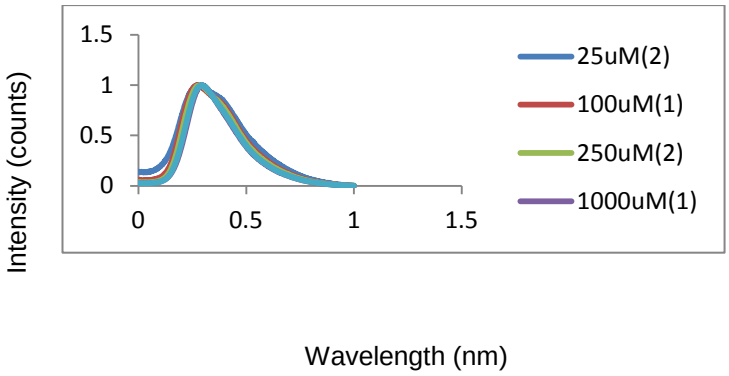



Figure.5.6 Normalised spectra for Ecgonine methyl ester at different concentrations

## Buprenorphine HCl

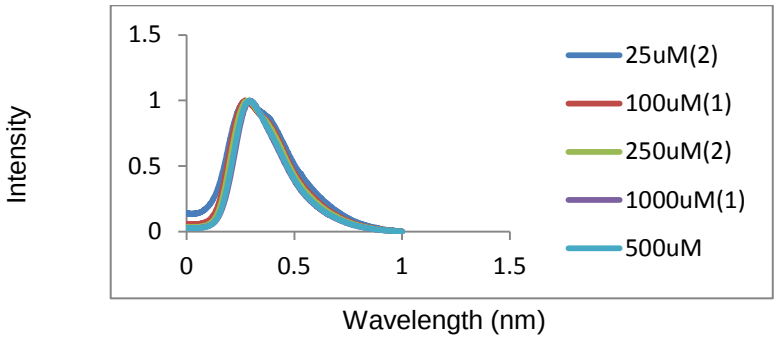



Figure.5.7 Normalised spectra for Buprenorphine HCl at different concentrations

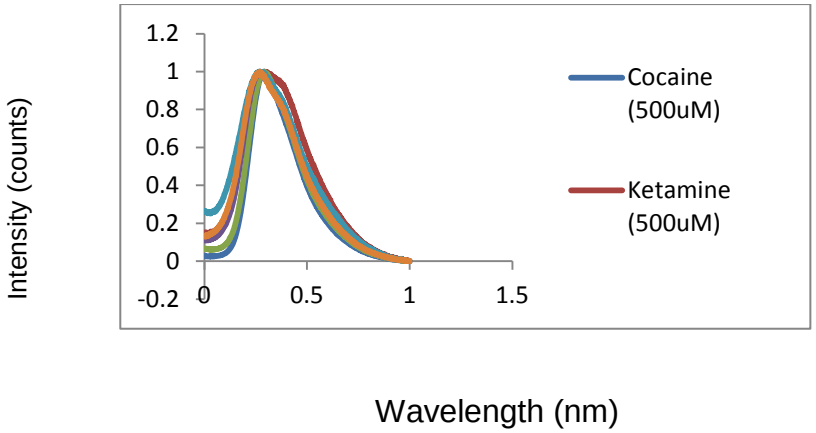



Figure 5.8 Normalised spectra for all comounds at 500uM concentration

## 5.2 Feature Selection

To adequately be able to discriminate between cocaine and non-cocaine spectra, it is desirable to be able to identify each spectrum by their unique characteristics. These unique characteristics also known as features of each spectrum asides from helping to uniquely identify the spectrum and thus the analyte, also serve as inputs to the neural network system.

From Chapter 2, it was explained that data may be fused in different ways one of which is decision in/decision out (DIDO). The decisions may be final sensor decision or feature decision on identification of target. It is the norm to reduce data from its raw form into features describing the data as inputs to a data fusion system. Sometimes, the raw data may be too many for the system resulting in what is usually termed as the 'curse of dimensionality' – the increase or jump in computational complexity and classification error for data with high amount of dimensions (Bellman, 1961) (Pechenizkiy et al., n.d.).

In this case, the large amount of data may be reduced to a set of features which aims to represent the signal in the same way as the entire measurement data represents it. Feature selection is one of the dimensionality reduction techniques (Liu, 1998). In principle, feature selection involves the transformation of a *d*-dimensional feature space pattern *y* by a mapping *f* to a pattern *x* of *m*-dimensional projected space, where *m*<*d* (Lerner et al., 1996) in such a way that an optimizing condition *J* is fulfilled.

$$x = f(y) \qquad (5.6)$$

Therefore, for a transformation *g(y)*, *f(y)* is the mapping which satisfies the condition (Devijver & Kittler, 1982) (Lerner et al., 1996),

$$J\{f(y)\} = \max_g J\{g(y)\} \qquad (5.7)$$

As described, the main aim of extracting features from a set of data can thus be narrowed down to reduction of the dimensionality since extraction of certain features from a wide variety implies reduction in the set of available features. Thus linear feature extraction can be described as locating a set of vectors which represent an observation and at the same time, reducing its dimensionality (Lee, et al., 1993). It is the process of mapping original measurements into a lower amount of features without excluding the key information describing the set of data (Guyon, et al., n.d) (Lerner, et al., n.d). Principal Component Analysis (PCA) is a popular method for selection of features from a multidimensional set of data especially in image recognition

(Sun et al., 2005). However, the PCA has limitations when it comes to spectral measurements as it only selects generalized features (Skurichina et al., n.d.). The aim of the feature selection process in this thesis is to find unique features which characterize the measurements into cocaine and non-cocaine data. The spectral band extraction technique is one feature selection technique which helps to identify these unique features.

### 5.2.1 Spectral Band Extraction technique

For many spectral measurements such as in this thesis, selection of features is done by finding unique disciminative band regions within a spectrum. Usually, these unique bands are identified via a study of the spectrum and physical background knowledge of the compound under study.

Over the years, researchers have proposed different algorithms to divide a spectrum into bands (Kumar et al., 2001) (Verzakov et al., 2004) (Skurichina et al., 2004) although in some cases, it is sometimes difficult to identify unique spectral regions as the information required for discrimination may be spread over a wide area of spectral features (Skurichina et al., n.d.). In some cases, the spectrum is simply divided into equal wavelengths, thus generating a high amount of features (Lowry & Isenhour, 1975) (Sutter & Jurs, 1997) whilst in another, unique features are selected based on spectra characteristics (Kumar et al., 2001). The technique used in this thesis involves the identification of ‘activity points’ within each spectrum as explained in the following section.

The spectral band extraction technique has achieved success in many researches such as in Kumar, *et al* (2001) who used a top down and bottom up algorithm technique to classify hyperspectral data,

#### 5.2.2.1 Extraction of features from Fibre-Optic Sensor data

The challenge presented by the fibre optic sensor for cocaine detection is that when prior knowledge of the concentration of the sample compound is not available (as is most likely the case in real life), it will be impossible to classify correctly. However, if the concentration is known *a priori,* classification proceeds easily. Therefore, in the case of non available prior information, more analysis of the data will be necessary to determine what analyte is detected. The features extracted and calculated were selected with the aim of achieving the best representation for each fluorescence spectrum.

Using the raw data as as a base, there are two thousand and fifty two (2052) raw data points corresponding to wavelengths vs intensity pairs. The high value of this dimensions implies that they cannot serve as input to a DF system without causing a dimensionality problem. It is thus imperative to select from these possible features a subset of features which also uniquely represent the compound. Rather than divide individual spectrum into equal bands using the wavelengths, this dissertation segments the spectrum using 'activity points'.

From the raw data plots of intensity against wavelength of cocaine samples (figure 5.1), what is obvious are what the author has tagged as "***activity points***" common to all the plots. These activity points are segments within

the plots that show a change in gradient (activity). For example, in Appendix A(i), with wavelengths between points 450 and 470, the gradient of each spectrum remains somewhat uniform. However, this changes as the wavelength increases with the gradient becoming steeper thus moving past an activity point to another segment. The activity points are used in segmenting the spectrum into sections between 450nm and 750nm (raw data values). Key features from the segments are: Intensity changes and slopes. From the intensity changes, four features were extracted and two features were extracted from the slopes. In addition, the peaks of the spectra were added as an additional feature.

These sections are then used in calculating the features as explained in the next section. Similar technique was used in (Sutter & Jurs, 1997), where the spectra were divided into ten equal time slice regions. However, in this thesis, using the systematic measure described, the spectra have been divided into six different sections.

Overall, 12 unique features were extracted –

- Peak
- Steepest Slope
- Maximum Negative Slope
- Most Positive change in intensity
- Average intensity in regions (there are six regions and therefore six values representing each spectrum)
- Most negative change in intensity

- Average change in intensity

The normalized values of these features were then used as inputs into neural network system to test their performance.

**Activity points** (normalized)

**Table 5.1 Activity points for normalized data**

| **Wavelength** | 0 - 0.0624 | 0.06381 - 0.14559 | 0.1470 - 0.2893 | 0.2907 - 0.5538 | 0.5552 - 0.8013 | 0.8026 – 1.0 |
|---|---|---|---|---|---|---|

## Plots

The plots below show the normalised plots for all compounds at 500$\mu$M. Figure 5.9(a) shows all the plots on a single graph. In figure 5.9(b), cocaine at different concentrations is also plotted. It can be seen that when the concentration is the same, it is easy to identify individual compounds. There is no way to identify cocaine (at concentrations lower than 1000$\mu$M) without further processing. I addition to providing a unique way to identify cocaine, this thesis also provides a means of improving th results from each feature decision by using data fusion techniques.

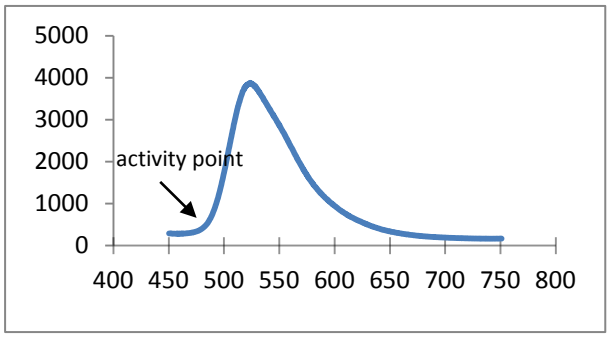



Figure 5.9 (a) Cocaine at 500$\mu M$ concentration showing activity point

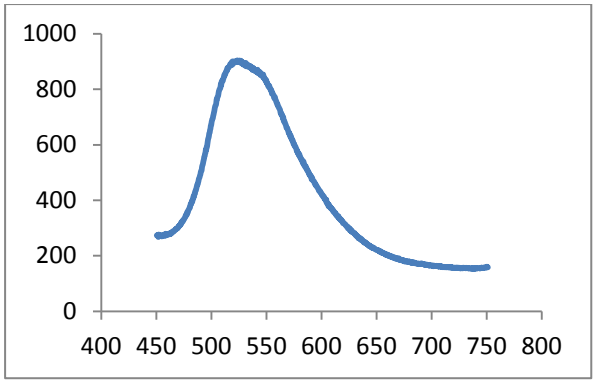



Figure 5.9 (b) Ketamine at 500$\mu M$ concentration

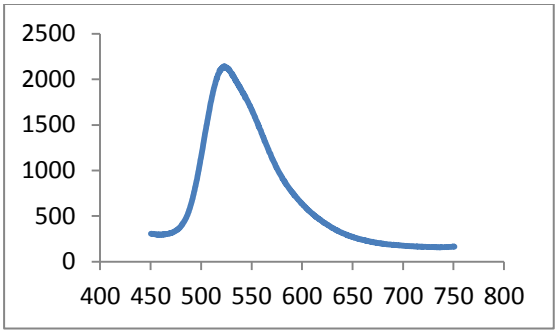



Figure 5.9 (c) Codeine at 500$\mu M$ concentration

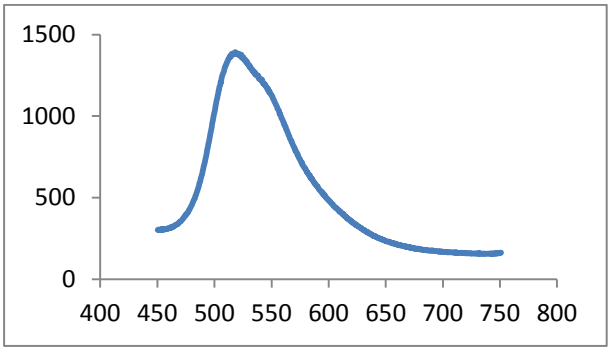



Figure 5.9(d) Amphetamine sulphate at 500$\mu M$ concentration

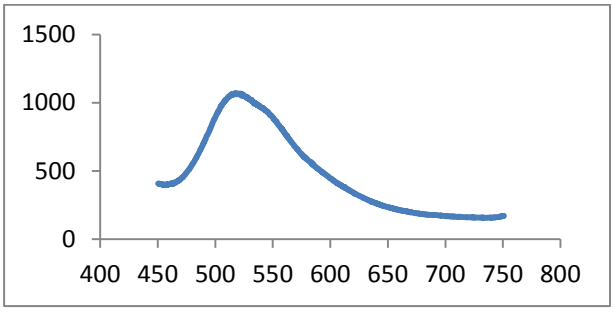



Figure 5.9 (e) Ecgonine methyl ester at 500$\mu M$ concentration

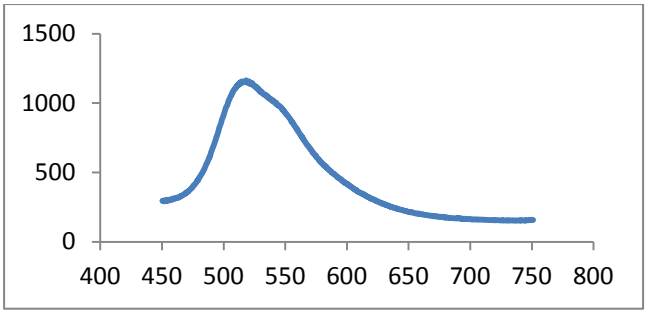



Figure 5.9 (f) Buprenorphine HCl at 500$\mu M$ concentration

Figure 5.9 Plots of Intensity vs Wavelength(nm) for various compounds at 500$\mu M$.

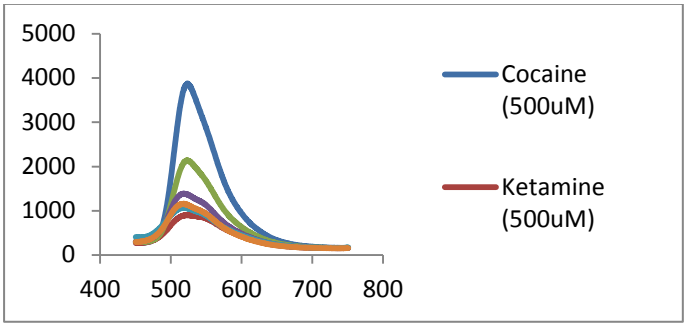



Figure 5.10(a) Plot of all compounds on same graph using raw data

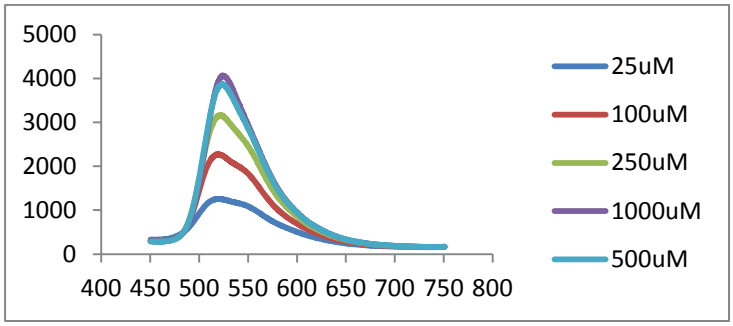



Figure 5.10(b) Plot of Intensity vs Wavelength(nm) for cocaine

Figure 5.10 Plots of Intensity vs Wavelength(nm) got (a) all compounds at 500uM and (b) for cocaine at different concentrations after normalization

## 5.3 Neural Network Validation

The previous sections looked at the feature extraction process and the features extracted. This section aims to validate the features using a k-fold

validation neural network system. The validation criteria are the R-squared[9], mse [10] and the correlation coefficient values. The results from the implementation of the ANN is shown and discussed.

### 5.3.1 Pre-processing

As in the case of sensors in general, in the extension of Fibre-Optic Sensor for cocaine detection, the main concerns is to ensure that the data collected is clean and devoid of noise (data pre-processing) and also to correct inconsistent values in the data (Ni, 2008). Correcting inconsistent values in data requires an understanding of the system and likely output of the sensor. For a fibre optic sensor, sources of inconsistent values include fluctuations in the optical source (Varghese et al., 2009). These can be corrected by using filtering techniques. Other pre-processing methods used in this dissertation will be discussed subsequently.

### 5.3.2 Neural Network Implementation

The multi layered feed forward neural network is one of the most successful and widely used artificial neural network architectures (Lyons, et al., 2000). There are three major sections: the input, hidden and output layers. The hidden layer can consist of one or more layers. The input layers receive input data which are usually characteristic features of the signal and sends out these inputs to the first hidden layer. Each neuron in the hidden layer receives input from every neuron in the preceding input layer. The neuron

[9] R-squared is a measure of how well the outcome of the network is described by the input variables
[10] mse is the mean squared error

then adds up all the inputs received and then compared against a threshold. This comparison is in fact a transformation based on non-linear transfer function (figure 5.11 below shows various types of transfer functions). Depending on the number of layers in the hidden layer, the above is repeated until the final result reaches the output layer.

The input to the output layer of the neural network is typically the output from the last hidden layer. This input for each neuron in the output layer (for layers with more than one neuron) is also then compared with a threshold and it then outputs a signal if it is greater than the threshold. If it is less than the threshold, no signal is output.

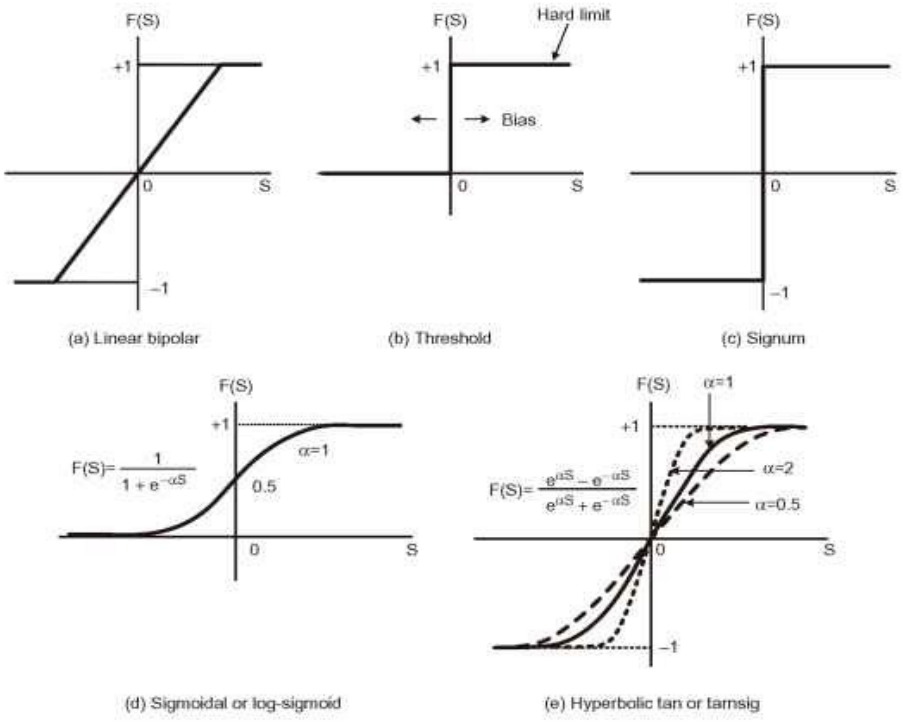



Figure 5.11 Different types of transfer functions (a) Linear bipolar (b) Threshold (c) Signum (d) Log-sigmoid (e) tansig

#### 5.3.2.1 Back Propagation

The back propagation algorithm is one of the best known neural network algorithms (Patterson, 1996), (Fausett, 1994). It uses the gradient descent method to search for the minimum of the error function in weight space. Once the minimum error has been found, the set of weights which produce this minimization is the solution of the learning problem. One important rule for back propagation to work is that the error function must be continuous and differentiable (Rojas, 1996). The minimum error which is also the difference between the actual output and the desired output is represented by

$$MSE = \frac{1}{n}(\sum_p \sum_i O_{pi} - T_{pi})^2, \qquad (5.12)$$

where *p* and *i* encompass all training and output neurons of the network. $O_{pi}$ and $T_{pi}$ represent the actual output and the desired output respectively.

The non-linear relationship between the input and the output parameters requires that an activation function is needed to appropriately connect the parameters of these layers. For this thesis, the sigmoid function is adopted as the activation function. This function is defined mathematically as follows:

$$s(x) = \frac{1}{1+ e^{-ax}} \qquad (5.13)$$

Where $a$ is the factor which determines the shape of the sigmoid function. As $a$ tends towards infinity ($a \rightarrow \infty$), EQ5.13 converges towards a step function

at the origin. The sigmoid function has been discussed in detail in a previous section.

Back propagation also requires that all data be normalized between zero and one [0,1]. The reason for this is because the output node signal is restricted to this range of values. The back propagation algorithm is a feed forward ANN and it adjusts weights by error propagation from the output to the input.

Initially, the weights of the network are chosen randomly and the back propagation algorithm used to calculate the errors where necessary. In this report, the following steps were followed in implementing the back propagation algorithm:

1. Initializing the weights and biases.
    a. The weights in the network are randomly initialized in the interval [0 1]
    b. The biases attached to each unit are also initialized within the same limits [0 1].
2. The training sample is then fed into the system together with matching target values.
3. The inputs are then propagated forward and the net input-output values for each unit of the hidden and output layers are calculated.
4. The difference from step 3 which is the error is back propagated to the system.
5. Based on the errors back propagated to the system, the weights and biases are updated

6. Steps 1 to 5 will then continue until the terminating conditions are met.

One run of steps 1 through to 5 is known as an epoch. Several epochs are thus run until a performance criterium is met. A criterium may be a specific amount of epochs to be run or a minimum error set for the system to meet or a combination of both. Once a performance criterium is met, the algorithm is stopped.

#### 5.3.2.2 Updating of weights

Step 5 above states that after the back propagation to the hidden layer stage, the weights of the network are updated. Everytime an input pattern is presented to a node, the weights attached to it are adjusted to match the desired response from that node. This process of pattern presentation and weight adjustment is repeated until the mean squared error is minimized and it is at that stage that the network is said to have learned the presented input patterns. To update the weights of the back propagation algorithm neural network, the iterative weight rule is (Amini, 2008):

$$\Delta W_{i,j}(n+1) = \eta(\delta_i O_i + \alpha \Delta W_{i,j}(n)) \qquad (5.14a)$$

Δ is the learning factor,

$W_{i,j}$ is the weight connecting node ito node j,

$\delta_i$ is the neuron error for output neuron i,

$\alpha$ is the momentum factor

The neuron error $\delta_i$, is given by

$$\delta_i = (T_i - O_i) * O_i * (1 - O_i) \qquad (5.14b)$$

A neural network system consists of three layers – the input layer, the hidden layer and the output layer. Each of these layers can have any number of neurons in the layer. In addition, the hidden layer can consist of more than one layer. A breakdown of a typical multilayer neural network has been discussed in chapter 2 of this thesis.

**Input Layer -** The input layer of a multilayer neural network represents the features which are fed into the system. For this network, there is one input layer with 12 neurons. The neurons represent the features from the previous section. Each set of features for each compound is fed into the system at a time. The input to a neural network aside the input layer neurons for a back propagation neural network is (Pao, 1989)

$$X_i = \sum W_{i,j} O_j + b_i \qquad (5.15\ a)$$

and the output a neuron is (Pao, 1989),

$$O_i = f(X_i) \qquad (5.15b)$$

where $W_{i,j}$ is the weight of the connection from neuron $i$ to $j$, $b_i$ is the bias and $f$ is the activation function.

**Output Layer -** For this system, there is one neuron in the output layer. The output for the neuron can be either 1 or 0, corresponding to whether the

system is reporting the presence of cocaine or not respectively i.e. a “1” output implies that cocaine has been detected while a “0” implies otherwise.

**Hidden layer -** The number of hidden layers is determined by the complexity of the system. For most problems, a single hidden layer will suffice. However, the number of neurons in the hidden layer is usually determined on a trial by error basis. In this thesis, fifteen (15) network architectures are developed ($N_1$, $N_2$ … $N_{15}$) with same number of input and output neurons (12, 2) but a varying number of neurons in the single hidden layer. The number of neurons in the hidden layer is varied from 5 to 20 increasing by a unit neuron for each network.

#### 5.3.2.3 k-fold validation

The k fold cross validation was tested with various values of k. The k classifier partitions the data into k sets. At every k stage, one set in turn is used as the test data and the remaining set(s) used as the training data. So for instance, if k = 3, the data set is divided into three, *A*, *B* and *C.* In stage 1, data set *A* is used to test the network and sets *B* and *C* are used as training data. In stage 2, data set *B* is used as the test data and *A* and *C* used as training data. This is repeated for stage 3. To select an optimum value for k, different values of k were used for a NN system with 12 inputs, the k-fold cross validation has the advantage that all the data sets are used as both training and test data.

#### 5.3.2.4 Training

The training of the neural network in this study is by supervised learning. Some other training techniques are as discussed in previous sections. Training in this case involves presentation of the features and target output to the network. In this case, there are 115 sets of data representing 59 non-cocaine data and 56 sets of cocaine data. In this report, the k-fold cross validation technique has been adopted where k = 5. This technique has the advantage of being superior to other algorithms especially for small data sets (Goutte, 1977). In using this technique, the entire 115 data was divided into five equal sets with twenty-three sets of data in each set. The data was then trained five times with four subsets of data and leaving out one subset each time to be used for validation and error computation. This is done for the different number of hidden units (between 5 and 20) and the mean squared error (mse) and regression ($R^2$) calculated. The result is tabulated below. The 'best' values for $R^2$ and mse values are the maximum and minimum values respectively for the number of neurons in the hidden layer.

The output of the neural network is trained to output a result of [1 0] for detection and [0 1] for non-detection.

#### 5.3.2.5 Error estimate

For every time the data is split into training and testing data, the error estimate ($e_i$) is calculated. This error is difference between the correctly

classified test data and falsely classified false data and the total error estimate after 5-fold validation is

$$\hat{e} = \frac{1}{N}\sum_{i=1}^{N} e_i, \qquad (5.16)$$

where N = 5. A confusion matrix showing the performance of each validation stage is computed at each fold.

## 5.4 Computation of feature values

The features below were calculated after normalisation of the raw data. The features were selected based on features which would best represent each spectra. At the end, twelve features were selected as explained below.

**Peaks (PK) -** After normalisation, the maximum (or peak value) intensity for all spectrum is unity. Therefore, the wavelength values at which these peaks occurred were selected as features as shown in the table below for one set of data (see Table 5.2).

**Table5.2 Table showing peaks of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | 0.2782 | 0.2784 | 0.2781 | 0.2576 | 0.2560 | 0.2574 |
| 100 µM | 0.2713 | 0.2713 | 0.2713 | 0.2505 | 0.2490 | 0.2504 |
| 250 µM | 0.2866 | 0.2864 | 0.2867 | 0.2658 | 0.2644 | 0.2658 |
| 500 µM | 0.2921 | 0.2921 | 0.2921 | 0.2713 | 0.2699 | 0.2713 |
| 1000 µM | 0.2921 | 0.2921 | 0.2922 | 0.2712 | 0.2700 | 0.2712 |

**Average change in intensity (AVI) -** The average change in intensity is the mean of the difference between intensity values corresponding to extreme wavelength values for each partition (see Table 5.3).

**Table 5.3 Table showing average change in intensity of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | -0.026 | -0.0470 | -0.0346 | -0.0404 | -0.0662 | -0.0456 |
| 100 µM | -0.0101 | -0.0293 | -0.0169 | -0.0229 | -0.0487 | -0.0281 |
| 250 µM | -0.0072 | -0.0261 | -0.0136 | -0.0196 | -0.0454 | -0.0246 |
| 500 µM | -0.0054 | -0.0250 | -0.0125 | -0.0186 | -0.0443 | -0.0236 |
| 1000 µM | -0.0067 | -0.025 | -0.0125 | -0.0186 | -0.0442 | -0.0236 |

**Steepest Slope (SS) -** The slope of a partition is the gradient at the mid-points of the partition. When the slope for all artitions are calculated, the steepest slope is the highest value of all the slopes (see Table 5.4).

**Table5.4 Table showing the steepest slope of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | 4.5143 | 2.5034 | 3.9747 | 2.7483 | 1.4224 | 2.1871 |
| 100 µM | 5.4937 | 3.5715 | 5.0431 | 3.8165 | 2.4907 | 3.2552 |
| 250 µM | 5.9126 | 3.9903 | 5.4620 | 4.2352 | 2.9095 | 3.6740 |
| 500 µM | 6.1401 | 4.2178 | 5.6894 | 4.4627 | 3.1369 | 3.9014 |
| 1000 µM | 6.1950 | 4.2571 | 5.7286 | 4.5020 | 3.1763 | 3.9406 |

**Maximum Negative Slope (MNS) -** After the slopes have been calculated as in *steepest slope* above, the maximum negative slope is the highest of all negative gradients calculated (see Table 5.5).

**Table5.5 Table showing maximum negative slope of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | -2.2753 | -1.6255 | -2.1560 | -1.9981 | -1.7982 | -1.9879 |
| 100 µM | -2.5228 | -1.8732 | -2.4034 | -2.2457 | -2.0458 | -2.2353 |
| 250 µM | -2.6315 | -1.9819 | -2.5122 | -2.3544 | -2.1546 | -2.3441 |
| 500 µM | -2.7310 | -2.0813 | -2.6118 | -2.4539 | -2.2541 | -2.4435 |
| 1000 µM | -2.7478 | -2.0914 | -2.6218 | -2.4640 | -2.2642 | -2.4537 |

**Most Positive Change in Intensity (MPCI) -** The change in intensity is calculated for each partition as before to give both positive and negative intensity changes (depending on the side of the spectrum the partition is). The most positive change in intensity is the highest value of all positive change in intensities (see Table 5.6).

**Table 5.6 Table showing most positive change in intensity of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | 0.6300 | 0.3564 | 0.5660 | 0.3912 | 0.1945 | 0.3115 |
| 100 µM | 0.7821 | 0.5085 | 0.7178 | 0.5433 | 0.3466 | 0.4633 |
| 250 µM | 0.8417 | 0.5681 | 0.7773 | 0.6029 | 0.4062 | 0.5228 |
| 500 µM | 0.8744 | 0.6004 | 0.8099 | 0.6353 | 0.4388 | 0.5554 |
| 1000 µM | 0.8819 | 0.6112 | 0.8205 | 0.6459 | 0.4494 | 0.5661 |

**Most Negative Change in Intensity (MNCI) -** As in the case of most positive change in intensity, after the change in intensity has been calculated for each partition, the most negative change in intensity is the highest of all negative changes in intensity or the least of all the changes in intensity (see Table 5.7).

**Table 5.7 Table showing most negative change in intensity of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | -0.5986 | -0.4278 | -0.5673 | -0.5256 | -0.4730 | -0.5231 |
| 100 µM | -0.6638 | -0.4929 | -0.6326 | -0.5908 | -0.5383 | -0.5881 |
| 250 µM | -0.6924 | -0.5214 | -0.6609 | -0.6195 | -0.5670 | -0.6166 |
| 500 µM | -0.7150 | -0.5476 | -0.6872 | -0.6456 | -0.5931 | -0.6429 |
| 1000 µM | -0.7230 | -0.552 | -0.6917 | -0.6500 | -0.5975 | -0.6473 |

**Average intensities in regions (AIR) -** For all the regions (6 regions), the mean of the intensity values in each region is also calculated. The values for average intensity for region 1 are shown in table 5.8.

**Table 5.8 Table showing most negative change in intensity of spectra at different concentrations**

| Concentration | Cocaine | Ketamine | Codeine | Amphetamine sulfate | Ecgonine methyl ester | Buprenorphine HCl |
|---|---|---|---|---|---|---|
| 25 µM | 0.1484 | 0.2755 | 0.1857 | 0.2399 | 0.3838 | 0.2682 |
| 100 µM | 0.0568 | 0.1839 | 0.0941 | 0.1484 | 0.2921 | 0.1765 |
| 250 µM | 0.0350 | 0.1621 | 0.0724 | 0.1265 | 0.2701 | 0.1547 |
| 500 µM | 0.0261 | 0.1531 | 0.0634 | 0.1177 | 0.2612 | 0.1458 |
| 1000 µM | 0.0260 | 0.1544 | 0.0647 | 0.1191 | 0.2625 | 0.1473 |

The above show the features calculated for each spectrum and which served as inputs to the neural network. The values for one set of data are shown in tables 3.2 to 3.8. In all, there are twelve (12) features representing twelve inputs to the network.

## 5.5 Determination of number of neurons in the hidden layer

Once the features have been identified and processed, the next step involves the development of the neural network model. The hidden layer is the middle layer of a 3-layer NN. It is important to find the optimum number of neurons in the hidden layer of the NN model. This is done in this dissertation using the k-fold cross validation.

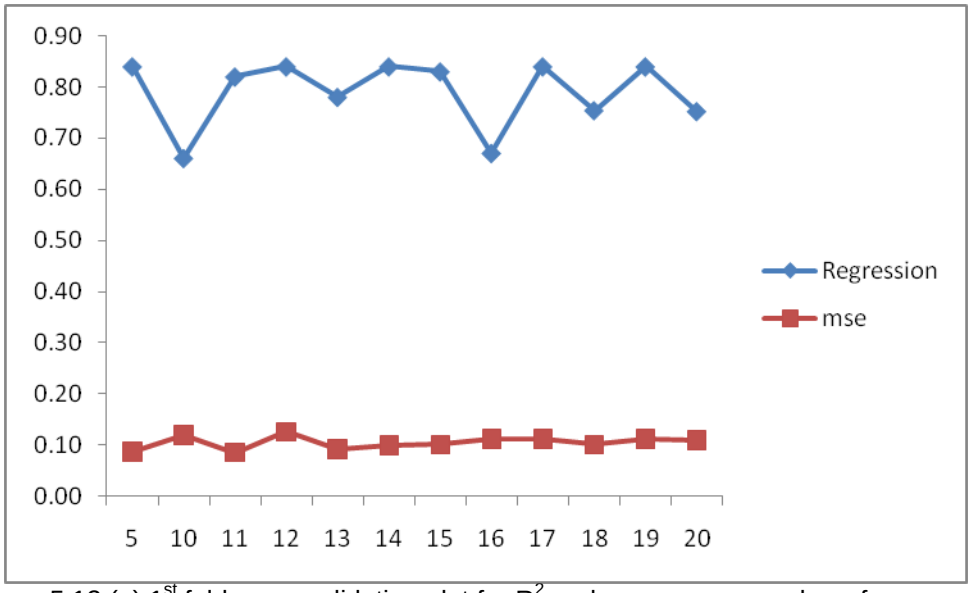



Figure.5.12 (a) 1st fold cross validation plot for $R^2$ and mse versus number of neurons in hidden layer

The set of data was divided into 5 subsets (A, B, C, D and E) as described in the previous section. k-fold cross validation involves selecting k-1 subsets of data as training data and one set of data as validation data. For the first

stage of the 5-fold cross validation, subsets A, B, C and D were used as training data and E subset used as validation data.

Figure 5.12 (a) shows the plot of $R^2$ and mean squared error which are used as performance criteria in this report, versus the number of neurons in the hidden layer. The $R^2$ values equivalent to 5, 12, 14, 17 and 19 hidden layers show the best values while hidden layer with 11 neurons is best in terms of the mse.

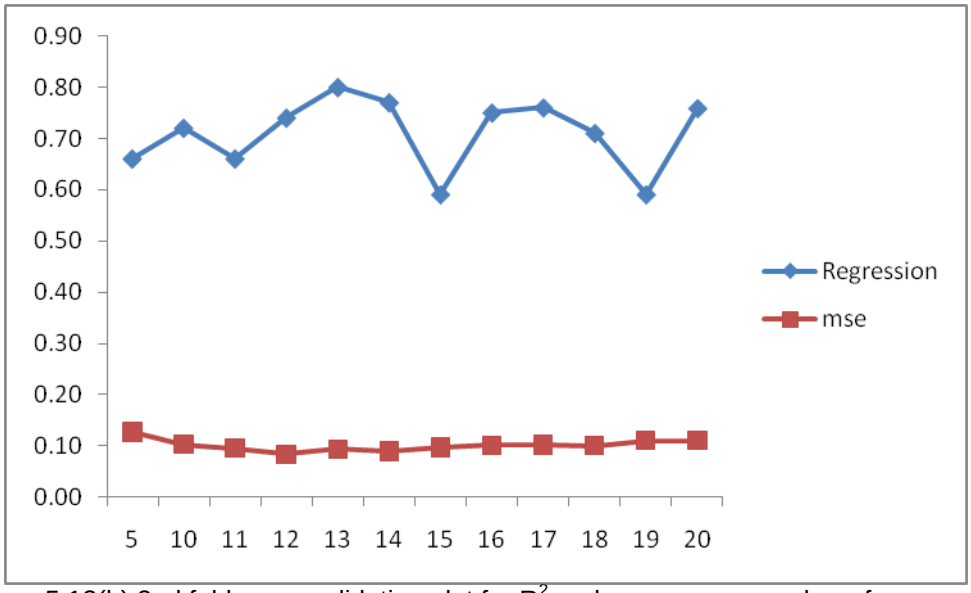



Figure.5.12(b) 2nd fold cross validation plot for $R^2$ and mse versus number of neurons in hidden layer

For the second stage, A, B, C and E were used as training data and D used as validation data. The result for the second stage of the 5-fold cross validation is as shown above in figure 5.12(b). Hidden layer with 13 neurons represent the best $R^2$ value while hidden layer with 14 neurons represent the best mse value with 0.0895 mean squared error.

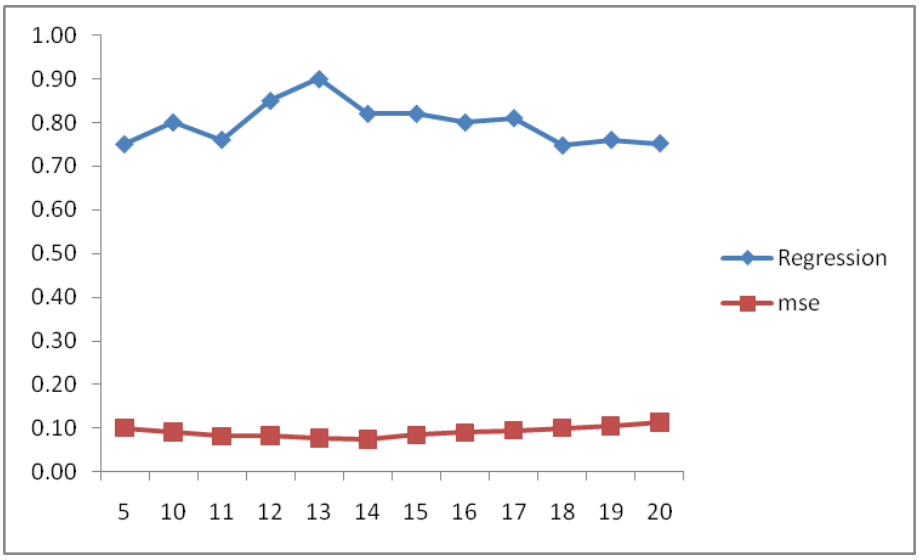



Figure.5.12(c) 3rd fold cross validation plot for $R^2$ and mse versus number of neurons in hidden layer

For the third stage of the of the 5-fold cross validation, subsets A, B, D and E were used as training data while C was used as validation data. The results are as shown in 5.12 (c) above. In this case, the plot shows best $R^2$ value when there are 13 neurons in the hidden layer and best mse value for when there are 14 neurons in the hidden layer.

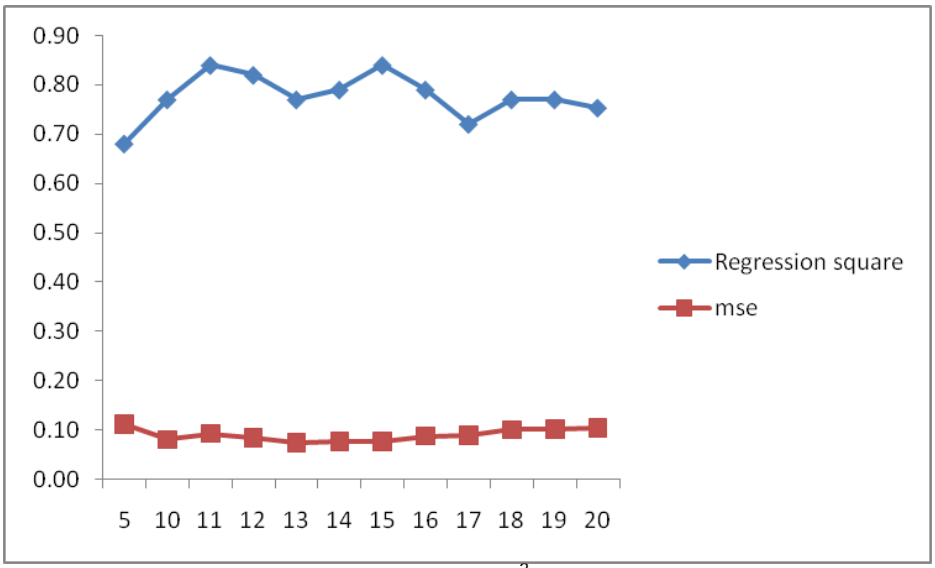



Figure.5.12(d) 4th fold cross validation plot for $R^2$ and mse versus number of neurons in hidden layer

In stage 4, subsets A, C, D and E were used as training data while subset B was used as validation data. Figure 5.12(e) shows the $R^2$ and mse plots against number of hidden layer neurons. The results show that hidden layer with 11 and 15 neurons give best $R^2$ results while hidden layer with 13 neurons give the best mse results.

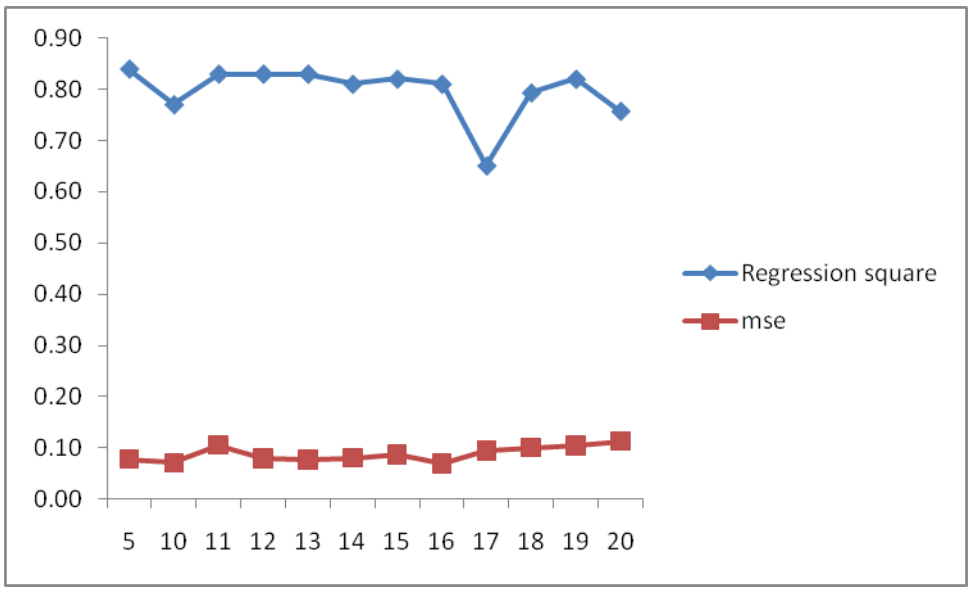



Figure.5.12(e) 5th fold cross validation plot for $R^2$ and mse versus number of neurons in hidden layer

The final stage of the validation process uses subsets B, C, D and E as training data and subset A as validation data. The results, shown in figure5.12(e) above shows that hidden layer with 5 neurons gives the best $R^2$ result while the lowest mse value is achieved when there are 16 neurons in the hidden layer.

Once the 5 stages for the 5-fold cross validation is complete, the mean of the mse and $R^2$ is calculated and plots against the number of neurons in the hidden layer is made. This plots are shown in figure 5.12 (f) and 5.12 (g).

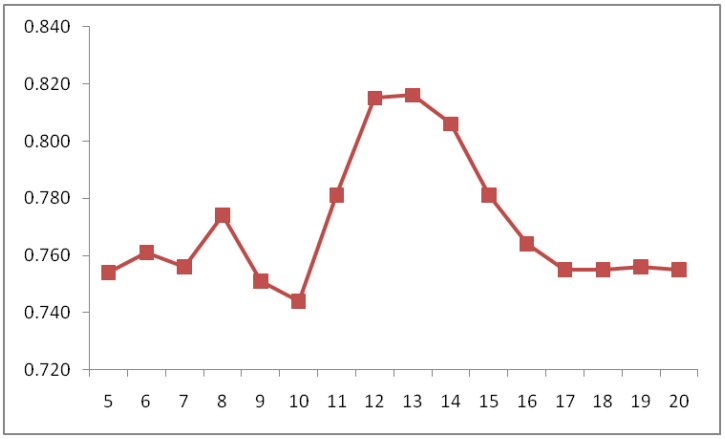



Figure.5.12 (f) Mean value of regression square for 5-fold cross validation against number of neurons in hidden layer

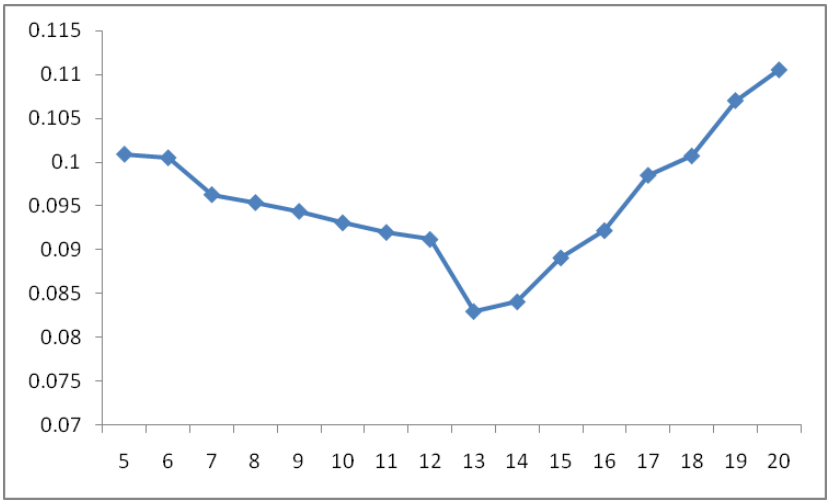



Figure.5.12(g). The average mse values for 5-fold cross validation against number of neurons in the hidden layer

From the mean mse and $R^2$ values as shown in figure5.12(f) and (g) above, it can be seen that the least mse occurs when there are 13 neurons in the hidden layer. In figure5.12, the highest $R^2$ values occur when there are 13 and 14 neurons in the hidden layer. As a result, 13 neurons are chosen as the ideal number of neurons in the hidden layer and the architecture for the

back propagation network is thus 12 – 13 – 1 representing 6 neurons in the input layer, 13 in the hidden layer and 1 in the output layer.

In addition to the number of neurons in the hidden layer, other parameters required for back propagation neural network include the learning rate and momentum. These parameters were chosen based on a trial by error method.

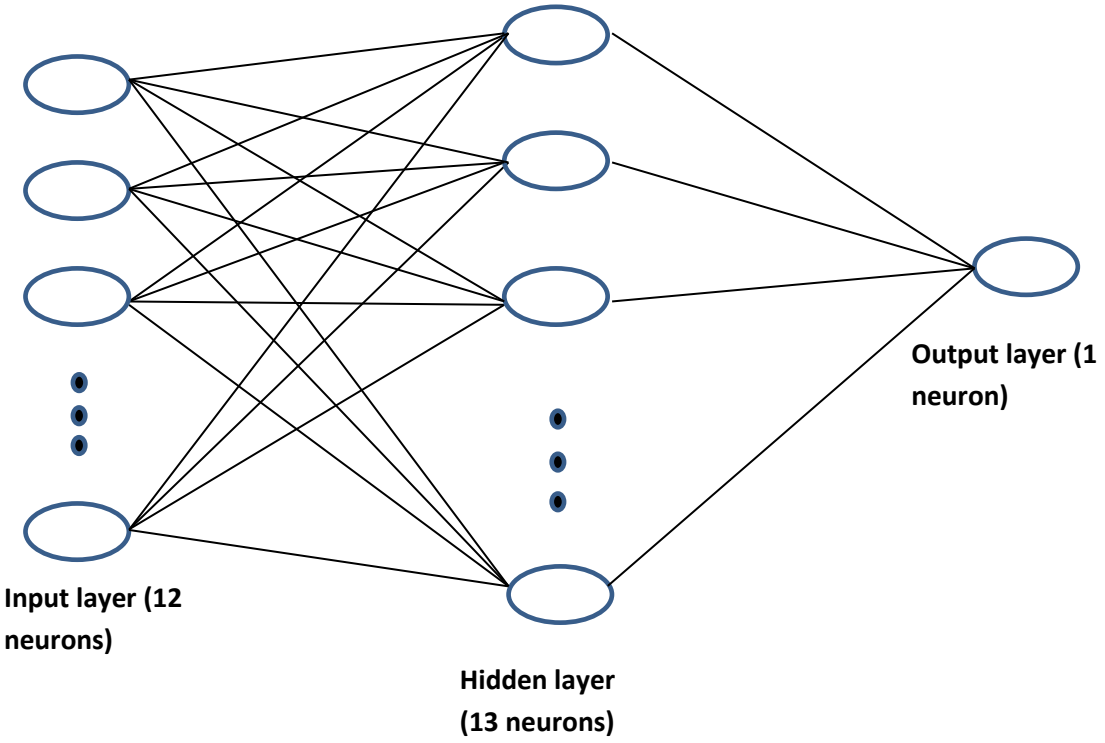



Figure.5.13 Final Neural Network architecture 12 – 13 - 2

## 5.6 Neural Network output results

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

**Test Mode**: Five-fold cross validation

Table 5.9 Confusion Matrix Representation for K-fold cross validation ANN for 1st and 2nd stages

| 312 iterations | | |
|---|---|---|
| 1st Fold cross validation<br>Confusion Matrix | | |
| **a** | **b** | **Predicted as** |
| 12 | 1 | a = cocaine |
| 1 | 9 | b = not cocaine |

**(a) Confusion matrix for 1st fold cross validation**

| 332 iterations | | |
|---|---|---|
| 2nd Fold cross validation<br>Confusion Matrix | | |
| **a** | **b** | **Predicted as** |
| 12 | 0 | a = cocaine |
| 0 | 11 | b = not cocaine |

**(b) Confusion matrix for 2nd fold cross validation**

The confusion matrix in table 5.9 shows the result of all 5 stages of the 5-fold cross validation process. The number of iterations which yield these results is also displayed. The values in the major diagonals show the correctly classified samples of both cocaine and not cocaine samples while the values off the diagonal indicate the confusion or incorrectly labelled samples. The false positive and true positive rates can be computed from the tables. For example, the true positive rate for the cocaine feature data is the ratio of correct cocaine prediction to total cocaine feature data in the data set.

The confusion matrix in Table 5.9(a) shows the result of the first stage of the 5-fold cross validation process. There are 23 test data consisting of 13 cocaine samples and 10 non cocaine samples. The table shows that the classifier correctly classified 12 cocaine features as cocaine and 9 not

cocaine features as not cocaine representing a detection rate of 91.3% and an error of 0.087.

In table 5.9(b), of the 23 sample feature data, 12 represent cocaine and 11 not cocaine data. The classifier at this stage performed better than in the first case detecting all 12 cocaine data as cocaine and all 11 not cocaine data as not cocaine representing 100% detection rate and 0 error.

Table 5.10 Confusion Matrix Representation for K-fold cross validation ANN for 3rd and 4th stages

| **315 iterations** | | |
|---|---|---|
| **3rd Fold cross validation Confusion Matrix** | | |
| **a** | **b** | **Predicted as** |
| 7 | 0 | a = cocaine |
| 0 | 16 | b = not cocaine |

**(a) Confusion matrix for 3rd fold cross validation**

| **433 iterations** | | |
|---|---|---|
| **4th Fold cross validation Confusion Matrix** | | |
| **a** | **b** | **Predicted as** |
| 8 | 0 | a = cocaine |
| 4 | 11 | b = not cocaine |

**(b) Confusion matrix for 4th fold cross validation**

Tables 5.10 (a) and (b) show the result of the third and fourth stages of the cross validation process. In 5.10(a) there is 100% classification rate for the 7 and 16 cocaine and not cocaine feature samples which also implies a 0% error. The iterations yielding these results are 315 and 433 iterations respectively.

The final stage, which is the fifth stage of the 5-fold cross validation process is shown in table 5.11 below. With 208 iterations at this stage, the ANN yields a classifier which detects 83.33% of the 12 cocaine samples and

correctly classifies 63.63% of the not cocaine samples. The error at this stage is thus 0.261. According to (5.17), the total error estimate is the average of all the error estimates. The error for the 5-fold cross validation is thus 0.104.

Table 5.11 Confusion Matrix Representation for K-fold cross validation ANN for 5th stage

| **208 iterations** | | |
|---|---|---|
| **5th Fold cross validation**<br>**Confusion Matrix** | | |
| **a** | **b** | **Predicted as** |
| 10 | 4 | a = cocaine |
| 2 | 7 | b = not cocaine |

Table 5.12 shows a summary of the results of the 5-fold cross validation. The summary table includes the True positive rates, the false positive rates, the precision and the area under the Receiver Operator Characteristics (ROC) curve.

**Table 5.12. Summary table of the results of the 5-fold cross validation ANN process.**

| | | | | | |
|---|---|---|---|---|---|
| Correctly classified instances | 103 | 89.57% | | | |
| Incorrectly classified instances | 12 | 10.43% | | | |
| **Detailed Accuracy by Class** | | | | | |
| | **TP Rate (%)** | **FP Rate (%)** | **Precision** | **ROC Area** | **Class** |
| | 87.5 | 8.5 | 90.72 | 0.881 | Cocaine |
| | 91.5 | 12.5 | 88.52 | 0.881 | Not Cocaine |
| **Weighted Avg (%)** | **89.55** | **10.55** | **0.961** | **0.933** | |

As can be seen in table 5.12, the percentage of incorrectly classified data (cocaine and not cocaine samples) is 10.43%. A look at this percentage will indicate that the ANN system has performed well. Looking at the True Positive rate, the system detects 87.5% of the cocaine feature data correctly and 91.5% of the not cocaine feature data correctly while it incorrectly classified 8.5% and 12.5% of the not cocaine and cocaine data as cocaine and not cocaine respectively. The **precision** which is the **ratio of the true positive rate to the sum of the true positive rate and false positive rate** represents the probability that cocaine is detected given that the sample is actually cocaine feature data and in this case gives a 0.91 value.

The correlation coefficient, cc, is sufficient to estimate the accuracy of the NN process (Matthews, 1975). This is given by:

$$cc = \frac{PN - OU}{\sqrt{(N+U)(N+O)(P+U)(P+O)}} \qquad (5.17)$$

where P is the number of correctly predicted cocaine samples, N is the number of correctly predicted non-cocaine samples, O is the number of false-positives, and U is the number of false-negative predictions. The value of cc is in [-1, 1], where cc= 1 indicates perfect prediction. From the above, cc is calculated to be 0.8.

## 5.7 Determination of feature threshold

The twelve features selected and validated in previous sections are used as input into the data fusion system. Each of the features make decisions on cocaine detection or otherwise. The decisions are made based on an

empirically determined threshold. The process of determining the threshold is discussed in this section. It should be noted that every feature has its own unique threshold.

The threshold determination is an important part of feature selection. This threshold is used to separate data received from the sensor into cocaine/non-cocaine data. To determine the threshold, key issues considered include:

Low False alarm: The false alarm (also called false positive) refers to the features ability to correctly identify non-cocaine data. If it identifies non-cocaine data as cocaine data, this is regarded as a false alarm. In general, false alarms are not ideal for any system as it tends to reduce operator confidence in the system. Therefore, it is important that each feature decision results in as low false alarms as possible.

High true positive: This is also known as correct detection and is the features ability to correctly identify cocaine data and report them as such. Being able to correctly identify cocaine data that is, having a high true positive value is ideal and therefore a vital factor to be considered in selecting the threshold value.

Prior to selection of the thresholds, the data measurements are pre-processed as discussed in section 5.1. After pre-processing, using the MatLab software, a model was developed which accepts data measurements and extracts feature values from the incoming data thus separating data

measurements by their features. For each feature, sample thresholds are used to divide the data into cocaine and non-cocaine data. The true positive rate (TPR) and the false alarm rate (FAR) are then computed for each threshold. The 'best' threshold is the one which gives a balance of low false alarm and high true positive rates. In figure 5.14 below, a normal distribution curve is produced for cocaine and non-cocaine data. The outline in green is for cocaine data while the outline in red is for non-cocaine data. These curves give a visual aid towards determination of a threshold. As can be seen below, the curves overlap between data values of about -0.04 and about 0.01. The overlapping area suggests areas of conflict. Depending on where the threshold is located, this area will determine the false alarm rate and the false negative rate (false negative are data points which are truly cocaine but which the system detects as non-cocaine. They are also known as missed detections).

**Feature1 (Average change in intensity)**

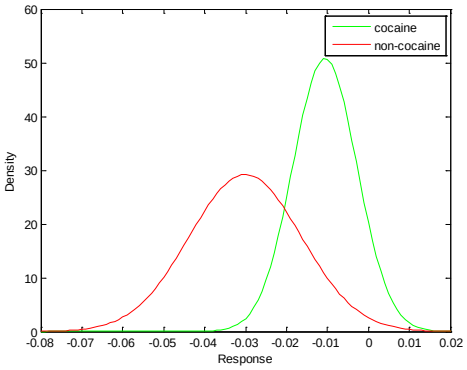



Figure 5.14 Normal distribution curve for classification using feature 1

**Table 5.13 TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| **-0.015** | **0.8** | **0.15** |
| -0.025 | 0.85 | 0.61 |
| -0.03 | 1 | 0.76 |

Taking the red outline (non-cocaine), if the threshold is set for example at say, -0.015, every data point detected to the left of the threshold, i.e. values below -0.015 will be correctly identified as non-cocaine. However, for the area under the curve but to the right of the threshold (i.e. above -0.015), values in this range will be false classified as cocaine and is the false alarm rate. Similarly, for the green outline, values to the right of the threshold value of -0.015 will be correctly identified as cocaine and measurements to the left of the threshold will be falsely identified as non-cocaine (missed detections). Threshold values must be maintained for both cocaine and non-cocaine data and cannot be changed in between. It is therefore imperative that a threshold value which gives optimum values of false alarm and true positive rates should be selected.

With feature one (Average Change in Intensity), the selected thresholds and their corresponding TPR and FAR values are shown in table 5.13. it can be seen from this table that a threshold of -0.03 will give 100% true positive rate but a high false alarm rate of 76%. Although the TPR is acceptable, the FAR is not acceptable and so this threshold value is not acceptable. Using the threshold value of -0.025 again gives a ‘satisfactory’ TPR value of 85% but a non-satisfactory FAR value of 61%. Finally, using a threshold value gives a TPR/FAR pair of 0.8/0.15. Compared other threshold values, this is the best pair based on low FAR and high TPR values therefore, the threshold selected is -0.015.

The following thus represents an algorithm for selection of threshold for features for cocaine detection:

Step 1: Collect set of data

Step 2: Set a pre-determined TPR/FAR pair to serve as reference

Step 3: Select a pseudo-random threshold value which would normally be a midpoint of the data collected in Step 1.

Step 4: Calculate TPR and FAR values based on this threshold value

$$TPR = \frac{\text{Number of cocaine data identified as cocaine}}{\text{Total number of cocaine data}}$$

$$FAR = \frac{\text{Number of non-cocaine data identified as cocaine}}{\text{Total number of non-cocaine data}}$$

Step 5: Compare TPR/FAR values in Step 4 with values in Step 2.

Step 6: If values above are satisfactory*, exit, if not repeat steps 3 and 4.

* Satisfactory implies close enough to the pre-determined values within a specified range.

The algorithm above was used in selecting thresholds for features 1 to 11 and the data is presented in Appendix C.

## 5.8 Chapter Summary

In Chapter 4, the raw data from the fibre optic sensor developed for the detection of cocaine was presented. This raw data was normalised in this chapter. The process of extracting twelve features used as inputs was analysed and the results also displayed. Once the features had been extracted, they were fed as inputs into a neural network system. We also

presented a method for deciding on the optimum number of neurons in the hidden layer and settled on a 12 – 13 – 1 network. Eventually, the result of this network on the input data using a 5-fold cross validation system has been displayed. The results show that the network gives a weighted true positive rate of 89.55% and a false positive rate of 10.55%. In addition, the probability of the system detecting cocaine from a cocaine sample is 90.72% and the probability of deciding not cocaine from a not cocaine feature sample is 88.52%. Finally, the correlation coefficient was calculated to be 0.8 which represents a good prediction. These probabilities show that the 5-fold cross validation system is a good classifier for cocaine using the fibre optic sensor.

Feature extraction and selection is an important stage in the pre-processing of data for data fusion. It is at this stage that the raw data is investigated for unique features which will efficiently represent the structure of the data. There are different feature selection techniques. Their use however, is dependent on the model. This chapter looked at the overall pre-processing stage which involves data normalization and feature extraction. One challenge of the fibre optic sensor data is its inability to identify individual analyte if the concentration is not known *a priori*. For the purpose of this thesis, this is not suitable as real life application in cargo screening does not give the luxury of having this prior knowledge. To solve this issue, the data is first normalized and then features are extracted. Normalization of the data was performed to remove the effect of the concentration dependent intensity.

# CHAPTER 6

# IMPLEMENTATION OF DATA FUSION TECHNIQUES

This chapter presents implementation and results obtained in the detection of cocaine using both the Bayesian and Dempster-Shafer algorithms for data fusion. The model used is an adaptation of that developed in chapter 4. A step by step performance of each feature in the classification process is presented and an algorithm for the Bayesian and Dempster-Shafer fusion of multi feature data over a time period is then developed. The performance of each feature in the classification process is then compared with the performance of the fusion process and the results presented at the end of the chapter.

## 6.1 Introduction

This chapter is divided into three sections. The first section details the steps involved in Bayesian statistics and then discusses the parameters involved in Bayesian statistics. The next section then looks at the features selected from

the previous chapter and their performance in the classification of data into cocaine and non-cocaine data. The final section involves the fusion of the decisions at feature level with a presentation and analysis of the results.

## 6.2. Bayesian Statistics

An overview of the Bayesian technique has been discussed in chapter 2 of this thesis. In this section, we will take a detailed look at the Bayesian technique and how it can be applied to improving results of detection.

The Bayesian fusion technique stems from the carried out by Thomas Bayes in 1760 and published in 1763 (Hall, 1992). He created a system of mathematical inference on how to predict the posterior probability distribution using the likelihood of an event happening and the prior probability. In simple terms, the 'Bayesian Inference' as it is called, uses what was previously known about and additional information to update/predict the likelihood of an event happening. It is the same way a blindfolded person may decide on the identity of a fruit he's holding based on prior knowledge of what fruits have the same shape with that which he/she is holding. As more information is supplied to him as regards the colour, texture, etc. of the fruit, he is able to perform an update and eventually provide an intelligent guess on what fruit it may be based on prior information and additional data.

Mathematically, the fusion of data using Bayesian inference can be represented as follows. If the aim is to detect the presence or absence of say 'E' given "$x_1$" and "$x_2$" as data received from one sensor over two time

phases or simultaneously from two sensors and assuming that the data retrieved from independent sources, then according to Bayes,

$$p(E|x_1, x_2) = \frac{p(x_2|E,x_1)\cdot p(x_1|E)\cdot p(E)}{p(x_1)\cdot p(x_2)} \qquad (6.1)$$

where

$p(E|x_1, x_2)\ is\ the\ posterior\ probability\ also\ called\ the\ a\ posteriori\ probability\ and$

$p(x_i) is\ the\ prior\ probability\ for\ i = 1,2$

$p(x_2|E) is\ the\ probability\ that\ the\ event\ E\ will\ provide\ x_2\ data.$

where we have more than two sensors, the equation can be expanded as follows

$$p(E|x_1, \dots, x_n) = \frac{\prod_{i=1}^{n} p(x_i|E)}{\prod_{i=1}^{n} p(x_i)} * p(E) \qquad (6.2)$$

Suppose however, that there is more than one possible state. That is, the sensors can detect for more than one substance as in this case, cocaine and some other compounds therefore, say there are *k* numbers of possible compounds, then Bayes describes two rules to determine how to make a decision – *Maximum a posteriori* probability rule and *Maximum likelihood* Rule.

### 6.2.1 Maximum *A Posteriori* Probability Rule

The maximum a posterior rule for making decisions dictates that the event with the highest joint probability value is selected when there are multiple a

posteriori probabilities to choose from. In the case of EQ.6.2, assuming the sensors can detect more than one compound *E*, for all $E_i$*s,* the highest value of $p(E_i|x_1, \dots, x_n)$ when all *E*s have been calculated for individual sensors is selected (Soriguera, 2011).

$$E_k = \ arg\ max_{1<i<k}\{p(E_i|x_1, \dots, x_n)\} \qquad (6.3)$$

#### 6.2.1.2. Maximum Likelihood

In the case of the maximum likelihood rule, like its name suggests, the decision is aligned towards the event with the highest likelihood function. i.e.

$$E_k = \arg max_{1<i<k}\ \{\prod_{j=1}^{n} p(x_j|E_i)\} \qquad (6.4)$$

Accordingly, Soriguera (2011) maintains that both the maximum a posteriori and maximum likelihood rules converge to the same decision when the prior probabilities are equal i.e. when *p(E) = 1/k*. This situation usually arises when there is no prior knowledge of the system and is also known as the *principle of indifference*. This will be explained further in the next section.

The maximum a posteriori probability rule is utilised in this thesis.

### 6.2.2. Assigning Probabilities

Bayesian inference as explained in previous sections uses new evidence/data and prior probability to compute the posterior probability. From equations 6.1 and 6.2 above, it can be seen that the posterior is dependent on the conditional probabilities P(xi|E) that xi occurs given E and on the prior

probability p(E). Therefore, key to any Bayesian application is the ability to evaluate the conditional probability and the prior probability. The process of evaluating these probabilities in general and also as used in this dissertation is shown in subsequent sections.

#### 6.2.2.1 Conditional probability

Let *X* be a random variable with discrete probability distribution *p* dependent on parameter theta. The likelihood function of theta given that *x* of *X* occurs is given as follows:

$$p(x|\theta) = p_\theta(x) = P_\theta(X = x) \qquad (6.5)$$

From equation 6.5, $p(x|\theta)$ is the likelihood probability

#### 6.2.2.2 Prior Probability

The difficulty in the selection of a prior is one of the 'weaknesses' ascribed to the Bayesian Inference (Hall, 1992). The prior probability or *a priori information* is the probability that expresses uncertainty about an event. It is the probability prior to the observation of new data. For example, if out of ten containers searched at a port, one contains a particular contraband, the probability of that a container drawn at random will contain that contraband is 0.1. This probability is the prior probability.

As shown in EQ.6.1, the product of the prior probability and the likelihood function before normalization gives the posterior probability. Thus the decision on what values to assign to the prior probability will greatly affect the

posterior probability. A biased prior will imply a non-correct posterior probability. The dependence of the prior probability on available information also then means that deciding on the correct value for a prior probability is a difficult task as prior information is usually not always available. Punska (Punska, 1999) suggests that where enough information is not available for the determination of a prior, the selected prior probability function should be a wide and flat function in comparison to the expected likelihood function. In general, depending on the situation, there are three ways of selecting a prior. These are:

1. Subjective (the user using his/her personal degree of belief)
2. Objective and Informative priors (the use of past data from experiments) which can be from statistical hypothesis tests such as the chi-squared (Osoba et al., 2011)
3. Objective and Un-informative priors (assigning equal priors to the events e.g. Jefferys prior, uniform distribution).

#### 6.2.2.3 Informative Priors

When there is enough information to determine the exact prior information about an event, the prior information determined from such is said to be informative. Usually, a normal distribution is used to represent informative priors (Figure 6.1). In addition, where the form of the present model is identical to the form of the previous model, then the posterior probability from the previous model may be used as the prior probability for the present model. This has the effect of improving on the precision of the posterior

distribution with more update from previous models as long as the condition of similarity is maintained.

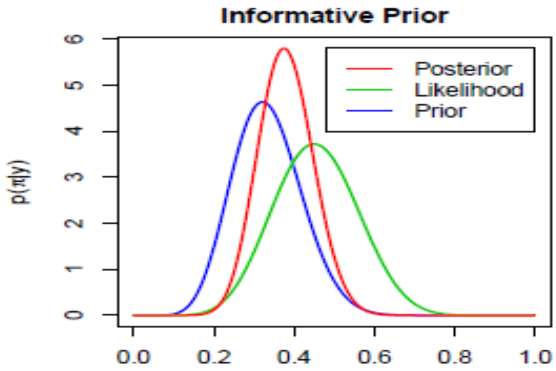



Figure 6.1 Posterior, Likelihood and Prior probabilities for informative priors

#### 6.2.2.4 Non-Informative Priors

In many common situations, little or nothing is known about the 'past' of the system. If nothing is known about the system, a non-informative prior distribution is used (Figure 6.2). The aim of non-informative priors is to reduce the effect of the prior on the final posterior probability thus allowing the posterior to be solely dependent on the data.

Obviously, the prior describes the extent of one's knowledge about the values of the parameter before examining the data. It is not in all situations where prior information is available and can be quantified. In cases where prior information is not available, it is advisable to choose a prior probability function which is flat relative to the expected likelihood function.In iterative fusion cases where the fusion process involves fusion of data over time, the effect of starting with a non-informative prior is balanced by updating each

step with the posterior from the previous iteration (*k-1*) serving as the new prior at the present step (*k*).

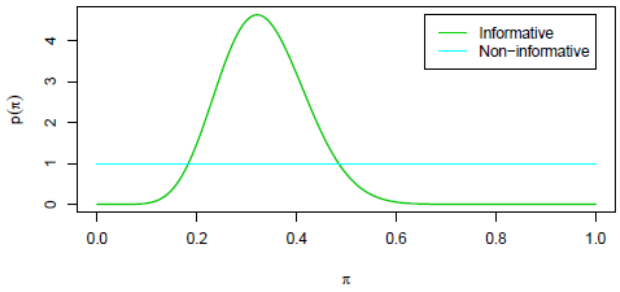



Figure 6.2 Informative and non-informative priors

**The Bayesian detection architecture**

In this dissertation, an optical sensor developed in the Chemistry Department at City University London by Nguyen (Nguyen et al., 2010), is used to collect data and to test the data collected for the presence or otherwise of cocaine. Then, as described in previous chapters, the spectrum plotted by the data collected can only be interpreted to be from a cocaine sample or not if the concentration of the sample tested is known *a priori*. In the laboratory, this may not be a problem as samples are usually labelled with the respective concentrations. However, in real life, the concentration of the sample is not usually known and it is thus important that the sensor be able to tell if the sample is cocaine or not without knowing the concentration of the sample. In Chapter 6 of this dissertation, a neural network architecture was developed to work as a feature selection model. The network successfully detected the presence and otherwise of cocaine by extracting features from the data input into the network.

The concept of Data Fusion has already been explained in Chapters 1 and 2 and Bayesian algorithm has been explained in previous sections of this chapter. In the following sections, the Bayesian data fusion approach and algorithm (as explained in Chapter 4) will be implemented and results will be displayed.

**Table 6.1 Confusion matrix**

| | | **Predicted Interaction** | |
|---|---|---|---|
| | | **YES** | **NO** |
| **Real Interaction** | **YES** | True Positive/True Detection | False Negative/Missed detection |
| | **NO** | False Positive/False Alarm | True Negative/ |

### 6.2.3 Data Fusion using Bayesian Inference

Data fusion is the combination of data from multiple sources with the aim of achieving results which will be more accurate in terms of lowering of false positives and/or improving on true positives compared to results obtained from individual sensors. In addition, data can also be fused from a single sensor over time. In this case, successive sets of information are added together over time. In this dissertation, we have employed two methods of fusion of data. One method analyses information from two similar sensors independently and fuses their information together while the second method involves analysing and combining information from a single sensor over time.

#### 6.2.3.1 Sensor data fusion

Also known as single sensor tracking or filtering (Koks & Challa, 2005), this implies the fusion of successive information or sensor measurements over time. The algorithm used for single sensor tracking in this dissertation is as follows:

1. Sensor collects data ($X_t$) at time *t*
2. Data is pre-processed (normalising and cleaning)
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 decision made by each feature is calculated
5. Sensor collects ($X_{t+1}$) data 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 decisions 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* is the posterior probability from time *t*.

The prior at the first stage of the fusion process is non-informative. This is because there is no prior information available. In the application of the sensor in cargo screening, future works should examine the possibility of the

operator determining a prior based on other information such as origin of cargo, history of shipper, etc. In addition, this information, if exhaustive, can be used to automatically generate a prior distribution based on the data described above. For this research though, the prior probability at the initial stage is non-informative. However, when the posterior is calculated and a decision made, there is still enough adequate 'power' at the disposal of the operator to decide if the container should be searched or not based on the decision of the sensor system. The posterior probability after the first stage is assigned as the prior probability in the second stage.

The block diagram of figure 6.3 represents the fusion process used in this research. $S1_t$ and $S1_{t+1}$ represents the same sensor but with data supplied at times $_t$ and *t+1*. Data is collected from the sensor at time *t* and a decision is made based on the fusion of individual decisions based on features extracted from the data. A probability matrix containing the true positive rate (TPR), false alarm rate (FAR), True negative and false negative rates is also created. Depending on the resulting probability matrix values, the operator decides if the decision made by the system is satisfactory. If it is, the process ends here however, if not satisfactory, the operator can then take another data sample running the process again (Steps 5 to 8) above and the result fused with the posterior probability resulting from the first data sample. The system thus uses this fusion process to provide a decision on the detection or otherwise of cocaine (see figure 6.3).

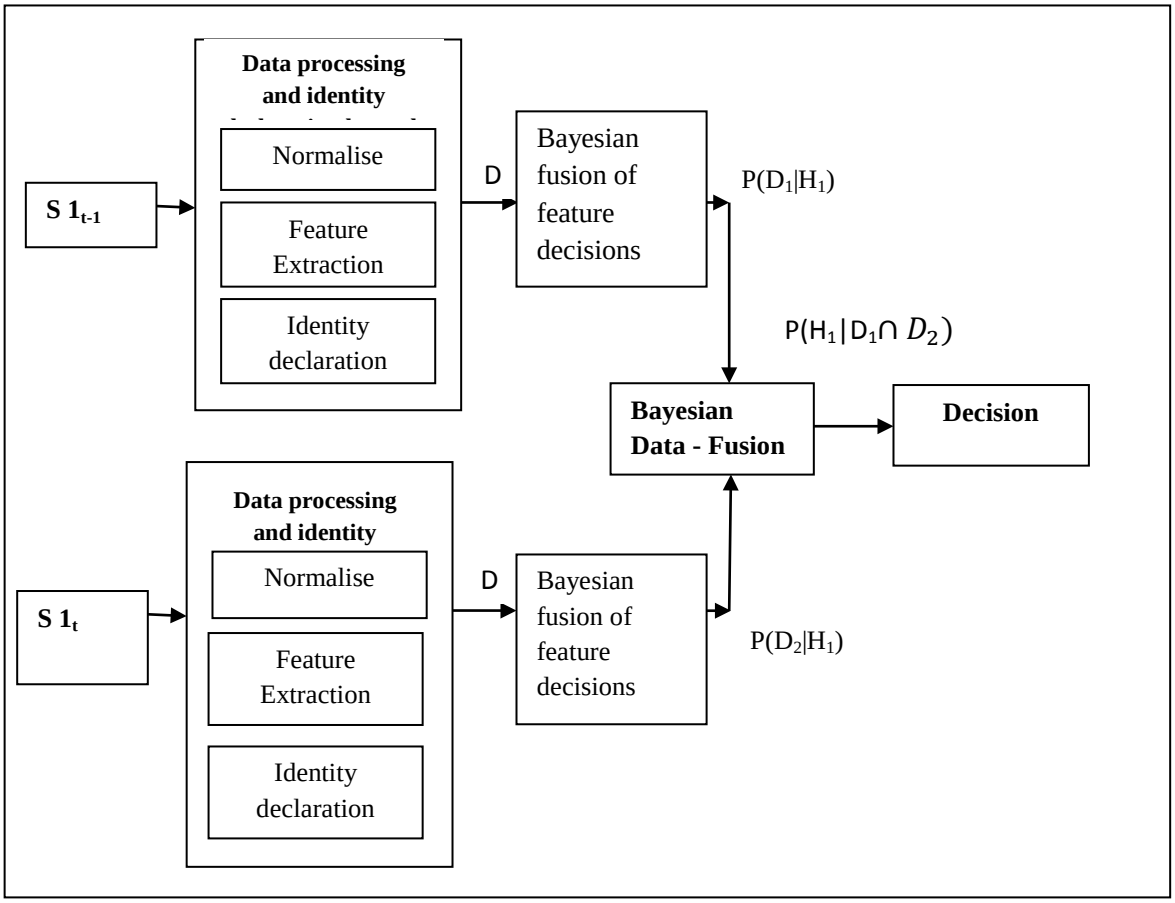



Figure 6.3 Block diagram for the data fusion model

## 6.3 Classifier Output for each feature

The results are displayed in forms of tables and figures (see Appendix C). The figures for each feature data show the discrete probability distribution for each feature data and both cocaine and not cocaine samples. From these, the spread of the data for each sample can be visualised.

The result also shows a figure of the normal distribution approximation for the data. The *central limit theory* states that for a sufficiently large data set of independent random variables with finite mean and variance, the posterior probability distribution can be approximated by the normal distribution (Clarke & Barron, 1990). In this section, using the each of the twelve features

used in this dissertation, the author has plotted normal distribution plots for cocaine and non-cocaine data. In addition, the confusion matrices containing the false positive and true positive rates are presented on tables in addition to the varying true positive and false positive values for varying threshold. The 'best' threshold for each feature is then selected. This best is based on a balance between high true positive and low false alarm. It should be noted that the confusion matrix values are calculated directly from the data. The normal distribution figure shown is only displayed to ease the explanation on how the confusion matrix values are evaluated.

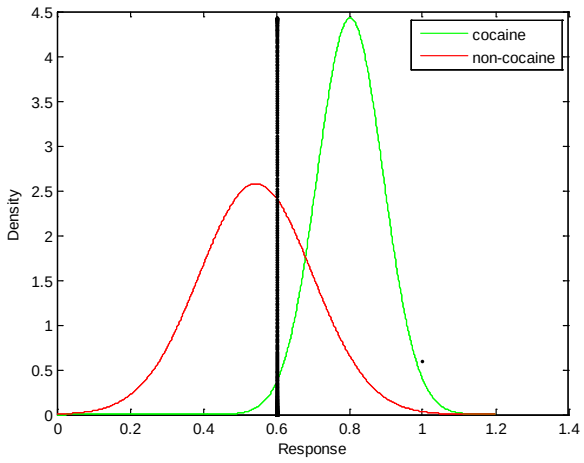



Fig. 6.4 Normalised sample cocaine and non-cocaine data

Figure 6.4 above represents the probability density function (pdf) for one of the twelve features extracted from the data supplied by the optical fibre sensor plotted using the Matlab software. The green plot represents feature data from cocaine sample while the red plot represents feature data from non-cocaine sample. The black straight line that runs from 0.6 upwards is the

decision line or **threshold**. Values to the right of the decision line indicate presence of cocaine while values to the left indicate non presence of cocaine. Any value under the green plot to the right of the decision line interprets as **true positives** while any values under the green plot but to the left of the decision line indicates **false negatives** or missed detection. For the red plot, all values to the right of the threshold and under the red plot are feature data classified for cocaine detection but are apparently wrong decisions and are known as **false positives** or false alarms. The values to the left of the threshold under the red plot are decisions correctly classified as non-detections and are known as **true negatives**.

From the above, depending on the position of the threshold, the posterior probability for all the 12 features is calculated. In addition, the receiver operator characteristic (ROC) curve which shows the relationship between the true positive and false positive rates is also developed. Appendix B shows the ROC curve for all the selected features.

To evaluate the performance of each feature in detection or non-detection of cocaine, we have divided the data into training and testing data in the ratio 0.7:0.3 respectively. Of the 30% test data, 53% are not cocaine data while the remaining 47% are cocaine data. These were all ‘pseudo’ randomly selected (it was ensured that both sets of data contained at least one of each type). The training data is used to evaluate the threshold or cut-off point and using that threshold, the test data is inputted to assess the performance of threshold chosen and overall feature (see Appendices C and D).

### 6.3.1 Results

**Test Data**

The test data consists of data that were not used during training. There are a total of 39 sample data consisting of 18 cocaine data and 21 non cocaine data. The individual features were used in detecting for the presence or otherwise of cocaine and a summary of the results in terms of correctly and incorrectly classified data is shown in table 6.2 below. The confusion matrix performance for each feature is presented in Appendix D.

**Table 6.2 Percentage classification rates for all features**

| Feature | % correctly classified | %wrongly classified |
|---|---|---|
| AVI | 90 | 10 |
| SS | 87 | 13 |
| MNS | 79 | 21 |
| MPCI | 87 | 13 |
| MNCI | 87 | 13 |
| PK | 74 | 26 |
| AIR1 | 87 | 13 |
| AIR2 | 87 | 13 |
| AIR3 | 77 | 23 |
| AIR4 | 79 | 21 |
| AIR5 | 82 | 18 |
| AIR6 | 87 | 13 |

While feature 12 aligns with previous performance rates for features 2,4 and 5, feature 11's 82% correct classification rate is third behind performances of

features 2, 4, 5, 7 and 8 with 87% and feature 1 with 90% correct classification rate.

Overall, in terms of correctly identifying which samples are cocaine and which are not, feature 1 performs best with 90% correct classification rate. Features 2, 4, 5, 7 and 8 follow second with 87% correct classification rates while feature 11 follows in third position with82% correct classification rate.

**Table 6.3 Summary confusion matrix table for both training and test data**

| Feature | Training Data | | Test Data | | AUC (%) |
|---|---|---|---|---|---|
| | FAR | TPR | FAR | TPR | |
| AVI | 0.15 | 0.80 | 0.05 | 0.83 | 87 |
| SS | 0.0976 | 0.70 | 0.095 | 0.83 | 91.46 |
| MNS | 0.1463 | 0.80 | 0.05 | 0.61 | 87.56 |
| MPCI | 0.1463 | 0.80 | 0.095 | 0.83 | 90.49 |
| MNCI | 0.1951 | 0.80 | 0.095 | 0.83 | 86.34 |
| PK | 0.2927 | 0.6250 | 0.05 | 0.61 | 73.96 |
| AIR1 | 0.1951 | 0.80 | 0.095 | 0.83 | 91.22 |
| AIR2 | 0.1707 | 0.80 | 0.095 | 0.83 | 90 |
| AIR3 | 0.1463 | 0.60 | 0.095 | 0.61 | 86.83 |
| AIR4 | 0.4390 | 0.80 | 0.24 | 0.83 | 79.02 |
| AIR5 | 0.0976 | 0.60 | 0.05 | 0.67 | 82.44 |
| AIR6 | 0.2927 | 0.80 | 0.095 | 0.83 | 83.51 |

The tables above show the summary of results from the feature classification above. The false alarm and true positive values for each feature

classification is shown for both training and test data. All features have varying FAR/TPR pair performances. As stated earlier, the overall aim of this system is to reduce the FAR with as high as possible TPR value. A close look at the table will show that Features 1, 2, 4, 5, 7, 8, 10 and 12 perform well in terms of test data TPR values with approximately 83% success rate. However, of all the 7 features with good TPR rate, features 1 and 11 have the best FAR value of 0.05. Again, with a FAR value of 0.05, features 3 and 6 provide low false alarm rates but their corresponding true positive rates of 0.61 respectively make them poor classifiers. Finally, the area under the curve values represents the area under the receiver operator characteristics curve. This curve is the plot of false alarm rates against true positive rates with varying thresholds (Appendix C). The ideal curve has an area of 100% with a sharp corner on the far left. The farther away from the corner the curve tends, the lower the area of the ROC curve and thus the lower the TPR/FAR acceptable pair rates.

Therefore, with a TPR/FAR pair of 83% and 5%, feature 1 is the best classifier from all 12 features. Overall, feature 1 gives a correct classification rate of 90%. In addition, feature 1 also gives a good AUC curve value of 87%.

### 6.4 Fusion of feature classification data

The results above give the performance of individual features in the detection of cocaine. Although the feature values are calculated from a single sensor,

they could also be seen as extracted from different sensors. Each feature performs with varying results depending on the threshold set for the feature.

From the two tables, we can see that features 1, 2, 4, 5, 7, 8, 10 and 12 give generally good TPR results for detection of cocaine while features 1 and 11 give good FAR results. The best feature classification has been given as feature 1 with TPR-FAR rates of 0.83 and 0.05. In general individual features do not give satisfactory results for detection or non-detection. To achieve a higher accuracy, the results from the individual features are fused together and Bayesian inference is used in this chapter as the preferred method for fusing the data.

### 6.4.1 Feature Fusion

The system developed for this project involves a robot ferret carrying the sensor(s) and moving around in a container. The sensor(s) gathers data over time and analyses the data for detection of cocaine in real time.

When the sensor is used to test an analyte, it generates a spectrum which is a plot of the wavelength versus intensity of the tested analyte. The spectrum then pre-processes the data and extracts features as explained in chapter 3. Using the steps described in the immediate previous section, the posterior probability of cocaine detection and non-detection for each feature is declared. The results declared by each feature are independent of each other as the only common factor between all features is the state. The data is analysed at every time t with the posterior probability from the previous iteration working as the new prior probability as given by EQ 6.6(a) below (Pangop et al., 2003).

Expanding on EQ.6.1, given an observation $P_{obs}$, the probability that it is a cocaine feature is given by

$$P(C|X_{obs}) = \frac{P(X_{obs}|C)*P(C)}{P(X)} \text{ and} \qquad (6.6a)$$

$$P(\sim C|X_{obs}) = \frac{P(X_{obs}|\sim C)*P(C)}{P(X)} \qquad (6.6b)$$

The $P(C|X_{obs})$ is the posterior probability of declaring for cocaine given the observation $X_{obs}$ , $P(X)$ is a normalizing factor that ensures that the sum of the *a posteriori* probabilities sum up to unity and is 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.

**Table 6.4 Posterior probabilities of features 1 - 12**

| Features | $P(T^+\|C)$ | $P(T^+\|\sim C)$ | $P(C\|T^+)$ | $P(\sim C\|T^-)$ |
|---|---|---|---|---|
| AVI | 0.71 | 0.16 | 0.8161 | 0.1839 |
| SS | 0.59 | 0.05 | 0.9219 | 0.0781 |
| MNS | 0.68 | 0.17 | 0.8000 | 0.2000 |
| MPCI | 0.72 | 0.09 | 0.8889 | 0.1111 |
| MNCI | 0.74 | 0.20 | 0.7872 | 0.2128 |
| PK | 0.71 | 0.30 | 0.7030 | 0.2970 |
| AIR1 | 0.83 | 0.19 | 0.8137 | 0.1863 |
| AIR2 | 0.78 | 0.13 | 0.8571 | 0.1429 |
| AIR3 | 0.75 | 0.14 | 0.8427 | 0.1573 |
| AIR4 | 0.81 | 0.37 | 0.6864 | 0.3136 |
| AIR5 | 0.53 | 0.11 | 0.8281 | 0.1719 |
| AIR6 | 0.76 | 0.23 | 0.7677 | 0.2323 |

When the sensor detects and takes sample data, the twelve features are evaluated from the spectrum. Using the threshold values introduced in section 6.3, each feature gives a declaration of cocaine or not cocaine with probability given in table 6.4 above at time *t*. For example, if the peak of a normalized data is greater than 0.28, feature 6 (PEAK) declares that the tested compound is cocaine with a probability of 0.71. The sensor then collects another set of data at time t+1 and analyses it giving a declaration of cocaine or not cocaine with likelihood probabilities. The new posterior is then calculated using this previous information.

For fused data, the process continues as before but at each stage, the fused decision based on combination of all feature decisions is extracted with corresponding posterior probability.

Figures 6.5 and 6.6 show the performance of all 12 features over 18 cocaine test data samples. They also include the posterior probability of the fused feature level decisions (in black). The posterior probabilities are plotted against no of samples which may also represent temporal change. Posteriors are used to update new data to calculate new posteriors. The very good performance is indicated by values closest to the highest point (probability = 1).

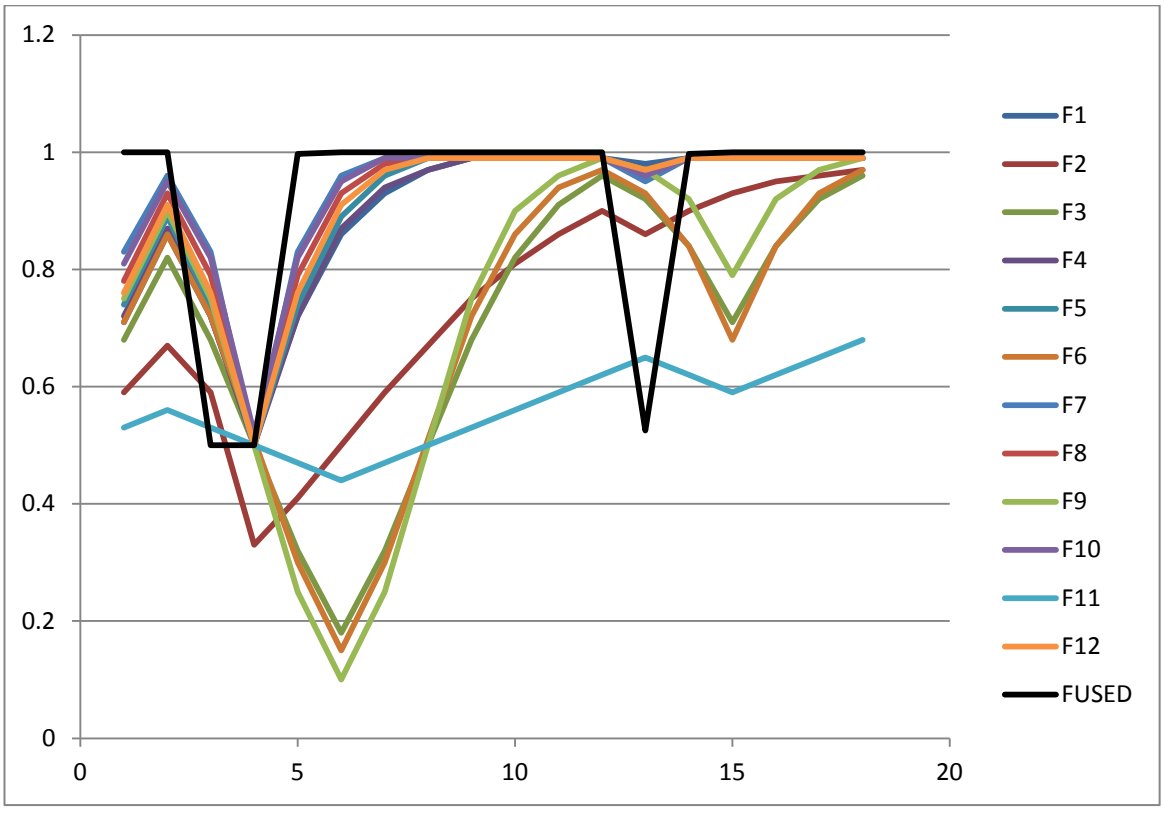



Figure.6.5 Posterior probability of feature classification being correct after feature has declared for cocaine $P(T^{+}|C)$.

As can be seen, fused posteriors (in black) give the best results for cocaine detection. Even at points 3, 4 and 13 where the individual features declare non-presence of cocaine, the fused data gives a posterior of over 50% probability which will warrant at least further investigation of the container. If the threshold probability of giving an alarm when cocaine is detected is set at 50%, then the fused decision will give a 100% true positive rate.

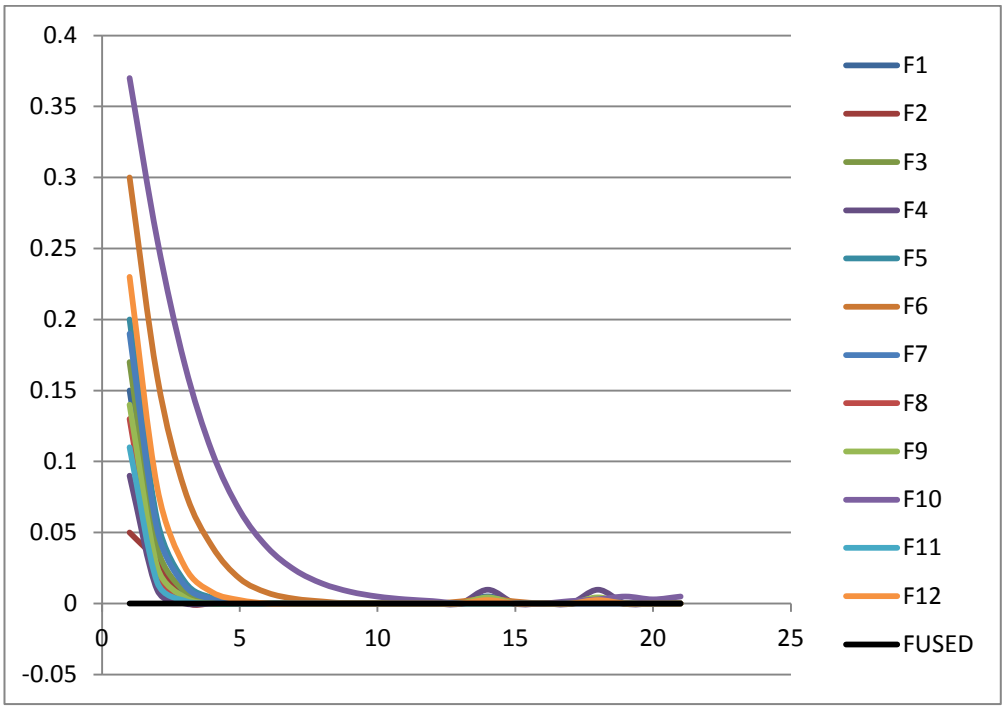



Figure. 6.6 Posterior probability of feature making wrong decision based on positive declaration for cocaine $P(T^{-}|C)$

Figure 6.6 shows the posterior probability for each sensor feature making a wrong decision after declaring for cocaine P(T-|C). The model is tested with 21 non-cocaine samples. Again, the probability of each of the sensors decreases to zero as time *t* increases. The ideal situation should be for each probability to tend to zero at every point in time. This is mirrored in the posterior probability for the fused data (shown in black).

**Table 6.5 Confusion matrix for fused feature data**

| a | b | Predicted as |
|---|---|---|
| 18 | 0 | a = cocaine |
| 0 | 21 | b = not cocaine |
| Correct classification = 100%<br>Incorrect classification = 0% | | |

The confusion matrix above represents results for fused feature test data using probability of 0.5 as threshold. As can be seen this yields in 100% prediction rate for cocaine. However, if a higher probability of 0.8 for instance is selected, the probability of a correct classification is reduced to approximately 89% with zero false alarms. Therefore, even when the prior threshold is increased from 0.5 to 0.8, there is no change in the false alarm rate even though the performance is slightly lower than the best feature classifier. Figure 6.7 shows the classification accuracies for individual features compared with the Bayesian Fusion.

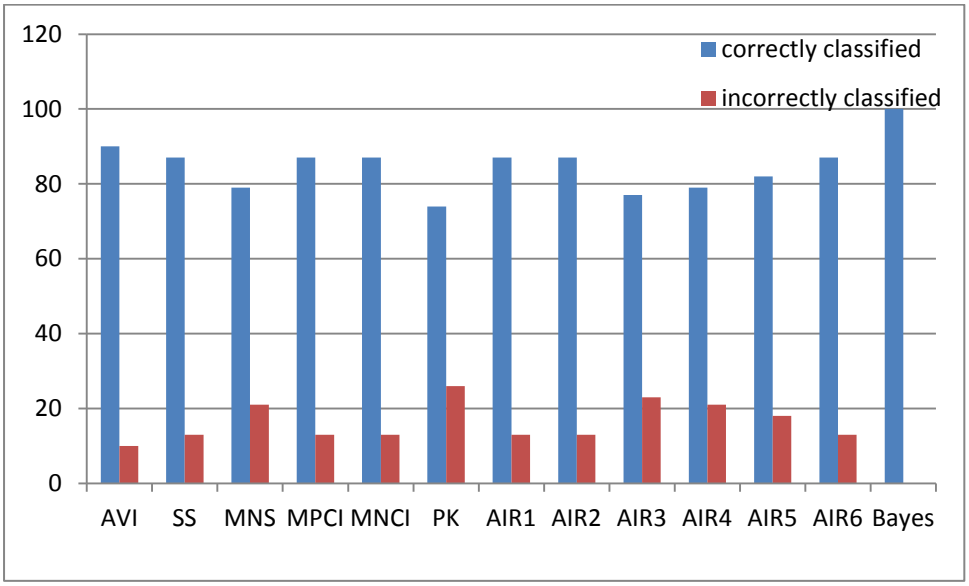



Figure 6.7 classification accuracies with features and Bayesian Fusion

## 6.5 Dempster-Shafer Implementation

Following the methodology described in chapter 3, this section describes its implementation using the Dempster Shafer technique. The task is to be able to identify from a previously unknown analyte, the presence or otherwise of cocaine given the individual feature judgements (belief assignments). The chapter takes a cursory look at the Dempster-Shafer technique and its relation to the Bayesian technique. The chapter will then look at the weaknesses of the Dempster-Shafer combination rule and attempts by researchers to correct these weaknesses.

### 6.5.1 Dempster-Shafer as a generalised Bayesian

The Dempster-Shafer method is often termed as the generalized Bayesian theory of subjective probability and it is also known as the theory of belief functions. It solves some of the major issues faced by the Bayesian method as highlighted in chapter 3. The Dempster-Shafer (D-S) belief functions provide a new method for assigning probabilities to sets rather than to mutually exclusive singletons (Sentz, et al., 2002). As opposed to the classical Bayesian theory of subjective probability where evidence is usually attached to a single event, the D-S evidence is associated to a set of events (probability intervals). Obviously, when the available evidence suffices such that probabilities can be assigned to a singles event, the D-S will then transform back to the classical Bayesian Inference method.

### 6.5.2 Dempster-Shafer Modifications

The steps to using the D-S technique have been dealt with in chapter 2. Key to these steps is the D-S combination rule. There are however, some weaknesses to this rule necessitating the introduction of modifications. The first weakness affects updating with new data (Seims, 2009). Due to the non-associative properties of the basic probability assignment, the D-S combination rule makes it imperative to recalculate the basic probability assignments affected by the system update. The second weakness is the occurrence of counter intuitive results (Seims, 2009) when the bodies of evidence to be combined have a high degree of conflict. This is a direct implication of the renormalization constant in the D-S combination rule which occurs as a result of not properly dealing with conflicts (Kari, 2002).

A classical medical example will suffice in explaining the second weakness (Sentz, et al., 2002):

Given two diagnoses by two different physicians for a particular patient with neurological symptoms, the first physician believes that the patient has one of meningitis or a brain tumour (probability of 0.99 and 0.01 respectively). The second physician believes the patient actually suffers from a concussion with a probability of 0.99 but accepts that he may actually have a brain tumour (probability of 0.01).

This implies that the patient may be suffering from one of meningitis, brain tumour or a concussion, i.e.

*m(X)* = {meningitis, brain tumour, concussion}

$m_1$*(meningitis) = 0.99*

$m_1$*(tumour) = 0.01*

$m_2$*(concussion) = 0.99*

$m_2$*(tumour) = 0.01*

From the D-S combination rule, EQ.4.12,

$$m_{1,2}(A_i) = \sum_{A=A_k \cap A_j} \frac{m_1(A_k) m_2(A_j)}{1 - \sum_{A_k \cap A_j = \emptyset} m_1(A_k) m_2(A_j)}, \qquad (6.7)$$

The denominator of this equation is known as the renormalization constant, K and its 'rationale' is to 'ignore' all conflicts in the systems by assigning them to the null set.

$$1 - \sum_{A_i \cap A_j = \emptyset} m_i(A_i) m_j(A_i) = 1 - K \qquad (6.8)$$

Thus, continuing for the example above,

1 - *K* = 1 – (*$m_1$(meningitis)$m_2$(tumour)* + *$m_1$(meningitis)$m_2$(concussion)* + *$m_1$(tumour)$m_2$(concussion))*

= 1 – (0.99 x 0.01 + 0.99 x 0.99 + 0.01 x 0.99)

= 1 - 0.9999

= 0.0001

Applying the combination rule (EQ.6.1) and normalizing with the value of *1 – K* above will assign a value of 1 to m(tumour) in effect saying that the patient undoubtedly has a brain tumour. Clearly, this is a counter-intuitive result as it negates the possible diagnosis of both physicians.

Over the years, researchers have come up with modifications to Dempster-Shafer's rule. The modifications focus on attempts to remove the weaknesses highlighted above. One of such modifications was introduced by Ronald Yager called the application of quasi-associative probability assignment to deal with the first weakness (Yager, et al., 1994).

#### 6.5.2.1 Yager's Associative Operators and modified Dempster-Shafer

A data fusion system needs to continuously receive data from sensors and as such, requires that there is a consistent update of results depending on the information deduced from the newest set of data. As stated in the previous section, the D-S basic probability assignment (bpa) is non-associative and thus the D-S combination rule has to constantly recalculate the bpas at every system update. To solve this problem, Yager (Yager, et al., 1994) introduced the quasi-associative operators/operations. Yager argues that an important feature of combination rules is that it must be able to update its system when new data becomes available. Below is an explanation of system updates.

Suppose we have a set of elements *W* and * is a binary operator. If $w_1$, $w_2 \ldots w_n$ are elements in *W*, associativity and commutativity are important properties for combination rules and implies that

$w_1 * (w_2 * w_3) = (w_1 * w_2) * w_3$ ------- associative (6.9)

$w_1 * w_2 = w_2 * w_1$ ---------- commutative (6.10)

Say the systems is to be updated with new data $w_{n+1}$, such that we now have

$w = w_1 * w_2 \ldots * w_n$ (6.11)

and $\bar{w} = w_1 * w_2 \ldots w_n * w_{n+1}$ (6.12)

then $\bar{w} = w * w_{n+1}$ (6.13)

(6.13) shows as update of the 'system' W for every new additional set of data $w_{n+1}$ and is only correct if the operator * is associative.

A common combination rule which does not update is the arithmetic averaging rule. With the introduction of new data, finding the average of the new data and the average of the previous data will not yield in the correct average of the entire system. That is if we have a set of data

A = a1, a2, … an

The arithmetic average is

$E = \frac{a_1 + a_2 + \ldots + a_n}{n} = \frac{1}{n} \sum_{i=1}^{n} a_i$ (6.14)

Introducing a new data point $a_{n+1}$, the new average

$$E_{new} \neq \frac{E + a_{n+1}}{n+1} \text{ i.e. } \neq \left[\frac{1}{n}\sum_{i=1}^{n} a_i\right] + \frac{a_{n+1}}{n+1} \qquad (6.15)$$

thus showing that the arithmetic average is not associative. To resolve this, Yager introduced a *quasi-associative* operator which responds to system update the same way an associative operator responds. To highlight this, suppose " ■ " is a quasi-associative operator, then, according to Yager (Yager, et al., 1994),

$$w_1 \blacksquare w_2 \ldots \blacksquare w_n = T(w_1 * w_2 \ldots * w_n) \qquad (6.16)$$

where *T* is a normalization mapping.

Applying this to the arithmetic averaging rule, the new data point is added to the sum of the original set of data points and then divided by the total number of data. We thus have from (6.16) above, first we add the new data point $a_{n+1}$,

$$S = a_1 + a_2 + \ldots a_n + a_{n+1}$$

Then we divide by the total number of data points i.e.

$$E_{new} = \frac{1}{n+1}\sum_{i=1}^{n+1} a_i$$

Introducing the normalization mapping, *T*, we then have

$$E_{new} = T\left(\sum_{i=1}^{n} a_i\right) \qquad (6.17)$$

Thus, using a sub-associative operator, the non-associative averaging operator has been converted to a quasi-associative operator.

Finally, applying the quasi-associativity to Dempster's combination rule, recall from (6.7) that the Dempster's rule is

$$m_{1,2}(A_i) = \sum_{A=A_k \cap A_j} \frac{m_1(A_k) m_2(A_j)}{1 - \sum_{A_k \cap A_j = \emptyset} m_1(A_k) m_2(A_j)}$$

Without the normalization, *1 – K*, we have (Seims, 2009)

$$\sum_{A=A_k \cap A_j} m_1(A_k) m_2(A_j) = q(A) \qquad (6.18a)$$

Yager's combination rule or modified Dempster's combination rule is thus now

$$m(A) = \frac{q(A)}{1 - q(\emptyset)} \qquad (6.18b)$$

$$q(\emptyset) = \sum_{A_k \cap A_j = 0} m_1(A_k) m_2(A_j) \qquad (6.19)$$

or

$$m(A) = T(q(A)) \qquad (6.20)$$

*where q(A) is known as the "ground probability assignment", q(∅) is the degree of conflict and the normalization mapping T is the normalization of conflict.*

The modified Dempster's combination rule allows for update of data fusion systems using the Dempster's combination rule which was not possible with the Dempster's combination rule.

#### 6.5.2.2 Dealing with counterintuitive results

Another weakness of the Dempster's combination rule is the occurrence of counter-intuitive results as shown in the medical example above. A critical look at the this weakness using the medical example above would show that intuitively, considering that two experts have agreed on a an hypothesis assuming that both experts have the same degree of reliability, one would expect that the result should be a confirmation of individual decisions (Campos, et al., 2005). Considering that they were two experts with a common low level of belief in a hypothesis but high level of belief in two disjoint hypotheses. A logical explanation would be that as more experts place a high level of belief in individually differing hypotheses and a low level of belief in a common hypothesis, the uncertainty of the common hypothesis will decrease since all the experts believe in it while the uncertainty of the individual hypotheses will increase as number of experts increase.

The first step in trying to resolve this conflict as suggested by Yager is to replace the Dempster's basic probability assignment with his ground probability assignment as shown in (6.18a). The ground probability assignment has the same property as the basic probability assignment in that it is constrained between values of 0 and 1 (Seims, 2009).

With the ground probability mass *q(A)* obtained, the next step is to convert it to the basic probability mass. To do this, the 'ignorant frame of discernment' basic probability mass *q(∅),* is added to the non-ignorant null set ground probability mass *K*, to give (Wu, 2003)

With $m(\phi) = 0$ *then* (6.21a)

$$m_U(\Theta) = q(\Theta) + q(\phi) \quad (6.21b)$$

where $\Theta$ is the frame of discernment.

Note that by adding the conflict to the frame of discernment as seen in (6.21b), Yager rather than ignoring the conflict like Dempster (EQ.6.1), increases the degree of ignorance in the system therefore, if there are no conflicts in the system, Yager and Dempster's combination rules will yield the same results. If there are conflicts however, Yager's combination rule will yield a more meaningful result. With the medical example given in the previous section, Yager's combination rule yields that the belief that the patient is suffering from a tumour is 0.0001 and the degree of ignorance in the system is 0.9999 which is a more meaningful result (Yager, 1987).

#### 6.5.2.3 Inagaki's Unified Combination Rule

In his attempt to resolve the conflict issue and to have a single combination rule that can be used in every situation, Toshiyuki Inagaki created the Unified Combination Rule using Yager's ground probability mass. The combination rule can be explained with the following:

$$m(A) = [1 + kq(\phi)]\, q(A) \qquad A \neq \emptyset, A \neq \emptyset \quad (6.22)$$

$$m(\Theta) = [1 + kq(\phi)] \cdot q(\Theta) + [1 + kq(\phi) - k] \cdot q(\phi) \quad (6.23)$$

*where k, lies within the range*

$$0 \leq k \leq \frac{1}{1-q(\emptyset)-q(\Theta)} \quad (6.24)$$

What Inagaki has done is to convert the ground probability mass as defined by Yager (6.21a) to a basic probability mass using equations (6.22) and (6.23) as against (6.21b) used by Yager. The implication of this is that depending on the value of *k*, at the highest possible value of *k,* i.e. *k* $=\frac{1}{1-q(\emptyset)-q(\Theta)}$, Inagaki's unified combination rule aligns with Dempster's while at the lowest value of *k,* i.e. *k = 0*, it aligns with Yager's rule (Wu, 2003). In between, Inagaki believes that the optimal value of *k* is open to research.

The sections above have been able to highlight two main weaknesses in the D-S technique for data fusion. While alternate combination rules have been propounded by researchers to combat these weaknesses, two of them, developed by Yager have been discussed. Inagaki has gone ahead to propose a Unified combination rule with aims of aligning with both Yager and Dempster's combination rules depending on values chosen for the normalization *k.*

## 6.6 Application of Dempster-Shafer technique to cocaine detection

The details of the experimental set up used in this thesis have been explained in chapter 4. Feature extracting techniques used has also been explained in chapter 5. Twelve features have been extracted for decision making when detecting for the presence or otherwise of cocaine from an unknown substance. As explained in the previous chapters, each feature will

make a decision on whether there is cocaine, not cocaine or unknown. Therefore, the three hypotheses represented by the frame of discernment Ω,

$$\Omega = \{h_1\ h_2\ h_3\} \qquad (6.25)$$

and the corresponding power set

$$2^{\Omega} = \{\emptyset, \{h_1\}, \{h_2\}, \{h_3\}, \{h_1, h_2\}, \{h_1, h_3\}, \{h_2 h_3\}, \Omega\} \qquad (6.26)$$

Where $h_i$ are the possible hypotheses i.e. cocaine, not-cocaine and 'unknown' (could be cocaine or not-cocaine) for i = 1, 2.

The model used for the DST is adapted from the model developed in chapter 4 and is similar to the Bayesian adaptation (figure 6.8).

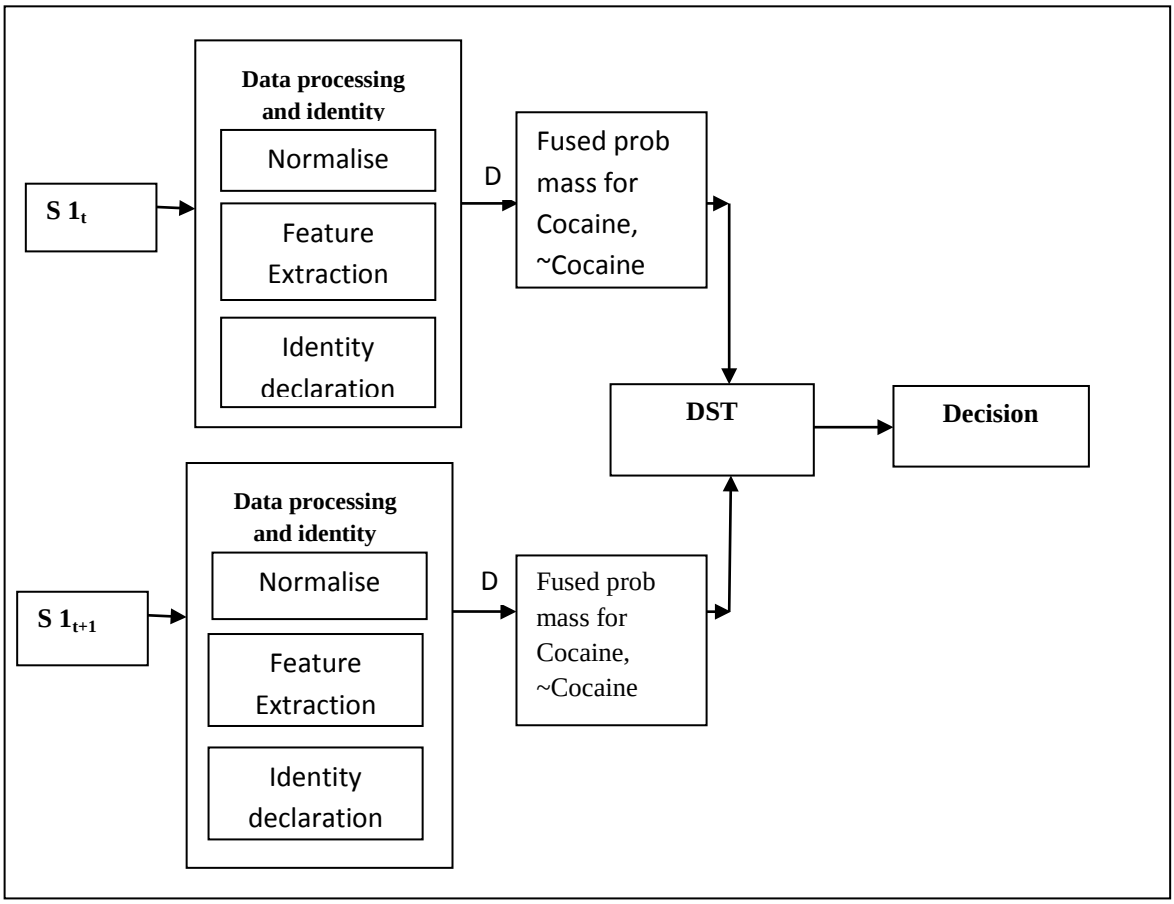



Figure.6.8 the DST model

The aim of the DS approach is to improve on the result from each feature based classification technique. These features are known as the primary classifiers. They assign probability masses to the propositions (i.e. cocaine, ~cocaine) and the masses are then combined using combination rules (described in earlier sections) and evidential intervals are then defined.

*Basic Probability Assignment (bpa)*

The bpa has been explained in previous sections. It Is also known as the belief mass and is the weight attached to knowledge known about features. The features identified determine the presence or otherwise of cocaine that is C or ~C. Dempster Shafer gives the opportunity for the system to output

an ‘unknown’ event that is (C, ~C). The belief masses for the features are as defined below:

*Cocaine*

Cocaine belief mass is defined as correct cocaine prediction as given in equation 6.27

$$m^i(C) = \frac{A_{cc}^i}{N_{tc}^i} \qquad (6.27)$$

where $m^i(C)$ is the cocaine belief mass from feature *i*; $N_{tc}^i$ is the total number of correct cocaine data from feature *i*; $A_{cc}^i$ is the correct cocaine decisions from feature *i*.

*Non-Cocaine*

Non-cocaine belief mass is defined in terms of non-cocaine prediction as given in equation 6.28

$$m^i(\sim C) = \frac{A_{nc}^i}{N_{tn}^i} \qquad (6.28)$$

where $m^i(\sim C)$ is the non-cocaine belief mass from feature *i*; $N_{tn}^i$ is the total number of correct non-cocaine data from feature *i*; $A_{nc}^i$ is the correct non-cocaine decisions from feature *i*.

*Unknown belief mass*

The unknown belief mass is defined as the wrong cocaine and non-cocaine decisions as shown in equation 6.29.

$$m^i(C, \sim C) = \frac{A^i_{cw} + A^i_{nw}}{N^i_{tc} + N^i_{tn}} \qquad (6.29)$$

where $m^i(C' \sim C)$ is the unknown belief mass from feature *i*; $A^i_{cw}$ is the number of wrongly classified cocaine data from feature *i*; $A^i_{nw}$ is the number of wrongly classified non-cocaine data from feature *i*;

**D-S for tracking**

As previously mentioned, the fusion performed in this dissertation is in two stages – the first stage is the fusion of decisions across features while the second stage is the fusion of over time or tracking. Unlike the Bayes rule which accommodates time evolution, a first glimpse of the D-S rule will show that it does not accommodate evolution over time (Koks & Challa, 2005). However, this can be allowed for by extending the application of the D-S equation stated in (6.7) this time allowing the sets $A_k$ and $A_j$ to refer to new and old data rather than data from two different sensors (Zou et al., 2000) (Koks & Challa, 2005).

## 6.7 Decision making and Results

Using the individual feature classification estimates as baseline the Dempster-Shafer combination of feature probability estimates were tested on the same experimental test data for comparison. Equations 6.27 to 6.29 show the equations used in calculating the belief masses.

The bpa are used to assign mass values for each feature based on the feature's decision. For every test data, all the mass values are combined to

obtain a total mass values for m(C), m(~C) and m(C,~C). The highest assigned mass value is chosen as the correct hypothesis.

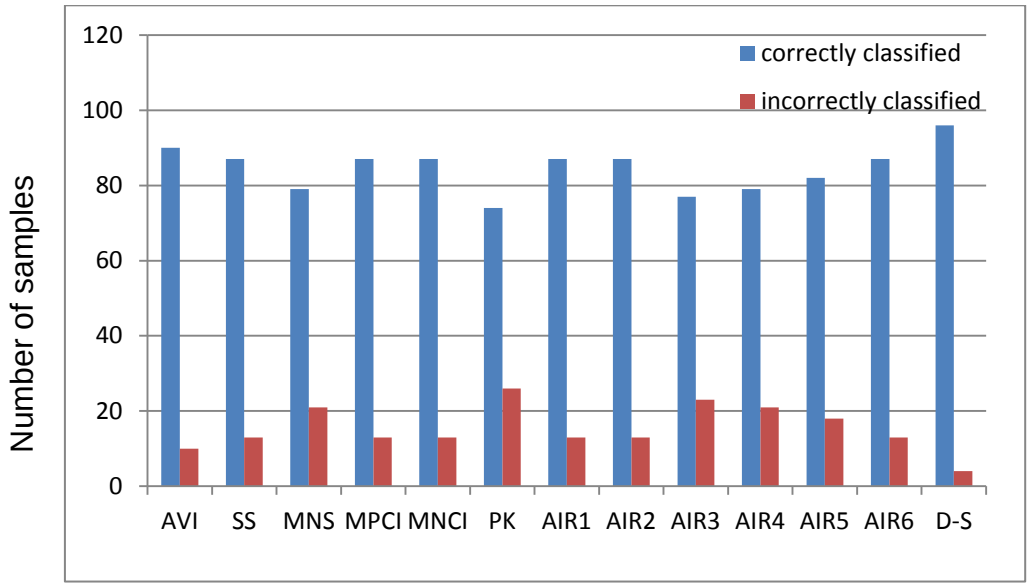



Figure 6.9 D-S classification accuracy compared with features.

The figure above shows the results of the classification accuracy of all features and that of the D-S. The result is an evaluation over two time periods. In the first time period, the masses are evaluated as explained earlier in this section with corresponding confidence interval. Time period two follows the same process as time period one and the confidence intervals of both periods are then combined. The accuracy in terms of number of correctly identified cocaine/non-cocaine samples are the same at both levels but with increased confidence intervals.

In Figure 6.9, Feature1 (AVI) performs better than all other features in terms of correct classifications but the D-S performs even better than feature1. The results are based on the same 39 test data used for the Bayesian technique. The D-S was able to correctly classify 95% of the data correctly identifying

17 as cocaine and 20 as not cocaine. Interestingly, the non-identified cocaine and non-cocaine data were classified as non-cocaine and cocaine respectively while no data set classified as unknown. This will most likely be due to the fact that the hypotheses involved in this dissertation are mutually exclusive. This shows that the D-S Technique's theoretical advantage over the Bayes technique in making room for a value for 'unknown' situations does not come into play in this application. Moreover, in reality, an *unknown* decision will generally yield the same response as a positive response/false alarm which is a manual search of the container. This means that for the purpose of this dissertation, this ability of being able to decide for unknown situations does not mean an advantage.

## 6.8 Comparison of results

Once a justification for the technique adopted has been shown, the rest of the project involves the display of the results to the operator. This chapter will outline the development process for the interface built to display results in this dissertation with an overview of the ferret robot itself.

As a reminder, the two techniques have been detailed earlier in this chapter. These are the, Bayesian and Dempster-Shafer techniques.

**Table 6.6 Comparison of feature classifiers and fusion classifiers.**

| | Performance | |
|---|---|---|
| | **% correctly classified** | **% wrongly classified** |
| **Feature 1** | 90 | 10 |
| **Feature 2** | 87 | 13 |
| **Feature 3** | 79 | 21 |
| **Feature 4** | 87 | 13 |
| **Feature 5** | 87 | 13 |
| **Feature 6** | 74 | 26 |
| **Feature 7** | 87 | 13 |
| **Feature 8** | 87 | 13 |
| **Feature 9** | 77 | 23 |
| **Feature 10** | 79 | 21 |
| **Feature 11** | 82 | 18 |
| **Feature 12** | 87 | 13 |
| **Bayes** | 100 | 0 |
| **DST** | 95 | 5 |

Table 6.6 shows the performance of both the Bayes and DS techniques on the test data. It also includes individual performance of each feature. It can be seen that the Bayesian gives the best result given a 0.5 prior threshold. Even when this threshold is increased to represent little prior information about the container being investigated, the correctly classified data falls to about 89% which though is less than the D-S 95%, still performs better than the best feature value. In addition, achieving a zero false alarm rate at this threshold performs better than the 5% false alarm rate of the D-S technique.

In summary, Wu (2003) gives the following as situations where Bayesian or Dempster-Shafer technique may be more appropriate for sensor data fusion (see table 6.7):

**Table 6.7. Where to use Bayesian/Demspter-Shafer (Wu 2003)**

| Use Bayesian | Use Dempster-Shafer |
|---|---|
| All hypotheses are mutually exclusive i.e. if P(A) = p, then P(B) = 1 – p where A and B are complements | Hypotheses may have overlapping hypothesis |
| Prior probability distributions are known. | Prior probability distribution is unknown, |
| Data measurements are easily evaluated by their probabilities | Data measurements partially correlate to probability distributions, ignorance needs to be accounted for |
| Joint probability distribution is known, or data measurements are conditionally independent | Joint probability is not known, observations are independent |
| Direct evaluation of probabilities helps in maximizing expected utility | Difficulty in relating evidence with probability distribution, thus weak in decision-making support |

## 6.9 Human Computer Interface

The user interface is the interface between the operator and the DF system. The results are displayed for the user to interpret. This chapter will highlight issues to be considered in the HCI. It will also give a brief description of the ferret robot and its key features.

### 6.9.1 The Human – Computer Interaction

Displaying the results of a data fusion implementation is critical for the user. The human/computer interface (HCI) provides tools for the human operator

to control and guide the fusion process. It is also through the user interface that the results of the fusion process are presented to the operator.

The role of the user in a data fusion system (level 5) is varied and dependent on the system. The user can play a very important role by being the one who determines what is of interest and what information is needed to support the data acquired (Blasch, 2009). However, in many cases, the User could just be the operator whose responsibilities are limited to interpreting information, facilitating and reporting information. In this thesis, the user assumes the latter role. Although in many cases, the operator is trained for the task and his role is well defined (Blasch, 2009), it is important that the user interface be well designed such that it is easy for the operator to use. The challenge therefore is how to present the complex data to the user in such a way that he will be easy to retrieve and manipulate information (Mandiak et al., 2005). In addition, for many fusion processes, post fusion reports may also be generated by the interface and stored or passed on to relevant authorities.

In this thesis, the user interface serves two main purposes. The first purpose is to acquire and manipulate data from the sensor(s) and the second purpose is to provide a means to control the ferret robot which will serve as the platform which will carry the sensors. Figure 6.10 below shows the robot which was designed and built by the team at The University of Sheffield, UK.

Development of the User interface was done using Microsoft Visual Basic.net – an object oriented software program implemented on the .NET framework.

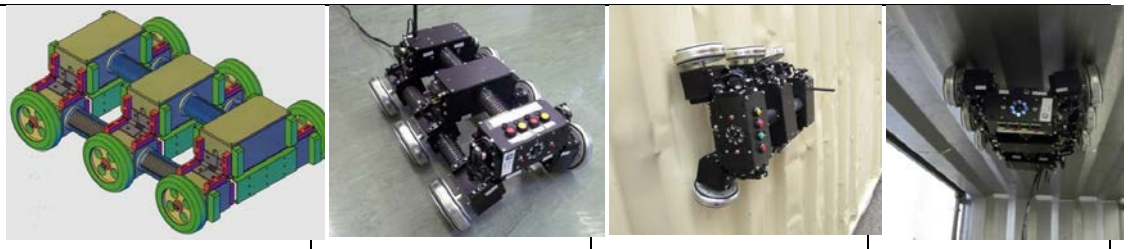
The approach used in the design of the interface allows a direct interaction between the operator and the robot. This direct interaction involves the

operator deciding when to move the robot, in what direction and when to stop the robot from moving. His decision will be based on a number of factors including visual information (the operator is able to see what the robot can see via two cameras placed in front and behind the robot) and sensor information (based on the decisions on the presence or non-presence of cocaine, the operator will decide whether to stop the robot or not).

### 6.10 The Ferret Robot

The ferret robot is shown in figure 6.10 below. Its main features are the body, its wheels and the sensors.

The robot's body is divided into three sections. The front and back sections contain the cameras which act as the robot's 'eyes' and provide visual data for the operator via the user interface. The other cocaine detecting sensors are placed on the front panel of the robot. The middle section carries the control panel of the robot and it is this control panel that communicates directly with the computer where the user interface is installed.



| (a) 3D Model | (b) Moving on level ground | (c) Wall Climbing | (d) Ceiling Climbing |
|---|---|---|---|

Figure.6.10 showing a 3D model of the robot and different moving positions

The wheels of the robot are made of magnetic material. This enables the robot to be able to mount on metallic surfaces and attach itself to the surface whilst moving around the container.

The sensors are the camera and fibre optic sensors which are the parts of the robot that interacts with its environment. The cameras gather visual data for the operator while the sensors collect air sample data for further analysis. The extraction of features and eventual data fusion analysis is performed on the host computer.

Communication between the host computer and the robot is by RS232 for now. There are plans to convert this to wireless technology in the nearest future.

### 6.11 User Interface development

Figure 6.11 shows the HCI designed for this project. The main controls for robot manipulation include the power button (1) and direction buttons (2). The power button turns the robot on and off while the direction buttons (left, right, up and down arrows) manipulate the motion of the robot. There is also a power switch on the robot which acts as the primary power switch. Visuals transmitted via the cameras on the robot can also be recorded and replayed via the buttons shown directly beneath the video screen.

Once the robot is switched on (first by the main power switch on the robot and then via the power button on the interface), the video screen is also automatically turned on. It gives the operator the opportunity of seeing the

robot's environment and then determining in what direction to move the robot. Once the robot begins to move, the operator can also activate data acquisition by the sensors by clicking on the 'acquire data' button. This gives the sensors attached to the robot to begin to acquire data. As data is acquired, the system also begins to process the data (feature extraction, detection decision, etc) as explained in previous chapters. If a substance is detected, the system gives an alarm in form of a red light beside the 'cocaine meter'. The meter is a score of probability of positive detection of cocaine ranging from 0 to 100. The red circle in front of the meter changes colour from green (no cocaine) through yellow to red. The yellow and red colours are dependent on the probability value. Between 0 and 50, the colour flashes yellow but when this value exceeds 50, the colour changes to red.

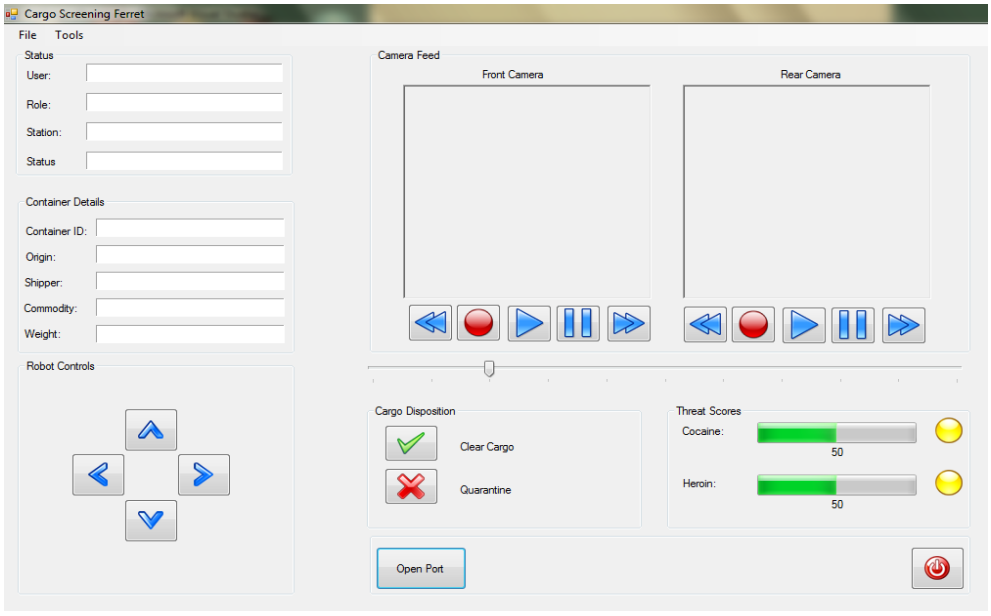



Figure 6.11 Human/Computer Interface

### 6.11.1 User Interface features

*Power Button*

The red power button at the bottom right of the interface will turn the robot on and off.

*Sensor Power Button*

The sensor start and stop button opens and closes the port connecting the sensor via the robot's circuit board to the computer.

*Direction buttons*

The direction buttons are the four (up, down, right and left) buttons used in controlling the movement robot inside of the container. There is an auxiliary joy stick which can be used instead of the direction buttons.

*Visual Display Windows*

There are two visual display windows. Each one of them relays the feeds from the cameras placed in front and behind the robot real time.

Report Screen

The report screen displays the results of the fusion process when the operator demands for it. Basic information include, name of shipper, whether the container is cleared or not (determined by the operator), origin of container.

*Threshold Slider*

The threshold slider is available in steps of 0.2. The operator decides the threshold at the start of the process (before the sensor port is opened) and cannot be changed until the container is cleared (or otherwise). The slider is designed in such a way that slider values are locked to 0.1.

*Camera controls*

The camera controls are typical controls for a video feed. The record button allows for recording of feed, the play, pause, rewind and forward buttons allows for playing, pausing, rewinding and forwarding of recorded feed.

Cocaine Detection Alarm – the cocaine detection alarm is a glowing yellow button which turns red when cocaine is detected. It works together with an horizontal display which moves based on the final posterior probability. The visual alarm is triggered once this probability exceeds the pre-set prior.

In the design of the interface, the writer was guided by the nine usability criteria as given by Blasch (Blasch, 2000). These criteria are –

Appropriate functionality – The interface must meet the aims for which it was designed.

Visual Clarity - displayed information should clear and well organised, easy to read, unambiguous and should enable users to find required information, draw the user's attention to important information and also to allow the user to see where information should be quickly and easily entered.

Consistency – the looks and workability of the interface must be consistent to make it predictable by the operator.

Compatibility – the user interface should meet existing user interface user conventions making it easier for operators to navigate, understand and interpret controls on the interface.

Informative feedback – the operator should be clear about the steps they are taking and steps they will be taking while operating the interface.

Explicitness – the use of controls and overall way the system works should be clear to users. The interface is designed in such a way that the operator and anyone else can easily understand what the system does. The interface is also carefully labelled to guide operators.

Error Prevention and correction – occurrence of operator's errors should be minimised as much as possible and when they occur, operators should be able to check their inputs and thus correct potential errors before implementation. In the Ferret Robot User interface, the input required by the system includes data of the cargo being checked and the threshold set. For the former, there is an in built spell check system which suggests possible spelling corrections for the operator. The slider is used to select a threshold and has a 'lock' in system that locks to every 0.1 threshold value to reduce errors.

User guidance, usability and support – operator must have an informative and easy to use guidance to help him understand the use of the system. A relevant easy to use help system is provided with this interface to help the operator manoeuvre around the system. Access to this help guide is via the tools menu – Tools --> Help.

Flexibility and Control – the user interface must be flexible and easy to control. From the interface screenshot shown in figure 6.11, it would be

noted that the design is such that anyone can easily use it. The buttons are well labelled in clear plain language to enable easy interpretation.

### 6.12 Chapter Summary

In this chapter, the Bayesian and Dempster-Shafer techniques have been implemented into the model developed in chapter 4 accompanied by background explanations of the Bayes and DS theories and how they relate to data fusion. The chapter has also looked at the parameters involved in Bayesian statistics including likelihood, prior and posterior probabilities. The chapter then then went on to investigate the classification strengths of various features extracted from spectra to test for cocaine. A confusion matrix was extracted from the performance of each feature detailing its True positive, false positive, true negative and false negative values. In addition, a posterior probability for making correct and incorrect decisions based on the feature declaring for cocaine was also computed and the results displayed on a table. The posterior probabilities of the fused data was also compared with that of each of the features on a graph and again the fused data showed improved performance over performance of each of the features. Overall, using Bayesian algorithm, this dissertation has shown that fusion of data from multiple features presents a better result than each of the features respectively.

The chapter also described the Dempster-Shafer technique for combination of data from multiple features and then going on to discuss the strengths and weaknesses of this technique and various extensions as described by Yager

and Inagaki's modification of the combination rule. The implementation of the D-S technique gives an improved result compared to the individual feature accuracies. However, a comparison of both the Bayes and D-S shows that the Bayes accuracy is higher than that of the D-S at a certain threshold and even when the threshold is increased to make room for unavailable prior information, the false alarm rate of the Bayes still performs better than that of the D-S.

In the concluding parts of this chapter, the HCI development process was described along with a brief description of the ferret robot and its features. The HCI development follows some guidelines or usability criteria as given in literature and these are also highlighted.

# CHAPTER 7

# CONCLUSION

## 7.1 Summary

In traditional sensor fusion systems, the focus is usually aimed towards how to detect or evaluate specific target parameters improving on reliability and accuracy. Looking at this directly, perhaps the cheapest way of achieving this improvement on accuracy and reliability is by adding measurement redundancy which in turn can be achieved in one of two ways (Wu, 2003); either increasing the number of sensors to measure the same parameters or using a single sensor to take multiple measurements over time. In other words, sensor fusion can be achieved either temporal that is taking measurements over time or spatially with the use of multiple sensors (Hall, 1992). In addition to this, sensor fusion can be achieved by a combination both spatial and temporal data using of multiple sensors measuring similar

parameters over time. Usually before sensor fusion is performed, the raw data is pre-processed before fusion (Dey et al., 2000) (Wu, 2003).

Commonly used fusion techniques include Bayesian fusion and Dempster-Shafer techniques.

Fibre Optic sensors are developed using optical fibres as the sensing element. Fibres can be Step index multimode, Graded Index Multimode or Single mode. Optical fibres have the advantage of being small in size and economical. Fidanboylu *et* al (2009) have categorised fibre optic sensors into three based on their applications – Physical (temperature, stress), Bio-medical (measurement of blood flow, glucose content and so on) and Chemical sensors (usually used for spectroscopic, gas analysis and pH measurement). In addition, fibre optic sensors can be one of four types – Intensity based, Wavelength Modulated, Phase modulated and Polarization modulated fibre optic sensors. A kind of wavelength modulated fibre optic sensor called the UV multimode fibre-optic fluorescence sensor is used in this dissertation and is based on a MIP[11] containing fluorescence in moiety as the detector which signals the presence of the analyte in the compound.

The detection of cocaine using fibre optic sensors faces a major challenge. The challenge is that while the sensor can detect cocaine as required, it will simultaneously detect for multiple of other substances. When the concentration of the sample is known, it is possible to uniquely identify the

[11] MIP – Molecularly Imprinted Polymers are synthetic receptors that can be synthesized for a variety of target molecules

individual compounds from the sample. However, in real life, this concentration is unknown and as such a means of detecting cocaine from sample compounds without *a priori* knowledge of concentration of compound is required.

In this thesis, a methodology to address the challenges posed by the unique identification of cocaine when there is no *a priori* information on the concentration of the compound was recommended. The methodology uses a bottom up approach to meet the challenges.

In the first instance, the raw data extracted by the sensors is pre-processed and then normalised to achieve a common spatial reference for the data. Cleaning the data involves removing outliers and noise generally from the raw data. This stage is the level one processing stage represented in the JDL model.

The next step involves identifying unique characteristics which identify cocaine samples from non-cocaine samples. The identification of these unique features is a key process in the data fusion process. Features are basically a representation of the raw data and helps give a reduced data set. Features must however accurately and concisely represent the original information in the raw data (Hall, 1992). In addition to helping to reduce data, a feature extraction process also ensures that only a minimum data storage requirement is needed. For the feature extraction process in this dissertation, the raw data was analysed to find characteristics which will correctly model the raw data and aid the identification process. A number of features were

identified based on ‘activity points’ from the intensity count and wavelength graphs, plotted using the raw data. The features were tested and validated with real data taken from the fibre optic sensor tested in a laboratory at City University, London. The data was divided into two sets with training and test sets. The training set was used to train the features and the test set which was not part of the training set was used for validation. The results were positive and showed the features were a good fit to uniquely identify cocaine sample from non-cocaine samples. Once the features have been extracted, they are passed as inputs into the data fusion model.

Sensor fusion is mainly seen as a method of transforming data from its raw state into a more intelligent state with some level of abstraction (Wu, 2003). There are three major types of fusion architectures (Hall & Llinas, 1997) – **Centralised fusion with raw data** which basically uses either raw data or derived data from multiple sensors to make a decision on the state of an entity, **Centralized Fusion with feature vector data** which fuses features extracted from raw data in a central fusion process and **Autonomous fusion** architecture where the extracted features from raw data make individual decisions which are then passed on to a fusion process and an identity is declared. The fusion process in this case can be implemented using a wide range of methods including Bayesian Fusion and the Dempster-Shafer fusion algorithm. The bottom up methodology adopted in this dissertation is an extension of the autonomous fusion architecture. Consideration of the goal to be achieved was the background focus in developing the architecture. The

overall goal is to detect the presence or otherwise of cocaine in a cargo using optical fibre sensors. The sensor is carried to the container using a ferret robot which is remotely controlled from a safe distance. The entire processing and fusion of the data received from the sensor is displayed on a user interface. In addition, the interface is used to control the ferret. Researchers have advised that the fusion process be as close to the raw data as possible to avoid information loss even though there is a cost of increased computation workload that comes with this as shown in Table 7.1 (Hall, 1992):

**Table 7.1 Suggested proximity between raw data and fusion process**

| | **Relative Computational Requirement** | **Required Communication Bandwidth** | **Estimation Accuracy** | **Processing Complexity** |
|---|---|---|---|---|
| Increasing proximity from raw data  | High | High | High | High |
| | Medium | Medium | Medium | Medium |
| | Low | Low | Medium | Low |
| | Very High | Very High | High | High |

The qualitative nature of the above table shows that the methodology chosen allows a sacrifice of estimation accuracy for improved bandwidth, computational requirements and processing complexity. Quantitatively however, as shown in the results in previous chapters, this method leads to very high result accuracy.

This methodology interacts with the database although it does not state how this interaction should be implemented. However, in the implementation programme designed for the thesis, the database is connected to the system via the ODBC interface. The database and user interface are all on the platform system. The user interface is designed using visual basic programming language which gives room for future development. The database of a fusion system is mainly required for data input (receiving raw and feature data), storage (a priori data), archiving.

The user interface facilitates the display of information from the processing unit. This information must be received by the user on time and without complexities. As designed, it also gives the user the ability to manoeuvre the ferret robot via a joystick or direction buttons on the interface. The operator will also determine when to open ports to receive data from the sensors. Prior to opening the port, the operator can decide the sensitivity of the system by deciding a threshold for probability of detection for the DF system. As shown earlier, at 50% threshold, the system gave a 100% performance based on test data used. The lower the threshold, the higher the sensitivity of the system. Operators can base threshold values on *a priori* information on cargo being investigated. If the cargo is from a known shipper, the threshold can be set to a higher value (recommended value not higher than 60%). Unknown shippers should have a lower threshold and cargo from suspected shippers an even lower threshold.

Once the system starts receiving data, processing the data is done by the system without any input from the operator with the operator monitoring the movement of the robot via two screens on the interface showing visuals from the front and back cameras. The results are then displayed on the interface for the operator who can decide to produce a report detailing the process including shipper information as entered by the operator at the beginning of the process.

## 7.2 Recommendation to industry

### 7.2.1 Cocaine Detection Implementation Summary

Using the methodology summarised in section 7.1, the raw data was tested with two data fusion techniques – Bayesian and Dempster-Shafer (DST). The results from the tests show that the Bayesian gives better results in terms of the parameters identified and defined in Appendix C - false positives, false negatives, true positives and true negatives than the DST.

### 7.2.2 Successful Implementation

This thesis addresses the challenges of a fibre optic sensor in the detection of cocaine from a sample. Specifically, a generalizable architecture for fusion is suggested using extracted features and a sensor fusion algorithm is implemented. Ideally, the methodology suggested is a bottom – up methodology which shows great promise in overcoming the sensor limitations and in addition, providing an opportunity for additional sensors to be used. The additional sensor could be a similar fibre optic sensor for

cocaine detection or another sensor for detection of contraband. In either case, the additional sensor(s) will work independent of other sensor(s). The results are fused based on a logical OR fusion node. The algorithm implemented uses the Bayesian theory as a fusion technique. The decision of the system is accompanied by a probability value which if it exceeds a threshold pre-set by the operator, decides for positive detection. This will let the operator have an input into the system. This input could be based on what is known as prior information.

Artificial Neural Networks technique was also used in validating the features selected in this thesis. Thus introducing these techniques into cocaine detection analyte detection using a fibre optic sensor and the feature selection and validation are key contributions of this thesis. In addition, a user interface which gives the operator the ability to remotely control the robot and also visualize robot's movements within the container while watching out for sensor results was also developed in this dissertation.

## 7.3 Limitation of the Research

In summary, this thesis has succeeded in developing a data fusion system for cocaine detection in a cargo container using a ferret robot which carries the sensor to the container. This system as it stands gives a high enough true positive rate and a low false positive rate to improve operator confidence. However, cocaine is not the only contraband which threatens UK borders (and other borders around the world). Heroin, and perhaps importantly, Improvised Explosive Devices (IEDs) are all threats. For some, there are

existing sensors which work in detecting these substances. In cases where sensors are not available, sensors could be developed to detect them. The model developed in this dissertation makes room for additional sensors to be added. The sensor may either detect for cocaine or other contrabands. In the case of the former, a fusion node working with a logical OR may be added to give the system flexibility of giving of a positive detection alarm when either of the sensors decides so. If an AND logic is used for the former, it will make the system more reliable in terms of lower false alarms but may affect the true positives since it needs the two sensors to decide positive to give a positive detection. However, for the latter, an OR logic fusion node will be added. Each of the sensors will have serial and independent processing units. A challenge may come up in terms of sensor data collection times. This is a factor that needs to be decided if additional sensors are used.

Another area for improvement is in the threshold set by the operator at the start of the process. This threshold for this system is set by the operator. However, in future, it could automated based on parameters such as origin of cargo, prior information on shipper's integrity. A rating system may be developed for shippers whereby the system only needs the shipper's rating to automatically determine a threshold.

## 7.4 Conclusion

In conclusion, this dissertation has developed a model to implement data fusion for single sensor detection of cocaine in cargo containers coming through the borders. The sensor output was analysed and processed and

then inputted into a data fusion model. This model has been developed in such a way that additional sensors for detecting other substances can be easily integrated in future. The Dempster Shafer and Bayesian techniques were investigated with the Bayesian technique showing better results in terms of correctly detected data. The overall result is thus an output which provides meaningful information and thus instils confidence in detection decision. This result is displayed in a visual user interface. The dissertation is an important part of a multi-disciplinary project tagged the Cargo Screening Ferret project sponsored by the UK Home Office and EPSRC[12].

[12] Engineering and Physical Sciences Research Council

# References


Akiwowo, A.O. & Eftekhari, M., 2010. Multisensor Fusion application for Cargo Screening - A Bayesian approach. In *2nd International Conference on Computer Technology and Development*. Cairo, 2010. IEEE Explore.

Amini, J., 2008. Optimum Learning Rate in Back-Propagation Neural Network for Classification of Satellite Images (IRS-1D). *Scientia Iranica*, Vol. 15, No. 6. pp.558-67.

Drug Enforcement Administration., 2003. Cocaine Impregnated Silicone in Baseball Cap parts in Peru. *Microgram Bulletin*, December. Vol. 36. No. 12. pp.271 - 272.

Anon., 2009. *Border Patrol History*. [Online] Available at: "http://www.cbp.gov/xp/cgov/border_security/border_patrol/border_patrol_ohs/history.xml" http://www.cbp.gov/xp/cgov/border_security/border_patrol/border_patrol_ohs/history.xml [Accessed 31 July 2009].

Anon., 2009. *ITV News Report*. [Online] Available at: "http://www.itv.com/News/Articles/Heroin-smuggled-in-car-airbag-277806607.html" http://www.itv.com/News/Articles/Heroin-smuggled-in-car-airbag-277806607.html [Accessed 11 Augustus 2009].

Anon., 2010. *CBP Canine Disciples*. [Online] Available at: "http://www.cbp.gov/xp/cgov/border_security/canine/disciplines_2.xml" http://www.cbp.gov/xp/cgov/border_security/canine/disciplines_2.xml [Accessed 20 February 2011].

Anon., 2011. *United States Drug Enforcement Agency*. [Online] Available at: "http://www.justice.gov/dea/programs/forensicsci/microgram/mg0305/mg0305.html" http://www.justice.gov/dea/programs/forensicsci/microgram/mg0305/mg0305.html [Accessed 09 October 2011].

Anon., n.d. *Several activation or transfer functions of artificial neurons (Electronics And Motor Drives)*. [Online] Available at: "http://what-when-how.com/electronics-and-motor-drives/several-activation-or-transfer-functions-of-artificial-neurons-electronics-and-motor-drives/" [Accessed 01 July 2011].

Asheq, K., 2004. *Data Fusion in Sensor Networks*. New York: University of Buffalo - The State University of New York.

Beck, N.E. & Hale, J.E., 1993. Cocaine "Body Packers". *Br J Surg*, pp.1513 - 1516.

Bedworth, M. & O'Brien, J., 1999. The Omnibus Model: A New Model of Data Fusion? In *2nd International Conference on Information Fusion*. Sunnyvale, 1999.

Bellman, R., 1961. *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

Berger, J., 1980. *Statistical Decision Theory: Foundations, Concepts, and Methods*. New York: Springer-Verlag.

Bethel, R.E. & Paras, G.J., 1998. PDF Multisensory Multitarget tracker. *IEEE Transactions on Aerospace and Electronic Systems*. Vol. 34, No. 1. pp.153 - 168.

Bishop, C., 1995. *Neural Network for Pattern Recognition*. Oxford: Oxford University Press.

Blasch, E., 2000. Assembling an Information-Fusion Human-Computer Cognitive Decision Making Tool. *IEEE AESS*, pp.11-17.

Blasch, E., 2009. Introduction to Level 5 Fusion: The Role of the User. In Liggins, M.E., Hall, D.L. & Llinas, J. *Handbook of Multisensor Data Fusion: Theory and Practice*. Florid: CRC Press. pp.503 - 535.

Bosch, M.E., Sanchez, A.J., Rojas, F.S. & Ojeda, C.B., 2007. Recent Development in Optical Fibre Biosensors. *Sensors*, pp.797-859.

Boudjemma, R. & Forbes, A.B., 2004. *Parameter Estimation Methods for Data Fusion*. National Physical Laboratory.

Brown, W.M. & Swonger, C.W., 1989. A Prospectus for Automatic Target Recognition. *IEEE Transactions on Aerospace and Electronic Systems*, pp.401 - 410.

Campos, F. & Campello de Souza, F.M., 2005. *Extending Dempster-Shafer Theory to Overcome Counter Intuitive Results*. IEEE.

Carvalho, H.S., Heinzelman, W.B., Murphy, A.L. & Coelho, C.J.N., 2003. A general data fusion architecture. In *Architecture*., 2003. Citeseer.

Casti, J.L., 1990. *Searching for Certainty: What Scientists can know about the Future.* New York: William Morrow and Company.

Charniak, E., Riesbeck, V. & McDermott, D., 1979. Artificial Intelligence Programming.

Clarke, B.S. & Barron, A.R., 1990. Entropy Risk and the Bayesian Central Limit Theorem. *Annals of Statistics*.

Clark, L.C. & Lyons, C., 1962. Electrode Systems for continuous monitoring in cardiouvascular surgery. *Academy of Science*, pp.29-45.

Clement, V., Giraudon, G., Houzelle, S. & Sadakly, F., 1993. Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion Using a Multispecialist Architecture. *IEEE Trans. on Geoscience and Remote Sensing*, pp.779-91.

Concheiro, M. et al., 2007. Confirmation by LS-MS of drugs in oral fluid obtained from roadside testing. *Forensic Science International*, pp.156-62.

Crouch, D.J. et al., 1998. A comparison of ontrak testcup, abuscreen, ontrak, abuscreen online and GC/MS urinanalysis test results. *Journal of Forensic Sciences*, pp.35-40.

Daar, A.S., 2002. Top ten biotechnologies for improving health in developing countries. *Nat. Genet*, pp.229-32.

Dasarathy, B., 1994. Decision Fusion.

David, M., Andrea, B., Mariolino, D.C. & Dario, P., 2008. Tutorial 14: Multisensor Data Fusion. *Instrumentation and Measurement Magazine*, pp.24-33.

DeMarco, D.R., Elric, S.W., McCrae, D.A. & Lim, D.V., 1999. Rapid Detection of Escherichia coli O157:H7 in Ground Beef using a Fibre-Optic Biosensor. *Journal of Food Protection*, pp.711-16.

Derrick, D.C. et al., 2010. Border Security Credibility Assessments via Heterogenous Sensor Fusion. *IEEE Computer Society*, pp.41 - 49.

Devijver, P.A. & Kittler, J., 1982. *Pattern Recognition - A Statistical Approach*. NJ: Prentice Hall.

Devine, P.J. et al., 1995. A Fibre-Optic Cocaine Biosensor. *Analytical Biochemistry*, pp.216 - 224.

Dey, A.K., Mankoff, J. & Abowd, G.D., 2000. *Distributed Mediation of Imperfectly Sensed Context in Aware Environments*. Atlanta: GVU Technical Report.

Ellsberg, D., 1961. Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, pp.643-49.

Emase, 2007. *Shipping Containers*. [Online] Available at: "http://www.emase.co.uk/data/cont.html" http://www.emase.co.uk/data/cont.html [Accessed 11 Augustus 2008].

Erhard, A. & McGalliard, S., 2011. Advaces in Military Multi-Sensor Data Fusion Technology and Applications for Civilian Use. In *University of Pittsburgh Swanson School of Engineering Eleventh Annual Freshman Conference*. Pittsburgh, 2011.

Estebani, J. et al., 2004. *A Review of Data Fusion Models and Architectures: Towards Engineering Guidelines*.

Fan, C., Plaxco, K.W. & Heeger, A.J., 2005. Biosensors based on binding-modulated donor-acceptor distances. *Trends Biotechnol.*, pp.186-92.

Fausett, L., 1994. *Fundamentals of Neural Networks*. Prentice Hall: New York.

Fidanboylu, K. & Efendioglu, H.S., 2009. Fibre Optic Sensors and their Applications. In *5th International Advanced Technologies Symposium (IATS '09)*. Karabuk, TUrkey, 2009. IATS Karabuk University, Turkey.

Fidanboylu, K. & Efendioglu, H.S., 2009. Fibre Optic Sensors and their Applications. In *5th International Advanced Technologies Symposium*. Karabuk, 2009.

Fowler, K.R. & Schmalzel, J.L., 2004. The First Stage in the Measurement Chain. *IEE Instrumentation and Measurement Magazine*.

Galbraith, E. & Nguyen, H., 2009. *Optical Fibre Sensors for Cocaine*. Sheffield.

Goutte, C., 1977. Note on free lunches and cross-validation. *Neural Computation*, pp.1211-15.

Grabherr, S. et al., 2008. Detection of Smuggled Cocaine in Cargo Using MDCT. *American Journal of Roentgenology*, pp.1390-95.

Grattan, K.T.V. & Meggitt, B.T., 1999. *Chemical and Environmental Sensing*. Kluwer Academic Publishers.

Guyon, I. & Elisseeff, A., n.d. An Introduction to Feature Extraction.

Gysin, C. & Mills, J., 2008. *MailOnline News Report*. [Online] Available at: HYPERLINK "http://www.dailymail.co.uk/news/article-1093090/Ex-England-cricket-star-Chris-Lewis-smuggled-200-000-cocaine-Gatwick.html" http://www.dailymail.co.uk/news/article-1093090/Ex-England-cricket-star-Chris-Lewis-smuggled-200-000-cocaine-Gatwick.html [Accessed 11 Augustus 2009].

Haack, B. & Bechdol, M., 1999. Multisensor Remote Sensing Data for Land Use/Cover Mapping. *Computers, Environment and Urban Systems*, pp.53 - 69.

Hall, D., 1992. *Mathematical Techniques for Multisensor Data Fusion*. MA: Artech House Inc.

Hall, D.L. & Garga, A.K., 1999. Pitfalls in Data Fusion (and How to Avoid Them). In *2nd International Conference on Information Fusion*. CA, 1999.

Hall, D.L. & Llinas, J.A., 1994. A challenge for the data fusion community I: research imperatives for improved processing. In *7th National Summit on Sensor Fusion*. Albuquerque, 1994.

Hall, D.L. & Llinas, J., 1997. An Introduction to Multisensor Data Fusion. *IEEE*, pp.6 - 23.

Hall, D. & Llinas, J., 2009. Multisensor Data Fusion. In Liggins, M., Hall, D. & Llinas, J. *Handbook of Multisensor Data Fusion (Theory and Practice)*. Florida: CRC Press. pp.1-14.

Harris, C.J., Bailey, A. & Dodd, T.J., 1998. Multi-sensor Data Fusion in Defence and Aerospace. *Aeronaut J*, pp.229 - 244.

Haupt, K. & Mosbach, K., 2000. *Chem Rev*, pp.2495-504.

Haykin, S., 1998. *Neural Networks: A Comprehensive Foundation*. Prentice Hall.

Hergan, K., Kofler, K. & Oser, W., 2004. Drug smuggling by body packing: what radiologists should know about it. *Eur Radio*, pp.736 - 742.

Hernandez, A. et al., 1996. Fusion of Ultrasonic and Radiographic images of the breast. In *IEEE Ultrasonic Symposium*. Texas, 1996.

Hernandez, A.I. et al., 1999. Multisensor Fusion for atrial and Ventricular Activity Detection in Coronary Care Monitoring. In *IEEE Transactions on Biomedical Engineering*., 1999.

Jayalakshmi, T. & Santhakumaran, A., 2011. Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*, pp.1793-8201.

Jellema, R., 2011. Analysis of Cocaine Hydrochloride in wax. *Microgram*, pp.183 - 189.

Jitendra, R.R., 2010. *Multisensor Datafusion with MATLAB*. Florida: CRC Press.

Jung, C.C. et al., 2003. RAPTOR: A Fluoroimmunoassay-Based Fiber Optic Sensor for Detection of Biological Threats. *IEEE Sensors Journal*, pp.352 - 360.

Kari, S., 2002. Combination of Evidence in Dempster-Shafer Theory. *SAND 2002-0835*.

Keller, J., 2008.

Klein, L., 1999. *Sensor and Data Fusion Concepts and Applications*. SPIE Optical Engineering Press.

Klir, G.J., 1999. Uncertainty and Information Measures for Imprecise Probability: An Overview. In *International Probabilities and their Applications*. Ghent, Belgium, 1999.

Koks, D. & Challa, S., 2005. *An Introduction to Bayesian and Dempster-Shafer Data Fusion*. Edingburgh, Australia: DSTO Systems Sciences Laboratory.

Kumar, S., Ghosh, J. & Crawford, M.M., 2001. Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data. In *IEEE Transactions on Geoscience and Remote Sensing*., 2001.

Lee, C. & Landgrebe, D., 1993. Feature Extraction Based On Decision Boundaries. 15(4).

Lerner, B., Guterman, H., Aladjem, M. & Dinstein, I., 1996. Feature Extraction by Neural Network Nonlinear Mapping for Pattern Classification. In *ICPR'96*., 1996. IEEE.

Lerner, B., Guterman, H., Aladjem, M. & Dinstein, I., n.d. Feature Extraction by NN Nonlinear Mapping.

Liu, H., 1998. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers.

Liu, J., Yang, J.-B., Wang, J. & Sii, H.S., 2002. Review of Uncertainty Reasoning Approaches as Guidance for Maritime and Offshore Safety-based Assessment. *Journal of UK Safety and Reliability Society*, pp.63 - 80.

Lowrance, J. & Garvey, T., 1982. Evidential Reasoning: A Developing Concept. In *International Conference on Cybernetics and Society*., 1982.

Lowrance, J.D. & Garvey, T.D., 1982. Evidential reasoning: A developing concept. In *IEEE International. Conf. on Cybernetics and Society*. Seattle, WA, 1982.

Lowry, S.R. & Isenhour, T.L., 1975. A Feature Selection Technique for Binary Infrared Spectra. *Journal of Chemical Information and Modeling*, pp.212-15.

Luo, R.C., Chang, C.C. & Lai, C.C., 2011. Multisensor Fusion and Integration: Theories, Applications and its Perspectives. *IEEE Sensors Journal*, pp.3122 - 3137.

Lyons, W.B. & Lewis, E., 2000. Neural networks and pattern recognition techniques applied to optical fibre sensors.

Ma, B., 2001. *Parametric and Non-Parametric approaches for Multisensor Data Fusion*. Michigan: PhD Thesis.

Mandiak, M., Shah, P., Kim, Y. & Kesavadas, T., 2005. Development of an Integrated GUI Framework for Post-Disaster Data Fusion Visualization. In *International Conference on Information Fusion (FUSION)*. Ney York, 2005.

Mao, J. & Jain, A.K., 1995. Artificial Neural Networks for Feature Extraction and Multivariate Data Projection. 6.

Marks, A., 2007. Drug Detection Dogs and the Growth of Olfactory Surveillance: Beyond the Rule of Law? *Surveillance and Society*, pp.257 - 271.

Matia, F. & Jimenez, A., 1998. Multisensor Fusion: an Autonomous Mobile Robot. *Journal of Intelligent and Robotic Systems*, pp.129 - 141.

Matthews, B.W., 1975. Comparison of the Predicted and Observed Secondary Structure of T4 phagelysozyme. *Biochim. Biophys. Acta 405*, pp.442 - 451.

Maurer, D.E., Schirmer, R.W., Kalandros, M.K. & Peri, J.J., 2006. *Sensor usion Architectures for Ballistic Missile Defense*. Maryland: Johns Hopkins APL.

Mayer, C., 2010. *Cricketer caught smuggling cocaine*. [Online] Available at: "http://uk.news.yahoo.com/4/20090512/tuk-cricketer-faces-drugs-smuggling-tria-dba1618_1.html" [Accessed 12 Augustus 2010].

Meijer, R. & Bots, M.L., 2003. Detection of intestinal drug containers by ultrasound scanning: an airport screening tool? *Eur Radiol*, pp.1312 - 1315.

Mitchell, H.B., 2007. *Multi-Sensor Data Fusion*. New York: Springer Berlin Heidelberg.

Montagnana, M., Caputo, M., Giavarina, D. & Lippi, G., 2009. Overview on self-monitoring of blood glucose. *Clin. Chim*, pp.7-13.

Morgan, M.T. et al., 2006. Binding Inhibition Assay using Fibre Optic based Biosensor for the detection of Foodborne Pathogens. *Key Engineering Materials*, pp.1145-50.

Myers, J.W., Laskey, K.B. & DeJong, K.A., 1999. Learning Bayesian Networks from Incomplete Data using Evolutionary Algorithms. In *The Genetic and Evolutionary Computation Conference*. Orlando, 1999.

Nath, N. et al., 1999. A Rapid reusable Fibre Optic Biosensor for detecting cocaine metabolites in urine. *Journal of Analytical Toxicology*, pp.460 - 467.

Neumann, B., 2008. *Radiation Safety Training*. Synchrotron Radiation Centre, University of Wisconsin-Madison.

Nguyen, H., Sun, T., Grattan, K. & Hardwick, S.A., 2010. A fibre-optic chemical sensor for the detection of cocaine.

Ni, X., 2008. Research of Data Mining Based on Neural Networks. *World Academy of Science, Engineering and Technology*, pp.381-84.

Official Documents, 2009. *The Home Office Departmental Report*. [Online] Available at: "http://www.official-documents.gov.uk/document/cm75/7592/7592.pdf" [Accessed 11 Augustus 2009].

Oja, E., 1992. Principal components, minor components and linear neural networks. 5.

Osoba, O., Mitaim, S. & Kosko, B., 2011. Bayesian Inference with Adaptive Fuzzy Priors and Likelihoods. *Systems, Man and Cybernetics*.

Oxford English Dictionary 2005 OUP Oxford, UK.

Oxford, D., 2011. *cargo*. [Online] Available at: "http://oxforddictionaries.com/view/entry/m_en_gb0124620" \l "m_en_gb0124620" [Accessed 31 March 2011].

Pangop, L.N., Chausse, F., Lasmea, R.C. & Cornou, S., 2003. Asynchronous Bayesian Algorithm for Object Classification: Application to Pedestrian Detection in Urban Areas.

Pao, Y.H., 1989. *Adaptive Pattern Recognition and Neural Network*. Addison-Wesley Publishing Company Inc.

Patterson, D., 1996. *Artificial Neural Networks*. Singapore: Prentice Hall.

Pechenizkiy, M., Puuronen, S. & Tsymbal, A., n.d. FEATURE EXTRACTION FOR CLASSIFICATION IN THE DATA MINING PROCESS. *International Journal "Information Theories & Applications"*, pp.271 -278.

Punska, O., 1999. *Bayesian Approaches to Multi-Sensor Data Fusion*. Cambridge: University of Cambridge.

Rocha, L.M., n.d. Relative Uncertainty and Evidence Sets: A Constructivist Framework. *International JOurnal of General Systems*, pp.35 - 61.

Rojas, R., 1996. *Neural Network A Systematic Introduction*. Berlin: Springer-Verlag.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J., 1986. Learning Internal Representations by Error Propagation. In Rumelhart, D.E. & McClelland, J.L. *Parallel Distributed Processing:Explorartions in the Microstructure of Cognition*. Cambridge: MIT Press.

Sedat, J.W., Shuman, M., Agard, D. & Stroud, R., 2010. *Letter to Dr John P. Holdren*. [Online] Available at: "http://www.whitehouse.gov/sites/default/files/microsites/ostp/ucsf-jph-letter.pdf." [Accessed 15 January 2011].

Seims, S.J., 2009. *A Study of Dempster-Shafer's Theory of Evidence in comparison to Classical Probability Combination*. San Luis: Carlifornia Polytechnic State University.

Sentz, K. & Ferson, S., 2002. Combination of Evidence in Dempster-Shafer Theory.

Shafer, G., 1990. Perspective on the Theory and Practice of Belief Functions. *International Journal of Approximate Reasoning*, pp.1 - 40.

Skurichina, M., Paclik, P. & Duin, R.P., 2004. Selection/Extraction of Spectral Regions for Autofluorescence Spectra Measured in the Oral Cavity. In *Lecture Notes in Computer Science*. New York, 2004. Springer-Verlag.

Skurichina, M., Verzakov, S., Paclik, P. & Duin, R., n.d. *Effectiveness of Spectral Band Selection/Extraction Techniques for Spectral Data*. Delft: Delft University of Technology.

Smith, A.D., 2008. *Sensors may revolutionize everyday life*. [Online] Available at: "http://www.dentonrc.com/sharedcontent/dws/bus/stories/DN-sensors_06bus.ART0.State.Edition1.46a4a51.html" [Accessed 05 January 2012].

Smith, T.M. & Kelly, P.A., 1999. Random Sets Technique for Information Fusion Applied to Estimation of Brain Functional Images. *Proceedings of SPIE- The International Society for Optical Engineering*, pp.1158 - 1169.

Soriguera, F., 2011. *Highway travel time estimation with data fusion*. Barcelona: Catedra Abertis.

Sossai, C., Bison, P., Chemello, G. & Trainito, G., 1999. Sensor Fusion for Localization using Possibility Theory. *Control Engineering Practice*, pp.773 - 782.

Soumekh, M., 1999. Signal Subspace Fusion of Uncalibrated Sensors with Application in SAR and Diagnostic Medicine. *IEEE Transactions on Image Processing*, pp.127 - 137.

Stanley, M., 2010. *German Smuggled Heroin in Car Airbag*. [Online] Available at: "http://uk.news.yahoo.com/4/20090520/tuk-german-smuggled-heroin-in-car-airbag-dba1618.html" [Accessed 12 August 2010].

Staples, E.J. & Viswanathan, S., 2008. Detection of Contraband in Cargo containers using a High Speed Gas Chromatograph with Surface Acoustic Wave Sensor. *Ind. Eng. Chem*, pp.8361 - 8367.

Steinberg, A.N. & Bowman, C.L., 2009. Revisions to the JDL Data Fusion Model. In Liggins, M.E., Hall, D.L. & Llinas, J. *Handbook of Multisensor Data Fusion*. CRC Press. pp.45 - 67.

Steinberg, A.N., Bowman, C.L. & White, F.E., 1998. Revisions to the JDL Data Fusion Model. In *Sensor Fusion: Architectures, Algorithms and Applications, Proceedings of the SPIE Conference*., 1998.

Steinberg, A.N., Bowman, C.L. & White, F.E., 1999. Sensor Fusion: Architectures, Algorithms and Applications. In *SPIE*., 1999.

Stubbs, D.D., Lee, S.-H. & Hunt, W.D., 2005. Cocaine Detection of a Narcotic Using Surface Acoustic Wave Immunoassay Sensors. In *IEEE Sensors Journal*., 2005.

Stubbs, D.D., Lee, S.-H. & Hunt, W.D., n.d. Cocaine Detection Using Surface Acoustic Wave Immunoassay Sensors. In *IEEE International Frequency Control Symposium and PDA Exhibition*.

Sun, Z. et al., 2005. Using FCMC, FVS, and PCA Techniques for feature extraction of multispectral images. *IEEE Geoscience Remote Sensing Letters*, pp.108-12.

Sutter, J. & Jurs, P., 1997. Neural Network Classification and Quantification of Organic Vapors Based on Fluorescence Data from Fiber-Optic Sensor Array. 69.

Thomopoulos, S.C., 1989. Sensor Integration and Data Fusion. In Thomopoulos, S.C. *Sensor Fusion II: Human and Machine Strategies*.

Thomopoulos, S.C., 1989. Sensor Integration and Data Fusion. *Proceedings of SPIE 1198, Sensor Fusion II: Human and machine strategies*, pp.178 - 191.

Tromp, D., 2006. *Sensor Technologies for Border and Cargo Security*. MITRE.

TSA, T.S.A., 2011. *Frequently Asked Questions (Advanced Imaging Technology)*. [Online] Available at: "http://www.tsa.gov/approach/tech/ait/faqs.shtm" [Accessed 20 September 2011].

U.S. Department of defense, d.F.S.o.t.J.D.o.L.T.P.f.C., 1991. *Data Fusion Lexicon*.

UKBA, 2009. UK Border Agency Business Plan April 2009 – March 2012.

Varghese, B.P., Satish, J. & Madhusoodanan, K.N., 2009. Fiber optic sensor for the measurement of concentration of silica in water with dual wavelength probing. *Review of Scientific Intruments*, 81(3).

Velosos, M., Bento, C. & Pereira, F.C., 2009. *Muti-Sensor Data Fusion on Intelligent Transport Systems*. MITPortugal Transport Systems.

Verzakov, S., Paclik, P. & Duin, R.P., 2004. Feature Shaving for Spectroscopic Data. In *Lecture Notes in Computer Science*. Berlin Heidelberg, US, 2004. Springer-Verlag.

Vogel, H. & Brogdon, B.G., 2003. Search of Luggage, Cargo and Transport. In Brogdon, B.G. *A Radiological Atlas of Abuse, Torture, Terrorism, and Inflicted Trauma*. Florida: CRC Press. pp.309 - 316.

Vogel, H., Haller, D. & Laitenberger, C., 2006. Rontgendiagnostik des Drogentransports. *Arch Kriminol*, pp.1 - 21.

Wald, L., 2000. Data Fusion. In *EARSeL Symposium*. Dresden, 2000.

Waltz, E.L. & Llinas, J., 1990. *Multisensor Data Fusion*. Maryland: Artech House, Inc.

Warner, M.J., 1993. *Ann Intern Med*, pp.226 - 234.

Wei-feng, T., Li, Q. & Tong, S.J., 2002. Application of Neural Networks for Sensor Fusion in A Remote Sensing Satellite. In *IEEE International Symposium on Intelligent Control*. Vancouver, Canada, 2002.

White, J. et al., 1997. Identification of Multiple Analytes Using an Optical Sensor Array and Pattern Recognition Neural Networks. *Anal Chem*, pp.4641-48.

Wu, H., 2003. *Sensor Data Fusion for Context-Aware Computing Using Dempster-Shafer Theory*. Pennsylvania: The Robotics Institute Carnegie Mellon University.

Yager, R., 1987. On the Dempster-Shafer Framework and New Combination Rules.

Yager, R. & Kacprzyk, J., 1994. *Advances in the Dempster-Shafer Theory of Evidence*. John Wiley and Sons.

Yu, X. & Yi, L., 2011. Using personal glucose meters and functional DNA sensors to quantify a variety of Analytical Targets. *Nature Chemistry*, pp.697-703.

Zadeh, L.A., 1968. Fuzzy Algorithms. *Info Control*, pp.94-102.

Zou, Y., Ho, Y.K., Chua, C.S. & Zhou, X.W., 2000. Multi-ultrasonic sensor fusion for autonomous mobile robots. *Sensor Fusion: Architectures, Algorithms and Applications IV*, p.314–321.

# APPENDIX

## Appendix A

The figure below shows the intensity plots of raw data against wavelength for cocaine at five different concentrations. Cocaine at 1000uM has the highest wavelength peak followed by 500uM, 250uM, 100uM and 25uM respectively implying a proportional increase in concentration relative to wavelength peak. This trend is repeated for other compounds.

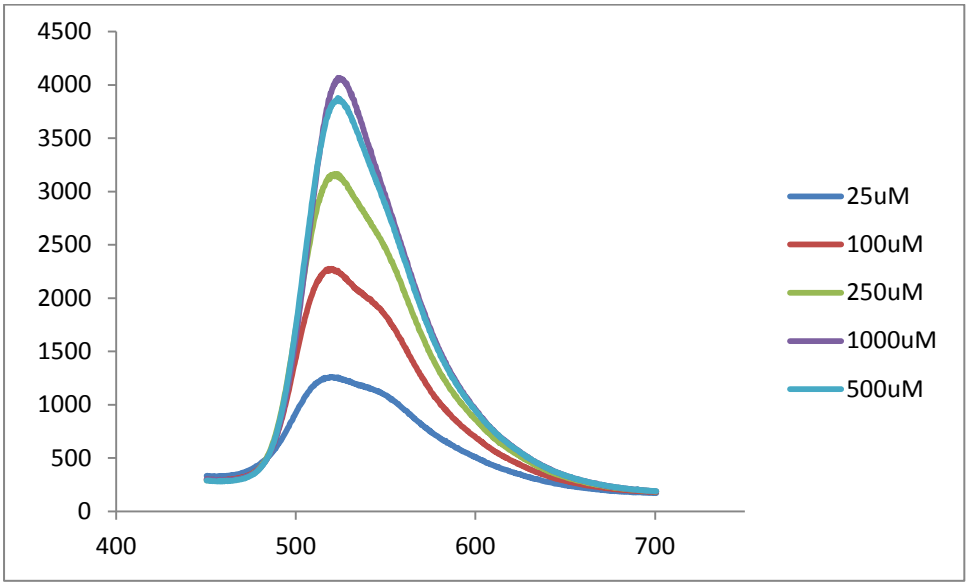



Figure A (i) raw data plot of intensity vs wavelength for cocaine at different concentrations

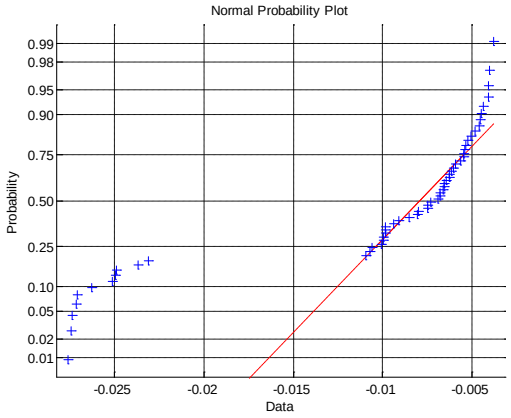
# Appendix B

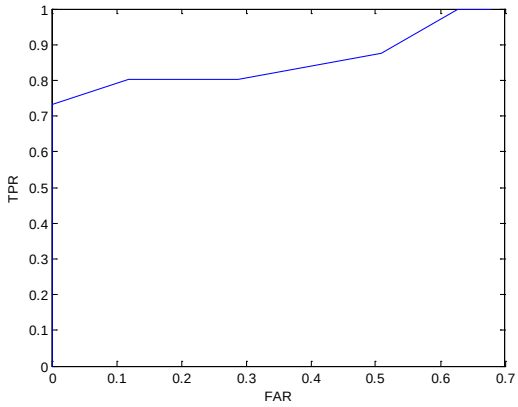
**Normal and Receiver Operator Characteristics plot for features**

The Receiver-Operator Characteristics (ROC) curve is a plot of false alarm rates (specificity) versus true positive (sensitivity) for every possible threshold value. It is also a technique to select appropriate threshold value based on trade-offs between the specificity and selectivity.

The area under the ROC curve (AUC) is used as a parameter to indicate the accuracy of the test. Ideally, the area under the curve should be 1 however, in real life, this is rarely so. Realistically, the closer to 1 the AUC is, the more accurate the result.

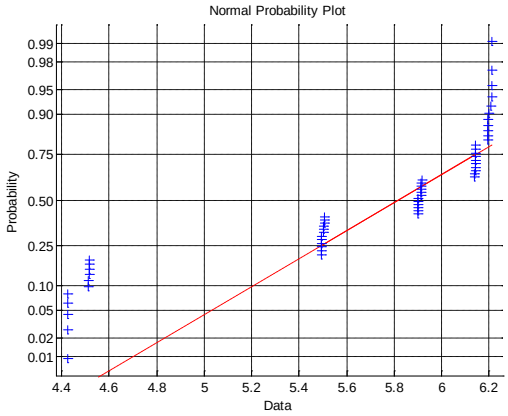
**Feature 1**



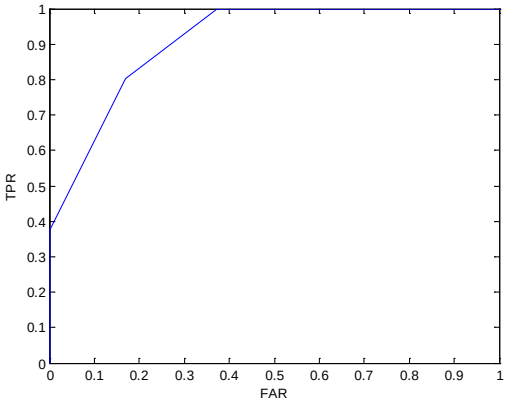




FigureB(i) Normal Plot for feature 1 FigureB(ii) ROC curve for feature 1

## Feature 2

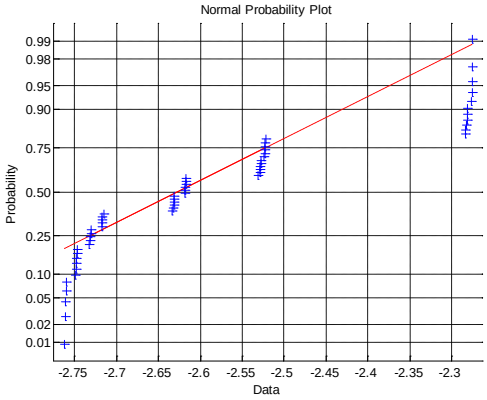



FigureB(iii)Normal Plot for feature 2

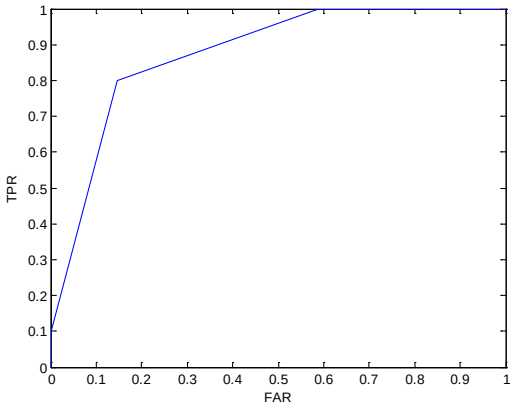



FigureB(iv) ROC curve for feature 2

## Feature 3

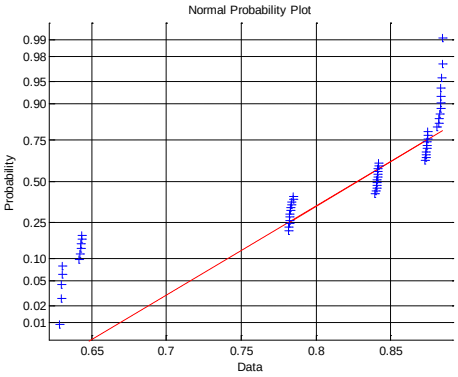



FigureB(v)Normal Plot for feature 3

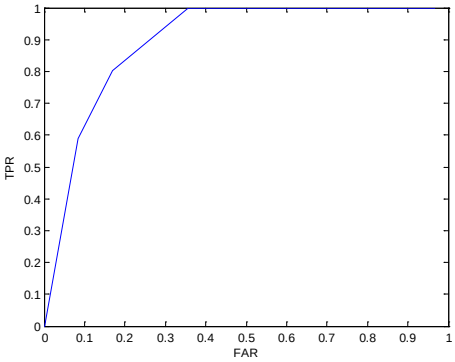



FigureB(vi) ROC curve for feature 3

## Feature 4

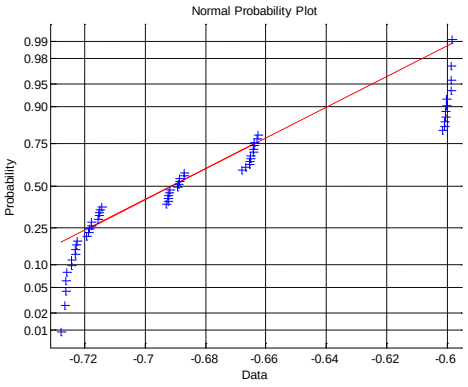



FigureB(vii)Normal Plot for feature 4

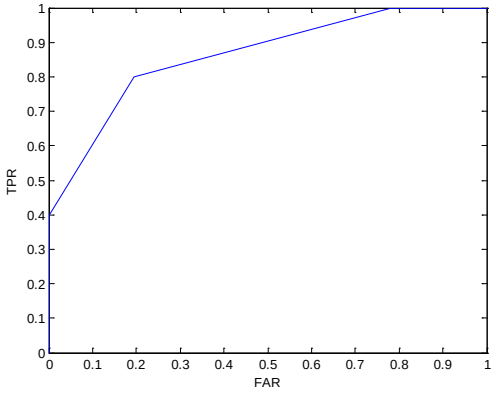



FigureB(viii) ROC curve for feature 4

## Feature 5

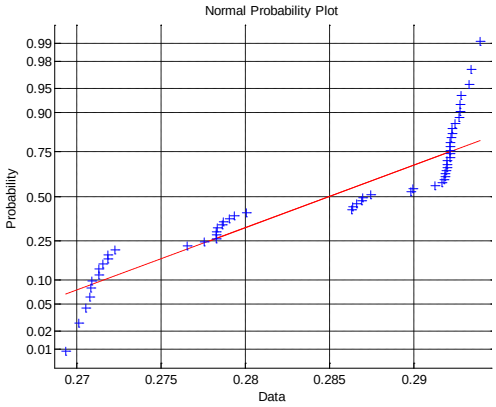



FigureB(ix)Normal Plot for feature 5

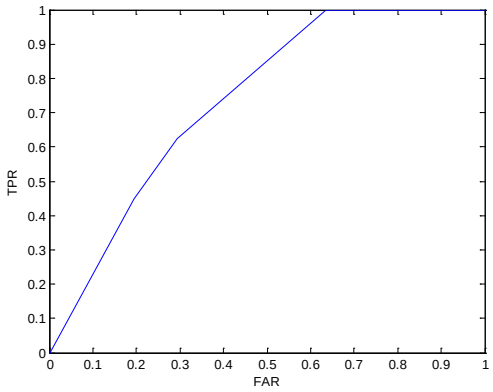



FigureB(x) ROC curve for feature 5

## Feature 6

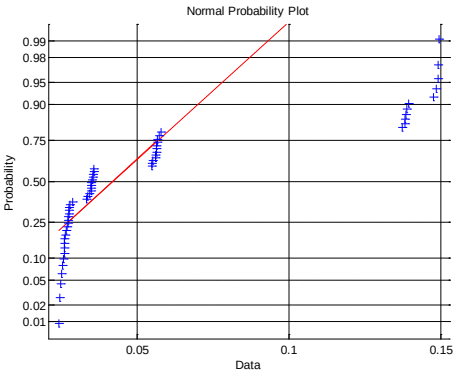



FigureB(xi)Normal Plot for feature 6

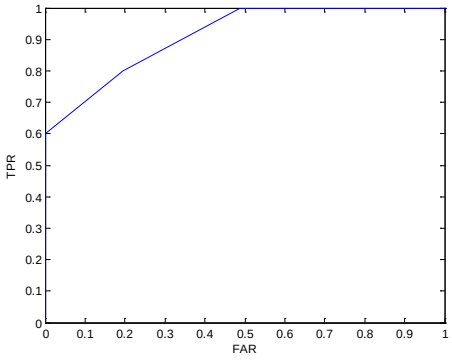



FigureB(xii) ROC curve for feature 6

## Feature 7

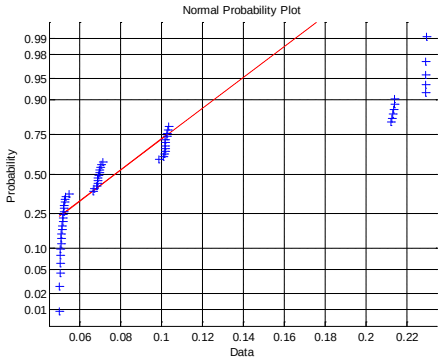



FigureB(xiii)Normal Plot for feature 7

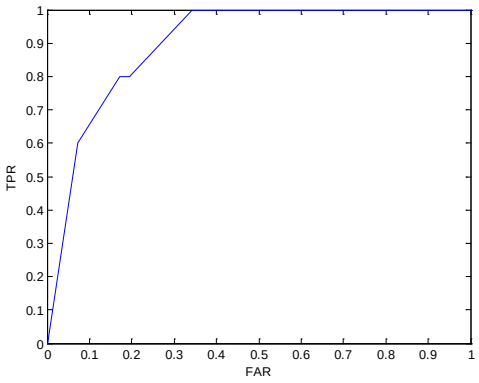



FigureB(xiv) ROC curve for feature 7

## Feature 8

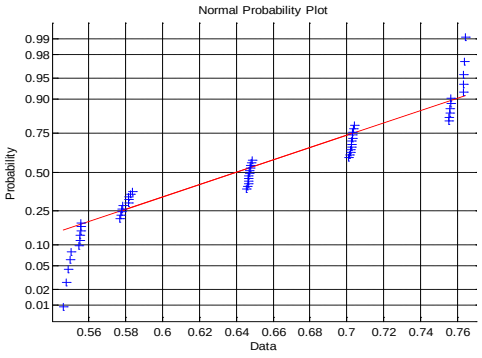



FigureB(xv)Normal Plot for feature 8

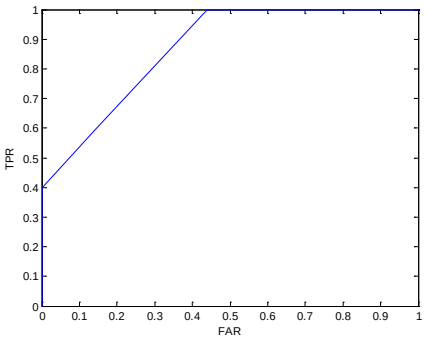



FigureB(xvi) ROC curve for feature 8

## Feature 9

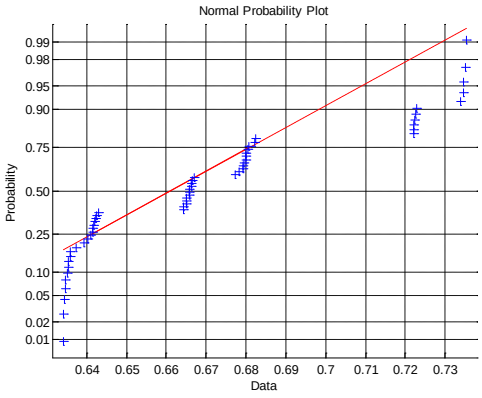



FigureB(xvii) Normal Plot for feature 9

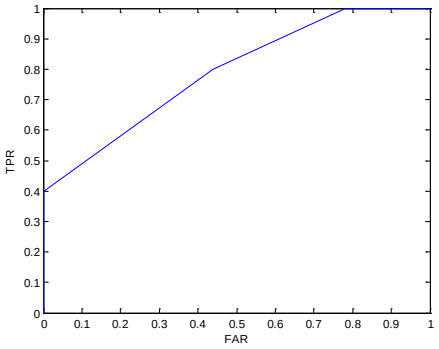



FigureB(xviii) ROC curve for feature 9

## Feature 10

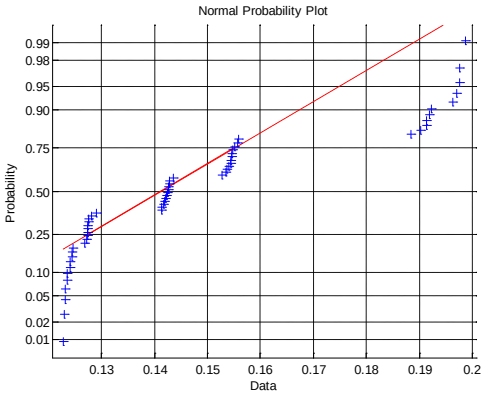



FigureB(xix) Normal Plot for feature 10

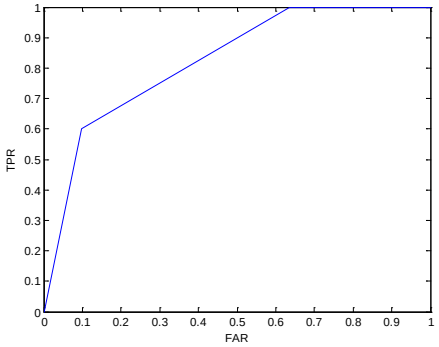



FigureB(xx) ROC curve for feature 10

## Feature 11

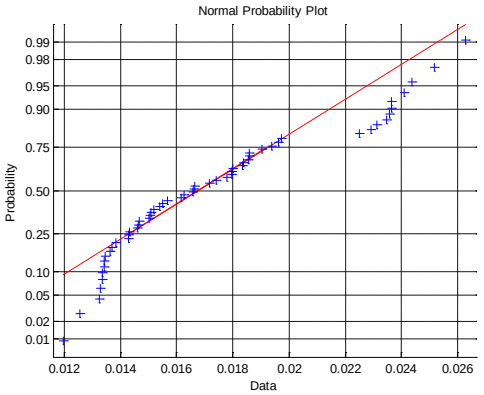



FigureB(xxi) Normal Plot for feature 11




FigureB(xxii) ROC curve for feature 11

## Feature 12




FigureB(xxiii) Normal Plot for feature 12

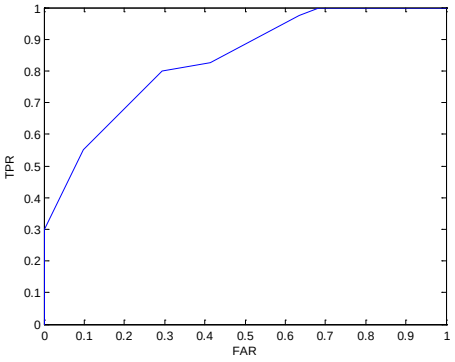



Figure B(xxiv) ROC curve for feature 12

# Appendix C

## Feature1 (Average change in intensity)

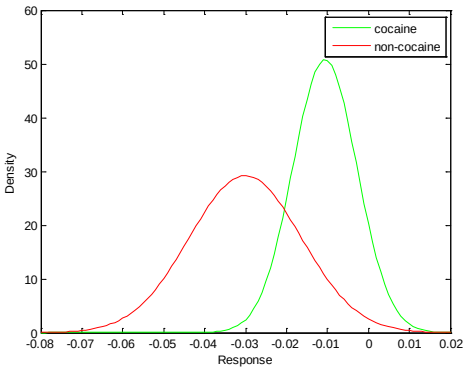



Figure C(i) Normal distribution curve for classification using feature 1

**Table C(i) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| **-0.015** | **0.8** | **0.15** |
| -0.025 | 0.85 | 0.61 |
| -0.03 | 1 | 0.76 |

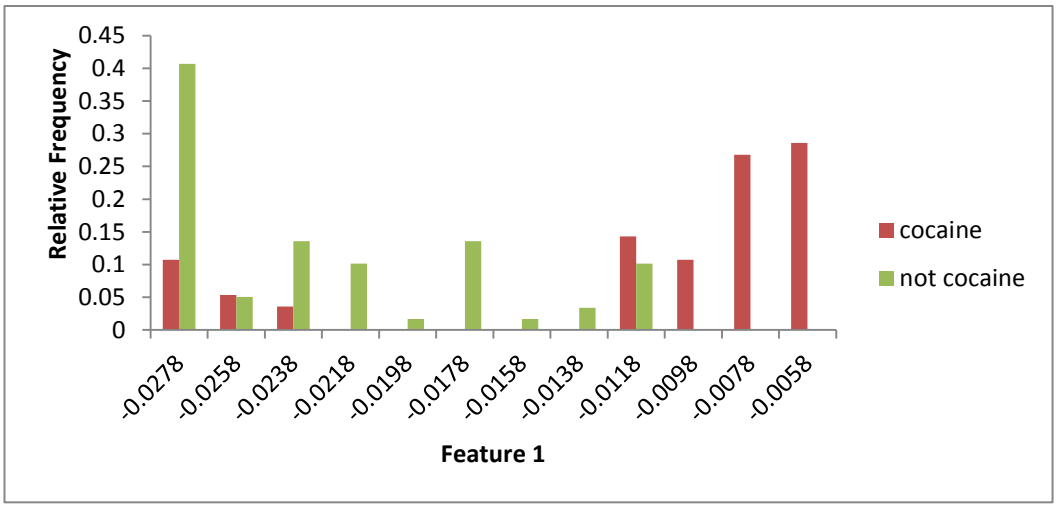



Figure C(ii) Column chart showing relative frequency of data for feature 1

Using -0.015 as the threshold, the feature misclassified 15% of the not-cocaine test data as cocaine and 19% of the cocaine data as not-cocaine.

## Feature 2 (Steepest Slope)

AUC = 91.46%

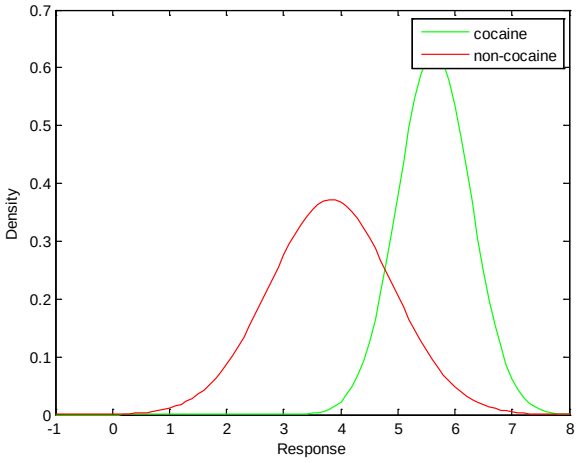



Figure C(iii) Normal distribution curve for classification using feature 2

**Table C (ii) TPR and FAR for varying threshold (Steepest slope)**

| Threshold | TPR | FAR |
|---|---|---|
| 2.0 | 1 | 1 |
| 2.5 | 1 | 0.9512 |
| 3.0 | 1 | 0.8780 |
| 3.5 | 1 | 0.7317 |
| 5.0 | 0.80 | 0.1951 |
| **5.5** | **0.70** | **0.0976** |

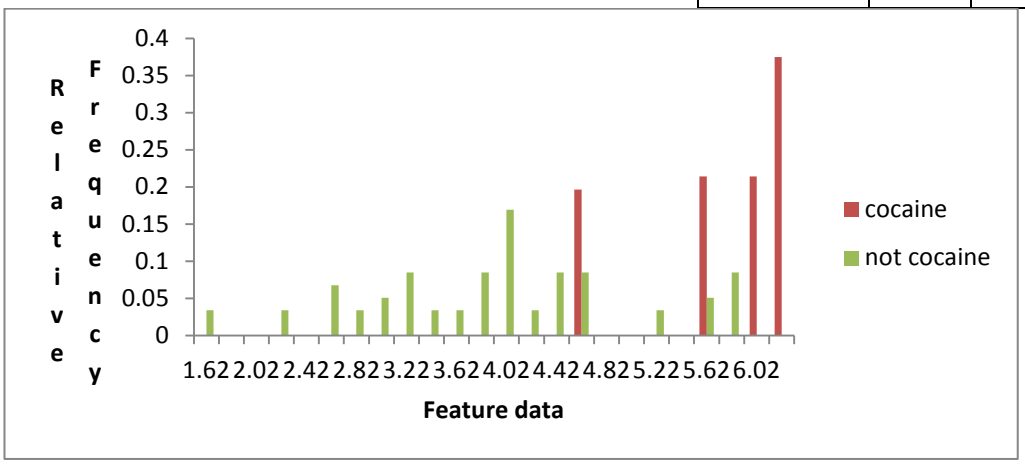



Figure C(iv) Column chart showing relative frequency of data for feature 1

Feature 2 involves classifying the analytes using the 'steepest slope' as the input. As can be seen from the table, the best threshold value is between threshold 4 and threshold 5 giving FAR and TPR pairs of (19.51%, 80%) and

(9.76%, 70%) respectively. The Area under the ROC curve is 91.46%. In this case, with a lower false alarm rate of about 9%, the threshold 5.5 is chosen.

Using 5.50 as the threshold, the feature misclassified 5% of the not-cocaine test data as cocaine and approximately 31% of the cocaine test data as not-cocaine.

**Feature 3 (Maximum negative slope)**

AUC = 87.56%

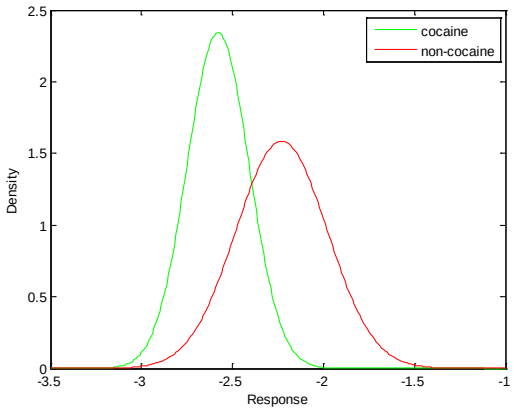



Figure C(v) Normal distribution curve for classification using feature 3

**Table C (iii) TPR and FAR for varying threshold (maximum negative slope)**

| Threshold | TPR | FAR |
|---|---|---|
| -1.5 | 1 | 1 |
| -1.75 | 1 | 0.9756 |
| -2.00 | 1 | 0.8780 |
| -2.25 | 1 | 0.5854 |
| **-2.50** | **0.8** | **0.1463** |
| -2.75 | 0.1 | 0 |

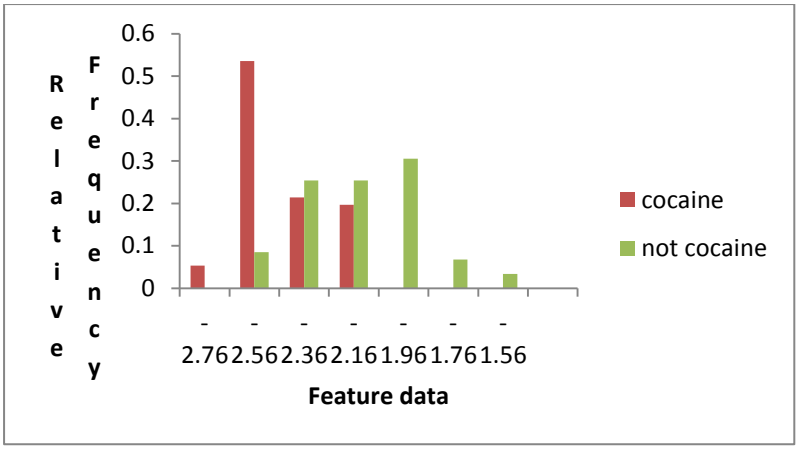



Figure C(vi) Column chart showing relative frequency of data for feature 3

Using -2.50 as the threshold, the feature misclassified 15% of the not-cocaine test data as cocaine and approximately 20% of the cocaine test data as not-cocaine.

Classification by the maximum negative slope gives the above figure and table. From the table C(iii), classification is done by values less than the threshold and the TPR and FAR values are shown on the table. With a target FAR value less than 0.5, the chosen threshold in this case is -2.50. The -2.25 threshold gives a good TPR but a poor FAR value compared to the target FAR value for all the features and is thus neglected. The area under the curve 88% is a good figure.

## Feature 4 (Maximum change in intensity)

AUC = 90.49%

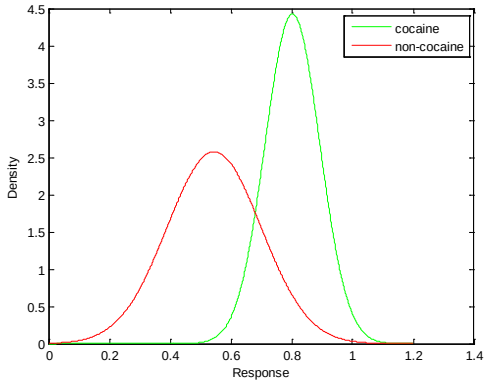



Figure C(vii) Normal distribution curve for classification using feature 4

**Table C (iv) TPR and FAR for varying threshold (maximum positive change in intensity)**

| Threshold | TPR | FAR |
|---|---|---|
| 0.35 | 1.0000 | 0.9512 |
| 0.55 | 1.0000 | 0.5854 |
| 0.60 | 1.0000 | 0.4146 |
| **0.75** | **0.8000** | **0.1463** |
| 0.80 | 0.6000 | 0.0976 |

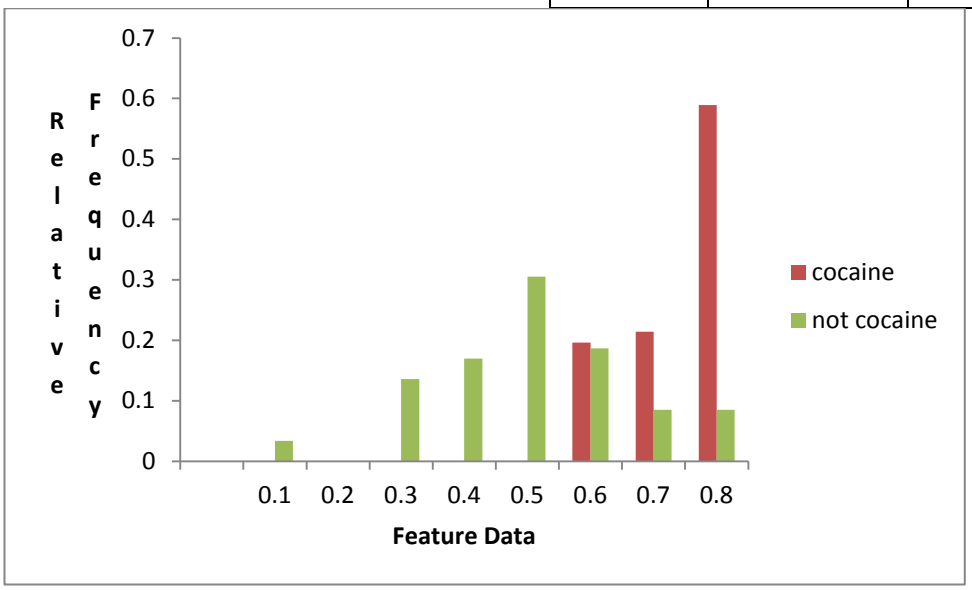



Figure C(viii) Column chart showing relative frequency of data for feature 4

Using 0.75 as the threshold, the feature misclassified 11% of the not-cocaine test data as cocaine and 19% of the cocaine test data as non-cocaine. As can be seen from the table above, with a false alarm rate of 41.5%, this feature will give 100% true positive rate using 0.6 as the threshold. Increasing the threshold to 0.75, the false alarm rate reduces to 15% with about 80% detection. Thus the 'best' threshold lies between 0.6 and 0.75. Choosing the former assures that there will be no missed detection but a high false alarm compared to the later. We have thus selected 0.75 as the best threshold in this case.

**Feature 5 (Maximum negative change in intensity)**

AUC = 86.34%

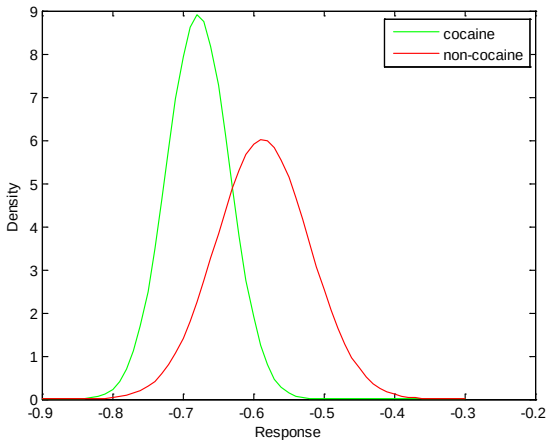



Figure C(ix) Normal distribution curve for classification using feature 5

**Table C (v) TPR and FAR for varying threshold (maximum negative change in intensity)**

| Threshold | TPR | FAR |
|---|---|---|
| -0.45 | 1 | 0.9756 |
| -0.50 | 1 | 0.9268 |
| -0.55 | 1 | 0.7805 |
| -0.60 | 0.90 | 0.4878 |
| **-0.65** | **0.80** | **0.1951** |
| -0.70 | 0.40 | 0 |

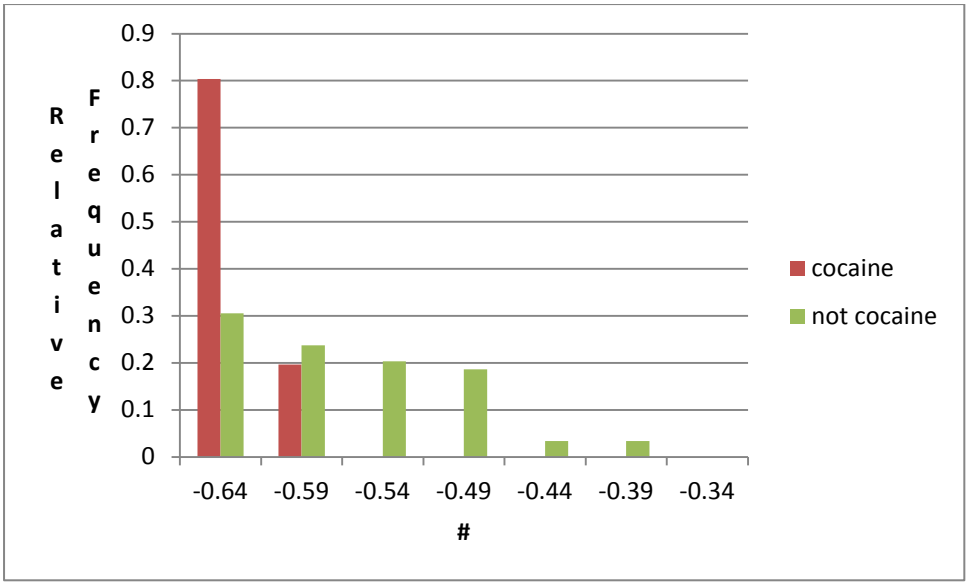



Figure C(x) Column chart showing relative frequency of data for feature 5

Using -0.65 as the threshold, the feature misclassified 10.5% of the not-cocaine test data as cocaine and correctly classified 81% of the cocaine test data as cocaine.

From observation of the normal distribution curve (table C(v), like feature 3, the positive detection are values to the left of the threshold. With this in mind, the threshold value of -0.6 and -0.65 act as the two possible thresholds with realistic TPR and FAR pairs. However, with a considerable lower FAR, the -0.65 threshold is chosen.

The AUC value of 86.34% confirms that this is a good test for detection of cocaine.

## Feature 6 (Peak)

AUC = 73.96%

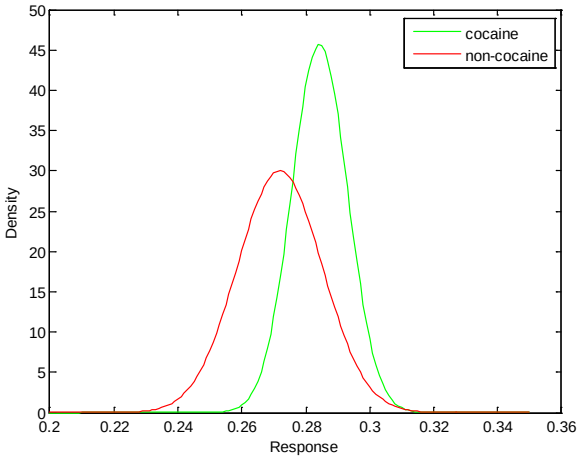



Figure C (xi) Normal distribution curve for classification using feature 6 (peak)

**Table C (vi) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.24 | 1 | 1 |
| 0.25 | 1 | 0.9512 |
| 0.27 | 1 | 0.6341 |
| **0.28** | **0.6250** | **0.2927** |
| 0.29 | 0.45 | 0.1951 |
| 0.30 | 0 | 0 |

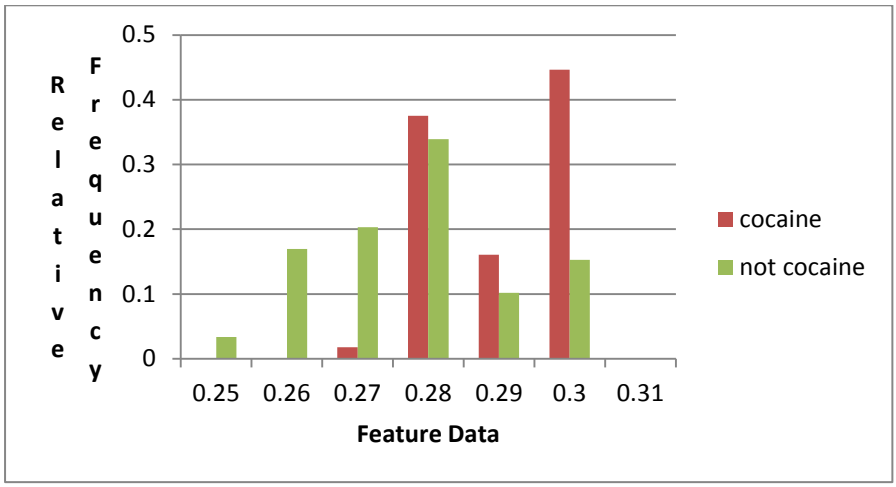



Figure C(xii) Column chart showing relative frequency of data for feature 6

Using 0.75 as the threshold, the feature misclassified 16% of the test non-cocaine data as cocaine and 44% of the test cocaine data as non-cocaine.

From table C(vi) above, we can infer that using the peak as classification feature, one can get as high as 100% true positive probability but with a corresponding poor false alarm rate of almost 63%. False alarm rates can be a nuisance to security agents and our aim is to have as low false alarms as possible within a satisfactory true positive rate. Therefore, a threshold with a lower false alarm is selected i.e. 0.28 with corresponding true positive and false positive rates of [0.6250 0.2927].

**Feature 7 (Average value of Region 1)**
AUC = 91.22%

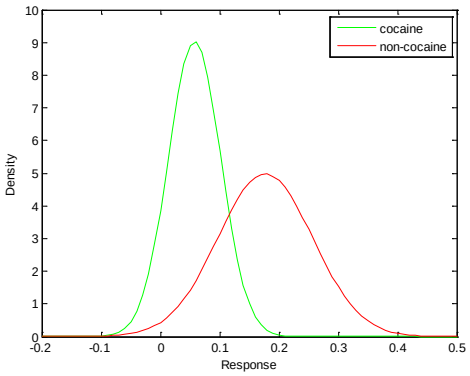



Figure C(xiii) Normal distribution curve for classification using feature 7

**Table C (vii) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.3 | 1 | 1 |
| 0.2 | 1 | 0.7805 |
| **0.1** | **0.80** | **0.1951** |
| 0 | 0 | 0 |

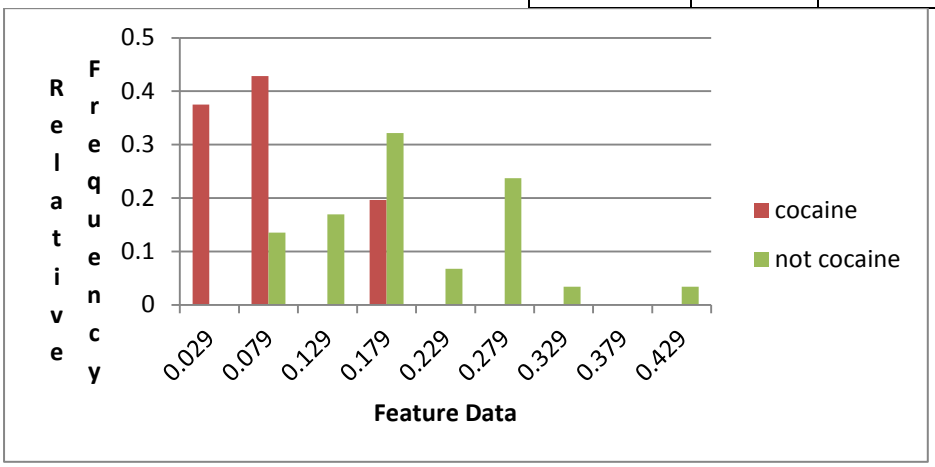



Figure C(xiv) Column chart showing relative frequency of data for feature 7

Using 0.1 as the threshold, the feature correctly classified 81% of the test cocaine data and 89% of the not-cocaine test data cocaine data and not-cocaine respectively.

From table C(vii), it can be seen that this feature gives as high as 80% true positive probability but with a corresponding false alarm rate of almost 20%. The AUC value of 91.22% confirms that this is a good test.

**Feature 8 (Average value of Region 2)**

AUC = 90%

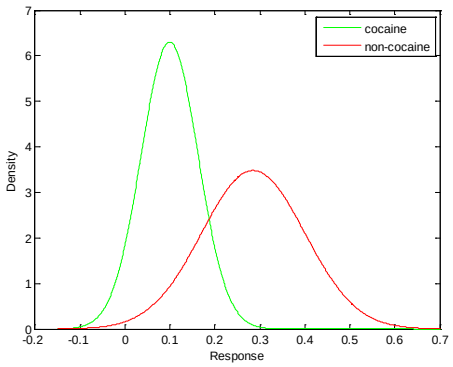



Figure C(xv) Normal distribution curve for classification using feature 8

**Table C (viii) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.45 | 1 | 1 |
| 0.40 | 1 | 0.9268 |
| 0.30 | 1 | 0.6829 |
| **0.15** | **0.80** | **0.1707** |
| 0.05 | 0 | 0 |

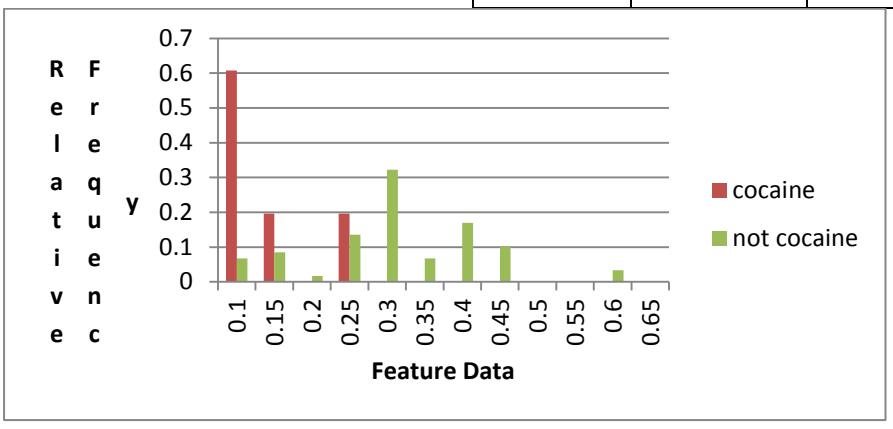



Figure C(xvi) Column chart showing relative frequency of data for feature 8

With 0.15 as the threshold, the feature misclassified 19% of the test cocaine data and 11% of the not-cocaine test data as not-cocaine data and cocaine respectively.

From table C(viii), for 0.15 threshold, the corresponding TPR is 0.8 and FAR is 0.17. The AUC value is 90%.

**Feature 9 (Average value of Region 3)**
AUC = 86.83%

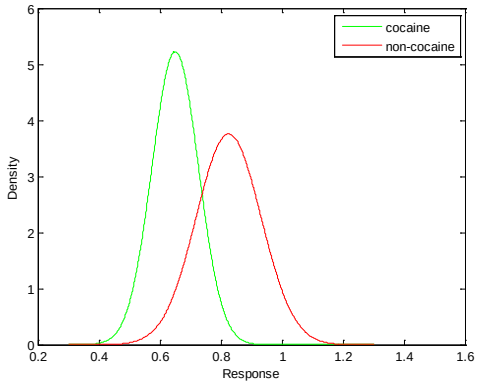



Figure C(xvii) Normal distribution curve for classification using feature 9

**Table C (xix) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.6 | 1.0 | 0.8293 |
| **0.7** | **0.60** | **0.1463** |
| 0.8 | 0.40 | 0 |

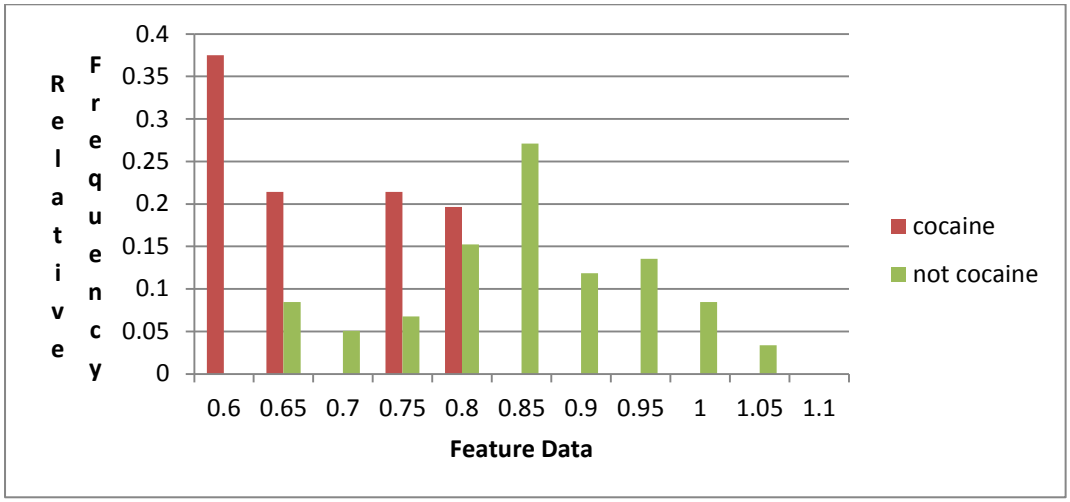



Figure C(xviii) Column chart showing relative frequency of data for feature 9

For feature 9, the AUC is 86.83% which represents a good value. For the selected threshold, i.e. 0.7, the TPR/FAR rates are 0.60 and 0.15 respectively. The false alarm rate is good enough but the true positive rate of 0.6 is just above average. However, for the test data, the feature misclassifies 10.53% of the test not-cocaine data as cocaine a high value of 44% cocaine data as non-cocaine implying a high missed detection rate for the test data.

**Feature 10 (Average value of Region 4)**

AUC = 79.02%

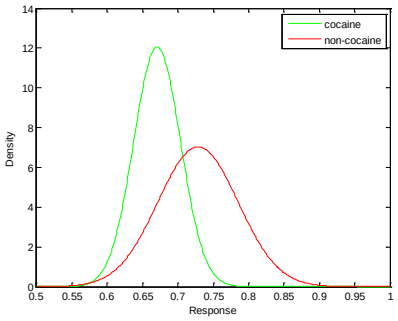



Figure C(xix) Normal distribution curve for classification using feature 10

**Table C (x) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.85 | 1 | 0.9756 |
| 0.75 | 1 | 0.7805 |
| **0.70** | **0.80** | **0.4390** |
| 0.65 | 0.40 | 0 |

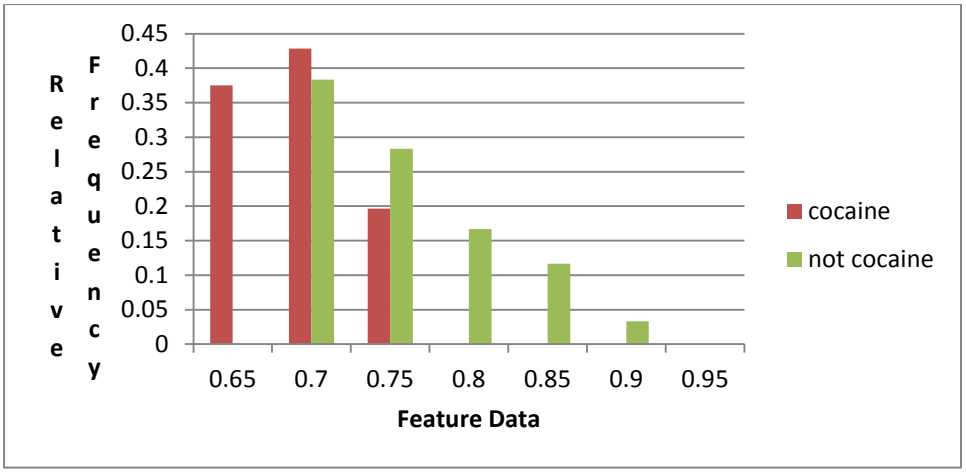



Figure C(xx) Column chart showing relative frequency of data for feature 10

With an AUC of 79.02% and high TPR value of 0.8 for the threshold of 0.7 chosen, this feature may at first sight seem like a good classifier. However, the FAR value of 0.44 is high but fits well relative to other possible TPR/FAR values. For the test data, the feature classifies 26.32% of the not-cocaine data as cocaine and 18.75% of the cocaine data as not-cocaine.

**Feature 11 (Average value of Region 5)**

AUC = 82.44%

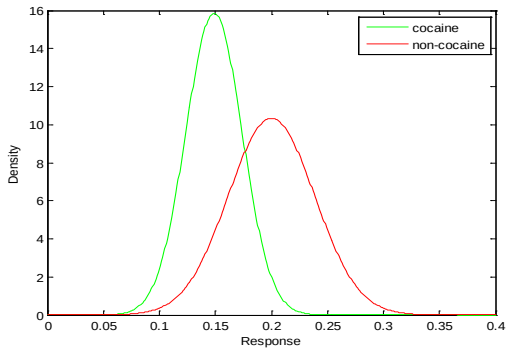



Figure C(xxi) Normal distribution curve for classification using feature 11

**Table C (xi) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.25 | 1.0 | 0.9268 |
| 0.20 | 1.0 | 0.6341 |
| **0.15** | **0.6** | **0.0976** |
| 0.1 | 0 | 0 |

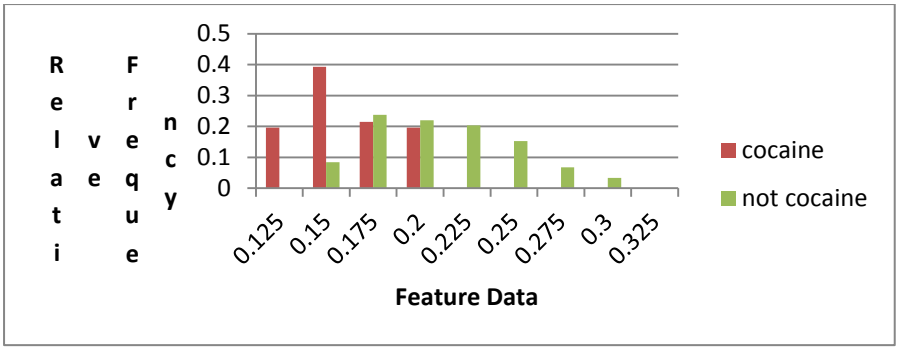



Figure C(xxii) Column chart showing relative frequency of data for feature 11

For the feature characteristics shown above, the values to the left of the threshold represent detection of cocaine. As the threshold moves to the left, the TPR reduces and the FAR also reduces. The 'best' threshold chosen representing a TPR of 0.6 and FAR of 0.0976 is thus 0.15. The area under the ROC curve is 82.44%.

The feature performs averagely when classifying the test cocaine data. It misclassifies 43.75% of the test cocaine data as not cocaine. It however improves on its classification when it correctly classifies approximately 95% of the not-cocaine data as not-cocaine.

## Feature 12 (Average value of Region 6)

AUC = 83.51%

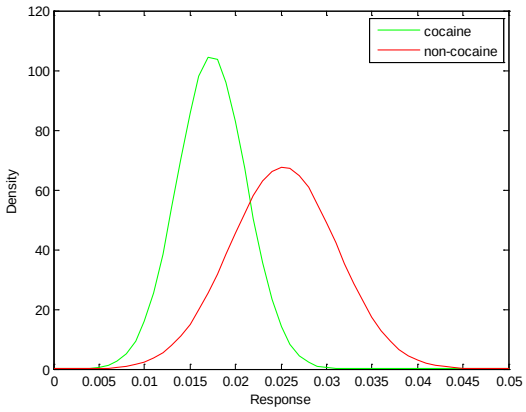



Figure C(xxiii) Normal distribution curve for classification using feature 12

**Table C (xii) TPR and FAR for varying threshold**

| Threshold | TPR | FAR |
|---|---|---|
| 0.0375 | 1 | 0.9756 |
| 0.0275 | 1 | 0.6829 |
| 0.0225 | 0.8250 | 0.4146 |
| **0.020** | **0.80** | **0.2927** |
| 0.0175 | 0.55 | 0.0976 |
| 0.015 | 0.3 | 0 |

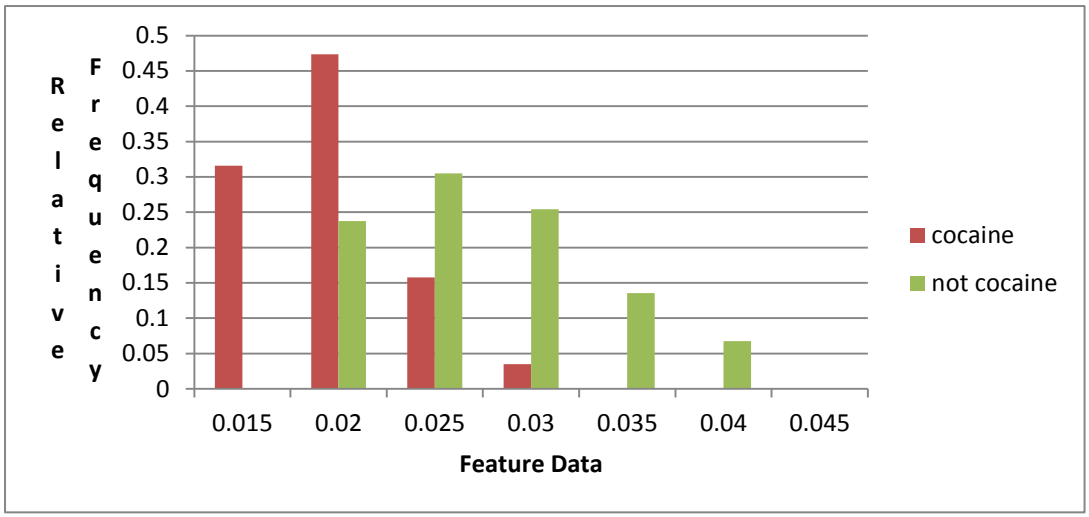



Figure C(xxiv) Column chart showing relative frequency of data for feature 12

The area under the ROC curve for feature 12 characteristics is 83.51%. The threshold chosen is 0.02 and it results in a TPR of 0.8 with a FAR value of 0.2927. The feature performs well with data not used during training. It correctly classifies 75% cocaine data as cocaine and 95% of not-cocaine data as not-cocaine.

# Appendix D

The tables below are the confusion matrix showing the performance of test data based on thresholds set using the training data for all 12 features.

**Table D(i) Confusion matrix for feature 1**

| a | b | Predicted as |
|---|---|---|
| 15 | 1 | a = cocaine |
| 3 | 20 | b = not cocaine |
| Correct classification = 90%% Incorrect classification = 10% | | |

**Table D(ii) Confusion matrix for feature 2**

| a | b | Predicted as |
|---|---|---|
| 15 | 2 | a = cocaine |
| 3 | 19 | b = not cocaine |
| Correct classification = 87% Incorrect classification = 13% | | |

The tables above show the confusion matrices for features one and two. As shown, feature one's performance on test data in terms of cocaine detection is same as that of feature two i.e. 83% true positive rate. However in terms of wrong classification of non-cocaine samples as cocaine, feature 1 performs better than feature two. Feature 1 wrongly classifies only 5%non-cocaine data as cocaine while feature 2 wrongly classifies approximately double that figure i.e.10%.

Below, feature 4 performs exactly the same way as feature two but feature three has a poorer performance in terms of cocaine classification. It wrongly classifies 7 cocaine samples as non-cocaine equating to about 39% even though its ability to detect non-cocaine samples matches that of feature one.

**Table D(i) Confusion matrix for feature 3**

| a | b | Predicted as |
|---|---|---|
| 11 | 1 | a = cocaine |
| 7 | 20 | b = not cocaine |
| Correct classification = 79% Incorrect classification = 21% | | |

**Table D(ii) Confusion matrix for feature 4**

| a | b | Predicted as |
|---|---|---|
| 15 | 2 | a = cocaine |
| 3 | 19 | b = not cocaine |
| Correct classification = 87% Incorrect classification = 13% | | |

So far, over all, feature 3 using the maximum negative slope has the worst performance in terms of cocaine detection. However, in terms of detecting for non-cocaine samples, the first four features have closely matching performances.

**Table D(i) Confusion matrix for feature 5**

| a | b | Predicted as |
|---|---|---|
| 15 | 2 | a = cocaine |
| 3 | 19 | b = not cocaine |
| Correct classification = 87%<br>Incorrect classification = 13% | | |

**Table D(ii) Confusion matrix for feature 6**

| a | b | Predicted as |
|---|---|---|
| 11 | 3 | a = cocaine |
| 7 | 18 | b = not cocaine |
| Correct classification = 74%<br>Incorrect classification = 26% | | |

Feature 5 above has the same performance with features 4 and 2 but a look at feature 6 shows an equally matching performance in terms of cocaine detection with feature 3 but a slightly poorer rate of classification of non-cocaine samples classifying only 86% non-cocaine data correctly.

**Table D(i) Confusion matrix for feature 7**

| a | b | Predicted as |
|---|---|---|
| 15 | 2 | a = cocaine |
| 3 | 19 | b = not cocaine |
| Correct classification = 87%<br>Incorrect classification = 13% | | |

**Table D(ii) Confusion matrix for feature 8**

| a | b | Predicted as |
|---|---|---|
| 15 | 2 | a = cocaine |
| 3 | 19 | b = not cocaine |
| Correct classification = 87%<br>Incorrect classification = 13% | | |

Features 7 and 8 have identical performance rates as features 2, 4 and 5 and feature 9 below performs equally with feature 3 in terms of cocaine detection but same as features 2, 4, 5 and 7 in terms non-cocaine classification.

**Table D(i) Confusion matrix for feature 9**

| a | b | Predicted as |
|---|---|---|
| 11 | 2 | a = cocaine |
| 7 | 19 | b = not cocaine |
| Correct classification = 77%<br>Incorrect classification = 23% | | |

**Table D(ii) Confusion matrix for feature 10**

| a | b | Predicted as |
|---|---|---|
| 15 | 5 | a = cocaine |
| 3 | 16 | b = not cocaine |
| Correct classification = 79%<br>Incorrect classification = 21% | | |

Feature 10 correctly classifies 83% of the test data as cocaine and 76% as non-cocaine having an overall correct classification rate of 79% and an incorrect classification rate of 21%.

**Table D(i) Confusion matrix for feature 11**

| a | b | Predicted as |
|---|---|---|
| 12 | 1 | a = cocaine |
| 6 | 20 | b = not cocaine |
| Correct classification = 82%<br>Incorrect classification = 18% | | |

**Table D(ii) Confusion matrix for feature 12**

| a | b | Predicted as |
|---|---|---|
| 15 | 2 | a = cocaine |
| 3 | 19 | b = not cocaine |
| Correct classification = 87%<br>Incorrect classification = 13% | | |

The performances of features 11 and 12 are shown above. While feature 12 aligns with previous performance rates for features 2,4 and 5, feature 11's 82% correct classification rate is third behind performances of features 2, 4, 5, 7 and 8 with 87% and feature 1 with 90% correct classification rate.