Neural networks for medical condition prediction: an investigation of neonatal respiratory disorder

Emma Annette Braithwaite

Thesis submitted for the degree of
Doctor of Philosophy
The University of Edinburgh
March 1998

Abstract

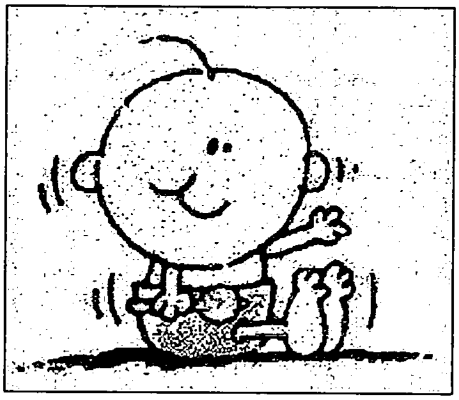
This thesis investigates how various signal processing techniques can be applied to diagnose problems in the medical domain. In particular it concentrates on breathing problems often experienced by premature babies who undergo artificial respiration. Medical Decision Support is an area of increasing research interest. It can underlie support for diagnosticians in everyday decision-making and can also alleviate habituation and fatigue - both of which can lead to errors, even in the most well-disciplined working environments. The neonatal intensive care unit (NICU) is a prime example.

Babies who are born extremely premature suffer from a number of conditions, in particular pulmonary (lung) function is not fully developed. These patients are placed on a variety of ventilators to assist them to breathe. Due to the extremely small size of the patient, problems which can occur to any ventilated patient are more common in neonates, for example blocked endotracheal tube and pneumothoraces. This thesis describes the investigation of techniques to be used as the core of a decision support device in Edinburgh's NICU. At present physiological signals are taken from the patient and archived, little diagnostic use is made of these signals and no investigation has taken place into their diagnostic relevance.

Within the scope of the work an investigation has taken place into the application area and some of its current problems have been identified. From these a physiological problem, respiratory disorder, was identified with characteristics which made it worthy of detailed study: it was extremely common, moreover expert knowledge and data about it already existed. With the current techniques the development of respiratory disorder is often missed or diagnosed too late. Signal processing techniques were evaluated with a view to applying them to predict the onset, or classify the development of, respiratory disorder, and a multi-layer perceptron network was chosen to perform as a classifier in the decision support tool. A number of tests were run which included an investigation of the efficiency of the chosen feature extraction techniques and the diagnostic relevance (with respect to the condition under investigation) of the signals being used to assist in diagnosis. Results show that at present the signals of greatest diagnostic relevance are not always used: a decision support device can be developed using a multi-layer perceptron classifier in combination with other signal processing techniques. The thesis also identifies other techniques where there is potential for improving the decision support tool's predictive and classification ability.

Declaration

I declare that this thesis has been completed by myself and that, except where indicated to the contrary, the research documented is entirely my own.



Acknowledgements

I could never have survived the past three and a bit years without support and help from huge numbers of people. Firstly my family who have jollied me along, told me to be patient and bought me the odd pint of beer (even though it isn't ladylike!), cheers. Talking of beer.... to everyone in the Friday night beer crew, Drew, Mu, Mike, Andy (x2), Robin, Vikki, thanks for reminding me that it was my turn to hand one of these things in. To all my friends both new and old who have put up with my anti-social tendencies over the past few months thanks for hanging in there.

On a slightly more serious note, within the Electrical Engineering department I would like to thank all staff and in particular my supervisors Alan Murray and Jimmy Dripps for their help and patience. At Edinburgh Royal Infirmary Andy Lyon, Neil McIntosh and Peter Badger cannot escape without a mention as they stopped me from forgetting what it was I was trying to do and let me pester them for three years with silly questions about babies. The EPSRC are acknowledged for supporting me financially and the IEE are thanked for supplying me with sweetie money for two of my three years.

To all in the Integrated Systems Group, Uncle Robin, Mike, Richard, Pete and all the other neural kiddies thank you for putting up with me through all my moods. Everyone who proof-read my ramblings or debugged my code, in particular Uncle Robin, Drew and Mike please congratulate yourselves on wading through it all. Finally Dave and Carole, thank you for not discussing work at home, I owe you! To anyone else who has helped and supported me and who I have been rude enough not to mention, two things; you know what my memory is like and errr cheers I couldn't have done it without you.

Contents

Glossary

Alveoli Air sacks in the lungs where gaseous exchange takes place

ANN Artificial Neural Network

ETT Endo-Tracheal Tube; pipe passed down throat administering air-mixture to lungs

FIO_2 Fraction of Inspired Oxygen; Measure of the proportion of the inspired air mixture which is oxygen

Foetus Unborn baby

Gestation period Time the foetus has developed in the womb

Habituation The process by which a person loses concentration when performing multiple similar tasks

HMM Hidden Markov Model

Homeostatic The control system by which the body maintains chemical functions

ICU Intensive Care Unit: Special Needs units for patients requiring more support than others.

Intravascular Within a vein

MAP Mean Airway Pressure

MLP Multi-Layer Perceptron; a type of ANN

Neonate A new-born baby

NICU A Neonatal Intensive Care Unit, supports babies which would have difficulty surviving on their own.

pCO_2 Partial Pressure of Carbon Dioxide; concentration of carbon dioxide in the blood

pO_2 Partial Pressure of Oxygen; concentration of Oxygen in the blood

Premature baby See Pre-term baby

Pre-term baby A baby which has been born at less than 37 weeks gestation

Pneumothorax A hole in the lung

Term Average gestation period of fully developed babies; usually 40 weeks

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Overview

Clinical diagnosis is an area in which no person (or machine) can be one hundred per cent accurate, one hundred percent of the time. For this reason new tools and new diagnostic techniques are being constantly developed. The aim of this thesis is to investigate the use of signal processing techniques for the development of diagnostic aids in the domain of premature baby monitoring. The intention is that the chosen signal processing techniques will capture underlying trends within the monitored signals and thus aid the clinician in the diagnosis and prediction of common physiological disorders.

1.2 Automatic medical diagnosis

The field of automatic medical diagnosis, as the title implies, encompasses the separate disciplines of medicine and automatic signal monitoring. This thesis must therefore address both clinical and signal monitoring issues. The aim is to combine the two approaches to produce a system which is acceptable to both clinicians and technologists.

Firstly the relationship between automated diagnosis and clinical techniques will be examined and then signal monitoring issues will be studied.

1.2.1 Clinical diagnosis

Currently the diagnosis of physiological conditions relies on the judgement and experience of clinicians, who in turn heavily rely on the use of monitored signals or symptoms. More efficient use of these monitored signals would enable diagnostic reasoning processes to be improved. Such improvement in the diagnostic ability of a clinician has obvious benefits:

- Increased level of patient care
- Prevention of the development of certain conditions
- Prevention of degradation of the condition of patients

One method of using the monitored signals more effectively is to use them as input to an automatic and autonomous monitoring system. With a clinician using such a system the benefits could be expanded to:

- Automated access to patient records to enable rapid diagnosis
- Reduced risk of habituation of staff
- Reduced number of false alarms in critical care environments
- Automatic control of life support equipment

There is however a further issue here. That is the perception of the patient of his/her treatment being controlled automatically and impersonally. Many people feel intimidated by computers and electronic devices and in the strange environment of a hospital or medical monitoring application the stress levels need to be kept to a minimum. The use of autonomous medical decision **makers** is therefore minimised. However, this is no hindrance to the development of diagnostic **aids** as they can provide a valuable service to the clinicians without decision making control being taken from them. Examples of these types of system range from assisting in the diagnosis of abnormal electrocardiograms [1,2,3] or cervical-smear tests, [4] to forming the basis of medical expert systems [5,6,7,8]. Patients are more prepared to accept this as their perception of the machines in these circumstances is that there is an expert using a machine, and ultimately decisions are being made by that expert. In some cases they believe that care has been improved by the device being present [9]. The ideal combination, therefore, seems to be a human specialist clinician assisted by a monitoring system which can advise him or her of any changes which have taken place while they have been visiting other patients. This combination carries the implication that the clinician must be able both to trust **and** understand the process by which the machine has reached its decision or diagnosis.

However, monitoring systems at their current performance are not capable of diagnosing many conditions. This is primarily due to inter-patient variability i.e. “normal” behaviour for one patient is often abnormal for others. This limitation is especially common in the aids which involve standard expert systems. A monitoring system must therefore be capable of taking this variability into account. It needs to adapt to the changing circumstances surrounding a particular patient. This can be as a result of changes due to clinical actions (for example medication) or to general improvements or deterioration in the overall condition of the patient. This can be broken down into two separate issues, speed of diagnosis and patient-specific diagnosis.

Speed of diagnosis

It is important to note that although the conditions being diagnosed may occur extremely rapidly they rarely occur instantaneously and there are often warning signs to presage them. For example, it is common knowledge that people may suffer from chest pains and tiredness a few minutes before certain types of heart attack [10]. This suggests that where such symptoms are present, an automatic monitoring system can be used to detect problems developing before the current diagnosis time.

Patient-specific diagnosis

Obviously this type of adaptable diagnostic system is a long way from being realisable except in its human form, in the brain of diagnosticians. However, it is felt that it may now be possible to develop a system which could be patient specific in much the same way as some speech recognition systems are speaker specific and which could adapt to the changing circumstances surrounding the condition of a patient. The aim is to design a system trained to recognise normal behaviour patterns for a particular patient and then to warn clinicians if the behaviour of a patient strays from these patterns.

However, this aim does not meet the full requirements of clinicians. Their ideal system is one which could be used on **any** patient regardless of their history. A trade-off must therefore be made between using a system which can recognise reliably the normal behaviour of a particular patient and one which can use a general model of various types of patient behaviour, i.e.. a generic system based on the behaviour of multiple patients. The decision of which system to use must be made once the performance characteristics, for example the false alarm rates, are known.

1.2.2 Signal Monitoring

Having discussed some of the clinical issues raised by automatic signal monitoring the thesis will now look at signal processing. Currently there are few signal processing methods used to perform the task of recognition of physiological signals in the medical domain. The majority of those that do exist have involved the investigation of the electroencephalogram (EEG) and the electrocardiogram (ECG) as both of these types of signals are easily obtained in sleep laboratories and other clinical monitoring areas. Techniques which have been used in conjunction with these signals and elsewhere in the medical domain include:

- Neural networks for sleep analysis [11], chest pain management in heart patients [12], biopsy analysis [4] and blood pressure analysis [13,7]
- Fuzzy sets for EEG analysis [14] and in alarm detection in monitoring systems [15,16]
- Prediction-based networks for ECG analysis in heart patients [17,18]
- Hidden Markov Models for information extraction in ECGs [19,20] and for the detection of cardiac arrythmia [21]
- Parametric methods for event detection in EEG signals [22]
- Expert Systems in machine monitoring [23,24]

Previous aids developed for this type of area have predominantly been machine driven [25,6,26, 27] in that they have been designed as machine controllers rather than patient monitors. Other systems which have been developed for this area have also used a single channel of diagnostic data and have attempted to identify irregularities within that [28,3,29,30].

Many of the techniques used in these applications were originally designed for use in other domains (see Figure 1–1) for example, Hidden Markov Models have predominantly been used to model phonemes in speech recognition [31]. In many cases it is possible to draw parallels between a signal or image derived from a non-medical application and one from a medical application. An example of this is the processing techniques used to determine the texture of wood or of synthetic aperture radar images which are now used on ultrasound images for the recognition of objects or areas of interest for example in breast biopsy and intra-vascular ultrasound images [32,33]. Diagnosis in the medical domain really means combining knowledge from a number of other domains and applying expert knowledge to achieve a decision.

It has been suggested by a number of people including Gorzalczany et al. [34] that neural networks may be used in combination with other signal processing techniques as medical expert systems. This approach presumes that a neural network can be trained to recognise the characteristics from a particular patient in much the same way as it can be trained to recognise the voice patterns of one speaker. The system can then "inform" the clinician of changes in the condition of a patient and decisions can be made based on this knowledge. This type of

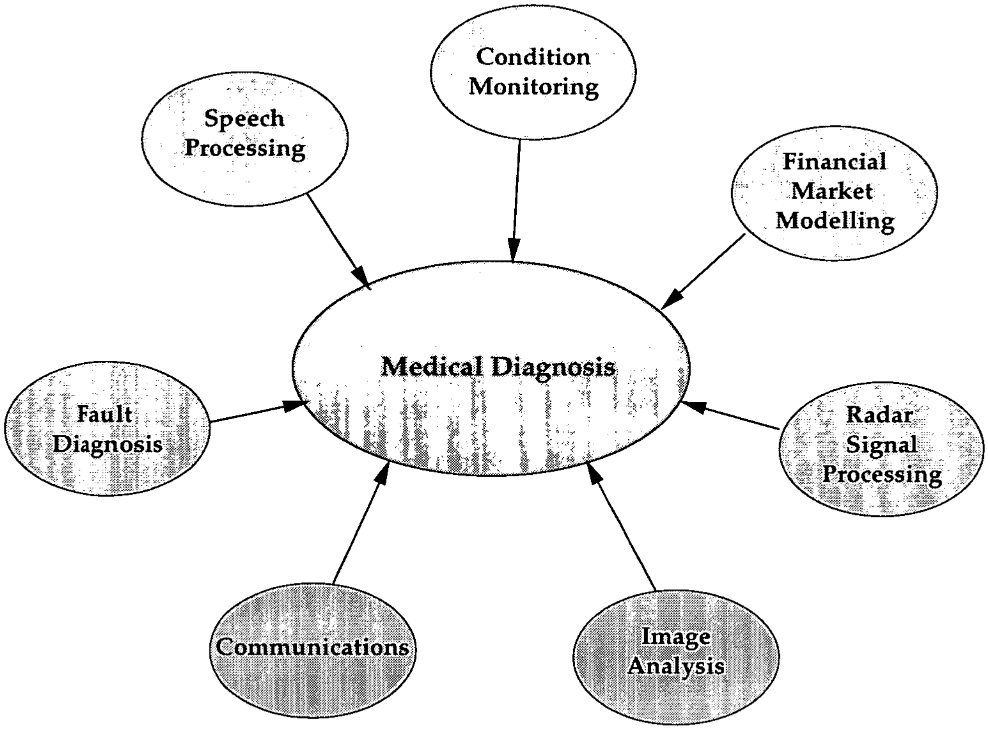


Figure 1–1: Influential domains for the sphere of Medical Diagnosis

system could be used to assist in the diagnosis of a number of conditions including heart monitoring, brain scan analysis, x-ray analysis and others [35]. Until recently, diagnostic systems were almost exclusively used where the recognition task involved numerous repetitions of a similar act such that the diagnosticians could become habituated. For example, cervical smear test evaluation [36]. Systems are needed which can reduce the risk of habituation *and* assist diagnosticians in circumstances where they might otherwise be working on a second by second basis. One of these areas of application is that of critical care monitoring.

Critical or Intensive Care Units attempt to sustain the lives of the patients within them by using a combination of support technology, for example ventilators and incubators or heated beds, monitoring equipment, and trained diagnosticians and care staff. Currently diagnoses are made almost entirely on a second by second basis, although this is largely due to the application area rather than choice. Clinicians would like a system that could monitor a patient continuously and, over time, assist in making diagnostic decisions by suggesting possible reasons for the current behaviour of the patient.

1.3 Aims of this study

Therefore the motivation behind the work described here was to determine whether it is possible to assist in the diagnosis of certain conditions by using an automatic recognition and monitoring system. In particular the work was to involve the recognition of conditions commonly found within a neonatal (new-born baby) intensive care unit (NICU) as Edinburgh has over 2000 records which have been collected from its NICU and a study can be made using these. These aims can be explicitly stated as;

- Can a diagnostic aid be produced for an NICU?
- Can an Artificial Neural Network be used as the heart of such a diagnostic process?
- Can the diagnostic aid have predictive ability?
- Is expert knowledge of particular relevance in this application area?

Certain constraints were, however, placed on the development of the system and these were;

- The system must not place further stress on clinicians or carers.
- The diagnostic processes of the system must use clinical knowledge to increase the acceptability of the system to all users.
- The system must be capable of enhancing the level of patient care.
- The system must not use an increased number of physiological signals than are already measured as part of the current monitoring process.

Given these constraints the aim of the work was therefore to develop a system which could assist in the automatic recognition of common clinical problems using non-invasive techniques and currently monitored and recorded physiological signals and records.

1.4 Impact Areas

The work described here has an impact on a number of important research areas. Although it investigates the possibility of using a neural network to diagnose the onset of certain clinical conditions in new-born babies, it is not only applicable in this extremely narrow domain. Staying within the domain of critical care monitoring it is obvious that any system which has been developed for the new-born baby unit must also be transferable to its "big brother" paediatric and adult special care units.

This is not however, where the influence of the work ends. Many of the techniques which have been applied here have used knowledge drawn from a number of other domains. This "knowledge transfer" may also occur in other directions, for example, in other fault diagnosis, or condition-monitoring problems where time-series are currently used to assist in the monitoring of the state of the system. The work will be of particular interest in application areas where faults are known to develop over a period of time and these may include the diagnosis of conditions ranging from the detection of faulty heart valves to component wear in electrical plant.

1.5 Thesis Plan

This thesis will investigate the possibility of developing a diagnostic aid for use in a critical care environment. The work described has investigated the possibility of using a number of different medical signals concurrently to diagnose changes in the condition of a particular patient by incorporating expert diagnostic knowledge into the design of the system.

Chapters two and three will describe the particular application area on which the work concentrates. They will outline both the clinical and non-clinical problems involved in the application area and the specific problems investigated here.

Chapter four follows these themes and explicitly describes the particular problem of interest. The following chapter describes the methodology behind the development of the system and the expert medical knowledge used to drive this design. The premise is maintained that this type of diagnostic aid must be patient-driven, and adaptable, and incorporate expert medical knowledge. In chapter 6 the prototype system is tested and some results are presented. It discusses using a variety of signals to diagnose particular physiological conditions and compares the results of these. Finally the last chapter discusses the implications of the work described and makes suggestions for further work in this field.

1.6 Summary

To summarise, medical clinical diagnosis is an extremely complex area which is constantly developing. This account will concentrate on a small area of this domain, the prediction of physiological conditions in a critical care unit, in this case a new born baby unit. Standard physiological signals and a neural network approach will be used. It will also investigates the effect that using particular combinations of signals has on the predictive capability of the developed system.

Chapter 2

Neonatal Intensive Care monitoring: A clinical view

2.1 Overview

As this thesis is based on monitoring of physiological signals from a neonatal intensive care unit (NICU) this chapter will describe the operation of such a unit from the perspective of an Electrical Engineer. It discusses the need for the NICU and how the treatment, and patients, within them vary from adult intensive care units. Common NICU problems, and the methods used to diagnose them, are described and the difficulties faced in diagnosing them are discussed.

2.2 The Neonatal Intensive Care Unit

Neonatal intensive care units are a unique type of intensive care unit with many similarities to adult intensive care units (ICUs). They differ not only in the size of the patients which they treat but also in the problems which they face. Adult and paediatric ICUs tend to treat patients who have suffered from some form of accident or who are undergoing post-operative monitoring whereas NICUs also treat patients who have developmental or physiological problems. All ICUs are designed to sustain the lives of the patients being treated in them, this maintenance can take a number of forms from artificial respiration to feeding and observation.

Neonate is the term given to a new-born baby regardless of the gestation period (the time spent in the womb). The normal foetus takes approximately thirty-eight weeks to develop fully, after this time the foetus can be born and is capable of breathing air and of surviving without artificial aid. In 1935 the World Health Organisation defined a premature baby as one born after less than thirty seven weeks gestation [37]. There was however a problem with this definition as it is extremely difficult to accurately gauge the gestation period of a particular neonate. Therefore another descriptor was introduced; if the neonate weighed less than 2500g (5lb 8oz) at birth it would be described as premature in terms of its development, this descriptor was adopted in Britain in 1938 [37].

In the first twenty-eight weeks *in-utero* (in the womb) the foetus' basic body structure forms, i.e at twenty-eight weeks the unborn baby is extremely small, weighing approximately 1000g, but nonetheless is recognisable as a baby. In the weeks following there is rapid tissue growth and the organs of the foetus continue developing. In 1960 [37] it was assumed that a baby which was born after twenty-eight weeks gestation was capable of survival if given environmental support. Any gestation period less than this required high levels of support which were very rarely successful.

The first neonatal special care unit was formed in Bristol (UK) in 1938 to support the lives of extremely immature neonates. It used a combination of oxygen therapy and warm beds, which to the modern-day neonatologist would look archaic. However, many of the techniques which were pioneered in these early units are still in use in modified form today.

The NICU aims to support babies which have been born extremely premature or cannot survive without assistance e.g those with a congenital heart defect awaiting treatment. The majority of the babies a neonatal ICU treats are of the former type. It is now possible to bring neonates born after as little as 23 weeks, see Table 2–1, gestation and weighing about 500 grams (the average is approximately 3500 grams) to term [9,38], i.e. to full development to allow them to leave the unit. Term is defined as being the average gestation period of approximately forty weeks.

Completed weeks of gestation at birth	**Survival**
21 weeks and less	0%
22 weeks	<<1%
23 weeks	5-25%
24 weeks	40-60%
25 weeks	50-80%
26 weeks	80-90%
27 weeks	> 90%
30 weeks	> 95%
34 weeks	> 98%

Table 2–1. Estimates of survival for live born infants in NICUs in the 1990s (source: University of Wisconsin Medical School)

As with any clinical situation each patient being treated is unique. Unlike their "big brother"

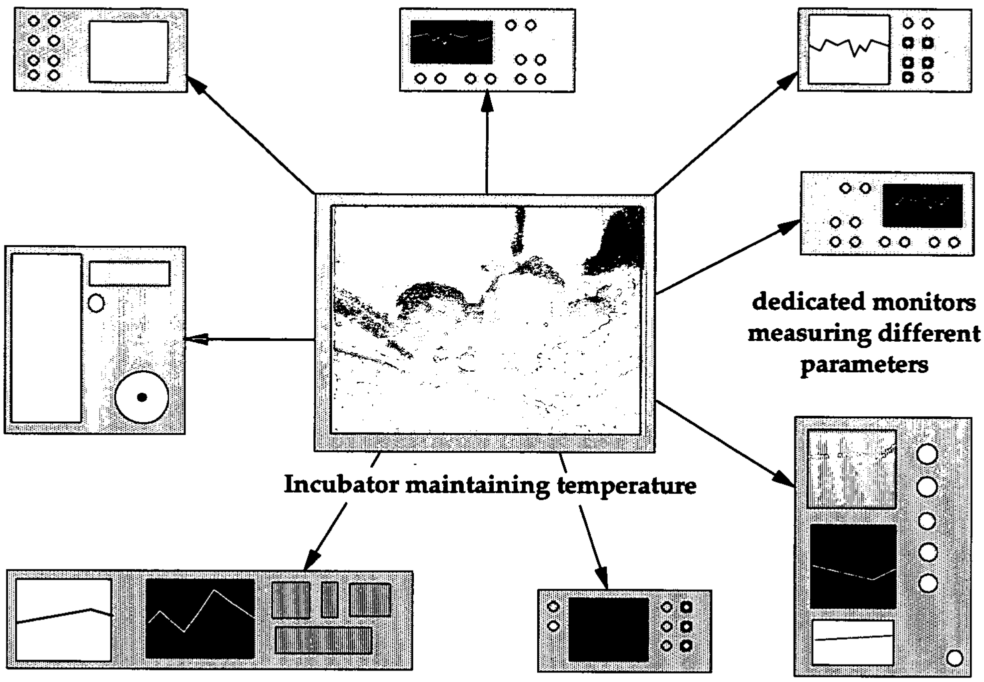


Figure 2–1: A schematic of the typical structure of an NICU bed

ICUs neonatal ICUs often have to deal with a combination of problems that would never be seen in older patients. This adds to difficulty of diagnosing and treating the patients within the unit as staff must take into account all the possible repercussions of a particular therapy regime.

2.2.1 NICU structure

The typical structure of an NICU is very simple. It consists of a number of incubators, of different types, each of which is linked to a cot-side monitor. The incubators maintain temperature conditions similar to those found inside the mother's womb and this temperature can be decreased as the condition of the patient improves. There are two main types of incubator; the open and the closed. The open type is used for older patients. The closed is used for patients who require greater homeostatic support, for example those requiring specific humidity levels. The homeostatic system is the control system by which the body maintains chemical functions. The incubators are used as beds for the patient who is then connected to a number of measuring devices which collect information about the current condition of the patient (see Figure 2–1).

Each cot is monitored continuously both by the dedicated monitors and by the trained medical staff within the unit. These staff have been trained to deal with the type of problems which can occur with the development of newborn babies, both pre-term and term deliveries.

2.3 Role of monitoring in Intensive Care Units

Many people are familiar with the ICU, as it is portrayed on television or film, as being a place where many clinical problems occur simultaneously and everyone is in an extremely serious condition. The latter is obviously true otherwise the patient would not be treated in this type of unit, however, the former stereotype of the ICU is a little different in reality.

Patients in an ICU spend the majority of their time in a stable condition, hence the term "serious but stable condition". These periods of stability are a fact of working life in an ICU and can lead to decreased levels of awareness, by the staff, of changes in condition of the patient.

The nature of an ICU leads to it being a highly stressful environment for both staff and visitors. All staff within an ICU are required to be alert at all times as the condition of a patient is capable of changing rapidly. There are also numerous monitors measuring different parameters for each patient. These are linked to threshold alarms which sound when a particular measurement strays out with pre-defined bounds. These threshold alarms are tailored to each patient but they do not take into account the current condition of the patient. This means that during certain periods these alarms can be triggered frequently without there being any evidence of a physiological problem occurring with the patient. This frequent sounding of alarms leads to habituation of both the experienced and inexperienced staff of the unit [39]. In Cropp's article [39] he discusses the ability of various ICU clinicians to differentiate between the different types of alarms within an intensive care unit. He proves that the less experience a clinician has in an ICU the less capable he/she is in differentiating between these alarms. This may seem obvious but Cropp also states that the frequency of alarms within a unit leads to increased stress levels and habituation as all alarms begin to sound the same or are ignored. This augments the stressful environment for patients [40], staff [41] and especially the visitors [42] who tend to believe that any alarm is signalling a problem. There have also been studies [43] which show that adult patients who have been released from ICUs, although they were under heavy sedation whilst patients there, complain of the noise. The assumption can be made that neonates are no less susceptible to this trauma therefore any reduction of false alarms would improve patient care and the working environment of the NICU.

Another problem faced within an ICU is that of "late" diagnosis of certain conditions. By "late" it is meant that the condition is diagnosed when it has reached a critical stage rather than before i.e. too late to prevent the condition from developing. The very nature of an ICU leads to this phenomenon as clinicians apply themselves to treating the ongoing condition of their patients. Most conditions develop over a longer period of time than the few minutes which the media portrays. In some cases the condition can develop over a number of hours or even days. This is especially true in the NICU as many conditions occurring are the ongoing deterioration of an extremely fragile body which is already being pushed to its limits.

Neonatal ICUs are unique in that the patients which they treat cannot communicate their condition, symptoms or feelings to the consulting physician. These clinicians rely on experience, the physiological signals of the patient which are measured and the appearance of the patient. The latter is in itself a problem as the patient is often shrouded in bubble-wrap to keep him or her warm. This renders the patient almost invisible. For these reasons the job of the neonatal ICU staff is even more stressful and difficult than that of the adult ICU or special care unit [9].

To summarise there is a need for better monitoring systems in the NICU to combat a number of problems. These include increasing the trust both staff and visitors can place in the system, reducing the noise levels in the unit and to diagnose certain problems earlier than is being done currently to avoid stressful and more expensive treatment.

2.4 Current monitoring system

The current monitoring system in place in the Simpson Memorial Maternity Pavilion (SMMP) at Edinburgh and in at least four [9] NICUs around Britain is called "Mary" and it is similar to monitoring systems used in many other intensive care units. It consists of a number of dedicated monitoring devices linked to a central display system. It is the display system in this instance which is called "Mary", see Figure 2–2. Each patient can be linked to a number of these monitors at any one time and they measure different physiological signals. These signals include heart rate, respiratory rate, blood pressure, gas concentrations and temperature.

Other measurements are also taken and logged depending on the treatment the patient is undergoing, for example, airway pressure is measured if the patient is on a ventilator. However, these signals are not logged by "Mary" in the same way as the ones mentioned previously. The standard signals which "Mary" accepts are stored on a second by second basis on the hard disk attached to the "Mary" personal computer. This allows the data from a patient to build up and for a historical log of the activity of the patient to be generated. This second by second log is kept for three days, after which time, due to the storage requirements needed, the data is archived by generating minute averages of the data and storing them in binary form. These archived records can be accessed at any time to enable inter-patient comparison and teaching aids to be produced. The non-standard measurements taken are also stored, however they are

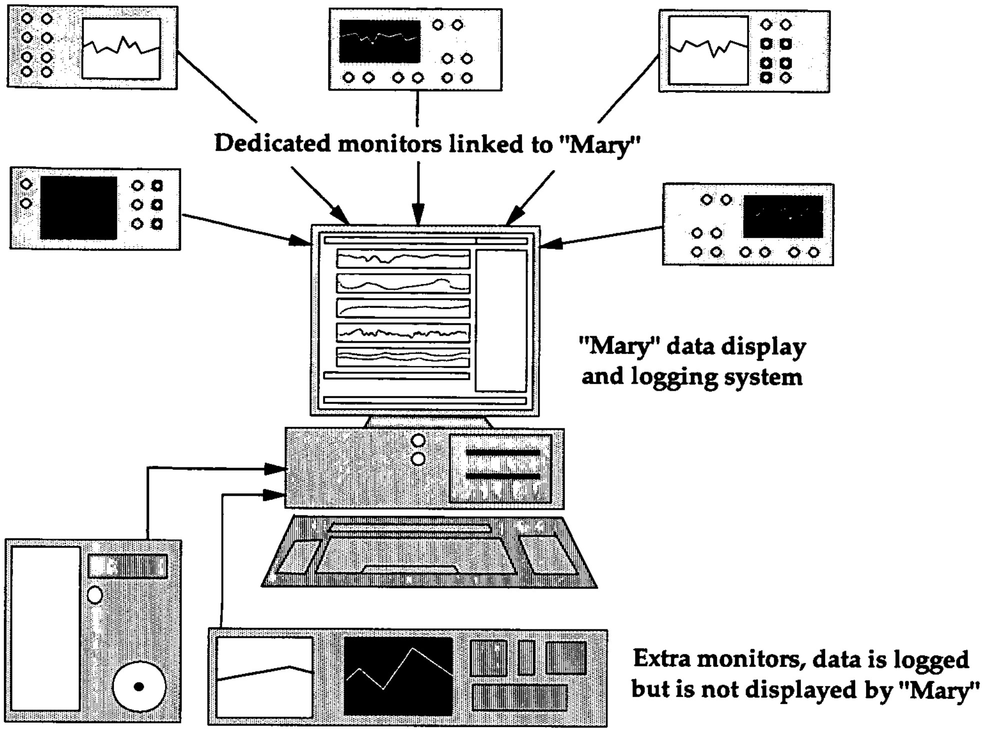


Figure 2–2: A schematic of the single cot "Mary" data logging system

stored in spreadsheet format and are accessible only by using the patient number and the date of interest. They are also not logged at the same time intervals as the "Mary" data. Therefore the combined use of the two types of data is difficult.

In addition to "Mary's" ability to store the raw physiological data in binary form it can also store the action or treatment that was used on the patient at a particular time. This would normally be seen in the right hand box shown in Figure 2–3. This box is often full of the patient record of what happened at a particular time, for example, "Morphine commenced 10mics/kg" (see Figure 2–4 for an example of a day's comments). The treatment is stored and the time noted when it is performed. These comment files are cross referenced to the data files and are also stored in binary format to compact them. It should be noted that the accuracy of these comment files is entirely dependent on the staff who enter them. It is possible to enter comments after treatment with the time at which treatment occurred, this time is determined by estimation on the part of the clinician or nurse. Some treatments or actions are never entered as the clinician is too busy or other situations arise so many uncertainties can creep in.

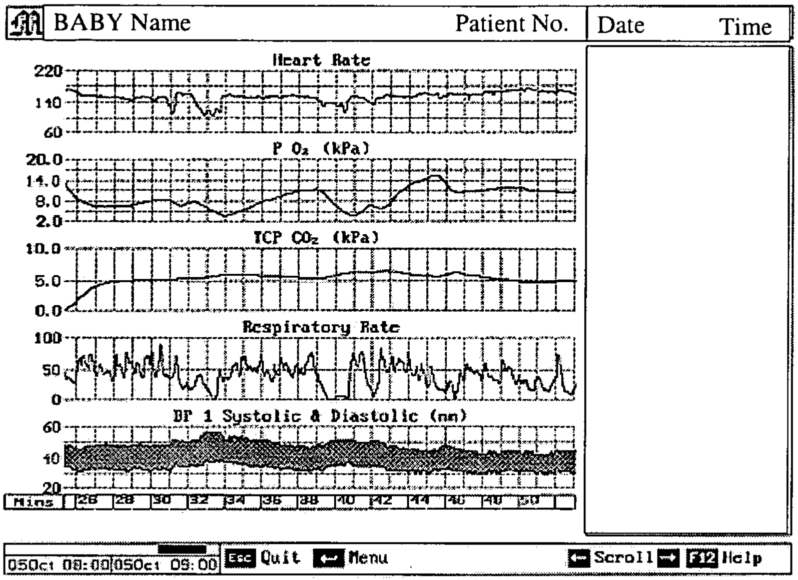


Figure 2–3: A typical "Mary" display screen

Patient a

160	Blood for gases
164	Bolus N-G feeds
203	Physiotherapy
207	ENDOTRACHEAL SUCTION
280	Change PO2 probe
400	REINTUBATED
496	All Care
557	X RAY, Change PO2 probe
604	Bolus N-G feeds
669	All Care
720	Bolus N-G feeds
787	Bolus N-G feeds
874	Bolus N-G feeds
951	Heel stab
1354	Change PO2 probe
1396	ENDOTRACHEAL SUCTION, Physiotherapy

Figure 2–4: An example of a typical day's diagnostic comments

2.5 Premature birth

Premature babies have not had the full amount of time to develop *in utero* (in the mother's womb). This means that there are certain conditions which are cared for within the NICU which are exacerbated by the problems of premature birth. The aim of the ICU is to attempt to bring these babies "to term" i.e. full gestation period in as non-invasive and non-stressful a manner as possible. Most babies which are born slightly premature are capable of survival on their own, however, babies which are born extremely premature require assistance to live and the NICU provides this. It aims to mimic, as far as it can, the conditions which the baby would experience *in utero* whilst treating the problems that occur due to the baby trying to survive in the hostile outside world. Currently a baby which is born more than 3 to 4 months (12 - 16 weeks) premature does not survive as its body can not cope with the outside world. These limits are always being stretched as new technology and treatment techniques come into common usage (see Table 2–2 [38,44]). Table 2–2 can be compared to percentage death rates in England and Wales in 1957 (see Table 2–3) [37]. It should be noted that the limits of this table extend to an approximate gestation age of 30weeks (3lb corresponds to approximately 29 weeks [37]), neonates born with a lower gestation age are collected together in the final category as at the time this table was published it was believed that neonates born at less than 28 week were non-viable (i.e. they could not survive).

Estimated gestation (weeks)	**Approx weight**	**1983**	**1985**	**1987**
Under 26	<1lb 8oz	79.4	102.9	236.8
26-27	2lb	407	528.9	606.1
28-31	<3lb 8oz	847	866.7	859.5
32-36	>4lb	974.7	981	983.7
37-41	>5lb 8oz	998.2	998.2	998.7

Table 2–2. Neonatal survival rates per 1000 live births (source: Neonatal intensive care in Scotland)

2.5.1 Problems of Premature Birth

Babies which are born pre-term suffer from a number of problems related to their incomplete development. These problems can be separated into two broad categories: those of instability of homeostatic control and immaturity of certain organs [45].

Approx gestation age	Weight	Distribution (%)	Deaths (%)	Survivors at 28 days (%)
<30 weeks	<3lb 4oz	12	45	33
>32 weeks	<4lb 6oz	18	9	82
<36 weeks	<4lb 15oz	20	3	93
37 weeks	<5lb 8oz	50	1.5	96
all cases	all cases	100	8	85

Table 2–3. Percentage deaths within 24hrs by weight groups of premature live births in England and Wales in 1957

2.5.2 Homeostatic control

The first type of problem which many babies who are born premature face is that they are incapable of maintaining the different control systems of the human body. Whilst the foetus is *in-utero* it is protected and cushioned from the harsh elements of the outside world. It is fed and supplied with its oxygen requirements through its umbilical cord. Its temperature is maintained by the control system of the mother and all the foetus' energy is concentrated on developing fully in this protected environment. If premature birth occurs, the systems which were previously cushioned by the mother's control systems have to be utilised by the neonate's. These are amongst the last systems to develop in the womb and therefore cause problems for the new-born baby. It must now attempt to regulate its own temperature, its glucose and calcium levels, and its digestive system. The ICU aims to help the neonate in its task and one of the methods of achieving this is to place the neonate in an incubator and to try to mimic the temperature and conditions of the mother's womb.

Despite the NICU's efforts the mortality rate of many of the extremely premature infants is high and a large contributor to this statistic is the fact that many of the organ systems of the neonate are immature and cannot cope in the harsh world outside the womb.

2.5.3 Underdeveloped organs

The organs of babies who are born extremely premature have not had the time to develop fully. This can lead to complications in their treatment after birth and often affects their development. Common problems of immaturity of certain organs or their related systems include [45]:

- immaturity of the liver: often leads to bleeding tendencies as the liver produces blood clotting coagulants.

- immaturity of the kidneys: the body is incapable of processing certain compounds and this can lead to problems.
- inadequate gastrointestinal function: this means that the patients cannot process and digest their food properly.
- immature bone marrow: too few red blood cells are formed and the patient has a higher risk of anaemia.
- respiratory system: this includes both immature lungs and breathing mechanisms.

All of these problems act in combination to provide an extremely complex task for the clinicians trying to treat the neonates. This thesis tackles the diagnosis and prevention of certain types of respiratory problems.

2.6 Diagnosis of common problems

"Mary" is currently used as a display and data-logging system in that it is capable of displaying a maximum of five graphs at any time on its screen. This means that a maximum of seven signals can be displayed at once. Multiple blood pressure traces can be displayed on a single graph (see Figure 2–3 where systolic and diastolic blood pressure is displayed). The selected signals can also be changed to allow the clinicians to view other parameters. This also leads to the problem that the correct signal for spotting a particular problem may not be displayed at a certain time.

Diagnostic processes therefore include all the available data and the visible condition of the patient. For example the patient may be turning blue and the Carbon Dioxide levels in the blood increasing. This may suggest that the patient is becoming cyanotic and that he/she require supplementary oxygen to help breathing. Clinicians frequently have to interpolate between the available data and what they think is happening.

2.6.1 Which measurements?

When a patient is admitted to the NICU she/he is immediately linked up to the cot-side "Mary" PC. The minimum number of probes (transducers) are attached to enable the clinicians to make diagnostic judgements without placing the patient under any undue stress. This in itself can lead to problems as in some cases parameters which would be used to diagnose certain conditions are not measured as they would involved the use of another probe and further damage to the delicate skin of the patient. Some probes can cause so much damage that the probe must be changed every couple of hours, for example those measuring partial pressures of Oxygen and Carbon Dioxide. Obviously which measurements are being taken affects the diagnostic processes that the clinicians use to determine the problem and their treatment of it.

2.7 Respiratory Disorder

Respiratory Distress Syndrome (RDS) or Respiratory Disorder (RD) can affect both adults and children alike. It is one of the most commonly occurring problems within an ICU and there are similarities between the treatment of adults and children. There are some differences in the triggers of RDS between the two categories of patients. In both adults and children triggers for RDS can include both viral and artificial stimuli [46], for example viral pneumonia or inhalation of toxic gases. Neonates are also adversely affected by the immaturity of their respiratory system [47]. The diagnosis of RDS is also more difficult in neonates as the symptoms are not always recognised and the patient cannot communicate them to the attending physician.

Until a foetus is approximately 30 weeks gestation [44] its lungs do not function properly. In these neonates the alveoli (air sacks of the lungs where gaseous exchange takes place) of the lungs are deflated and crumpled and the muscles surrounding the lungs are not strong enough to support the expansion and contraction needed for complete respiration. This means that the lungs can easily collapse. Also they do not produce the quantities of surfactant (lubricant) required by a full-term baby to allow it to breathe normally. RD is one of the most common causes of death in extremely premature infants as the respiratory system is one of the last systems to develop [37].

By careful monitoring of the gas levels within the blood of the patient it is possible to diagnose RDS in many cases. Unfortunately this diagnosis often takes place when the condition has become serious. Clinicians feel that it would be useful if there was a method of diagnosing RDS in neonates before the condition has advanced too far [1].

It should be noted that certain forms of RDS can be induced by the treatment that patients

[1]Personal Communication with Prof N McIntosh and Dr A Lyon at the NICU Edinburgh Royal Infirmary.

are given, in particular by the oxygen therapy or ventilation that many undergo. It is these types of RDS which the work described in this thesis is designed to prevent.

2.7.1 Oxygen Therapy

If a patient is incapable of breathing for themselves they are often given support in the form of an artificial respirator or ventilator. These mechanical devices supply oxygen directly to the patient by delivering a measured amount of oxygen in an air mixture. There are two types, positive pressure, and both positive and negative pressure [48] . The former operate in a single direction by forcing an air mixture into the lungs of the patient. On the "breathe-out" phase of respiration the lungs of the patient are allowed to naturally fall back into position, thereby pushing the waste products of respiration back into the atmosphere. The latter type of ventilator is the more commonly used in the NICU. It works by controlling both the inhalation and exhalation phases of the breathing cycle and supporting the neonate.

The air mixture that is supplied to the patient is applied to the lungs by use of a tube passed down the throat (trachea) of the patient (see Figure 2–5). This tube is called an Endotracheal tube (ET tube). It can be used either orally (through the mouth) or nasally (through the nose), and, as it can be imagined, the insertion of the ET tube is an extremely stressful experience for patients; adults and children alike. For this reason most patients undergoing artificial respiration are heavily sedated and this in itself can lead to diagnostic problems as some of the patients natural reaction mechanisms are damped.

There are two main types of bidirectional ventilator in use, those applying constant pressure and those which apply pulsed breaths, see Figure 2–6. The former preserves the integrity of extremely fragile lungs by maintaining a constant pressure in the lung cavity, the latter uses pulsed breaths and therefore places a much greater stress on the patients lungs and rib-cage as they are forced to continually expand and contract. The use of this pulsed type can often increase the risk of certain conditions occurring and can exacerbate the respiratory disorder which may already be present [49]. It may seem obvious that all patients should be placed on constant pressure respirators. However, there are other factors involved in this choice. If the positive pressure cycle of the inspired air mixture in the pulsed type of ventilator is allowed to exceed certain levels the fabric of the lungs can tear. Therefore the decision is made that those children which are seen to be strong enough with sufficiently developed lungs are placed on the pulsed type of ventilators and only the weaker patients use the constant pressure type.

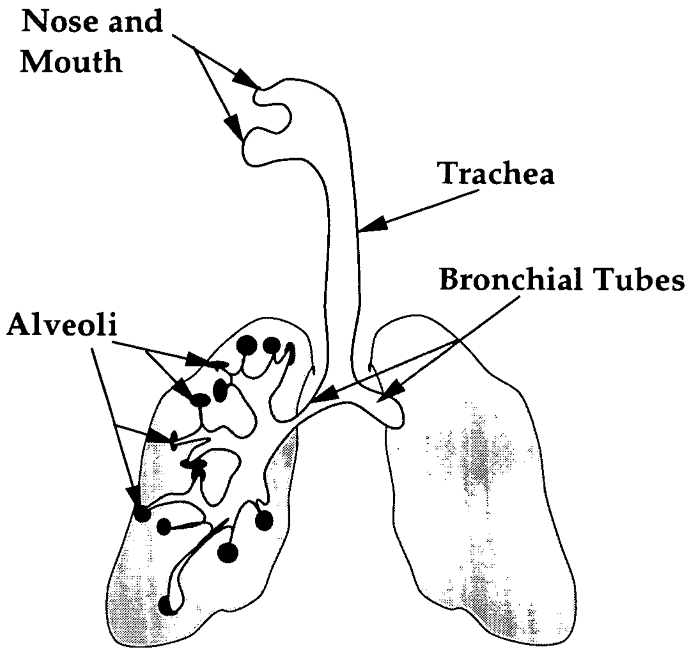


Figure 2–5: Basic anatomy of the respiratory system

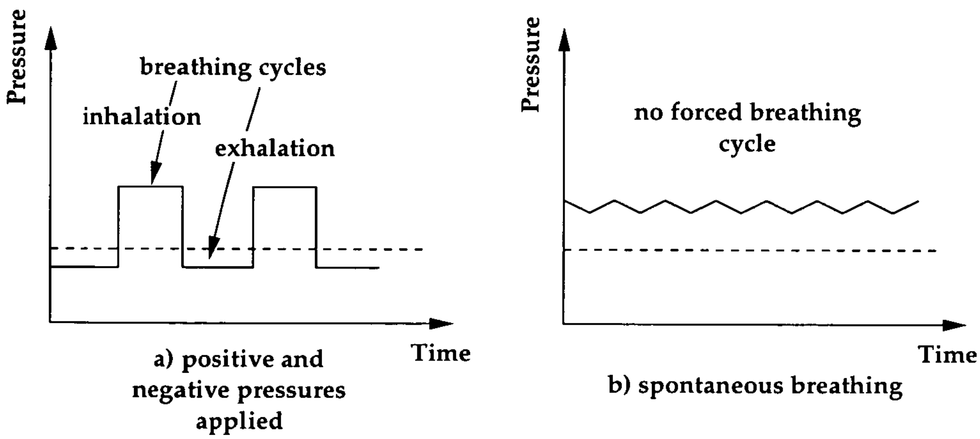


Figure 2–6: Graphs showing typical breathing patterns from bidirectional ventilators

2.7.2 Ventilator induced problems

There are some medical conditions which are associated with the use of ventilators used to maintain the respiratory function of a patient. Some of these conditions occur to adult and neonate alike whereas others are neonate specific, for example retinopathy of prematurity.

Retinopathy of prematurity

When oxygen therapy was first used in the care of low-birth-weight extremely premature infants (circa 1940 [37]) it was not known that high levels of oxygen in the air mixture being applied to these infants could cause blindness once the patient has been removed from the respirator. This condition occurs in neonates of less than 44 weeks gestation (i.e. it can occur to full and pre term babies alike when they are placed on ventilators at this early age) [47]. It is caused by an ingrowth of blood vessels into the vitreous humour of the eye after the ventilation treatment has been withdrawn. This ingrowth forms a mat of fibres behind the lens of the eye and it can ultimately lead to retinal detachment and blindness. The condition can be prevented by not allowing the percentage of inspired oxygen in the air mixture to rise above 40% (the percentage of inspired oxygen in air is normally 21%). Nowadays this problem is becoming less and less common. In Scotland in 1984 the proportion of patients suffering from retinopathy of prematurity was 1.8% [38], this is in contrast to proportions approaching 20% [37] in the United States in the 1960s.

However, there are still ventilator induced problems which cannot be prevented as easily as blindness.

Pneumothorax

A pneumothorax is a condition where the alveoli in the lungs , see Figure 2–5, have burst and there is an escape of air into the pleural (lung) cavity. This obviously reduces the lung capacity and can have serious affects on the gas exchange which takes place in the lungs.

This type of condition can be both spontaneous and induced, for example by ventilators. If the pressure of the air mixture being applied to the lungs is too great the alveoli burst and the pneumothorax occurs. Obviously this type of pneumothorax is preventable if it is realised that the positive pressure being applied to the patient is too great. However they are not often prevented as they tend to develop over long periods and diagnosis is made on a second by second basis. The precursors for a pneumothorax are similar to those of another type of ventilator induced disorder, that of the blocked endotracheal tube.

Blocked Endotracheal Tube

The tube which is passed down the throat of the neonate (or adult patient) must be "suctioned" regularly to prevent a build-up of secretions. These are a natural by-product of the breathing process. If this build-up is allowed to continue for too long the patient is starved of fresh Oxygen and cyanosis (Carbon Dioxide build-up in the blood) occurs. This condition is prevented by frequent suctionings of the ET tube to prevent full closure of the tube. However, this is not always completely successful and a blockage can occur. The only method of treatment at present is to re-intubate. This means that the old ET tube is removed and a new one is inserted. This can damage the throat. The condition is preventable in that there are, if spotted, precursors to its occurrence. There is usually an increase in the levels of Carbon Dioxide in the blood. Similarly to the development of the pneumothorax the blocking tube is often not diagnosed until it is complete. There is a perceived need to determine the partial blocking of a tube as it continues to accumulate mucus. This would prevent the patient undergoing the extremely stressful procedure of extubation and reintubation and would improve patient care and prognosis.

2.8 Diagnosis of Respiratory Disorder

The diagnostic processes used to determine the presence of RDS are currently very crude. Patients who are experiencing difficulties breathing use ventilators to assist them in this. To enable them to maintain the correct levels of various gas concentrations in their body clinicians can alter a number of parameters associated with the ventilator. These include

- Mean Airway Pressure
- Humidity
- Fraction of Oxygen

When clinicians are monitoring the condition of a patient they use a number of physiological parameters as a standard measurement. These include;

- partial pressure of Carbon Dioxide
- partial pressure of Oxygen
- Respiratory Rate

- Heart Rate

However, clinicians do not rely on all of these to determine if a patient is suffering respiratory difficulties instead they principally rely on the partial pressure measurements of Carbon Dioxide (pCO_2) and Oxygen (pO_2) levels. These give an indication of the efficiency of the gaseous exchange taking place within the lungs and how much Oxygen is reaching them. Typically the indicators for respiratory difficulties are that the Carbon Dioxide levels have risen significantly and the Oxygen levels have decreased. This can indicate either poor gas exchange due to a pneumothorax or a blocked endotracheal tube (ETT).

It should be noted that in many circumstances clinicians do not detect the changes in these two parameters and diagnosis is only made when the patient starts showing clear signs of distress. This is an area in which improvement could be made if a system were in place which would detect these changes and warn clinicians of possible problems.

Although clinicians rely heavily on the use of pO_2 and pCO_2 as diagnostic indicators for RD it is often not possible to place too much stress on the oxygen measurement as this can be manipulated. Carers can alter the levels of Oxygen in the air mixture being applied to the patient. A better measurement to indicate gas efficiency is to use both the inspired Oxygen (FiO_2) levels and those in the blood. Unfortunately this applied oxygen level (FiO_2) is not always taken and therefore clinicians continue to rely on the more traditional blood gas levels.

2.8.1 Measurement of pO_2 and pCO_2

The partial pressures of Oxygen and Carbon Dioxide are two of the most commonly measured signals from patients who are placed on ventilators. For that reason it is felt that more detail should be given on how these measurements are made and how the signals measured by the respective probes affect the signals which "Mary" logs.

These two signals are measured by a single transcutaneous (placed on the surface of the skin) probe which is heated up to a temperature of approximately 44°C. At these temperatures skin capillaries are at their maximum diameter and O_2 and CO_2 can diffuse through the skin and probe cell membranes into a solution where an electrochemical reaction takes place and a current is generated. This current is directly related to the gas concentrations in the blood [50]. The measurement of CO_2 and O_2 is made through a single probe. This however introduces problems as the measurement of Oxygen taken by this method is not necessarily the most reliable for diagnostic purposes as it varies in direct proportion to the fraction of inspired Oxygen

(FiO_2) in the air mixture. This FiO_2 measure can be altered by the clinicians. It is however the most commonly used and therefore results are often ambiguous.

2.8.2 Measurement of FiO_2

Despite being extremely useful for the diagnosis of certain respiratory conditions this measurement is often omitted in favour of the easily obtained pO_2 reading [2] which is generated as a by-product of the monitoring of pCO_2. FiO_2 is related to the current respiratory function of the patient by the relationship shown in 2.1 where the oxygenation index provides a measure of the efficiency of the respiratory process in the patient.

$$Oxygenation\ index = \frac{FiO_2 \text{ x } MAP}{pO_2} \tag{2.1}$$

MAP (mean airway pressure) is the amount of pressure being used to force the lungs of a patient to inflate. The oxygenation index can give a much better indication of the respiratory condition as pO_2 can be controlled to a certain extent by the carers. This signal is measured directly using an Oxygen analyser which is placed either in the ventilator circuit or in the incubator itself.

2.8.3 Summary

To summarise, the diagnosis of Respiratory Disorder involves the use of three physiological signals which are collected as part of the "Mary" monitoring system. These are blood gas concentration levels (pO_2 and pCO_2) and the fraction of inspired Oxygen in the air mixture generated by the ventilator (FiO_2). Clinicians feel that by using these three parameters it should be possible to predict ahead of the current diagnosis time the onset of respiratory disorder. Using the three different signals also permits rudimentary validation of the signals to take place. As pO_2 and pCO_2 traces are heavily correlated an outlier in one often appears in the other, the use of FiO_2 therefore acts to validate this region. An example of the type of these physiological signals is shown in Figure 2–7 where a single day's archived (minute-average) entry for the three gas level physiological signals for one patient is shown.

[2] FiO_2 measurements are not always logged as part of the *Mary* monitoring system and are therefore less easily obtained when data from *Mary* is being used.

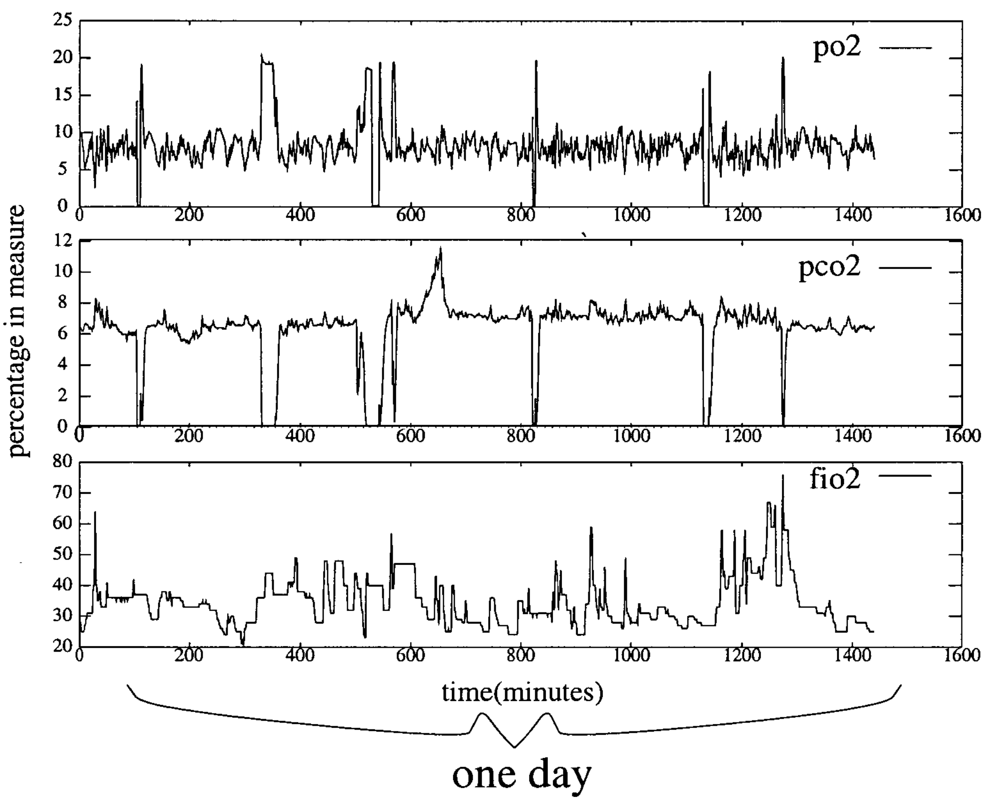


Figure 2–7: Graphs showing typical examples of physiological signals

2.9 Discussion

It is felt that any system which could increase levels of patient care and reduce the use of invasive techniques in treatment would be extremely beneficial. Clinicians feel that a method of achieving this could be to combine their expert knowledge and current monitoring system with other signal processing techniques to produce an early warning system for certain physiological conditions, in this case respiratory disorder. This thesis investigates the feasibility of producing this type of system.

2.10 Conclusions

The main difference between an adult ICU and a neonatal ICU is that in the adult ICU the patient tends to a certain extent to be able to communicate their symptoms whereas in the neonatal ICU the clinician is forced to diagnose cases from physiological signals and the visual condition of the patient alone.

There are also conditions which are unique to the NICU or are more prevalent there. The main one of these is respiratory disorder which is exacerbated by the poorly developed lungs of the neonate. This chapter has discussed the operation of the NICU and the problems which it faces in the diagnosis of conditions. In particular it has focussed on the development of Respiratory Disorder, RD, and suggested that technology may now be able to assist in its diagnosis.

Chapter 3

Monitoring and diagnosis: A technical view

3.1 Overview

In this, the second of two chapters dealing with the problem of monitoring within a neonatal intensive care unit the problem will be examined from a different angle; that where the final application area is ignored and analogous application domains are examined. This chapter can be split into two parts; systems which have been developed for domains where the application is analogous to the medical monitoring domain and the techniques used in those areas which might be useful to the actual application being investigated.

3.1.1 Medical diagnosis

The field of automated medical diagnosis is a small but expanding one. The benefits of introducing automatic monitors and diagnostic aids into a medical care environment are varied and some of them are described below;

- Reduction in habituation of diagnosing staff
- Greater opportunity for retrospective clinical training using patient's records
- More accurate record keeping

Until relatively recently most of the research carried out in the medical domain included work on human physiology to aid understanding of other problems. For example the Electroencephalogram (EEG) has been studied to enable clinicians to identify different patterns of brain signals and relate them to patient behaviour, for example sleep states [14,51]. As far as it is known little research has been carried out in the neonatal domain on which this thesis concentrates. To date the published work which has been carried out in this area has concentrated on the control of equipment within the neonate unit, for example control of ventilator flow [6,26] or on the elimination of false alarms [52]. Despite the lack of current research in this application area it is possible to draw parallels between it and other applications areas.

3.2 PART I: Perception of the problem

3.2.1 Introduction

In this section of the chapter a number of applications will be examined. The applications which are described, although upon first examination bear little resemblance to the clinical domain, have features inherent in them which are directly analogous to the application area under investigation.

On first investigation the problem is typical of many condition monitoring applications in that time-series signals are measured and if they stray out with specified bounds an alarm is sounded. The problem will therefore be treated as this type of application until further investigation has taken place.

3.2.2 Condition Monitoring

Traditionally systems are allowed to operate until they no longer operate above the required efficiency. For example a local water supply company may continue operation until its efficiency is below a certain level, for example the threshold could be that over 40% of the water used to supply homes in the local area is being lost through leakage. Obviously water leakage is a serious problem as it is not only wastes a valuable resource but it can also lead to further degradation of the piping system as non-insulated joints corrode.

In a condition monitored system different components of a system are monitored in the hope that when the system can no longer operate to its required efficiency it is possible to isolate the faulty section of the system. Essentially the type of system which is currently in place in the NICU is a condition monitoring one in that physiological signals are measured and if any of them stray out with pre-defined bounds an alarm is sounded. However there is no attempt to combine signals and alarms, for example if an alarm of a heart-rate monitor was triggered because the registered heart-rate was too low a validation check could be performed by checking the ECG or respiratory rate monitors and if those too showed problems an alarm would be sounded. It is therefore necessary that any new monitoring system designed for use in the NICU at the SMMP should take these details into account and only trigger an alarm when a combination of factors merits it. It is also necessary that the system be capable of isolating the problem area, for example the heart if the ECG drops, as this will speed up any diagnostic processes which the clinicians or carer is involved in. The means that the problem being tackled within the scope of this thesis is not a purely condition monitoring application, it is more a subset of this domain and it could be viewed as a fault diagnosis problem.

3.2.3 Fault diagnosis

As mentioned fault diagnosis attempts to, while monitoring a system, isolate the faulty component as the system breaks down. This does not necessarily mean that the problem itself is diagnosed, for example a gear meche has broken in a motor, it will rather identify the area which contains the specific problem, in this case the motor. In pure terms this means that the fault diagnosis problem is a combination of a condition monitoring problem which includes some classification of problems. Successful fault diagnosis requires a good knowledge of the system which is being monitored and to correctly identify faulty components even greater knowledge is required.

Classification

If classification of a faulty component is required as part of the fault diagnosis task then a large amount of knowledge about a particular system is required. For example information such as in a house a washing machine is less likely to fail than a light bulb therefore the probability that a light bulb is faulty is higher than that of the washing machine. There are many applications where classification of a problem has been required and these include satellite communication systems [53,54] [55], power systems [56,57], manufacturing systems [58], sonar signal classification [59], remote sensing [60], speech recognition [61,62,63,64,65,66] and some medical diagnosis applications [67,68,69].

In the cases where medical diagnosis has been performed it has tended to be performed on either images, for example cervical smears [4], mammograms [70] and Electrocardiogram (ECG) traces [3,20,28,71,72,73]. A typical classification system is designed to recognise patterns which indicate certain problems, these patterns can be in the form of dense tissue masses in the case of vascular ultrasound images [74] or the distances between the PQRS complex in an ECG waveform. In more general terms a classifier is a type of pattern recognition system which, using stored knowledge, can classify or categorise a specific problem. Pattern recognition techniques can also be used to predict the breakdown of systems, for example a tyre will leak before it goes completely flat.

Prediction

In many situations fault diagnosis is required to take place before the common fault occurs. For example diagnosis of a set of gears degrading before the gears fail. This type of monitoring system is in place to reduce the cost of running a system. These costs can be broadly divided into three categories depending on the application area;

- Economic: This includes the cost of replacement parts and the "down-time" as the system is repaired.
- Environmental: If systems are allowed to break down or the operation is allowed to degrade, in particular power systems, environmental damage can result, for example in the increase of sulphur dioxide in a filtering process.
- Health: If a system is allowed to degrade to a certain extent damage to its operator can result. For example in large machinery if insulators degrade the operator can suffer an electric shock.

It should be noted that of all the costs the economic is often the largest as in many systems it is relatively cheap to replace a small component but if that component is allowed to fail other parts may become damaged and the cost of repair of the system can escalate. For example in the case of a fluid power system if the fluid pump is allowed to break-down the entire system can be destroyed or badly affected [75].

In our application area the economic cost is analogous to treatment cost as prevention of a condition reduces the need for further, possibly invasive, more expensive treatment.

The purpose of a fault diagnosis system is therefore to monitor a complete system and to warn of impending problems, i.e. predict the breakdown of a certain part of a system before further cost is incurred from its complete breakdown. This type of system has been used in a number of different application areas which include satellite communication systems where if a fault occurs to the system irreplaceable data can be lost [55], terrestrial communication links where important information can become too degraded to permit retrieval [76], manufacturing systems where a component failure can lead to long periods of "down-time" and severe degradation in product, jet-engine starting to ensure optimal efficiency of the start-up procedure [77] and fluid power systems where pump failure can lead to system failure [78].

In our application it is necessary that the system can recognise the onset of the series of conditions known as respiratory disorder before the time that the diagnosticians currently recognise it. This can be thought of as part classification, part prediction as a prediction of the ultimate condition of the patient is being made while the system recognises, or classifies, the precursors of the condition. To date few people have looked at this type of early diagnosis in the medical domain and this thesis details the first time this type of system has been developed for neonatal intensive care.

3.2.4 Summary

On first investigation the application area of this thesis is extremely specialised, however, it should be obvious that many other applications have included factors which are directly analogous to the field under investigation. These application areas can be generalised to include the field of condition monitoring and its related subsets, see Figure 3–1. This means that though the field of research may be novel the techniques which can be applied there are not.

3.3 PART II: Techniques

3.3.1 Overview

In this part of the chapter techniques which have previously been applied to condition monitoring, and its associated subsets, will be examined and their relevance to the application area discussed.

3.3.2 Artificial Neural Networks

One of the most commonly used techniques applied to fault diagnosis and classification problems is that of the set of non-linear models known as artificial neural networks (ANN). Unlike most conventional non-linear modelling techniques little or no knowledge of the model underlying the system being investigated is necessary as ANNs are capable of adapting a preliminary model to "fit" the system. However, increasing the knowledge of the underlying model will increase the capabilities of the ANN to model that system.

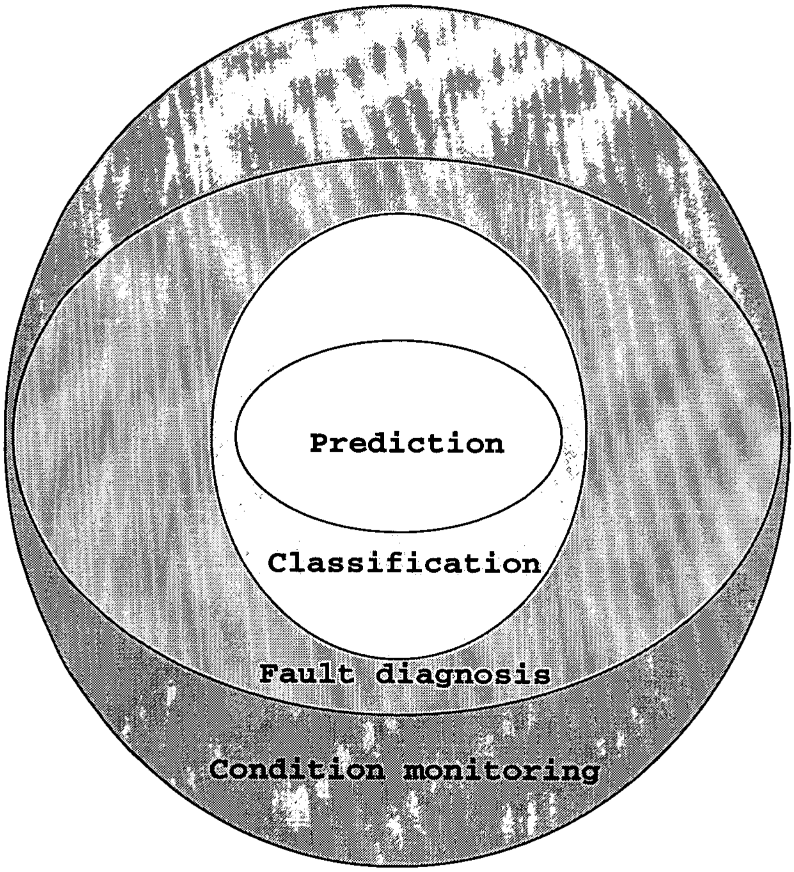


Figure 3–1: Condition monitoring and some of its sub domains

There are many different types of non-linear model which are collectively known as ANNs, however within the scope of this thesis only one type will be used; the Multi-layer perceptron (MLP) (see Appendix A), which is a supervised ANN. In general terms supervised ANNs are non-linear models which can be trained to recognise or classify new data based on the data it has been trained upon. In other words an ANN will classify previously unseen data in terms of its similarity to previously seen data. The ability of the ANN to classify new data correctly is therefore entirely dependent on the quality of its training data and how well this can be separated into the different classes of interest. The process of training the neural network is relatively simple: a set of training data is collected, this set must fully represent the classes (or in the case of fault diagnosis, the problems) requiring to be classified. Training can then proceed with the representative examples of problems being applied to the ANN. The ANN is also instructed as to the correct category to which the input data belongs. As training continues when the ANN incorrectly classifies the input data the ANN's connective weights are adapted so that the next time it sees the same training pattern it will either be correctly classified or the error between the correct and incorrect classifications will be reduced. At the end of training the aspiration is that the ANN will correctly classify all the training data and if presented with a set of test data (whose classes are known) it will correctly classify those as well. The training process can therefore be described as shown in Figure 3–2.

Medical diagnosis uses ANNs for a variety of tasks which include the diagnosis of abnormal ECG patterns [3,79,30], expert systems [34], detection of melanoma [80] and others [35]. They are especially suited to many medical applications as the underlying system is often poorly understood, for example the development of a disease can include a range of symptoms in different patients. In these cases the ANN models the system and some preprocessing is used to incorporate the expert knowledge of the system that exists.

3.3.3 Time-series analysis

Another technique which is commonly used in classification applications is that of time-series analysis. These techniques are directly relevant to the field under investigation as the data which is supplied by the clinical team is in time series format. However, time-series analysis is not only applied in the condition monitoring domain it is also used in a much wider variety of classification problems which range from speech processing [31,81,82] to outlier recognition in foetal heart-rate traces [83].

Time series analysis is also often used in combination with other classification techniques, for example in the pre-processing of input data before it is applied to an ANN. The more

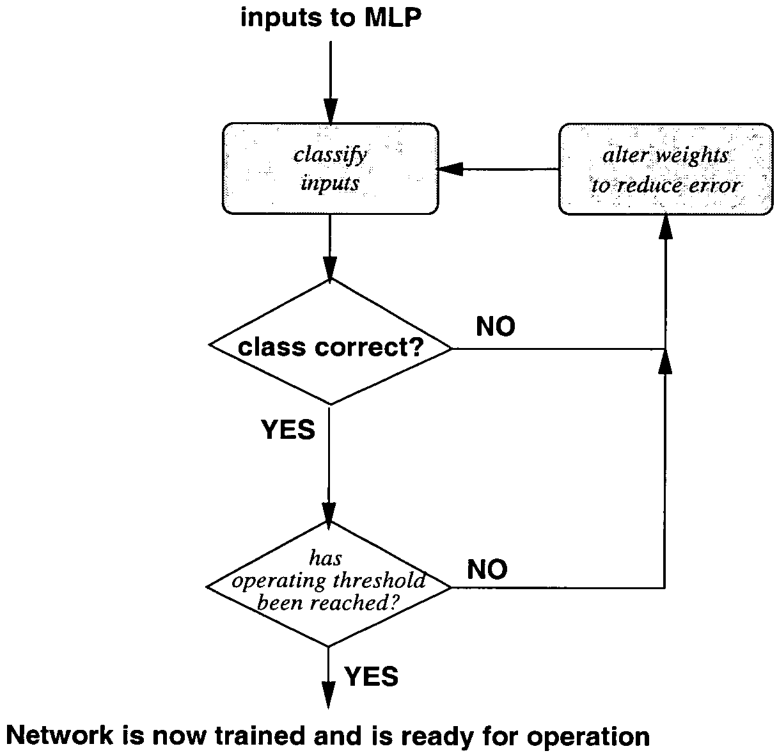


Figure 3–2: A schematic description of the training of a neural network

successful the pre-processing the greater the chance the ANN has of successfully modelling the underlying system being monitored. When used as pre-processing time-series analysis can be used for one of two reasons:

- improve data quality
- maximise information content in data

The first of these is carried out when there are large artifacts in the signal, i.e. sections of the signal which are not physically possible [83]. It is achieved by using some type of filter to remove the sections of the signal which are not of interest to the classification problem. In the second type of analysis the original data is changed in some way to maximise its information content. This type of analysis depends entirely on the application area and what information is known about the system. For example if it is known that important information is contained in the frequency spectra of the signal it is sensible to transform the raw data in the time domain into the frequency domain. In images this could mean looking at the texture of the image rather than just the black and white scan. Time-series analysis is an integral part of all classification problems as the better the analysis and hence the better the information being applied to the classifier the greater the likelihood is that the classifier will be able to discriminate correctly between classes.

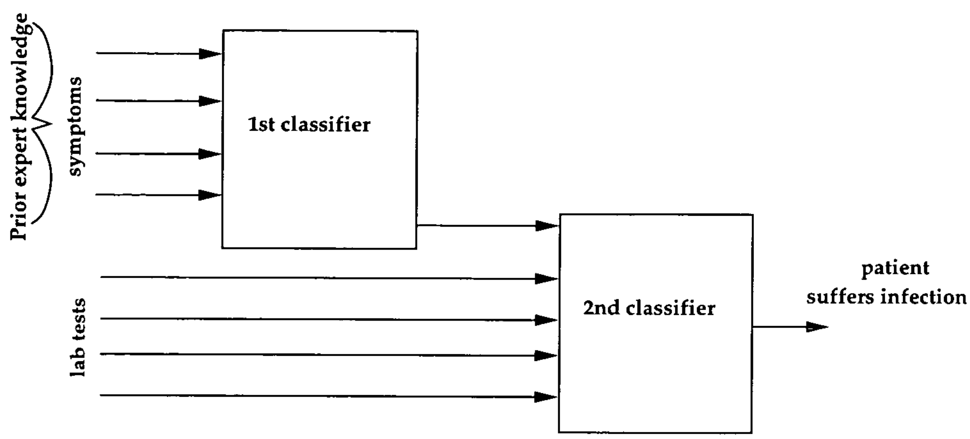


Figure 3–3: Using expert knowledge in system design

Time-series processing need not always be used in combination with an ANN, it is capable of classifying signals in its own right [84]. However, the situations where it is used as a classifying technique tend to be those where a lot of information is known about the system. In those cases where little is known about the underlying system it can only really be used in combination with other techniques. In the medical domain time series analysis is a necessary part of the classification process. It is in the time-series processing that the expert clinical knowledge can be used to maximise the accuracy of the classifier. Where many clinical judgements are made it is usually possible for diagnosticians to describe their diagnostic processes in a manner which can be adapted for use in pre-processing, for example if they look for trends within the data then the pre-processing should extract trends. An example of this type of prior knowledge being used in the design of a system is shown in Figure 3–3.

3.3.4 Hidden Markov Models

One technique which is commonly associated with both time-series analysis and neural networks is that of the Hidden Markov Model (HMM) (see Appendix B). This is an extension of the Markov chain approach to system modelling where a system can be described in terms of states, in the HMM the system is described in terms of unknown states which have known symptoms or observations. HMMs have until now been predominantly used in the field of speech recognition [85,64,65,86,66].

The purpose of using an HMM in any system is to try to incorporate temporal information into the model of the system. It is uncommon that a condition monitoring application can change instantaneously between two radically different states. The HMM approach attempts to incorporate this knowledge into the model by including intermediate states. Using an HMM allows the gradual breakdown of the system to be modelled and if an HMM is combined with a classifier it allows the classifier to take the previous states of the system into account. For example in the application on which this thesis concentrates, if breathing is becoming laboured (state one) there is an increased probability that the patient will become cyanotic (state two) rather than the patient breathing normally and suddenly becoming cyanotic.

This ability of the HMM to model systems where the outcomes or observations of the system (in our case symptoms) can be seen but where the system itself is too complex to model is ideal for use in speech processing [87,62,63] , medical applications [21,20,19] and condition monitoring applications [88,53]. To date it has been particularly successful in applications where specific conditions are to be diagnosed rather than entire systems being modelled. For example satellite communication systems where specific faults are to be diagnosed; motor failure, tracking failure etc. [55]. These types of systems have direct parallels with medicine where a specific condition requires diagnosis, for example various types of heart arrythmia [28] and the monitoring of glucose levels of a diabetic [89].

3.3.5 Summary

Techniques which have been developed for other application areas are applicable in the neonatal monitoring context. Some techniques are capable of being used individually as classifiers whereas others require more development (see Figure 3–4). In general it is possible to say that the greater the relevance to the problem being studied of the information being maximised the higher the chance of a classification system performing adequately.

3.4 Summary

This chapter has described the particular application area as perceived by a non-clinician regarding the problem for the first time in that similarities between the application domain and others are sought. Despite the initial appearances of this application involving an extremely specialised area it is possible to draw parallels between it and other application areas, in particular condition monitoring and its associated subsets of fault diagnosis, pattern classification and prediction. As it is possible to regard the particular application area of interest in this way it is possible that techniques which were originally developed for other domains can be used in the extremely specialised domain of neonatal monitoring.

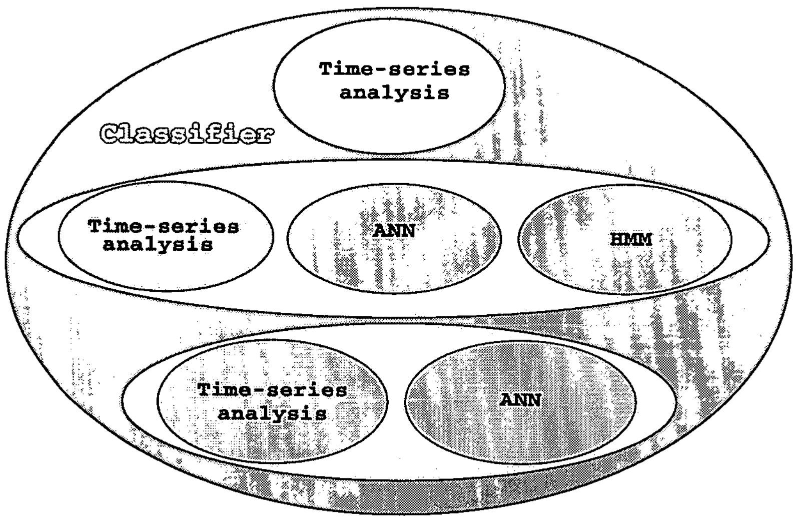


Figure 3–4: Classifiers and their sub-parts

Chapter 4

Condition monitoring system for neonatal intensive care

4.1 Overview

In this chapter the particular application area and project will be reassessed and described. As seen in the previous two chapters the application area is a particularly wide one in that there are many problems which exist there. However, within the scope of this thesis it has only been possible to isolate one particular problem for further investigation and this chapter will discuss it in greater detail.

4.2 Aim

The specific problem which has been investigated has been identified by clinicians for a number of reasons: It is

- extremely common
- occurs in adult ICUs but not with the frequency of the NICU
- and is often undiagnosed until further more expensive and invasive therapy is needed

The problem under investigation is whether it is possible to diagnose that a particular patient within the neonatal intensive care unit (NICU) will suffer from a form of respiratory disorder such as a blocking endotracheal tube or a pneumothorax before further therapy is required.

4.3 Materials

The data which are used in this investigation is retrospective data which has been collected as part of the current monitoring system which is in place in the NICU at Edinburgh (a typical days' data set is shown in Figure 4–1). This monitoring system allows physiological data to be cross referenced with patient treatment. The data is time-series analogue data which has been averaged every minute. It is binary-encoded to save space when it is archived therefore it requires to be decoded before analysis can take place. Once decoded the data are identical to that which the clinicians used in making their treatment decisions about the relevant patients. All patients chosen for the study were artificially respirated either because their lungs were not fully formed or because they were incapable of breathing efficiently unassisted. Expert knowledge of the conditions under investigation exists as often clinicians, after a case has been diagnosed, can in retrospect examine the time-series physiological data and detect the onset of the condition. In this study data from 21 patients on 51 different days were used. They all suffered from some form of respiratory problems during their stay in the neonatal intensive care unit.

The system which has been developed for this study is designed to detect the symptoms which are typical of onset of respiratory disorder.

4.4 Techniques

The classification system which has been developed uses a number of different techniques. It combines time-series analysis with a multi-layer perceptron neural network. Time-series analysis is used to incorporate expert knowledge into the classification process and the MLP neural network is used as a classifier. Temporal information is also included in the system within the time-series analysis section. Training patterns are formed from retrospective clinical data and classification is determined by the clinical diagnosis at the time of data collection.

Table 4–1 details how each part of the classification system was implemented. All Application specific code was written by myself and the generic code was written by Mike Smart of the Integrated Systems Group at Edinburgh University.

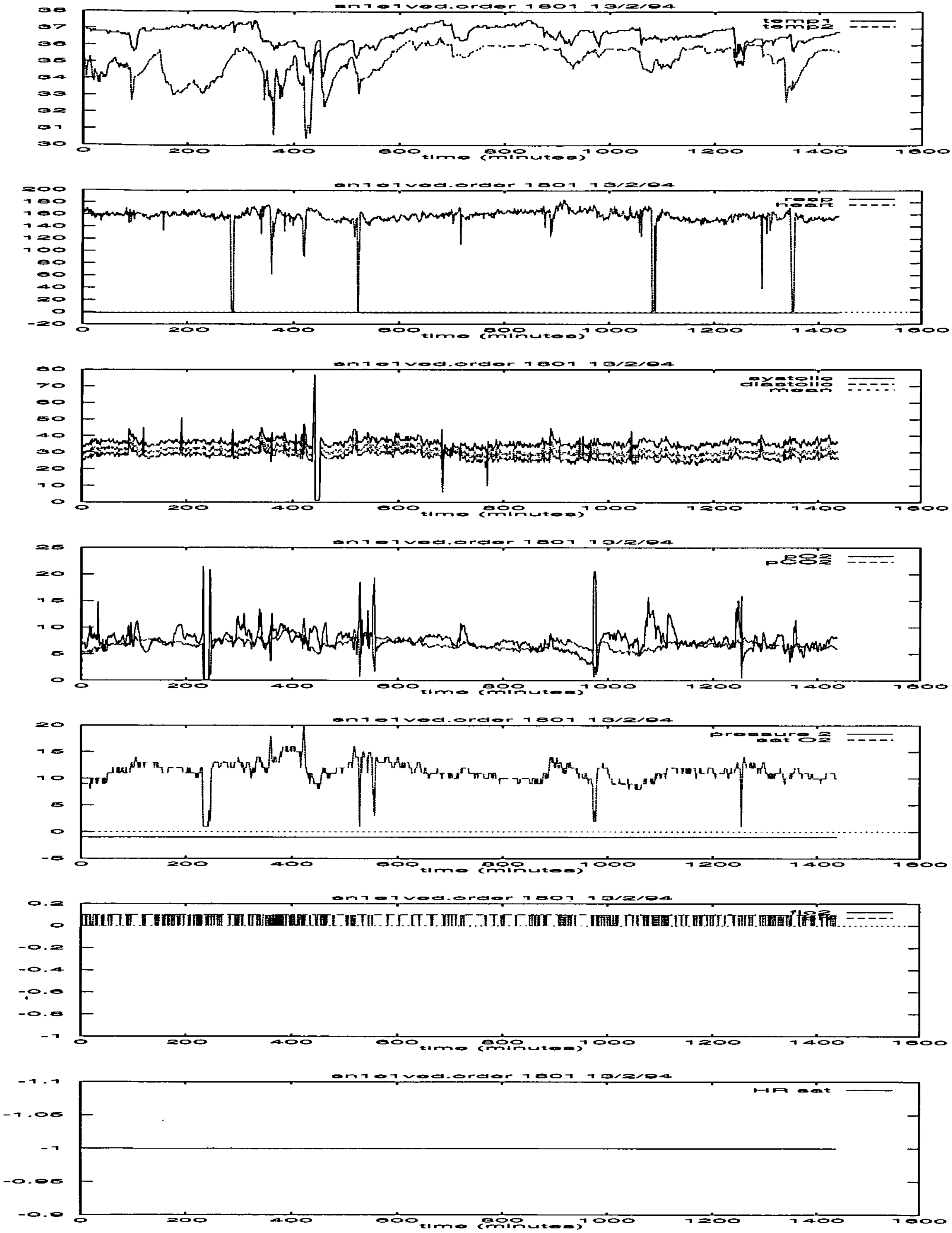


Figure 4–1: An example of a typical days data

System block	Software implementation
Binary Decoder	Application specific C-code
Filter	Application specific C-code
Signal Selector	Application specific C-code
Feature Extraction	Application specific C-code
Prototype classifier	$Neuralware^{TM}$ neural network simulator
Linear Classifier	Generic C-code
MLP Classifier	Generic C-Code

Table 4–1. Implementation methods for sub-sections of system

4.5 Ultimate objectives

It is hoped that ultimately the results of this study can be used to improve the level of patient care within both neonate and adult ICUs. This will be achieved by producing a diagnostic aid for respirated patient management and hence reducing the number of invasive and stressful procedures carried out. Diagnosis should take place before they are required. Less alarms will also be triggered which relate to the ventilator equipment as the system uses multiple channels in its classification and only if all channels are indicative of a problem will any warning be sounded. The system will act as a diagnostic aid which suggests to physicians that a problem may be developing.

4.6 Summary

This chapter has explicitly stated the aims and objectives of the work described in the latter part of the thesis. The work is described as a study into the feasibility of developing a system which could act as a diagnostic aid to physicians for artificially respirated patients. The system uses recorded physiological data to recognise the onset of respiratory disorder in this subgroup of neonatal patients.

Chapter 5
Design Methodology

5.1 Introduction

This section of the thesis describes development of the classification system used to monitor the onset of respiratory difficulties. It addresses the issues which have arisen during the design of the system and the motivation behind some of the decisions that were taken.

The decision was made that for any new system being developed it must be possible to explain to the end user, in this case the diagnostician, the reasoning behind the choices that have been made. For this reason the approach adopted has relied heavily on expert knowledge of the underlying processes happening to the patient at a particular time. The final system allows trained medical staff to validate the system output as they are aware of what particular patterns or symptoms the system has been developed to detect.

In this case it was decided that the classifier should be used to generate an output which gives an indication of the risk a particular patient has that he/she is developing respiratory problems.

5.2 Design Overview

Any system which is designed to classify or recognise patterns must be supplied with the best possible data to enable it to achieve its full potential. This is not always possible in real-world application areas as data is often corrupted by the data collection process itself. In this case the data is supplied by the neonatal intensive care unit (NICU) at Edinburgh. It is multi-channel, minute-averaged data archived using the current PC-based monitoring system, called *Mary* and the system developed must ultimately be capable of linking with this. The usual method for producing a classification system therefore involves three stages:

- Data Analysis
- Pre-Processing
- Classification

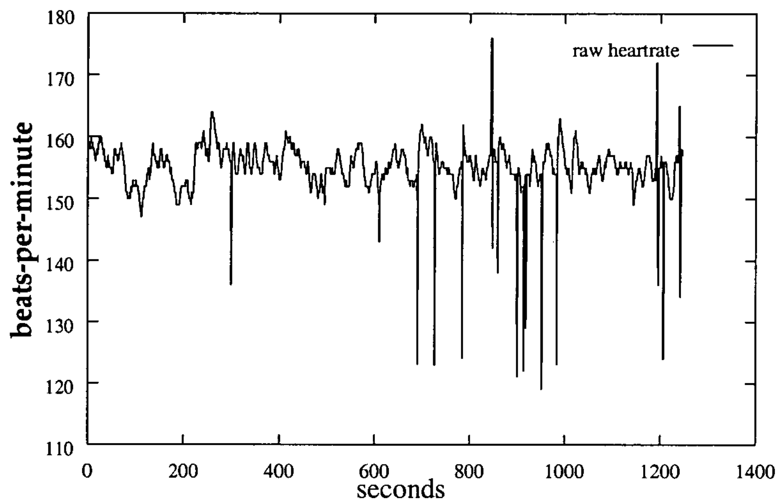
Data Analysis involves "getting to know" the data and its production. In this case the data is being produced by artificially respirated patients whose physiological signals are collected by purpose-built monitoring devices. The monitoring process introduces noise and artifacts to the signals, the noise is of no value in a diagnostic context and must therefore be removed. The case of the artifact, however, is slightly different.

5.2.1 Artifacts and outliers

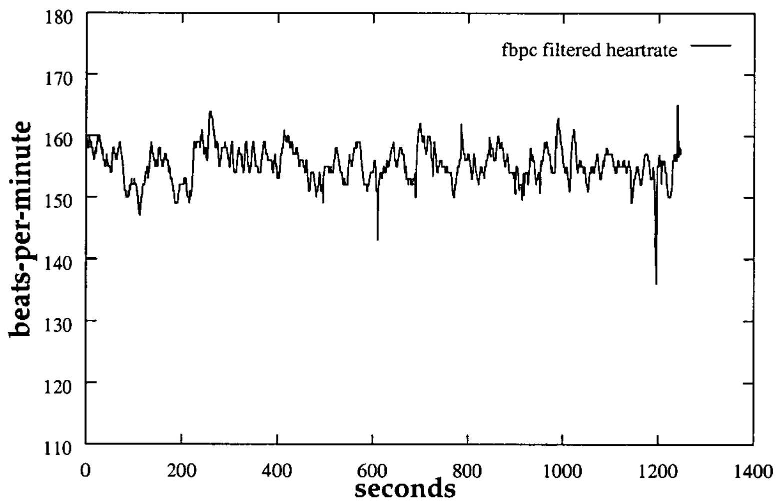
Artifacts or outliers are defined as sections of the signals which have been produced in error and hence are invalid. In this case they can be broadly divided into two categories; man-made artifacts and physiological artifacts.

Man-made artifacts are the result of the monitoring device being disturbed and logging an erroneous result. There can be a number of reasons for this ranging from movement of the wires connecting the patient to the monitor, or the monitor to the data logger, or large noise spikes in the signal. Most of these types of artifact can be removed easily using thresholding functions or band-pass filters that will only allow the region which lies within the range of interest to be examined. For example in the work of Bassil et al. [83] a forward-backward-predictor-corrector was used to remove outliers from a foetal heart rate signal. Figure 5–1 shows the results from their algorithm when it is applied to a neonatal heart rate trace. A common cause of this type of artifact in the neonatal trace is the probe being changed. This occurs every few hours as the transcutaneous probes placed on the skin of the patient are moved to prevent the skin from burning. This probe movement obviously leads to a change in the relative concentrations of gas being measured. An example of a probe change trace is shown in Figure 5–2. This type of artifact cannot easily be removed with thresholding as the values being generated during the artifact generation are "realistic" values. These artifacts must therefore either be removed using expert knowledge or included in the final system as they are a common occurrence.

Physiological artifacts are more difficult to deal with. This type of outlier is introduced by the patients themselves or by their treatment. The patient may experience a transient response to certain stimuli, for example when he/she is moved his/her heart rate may increase for a period of time. This type of outlier cannot be removed easily as, similar to the probe change, the



a) Raw neonatal heart-rate signal



b) Filtered neonatal heart-rate

Figure 5–1: Neonatal heart-rate trace before and after filtering

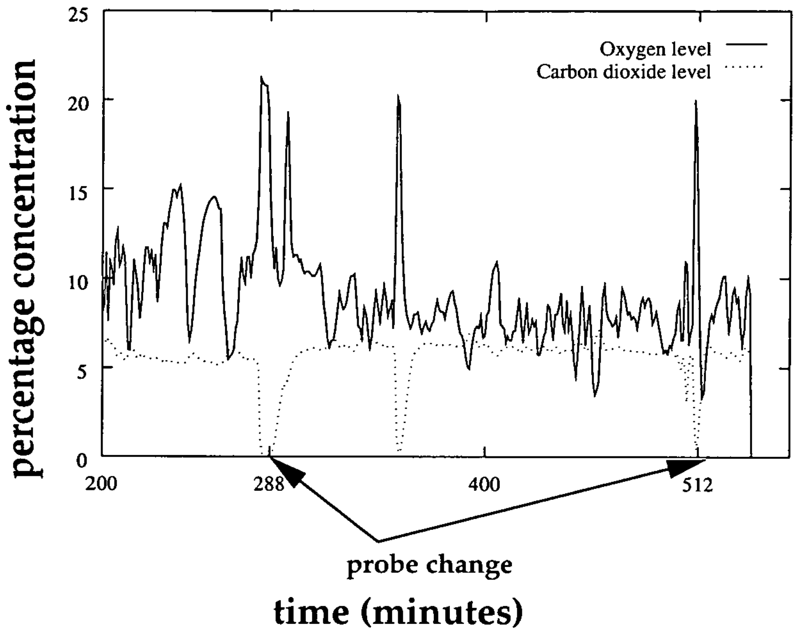


Figure 5–2: Gas levels during probe change

values measured are realistic in physiological terms, and could be of interest under certain circumstances.

In this application the majority of man-made artifacts and noise have been removed by filtering the signal. Probe changes and physiological artifacts remain, as these types of artifact are common in the signals generated by *Mary* (the PC based monitoring system currently in use at NICU in Edinburgh) and therefore the system being designed must either be able to cope with them, or learn to ignore them as artifacts.

5.2.2 System structure

The initial prototype for the complete classification system consisted of three distinct stages: filtering, feature extraction and classification as shown in Figure 5–3. The role of the individual stages is summarised below.

Filter

This involves the separation of the data of interest from the raw signal. This can mean removing or filtering certain elements from the raw signal, for example the high frequency noise and man-made artifacts.

Extract Features

This stage uses the filtered signal as an input. Further processing is performed on the data to convert it into a form that will maximise any ability of the classifier to diagnose the condition. For example, if all of the information required for a diagnosis is contained in the distances between the peaks in the ECG waveform, this feature extraction could involve converting the ECG into a single figure average heartrate measure [2] or into the R-R distances in the QRS complex [72] [79] [30]. The information which is extracted from the filtered signal and which is used to describe that signal is given the name features. It is in this section where expert knowledge of the underlying processes involved in the diagnosis of conditions can be applied as far as possible.

Classify

This is the final stage in the prototype system. In it, the information which has been extracted from the raw data in the prior stages is applied to some form of classifier. The classifier can be of any number of types. These can range from a neural network to an expert system or standard probability measures, of the likelihood of the condition occurring, given the information which the system has just received. In this instance it was decided that the classifier should give an output which reflected the probability that the patient was suffering or about to suffer from a particular condition. It was therefore decided that the classifier should produce an output which indicated whether the behaviour of the patient was typical of one class or another. These classes were defined as follows:

- CLASS 0 - no concern is warranted given the physiological data examined
- CLASS 1 - concern, something unusual is occurring which may indicate the development of respiratory disorder

The following sections detail the development of each of these stages in turn and explain the motivation behind the use of certain techniques.

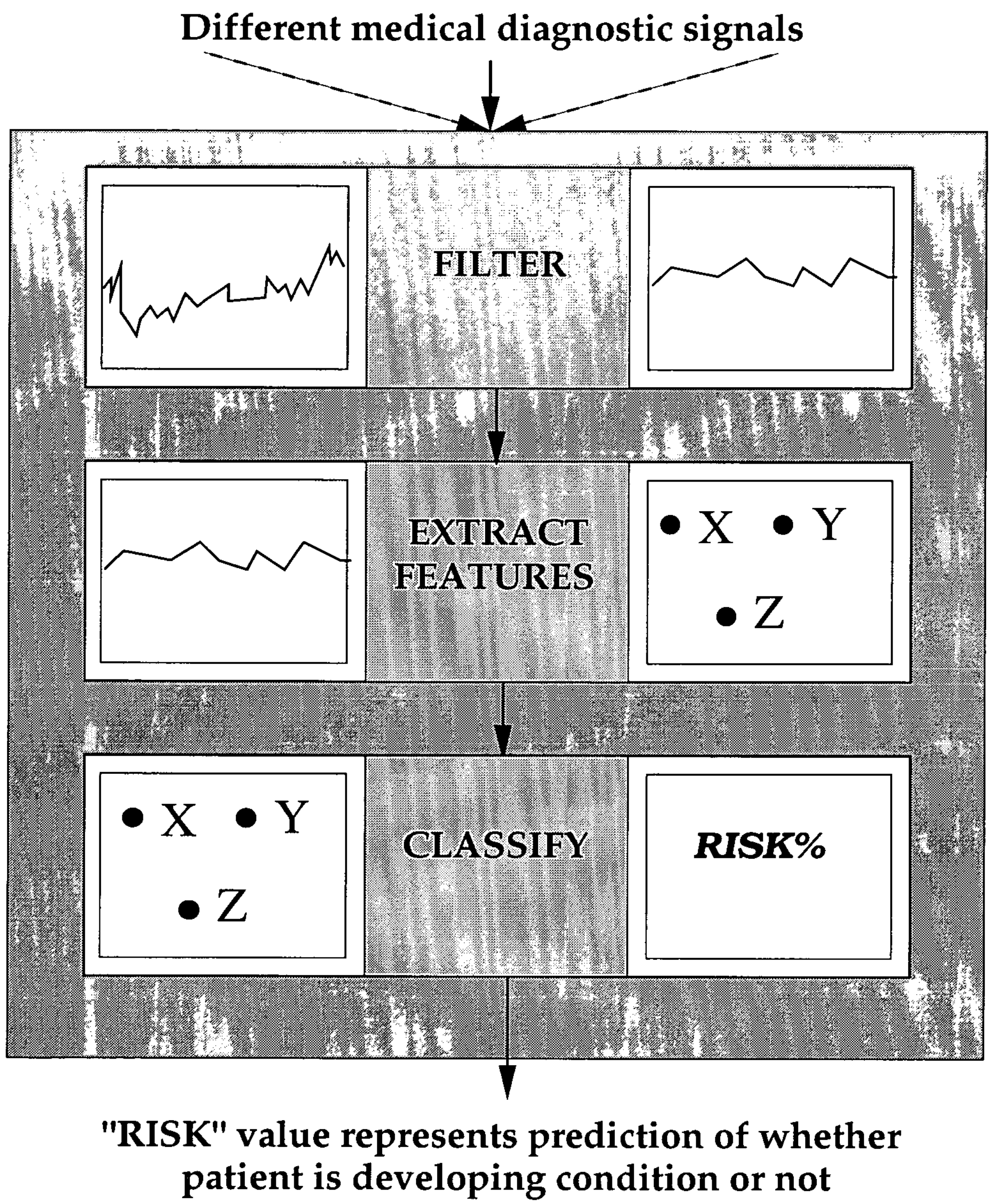


Figure 5–3: A schematic of a possible classification system

5.3 Data processing and analysis

In this section of the design, processing is carried out that will convert the data into a form which will maximise the ability of the classifier to discriminate between the classes being examined. It is at this stage that the data is examined and expert knowledge about it obtained.

5.3.1 Incorporation of Expert knowledge

For any type of diagnostic aid it is necessary to incorporate as much expert knowledge as possible. The reasons include:

- ease of operation by the end users. In this case these are non-technologists and therefore may be unfamiliar with the techniques and processes used
- greater understanding of the diagnostic process by the user, i.e. the use of techniques which are accessible to the user
- greater acceptability of the system to clinicians and medical staff
- ease of integration of new system into current monitoring system

In this specific application area there is little expert knowledge of the precursors or triggers of respiratory disfunction. However, clinicians often state after an "event" has occurred that they can identify the region of the archived physiological data where the problem started to develop. In particular this is true of the specific gas concentration levels which commonly exhibit trends in their behaviour which signify poor gaseous exchange in the lungs. These trends are summarised in Table 5–1.

Gas Type	Precursor
Carbon Dioxide pCO_2	increasing value
Oxygen pO_2	decreasing value

Table 5–1. Specific trends exhibited in the blood gas data

However, expert knowledge also tells us that the value of oxygen in the blood can to a certain extent be maintained artificially by increasing the oxygen concentration in the air mixture

(Fraction of Inspired oxygen FiO_2) being supplied to the neonate. A more complete table is therefore shown in Table 5–2.

Gas Type	**Precursor**
Carbon Dioxide pCO_2	increasing value
Oxygen pO_2	decreasing value
Fraction of Inspired Oxygen FiO_2	either stable or increasing

Table 5–2. Specific trends exhibited in gas concentrations

As the fraction of inspired oxygen in the air mixture (FiO_2) can also be used in the diagnostic process it was decided that all three measurements of gas concentration levels would be used and a study would be made of the effect of inclusion of FiO_2 on the ability of the classifier to identify the periods of abnormality in the raw physiological data.

Using these data and expert knowledge, techniques were employed which would extract both the trends and the temporal information from the raw physiological data. The first stage in this extraction process was to filter the data to remove the high frequency noise.

5.3.2 Filtering

In this stage of the signal processing the objective of filtering is to remove elements present within the raw signal that may "confuse" or inhibit later processing. In this application a simple low-pass filter was applied to the minute-averaged data archived by the current monitoring system, *Mary*. As the system is designed to be used on real-time data from *Mary* and the only data available for testing is archived, a method had to be found to maintain the time link between archived comment files and archived physiological data files. This was achieved by passing the raw data through the low-pass filter twice; once in the forward direction and once in reverse. This forward-backward filtering eliminated any time shift inherent in the filtering process (see Figure 5–4). Another effect of the filtering process is that the output of the filter is constrained within certain levels, i.e. it was thresholded. For example it is known that if the fraction of inspired oxygen (FiO_2) is below 20.98 % (the fraction in normal air) it is an artifact [90]. Therefore, on the first pass these artifacts are replaced by the preceding value. The window length (time) over which the mean of the raw data is found can also be varied but for this application it was chosen to be thirty minutes as clinical knowledge told us that trends were likely to be of greater length than this. Diagrams of the filtering process are shown in Figures 5–5 and 5–6.

An obvious problem with this approach is that the system cannot be run in real-time. However, if real-time processing is required one option may be to run the filter in the forward direction as the reverse pass is only required to maintain the time-correspondence between the medical record and the datafile. However, using this filter the time lag for each element of the signal cannot be guaranteed as this type of filter has a non-linear phase delay associated with it. This means that different elements (for example the higher frequency components) of the signal may be delayed by different amounts. Another option therefore would be to use a finite-impulse response (FIR) filter as these have a linear phase delay and therefore the lag introduced by this section of the system can be guaranteed for all of the signal components.

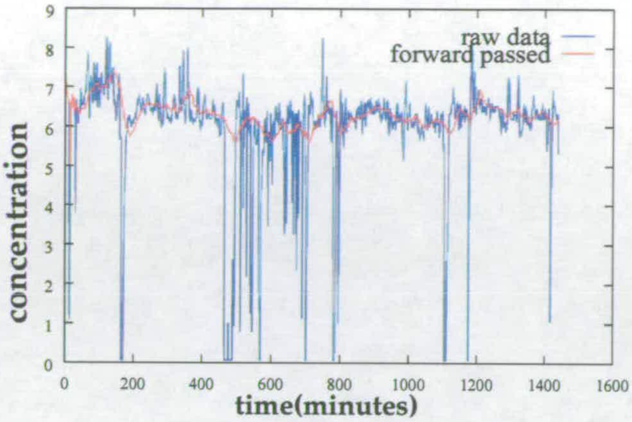
5.3.3 Feature Extraction

Once filtering has taken place, and the artifacts and outliers removed, the processed signal can be passed onto the next stage of processing. This stage is called feature extraction and its objective is to convert the filtered signal into a form which maximises the information content of its output. Expert knowledge tells us that Respiratory Disorder often develops over a number of hours and after its diagnosis it is frequently possible to identify trends within the physiological data which would have been indicators of RD. Tables 5–1 and 5–2 and Figure 5–7 shows Carbon Dioxide and Oxygen measures when a blocked ET tube was detected. As it can be seen Carbon Dioxide increases and Oxygen concentration decreases in the period preceding the diagnosis of the blocked endotracheal tube. The methods of feature extraction used within the scope of this thesis have attempted to incorporate this expert knowledge in both capturing some of the temporal aspect of the signal and in the processing of the trends.

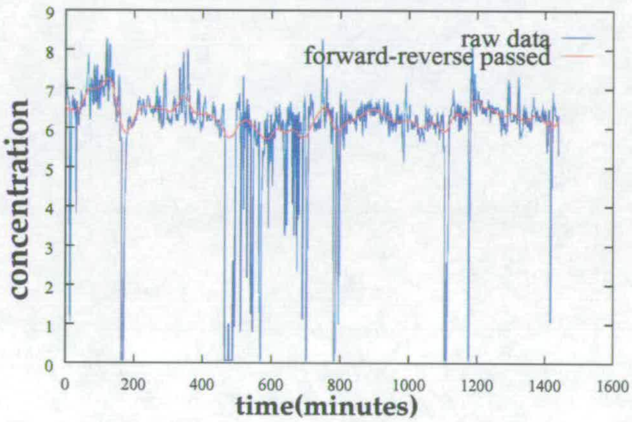
Two methods of including temporal information in the feature extraction have been developed and their results are compared later. These methods are designed to differ in the way in which temporal information is included.

Non-overlapping features

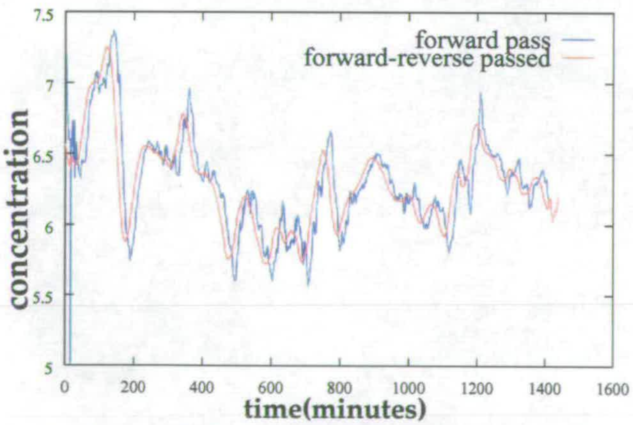
This method of feature extraction involves calculating the gradients of the filtered signal by separating the signal into non-overlapping, or contiguous, windows. The overall length of the combined windows is thirty minutes. This time interval was chosen as clinicians feel that a warning of thirty minutes before an "event" is a significant improvement on the current status of diagnosis. This interval can be extended or reduced by altering the time intervals used in the feature extraction process.



a) raw data and filtered data (forward pass)



b) raw and filtered data
(after both forward and reverse passes)



c) time shift in filtered data if only forward pass is used.

Figure 5–4: Single channel of data a) pre-filter b) post filter

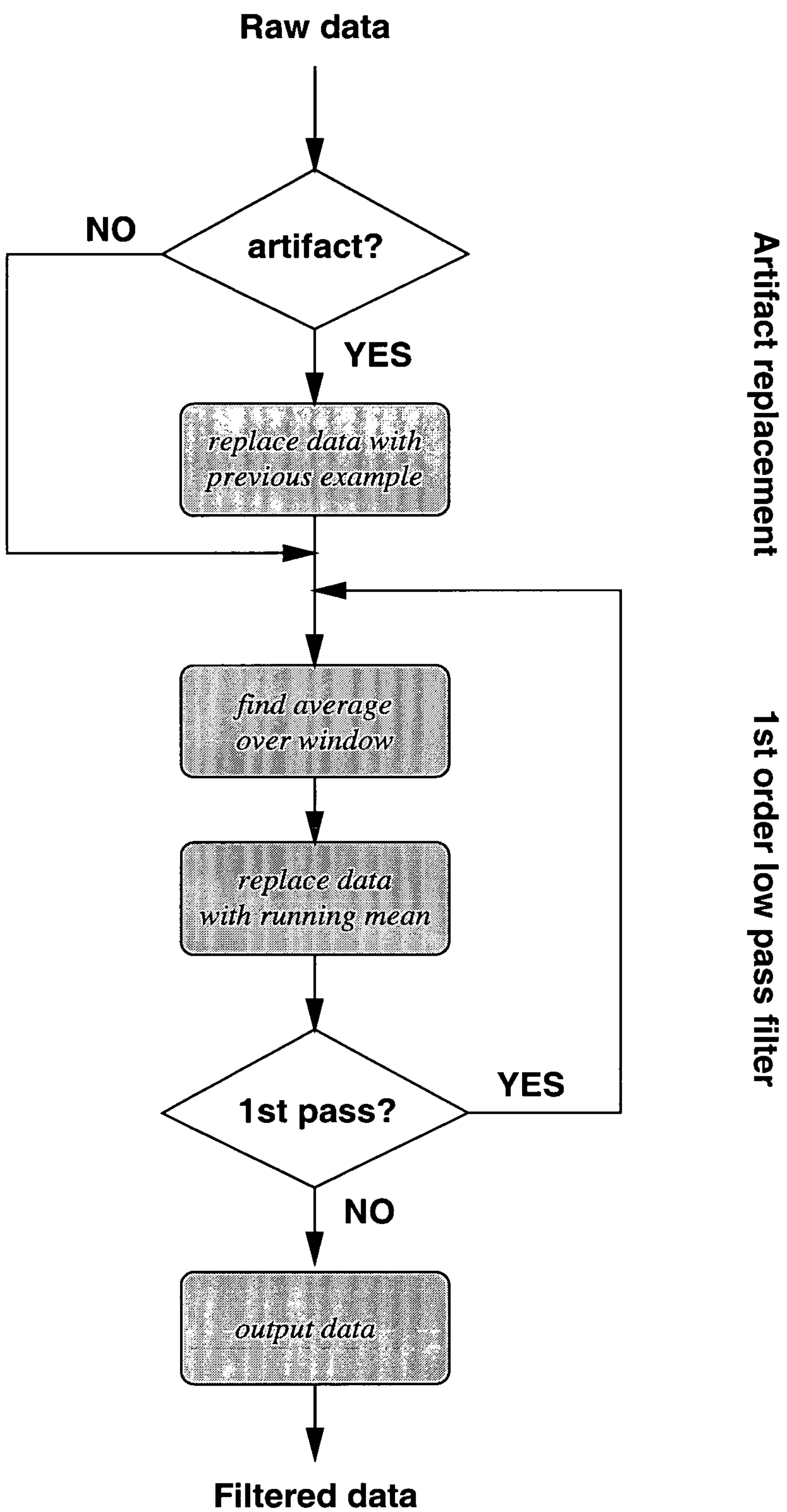


Figure 5–5: Artifact removal and low-pass filtering

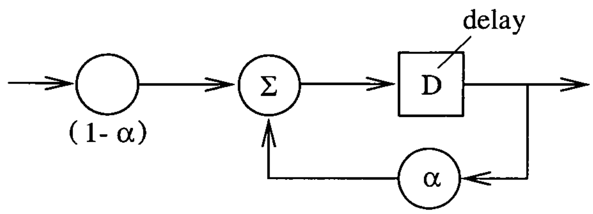


Figure 5–6: 1st order recursive mean estimator, low pass filter

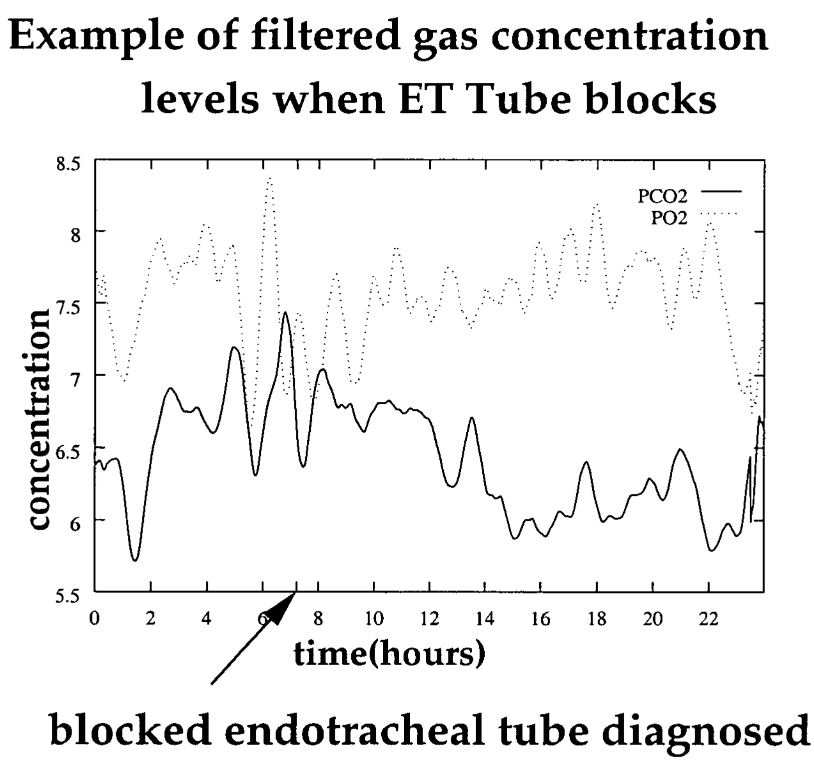


Figure 5–7: Transcutaneous gas levels from one day, one patient data including event

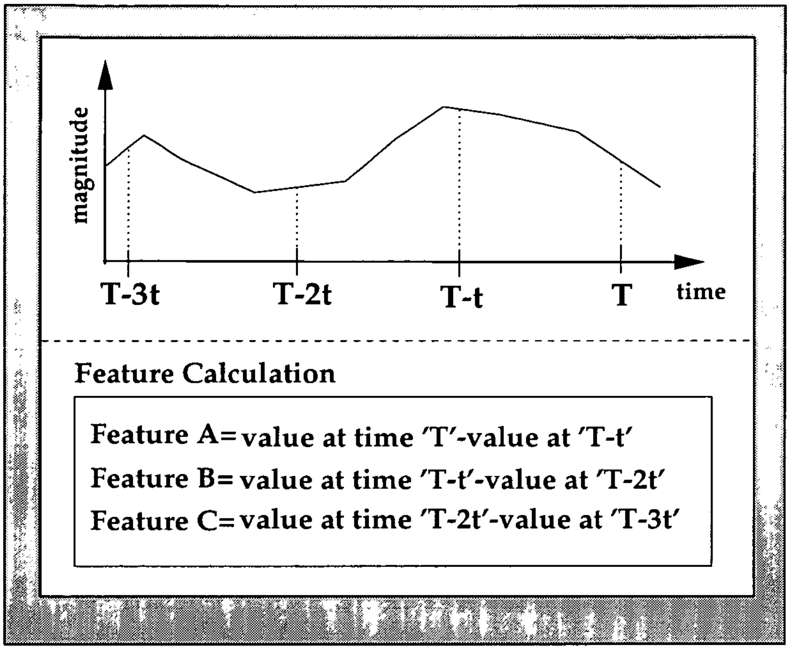


Figure 5–8: Feature calculation using contiguous, non-overlapping, gradients

The window of thirty minutes is subdivided into a number of contiguous smaller windows of equal length (see Figure 5–8). Gradients are calculated between the extremities of each of these small windows. In the example in Figure 5–8 three gradients are generated and these three values are used to describe the data at time T. This process is repeated for each of the channels of physiological data being examined and these features are then combined to form the input vector to the classifier.

The drawback of this technique is that certain types of trend which may presage RD, are not detected. These trends occur where the overall trend of the signal is different from the subtrends extracted, see Figure 5–9 for an example of this. The second technique for feature extraction attempts to take this problem into account.

Overlapping features

In this case the features are extracted in a similar manner to those of non-overlapping features by subdividing the complete window into a series of smaller windows. However, here the smaller windows are nested within the larger period of interest (see Figure 5–10). Gradients are calculated between the time of interest and one, two and three time periods prior to it. These

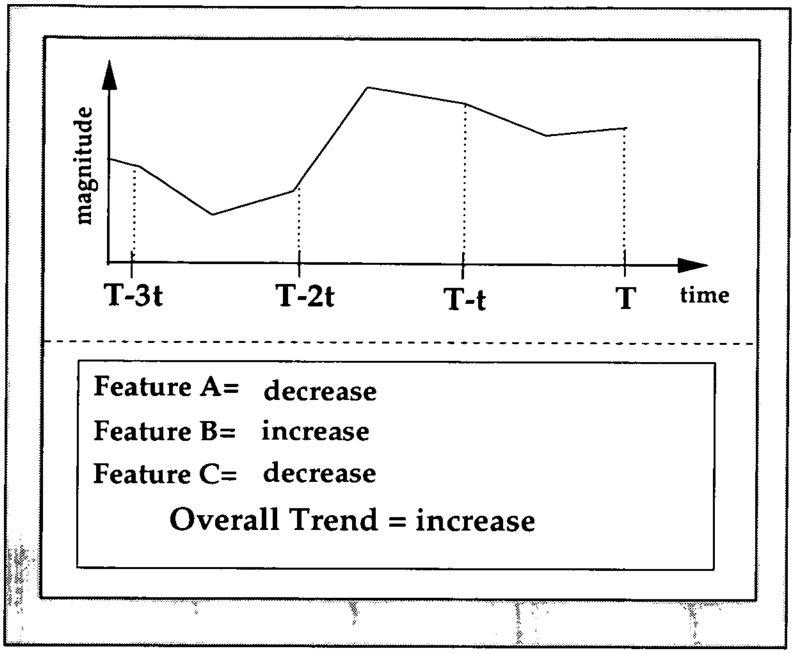


Figure 5–9: Problems with Technique 1

calculations are carried out for all signals and the combination of all gradients is used as the input to the classifier.

5.4 Classification

This stage of the process accepts the features which have been extracted from the selected signals and attempts to draw a decision boundary between the classes of interest. However, before this can take place a database of exemplars of the two classes must be formed.

5.4.1 Development of training and test sets

Using the features which have been extracted from the filtered data a database of training and test patterns is formed. This is done by choosing areas of filtered data as exemplars of periods where concern about the patient increased or where there was none present. The former were isolated using patients' records. Where a diagnosis of Respiratory Disorder was made, the preceding sixty minutes was assumed to indicate the presence of RD. This is a reasonable assumption as it is known that RD takes significant periods of time (for example up to four

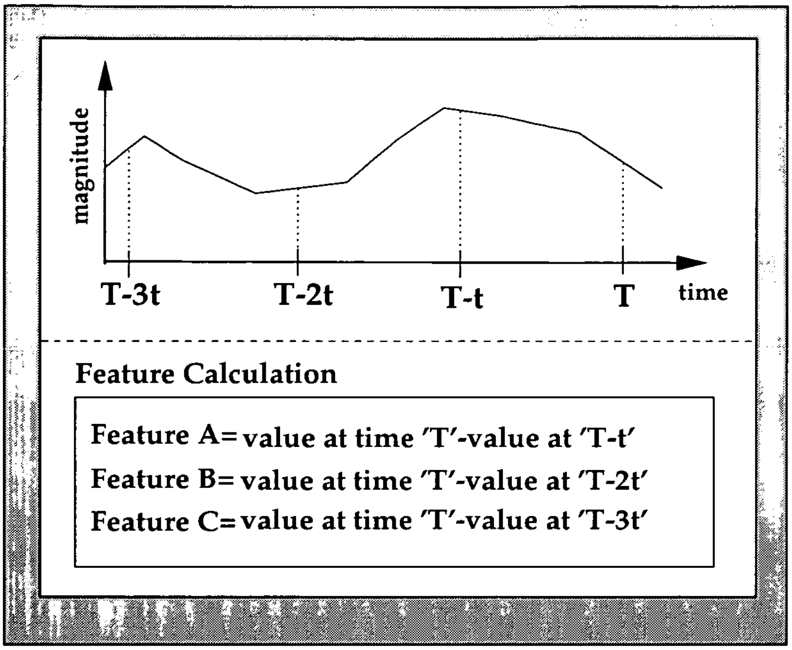


Figure 5–10: Feature calculation using overlapping, nested, gradients

hours [1]) to develop. One hour was chosen as it was felt that that if RD were developing signs would be seen over this shorter time period whereas if a four hour time period were used RD which developed over a shorter period may be missed.

The sections of signal chosen to indicate periods of little or no staff or clinical concern were periods where there had been no comment entered on the patient record for at least two hours. This was assumed to indicate that the patient was in a stable condition and that there was no treatment or activity taking place which was noteworthy.

Both of these assumptions are flawed in that total reliance is placed on the reliability of staff entering comments on to the computer. These problems can be separated into two types:

- Time a comment is entered
- Nature of the comments which are included

The first problem can be described as where a crisis has occurred priority lies with treating the patient. This means that often there is a delay in a clinical comment being entered. This problem has been combatted to a large extent by using the data-points in the sixty minutes

[1]personal communication from Prof N.McIntosh and Dr A.Lyon at Edinburgh NICU

prior to the event as exemplars. The second problem can be described as accurately finding periods of no concern to clinicians. Concern may not be noted although the patient may be placed under greater observation and preventative treatment may take place. The only means of eliminating this problem is to generate data which have been exhaustively annotated. Within the scope of this thesis this has not been possible but it is hoped that at some point it will be attempted as it is the only way to generate accurate period exemplars.

Classification of areas

Once the exemplary periods have been chosen they are classified in terms of the staff concern or clinical diagnosis. There are two classes defined; no concern (class 0) and concern (class 1). For example, where RD occurred during the preceding sixty minutes the event would be classified as class 1. This type of hard classification has its problems as clinical knowledge tells us that diseases do not progress from one state to another with negligible transition time. It is more likely that there is an intermediate state or a series of intermediate states which culminate in the state of highest concern. This type of intermediate state flow is similar to fuzzy class definitions used in some classification problems [15],[91], [34], [92], [93] and it is a technique which merits further investigation.

5.4.2 Single versus multiple patient recognition

In medical problems there is a recurring difficulty: that of the constituents of the data set. It is obvious that patients (and most systems) spend the majority of time in a stable condition requiring no concern. If a data set of training examples was generated from this "realistic" data the training and test sets would be heavily biased towards class 0, the stable condition. A method of combating this must therefore be used. There are several techniques for this which include:

- Train only on class 0. Class 1 vectors in the test set can be examined by looking at their dissimilarity to the class 0 standard. This is known as novelty detection

- Train on both classes and take the probabilities of occurrence into account

- Select exemplars of both class 1 and class 0 in equal proportions

It is the last technique that has been used here. However, one question remains. Should data from one patient or many patients be used for recognition? This application is similar to others [94] in that there are few examples of one class. The decision was therefore made that the

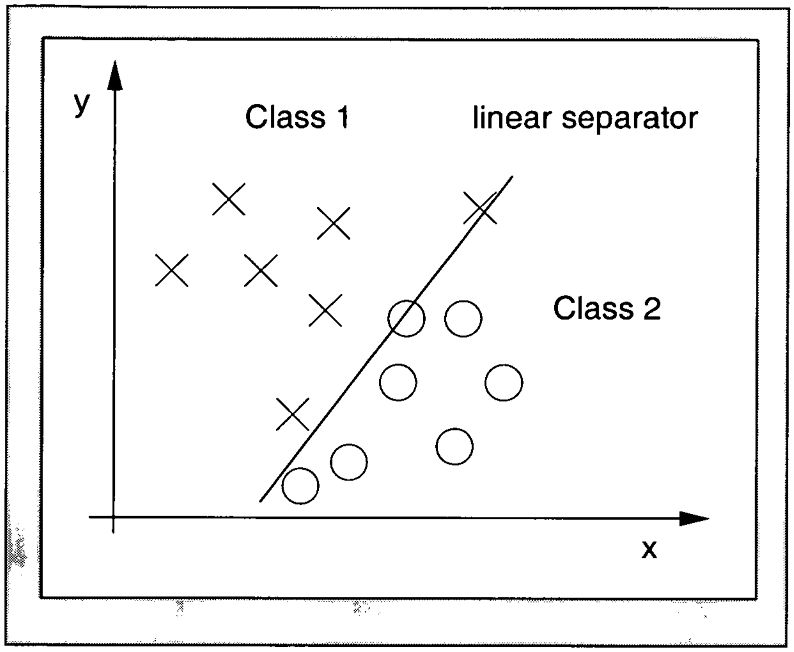


Figure 5–11: Example of linearly separable classes in two class problem.

examples of this class must be generated from a number of patients. There is, however, the opportunity to use examples of the stable class (class 0) from either multiple or single patients. Both techniques have been used here to investigate the feasibility of producing a multiple patient system rather than a single patient one. This is analogous to speech processing where it is easier to recognise one voice rather than a collection of voices [95] and would permit a generic system to be designed. This would mean that there would not be any re-training time required for each new patient who used the system.

Two classification techniques were used to enable comparisons to be made of the performance of the classifier. These were; linear classification and non-linear classification.

5.4.3 Linear Classification

One classification technique is to use a linear classifier which divides the pattern space (the area in which all data point lie) by a straight line or plane. All data points which lie on one side of this line, or plane, are classified as one class and while points on the other side of the line are assumed to be the other class. This type of classifier is used where the different classes being examined are linearly separable. Figure 5–11 provides an example of a two dimensional input, two class problem.

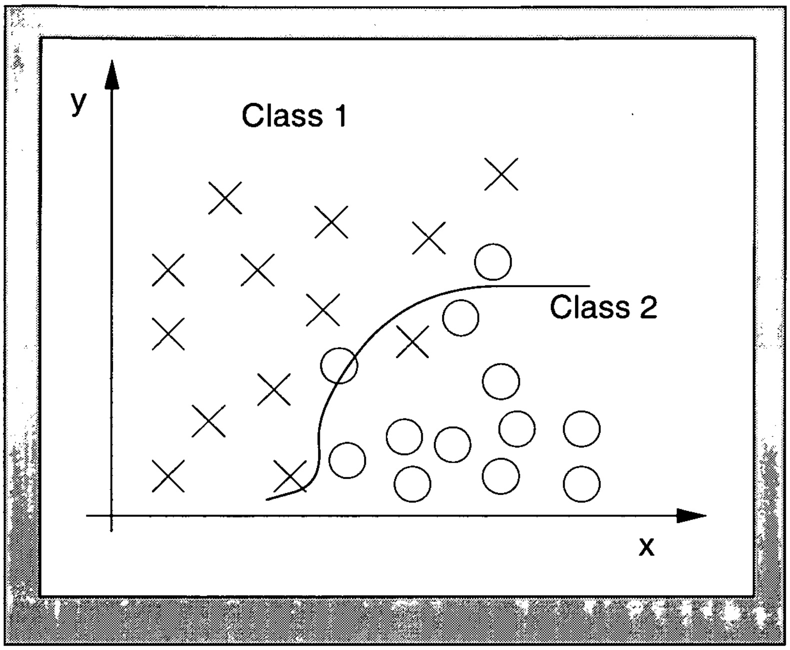


Figure 5–12: Example of nonlinear separation of two classes

However, not all classification problems allow linear boundaries to be drawn between classes and in these instances other classification techniques must be used. These are generally described as non-linear techniques as the decision boundary between classes is no longer a straight line. For an example of this type of problem see Figure 5–12. Medical signals fall into the category of non-linear problems as they cannot be modelled using standard linear methods [18]. One type of non-linear classifier which is used elsewhere in medical signal analysis is the artificial neural network [35] [79] [71].

5.4.4 Artificial Neural Network

Artificial Neural Networks exist in a number of forms; Multi-layer perceptrons (MLPs), Radial Basis Function networks and Kohonen networks amongst others [96],[97],[98]. For this particular application an MLP network was chosen to perform the classification function as they have previously been used for medical time-series classification [99],[72] and for condition monitoring applications [100], [75].

MLPs can model non-linear systems by learning the typical behaviour of the system and generalising when previously unseen input data is presented. They are trained, using supervised learning, to recognise different classes within input data space. MLPs rely on expert knowledge to define the classes of interest. They train using a series of iterative steps in order to minimise an error term. A more complete description of the operation of MLPs can be found in Appendix A.

The optimal architecture of the MLP was chosen by testing a number of networks and altering the number of hidden units until a maximal performance was reached. For each network architecture a series of ten tests was run using different training and test sets. This meant that if the network reached a local minima the effect would be minimised by the other test-runs in the series. The prototyping stage therefore involved the testing of 186 different MLP architectures and 36 different linear discriminant architectures.

5.5 Final system

Returning to Figure 5–3 it is now possible to complete the description of the classification system, (see Figure 5–13). An infinite number of possible classification systems exist however, the design chosen has aimed to:

- Use expert knowledge to enable the system to perform well, although given the assumptions made about the data this performance will not be optimal.
- Test the system on real data taken from the NICU at Edinburgh. This will enable the performance of the system to be evaluated on real clinical data and to permit judgement to be made about the applicability of the system in the NICU in its current form.
- Assess its performance- this evaluation will be carried out using a series of isolated exemplars and comparisons with other approaches, where possible, will be made.
- Permit discussion of its successes and limitations to be made and to enable suggestions for improvements and suggestions for further study to be made.

The system has therefore been designed to carry out prediction of a condition of a patient by using both current data and prior data which have been embedded in input vectors for a classifier. Prior to this, raw data is filtered and gross artifacts removed by using a first order recursive mean estimator. Inherent in this complete system is the expert knowledge of how Respiratory Disorder develops and the characteristics of its precursors in the physiological data examined.

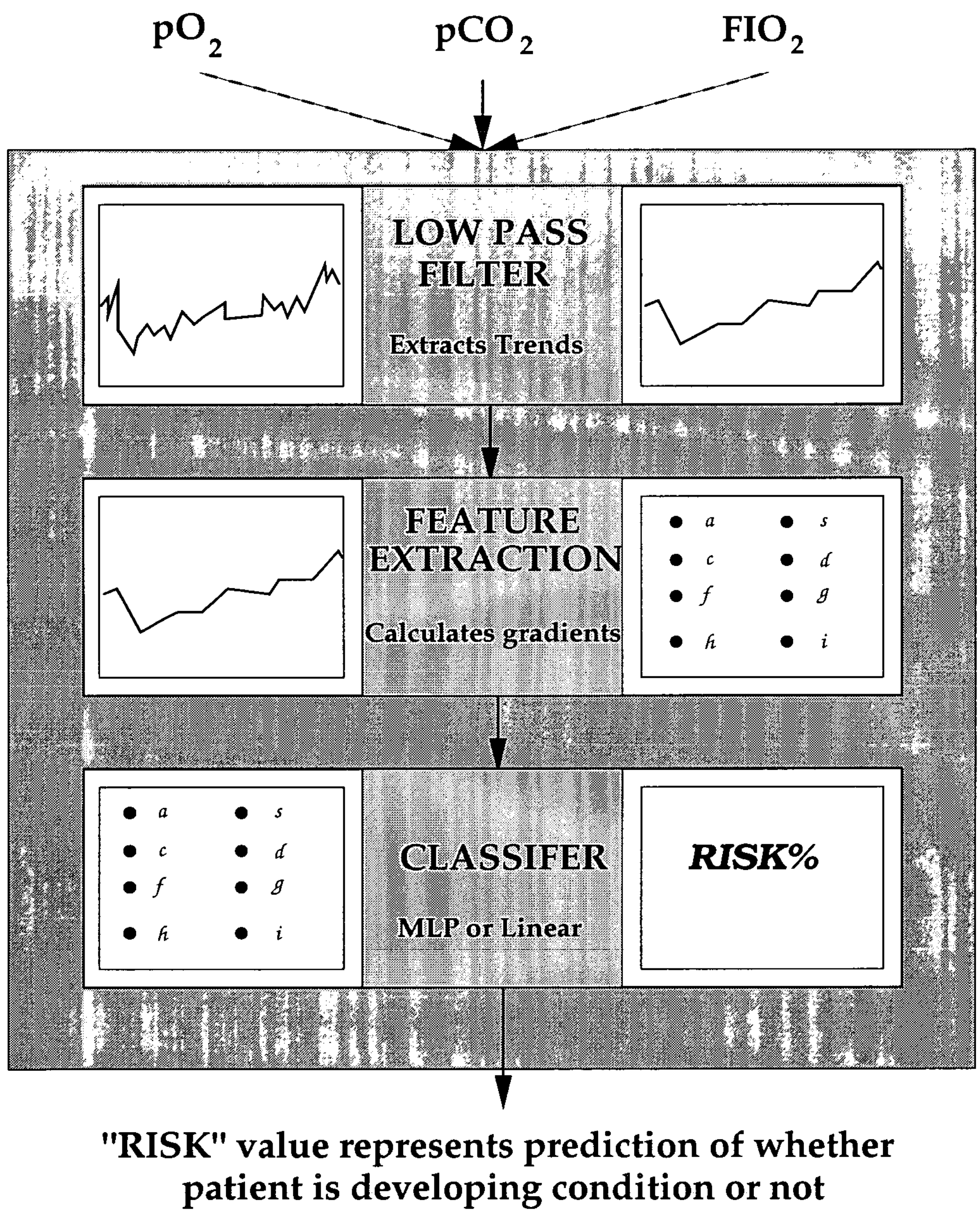


Figure 5–13: A schematic of the final system design

5.6 Conclusions

In this section the methodology of design has been described and the complete system detailed. The methodology has involved using expert knowledge in the early stages of signal processing to enhance the information deemed by clinicians to be vital in the diagnosis of respiratory disorder. Temporal information regarding the clinical history of the patient has been encoded as part of the feature extraction process for the non-linear classifier selected. Techniques which have been used have included linear signal processing techniques in combination with non-linear processing provided by a multi-layer perceptron neural network.

It is stressed that as the end-users of this system will ultimately be clinicians or care staff they must understand or be comfortable accepting the output of the classifier. In this case the method used to achieve this has involved including expert knowledge in the design of the system, not only in the method by which the raw signals are processed but also in the way that the temporal information regarding the history of the patient is incorporated.

Chapter 6

System Results

6.1 Chapter overview

In this chapter results from a number of different experiments will be described and their implications discussed. The results presented here are a subset of the simulation experiments run with the data obtained from *Mary*. In each case the architecture of the network was optimised for best classification performance and it is these results which are used to demonstrate the performance of the different simulations run.

The first section of this chapter will deal with how the performance of a classifier is evaluated. The last section details and discusses results which have been obtained from a number of tests and presents issues which have been investigated. These include;

- The use of Artificial Neural Networks for diagnosis,
- The relevance of different physiological measurements,
- The investigation of different feature extraction techniques,
- An investigation of the value of the inclusion of temporal information,
- An investigation of the methods of generation of training and test sets,
- The use of classifiers on complete days of physiological data.

Each of the tests has been chosen to determine the best techniques for use in this particular application area, and to enable suggestions and proposals to be made which will improve both patient care and classification accuracy.

In medical application areas there is an important issue that arises which is related to the accuracy of the system. That of false positive and false negative classifications. False positives are where an alarm is sounded when there was no problem event and false negatives are where an event has occurred which the system has missed. The latter is potentially extremely important as events may be life threatening and if they are missed patient care suffers and further problems may develop. If too many false positives occur this has a detrimental effect on the staff and visitors as stress levels rise with the number of alarms sounding, and habituation to the alarms.This means that the alarms may be treated casually when there are real problems. It is therefore necessary to monitor false negative and false positive levels and for judgements to be made as to which level the system should operate. Should it have a high false positive rating and potentially detect all events, or should the false positive rating be set low enough to reduce the effect of too many false alarms on staff and visitors?

In each of the cases presented in this chapter the results shown reflect the optimal (in terms of accuracy) architecture for that particular set of tests. The values shown are an average taken over ten runs for each architecture of each network. Approximately six thousand examples are used to generate the training, cross-validation, and test sets. Accuracy, however, is not the only performance measure available to us or of use in a medical application.

6.2 Performance measures

In order to decide on the most suitable network architecture some means of quantifying the performance of the classifier must be used. The most common measure for studying the performance of a two class problem is known as the confusion matrix: this is a 2 x 2 array that summarises the actual and desired output, as illustrated in Figure 6–1.

The confusion matrix provides a guide to the performance of the classifier. The matrix can further be used to calculate a number of different measures which further describe the performance of the system, see Figure 6–2.

The first of these measures is the classification rate or accuracy of the classifier, i.e. how many of the previously unseen examples does it correctly classify. This measure, although significant, does not fully describe the system performance. Other measures are therefore used:

- Accuracy: How many of the previously unseen examples does it classify correctly.
- Sensitivity: In a two class problem this is a percentage measure of how many examples of a particular class (in this case class 1) the classifier correctly identifies.

class	desired 0	desired 1
actual 0	63	13
actual 1	27	87

class	desired 0	desired 1
actual 0	*TN*	*FN*
actual 1	*FP*	*TP*

TP: True positives, how many of class 1 were correctly classified
TN: True negatives, how many of class 0 were correctly classified
FP: False positives, how many of class 0 were classified as class 1
FN: False negatives, how many of class 1 were classified as class 0

Figure 6–1: Confusion matrix for two class problem

$$\text{Accuracy} = \left(\frac{TN+TP}{TN+FN+TP+FP} \times 100 \right) \%$$

$$\text{Sensitivity} = \left(\frac{TP}{TP+FN} \times 100 \right) \%$$

$$\text{Specificity} = \left(\frac{TN}{TN+FP} \times 100 \right) \%$$

$$\text{Selectivity} = \left(\frac{TP}{TP+FP} \times 100 \right) \%$$

Figure 6–2: Performance measures for a network

- Specificity: This is a percentage measure of how many examples of class 0 are correctly identified.

- Selectivity: This is a percentage measure of the ratio of the number of class 1 correctly identified to the number of class 1 decisions made. This gives a measure of the false alarm rating.

Between them these measures describe the performance of a system on any test data which is presented to the network. The measures have different significance during the development and implementation stages of a system. During the development stage the application area is ignored and benchmarks are used to assess performance of the classifier. These benchmarks are accuracy and selectivity i.e. the number of previously unseen data it can correctly classify and the number of false classifications of a particular class. In implementing the system the application area must be considered and the costs associated with the system must be taken into account. A trade-off is made between detecting all the possible problems (high sensitivity), which would be the ideal scenario for the clinicians, and having a high false-alarm rate (low selectivity) which may be unacceptable in the environment. At this stage the classifier can be "tuned" to classify at the optimal performance for the particular application area.

All of these measures have previously been used in the work of Tarassenko [94] where the need for a selectivity measure in medical application areas is discussed. Tarassenko [94] suggests that of all the measures used, selectivity is the most significant, as specificity and sensitivity often disguise the actual operation of the classifier in its early developmental stages.

6.3 Investigation of the use of an ANN Classifier

One of the purposes of this work has been to determine whether a neural network can be used to classify a given medical condition. As explained previously a multi-layer perceptron network was chosen to classify pre-processed data. The performance of this method is compared with that of a simple linear classifier. This allowed conclusions about the type of problem being investigated to be drawn. In Table 6–1 a direct comparison is made between the performance of a linear classifier and that of a multi-layer perceptron classifier. In each case, or feature extraction process, being studied the same data set has been used and the results shown are taken from the best network for this dataset. It can be seen there is an improvement in accuracy in the case of every experiment. Table 6–1 only includes the results of the network where all possible physiological data and multiple patient information has been included. Figure 6–3 shows the improvement in performance for all sets of data investigated. In every case the accuracy of the

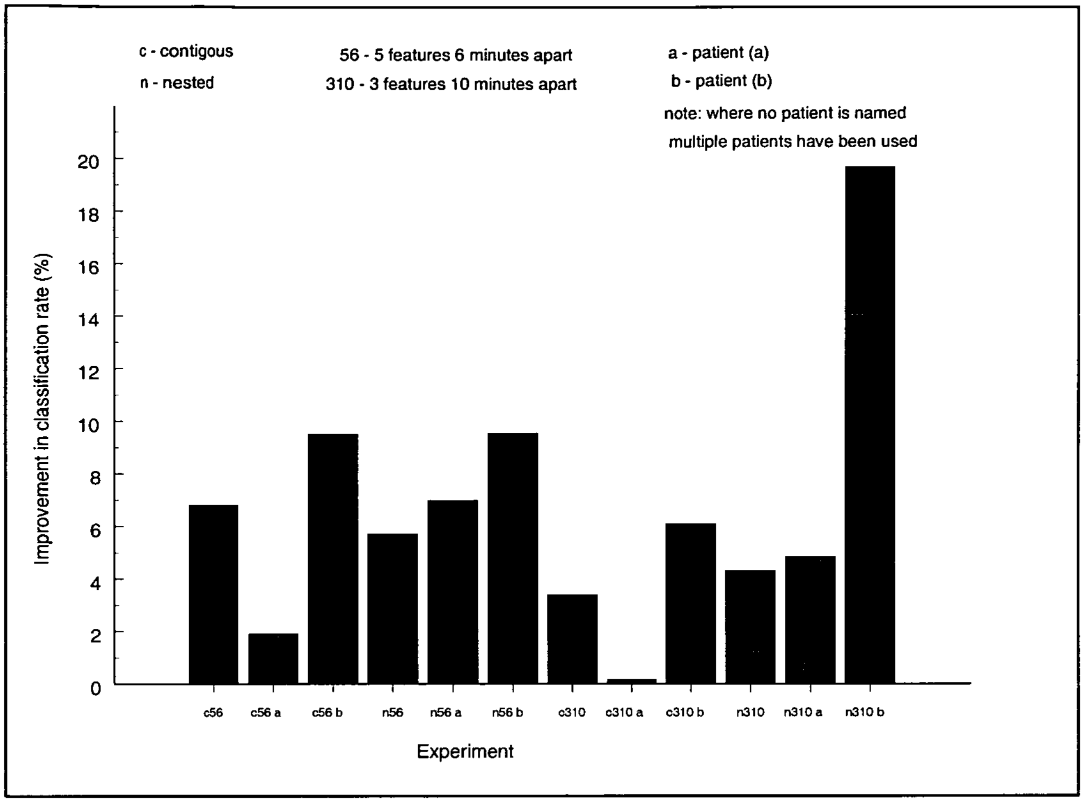


Figure 6–3: Percentage improvement in classification rate when MLP classifier is used in preference to a linear classifier

network was increased, in some cases by as much as 19.62%, the average increase across the tests which were carried out was 6.55% .

Feature extraction process	Linear Classifier	MLP Classifier	Change
Contiguous 5 features 6 minutes apart	63.85	68.18	6.78
Nested 5 features 6 minutes apart	62.17	67.32	8.28
Contiguous 3 features 10 minutes apart	65.65	67.86	3.37
Nested 3 features 10 minutes apart	64.76	69.04	4.28

Table 6–1. MLP classifier accuracy compared to that of a linear classifier (%)

6.3.1 Discussion

It can be seen from both Table 6–1 and Figure 6–3 that the use of the multi-layer perceptron classifier provides an improvement in the overall classification rate of the network. This suggests that the MLP is more suited to this problem. This could be due to a number of reasons:

- That the signals being classified are inherently non-linear and therefore the non-linear MLP classifier is able to model the underlying processes more successfully

- There are too few data available to permit a linear boundary to be drawn between the classes being investigated. This reason is dismissed given that the data sets used contain over 6000 examples.

The decision has therefore been made that, given the number of input vectors used (approximately 6000) to train the classifier, the non-linear classifier should be used. It yields better results for the conditions imposed on it. It is also suggested that, given the evidence stated in the literature by Weigend and Gershenfeld [18] that medical physiological time-series are non-linear, a non-linear classifier must always be used. From the evidence presented in this section the conclusion can be drawn that we now know that the data studied are non-linear.

Bearing this decision in mind it is now possible to further investigate the use of the ANN classifier to determine the onset of respiratory disorder in ventilation assisted neonates.

6.4 Investigation of diagnostic relevance

The current methods of diagnosis in the neonatal intensive care unit (NICU) rely entirely on clinical expertise. In this section the diagnostic relevance of various signals which may presage to respiratory disorder are investigated.

Two measures are commonly used to determine the pulmonary function of the neonate, these are, the levels of carbon dioxide and oxygen. In this section the effect of inclusion of another measure will be investigated. That measure is the fraction of inspired oxygen in the air mixture being applied to the lungs. It is well known that this has a direct effect on the oxygen in the blood [48].

In Figure 6–4 the accuracy of the networks is shown when different combinations of physiological measures are used as input to the system. Groups of three test-runs are presented along the x-axis; the first column in each is where all three physiological signals have been used as inputs to a classifier, the second is where the current standard signals (pCO_2 and pO_2) are used to determine the state of the patient, and in the third instance only signals from the pCO_2 and FiO_2 measures are used.

It can be seen there is a clear pattern in the results (Figure 6–4). In each group of three the central experiment is significantly lower than that of it neighbours. In some cases this reduction in accuracy performance is as great as 7.9 %. However, the question is raised; as to whether this improvement in accuracy is achieved at the expense of some other operating characteristic of the system. In Figures 6–5, 6–6 and 6–7 it can be seen that there is little clear pattern in the effect of the inclusion of FiO_2 on the three other measures of performance. However, Table 6–2 gives more details and shows that on average there is an improvement in performance for all characteristics if FiO_2 is used either in combination with the standard measurements or as a substitute for the oxygen measure currently used.

The percentage changes shown in Table 6–2 were calculated are shown in equations 6.1 and 6.2.

$$FiO_2\,substituting\,for\,pO_2 = \frac{(pCO_2 + FiO_2) - (pCO_2 + pO_2)}{(pCO_2 + pO_2)} \text{ x } 100\% \tag{6.1}$$

$$FiO_2\,in\,addition\,to\,pO_2 = \frac{(pCO_2 + pO_2 + FiO_2) - (pCO_2 + pO_2)}{(pCO_2 + pO_2)} \text{ x } 100\% \tag{6.2}$$

Measure	FiO_2 included	FiO_2 substituted for pO_2
Accuracy	7.9	5.88
Sensitivity	8.84	8.89
Specificity	9.94	9.7
Selectivity	11.27	10.58

Table 6–2. Percentage improvements in the classification rate when FiO_2 is included in the data set

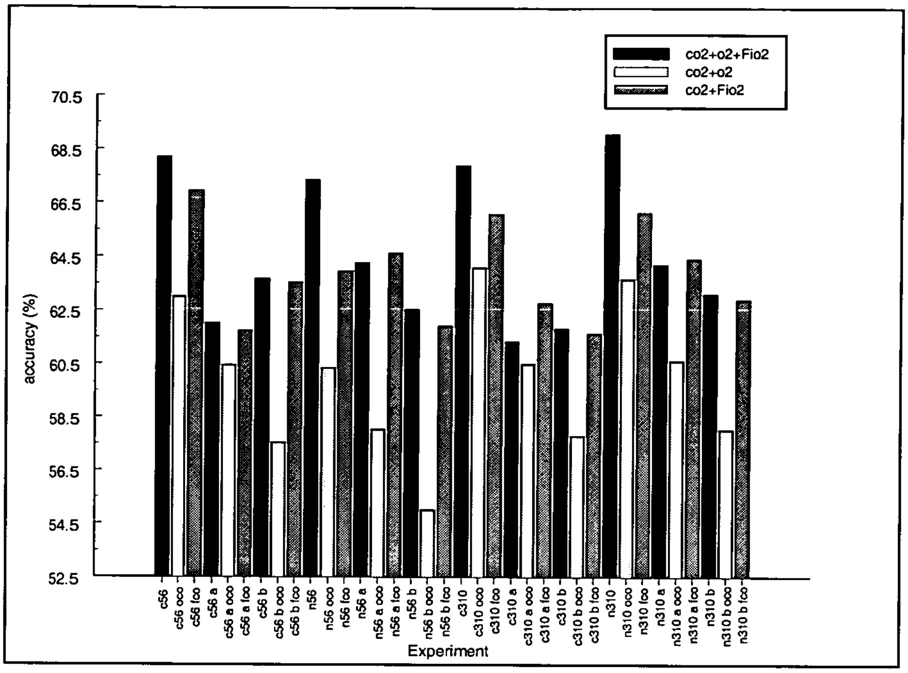


Figure 6–4: Comparison of accuracy rates when different physiological signals are used

6.4.1 Discussion

From the results shown in Table 6–2 it is possible to conclude that the inclusion of the fraction of inspired oxygen in the classification process improves the performance of the classifier. Most noticeably the accuracy and selectivity of the classifier are improved. This tells us that when FiO_2 is used the proportion of correctly predicted events increases and obviously has implications for medical practice. The suggestion can therefore be made that where possible FiO_2 should be included in any decision-making process which relates to the diagnosis of the types of respiratory disorder investigated here. Its inclusion either in combination with the current standard measurements or combined with the carbon dioxide concentration in the blood, has a marked effect on the ability of the decision support device to categorise the event with which it is presented.

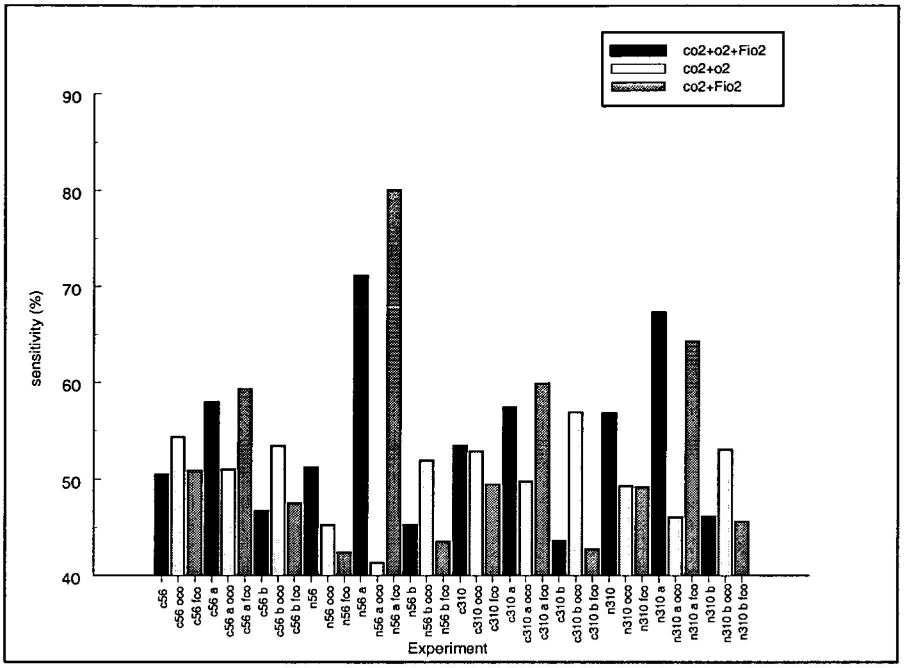


Figure 6–5: Comparison of sensitivity rates when different physiological signals are used

6.5 Investigation of feature extraction techniques

In this section the impact of different feature extraction techniques on the performance of the classifier will be investigated. Ultimately four types were tested. The method for calculating these is shown in Figure 6–8. These can be grouped broadly into two areas of investigation.

- The use of different methods of including temporal information in the extracted features
- The use of different numbers of features and subdividing time intervals used

Two techniques used for the extraction of features were investigated, contiguous (non-overlapping) extraction, and nested (overlapping) extraction. The second technique examined the effect of altering the subdivisions used to extract features from the physiological signal. This investigated how the information was extracted from the data, i.e. is it necessary to maximise long or short term trends in the raw data? The period of time used for inclusion in the feature extraction process was thirty minutes. In one example this was subdivided into five windows of six minutes, and in the other three windows of ten minutes.

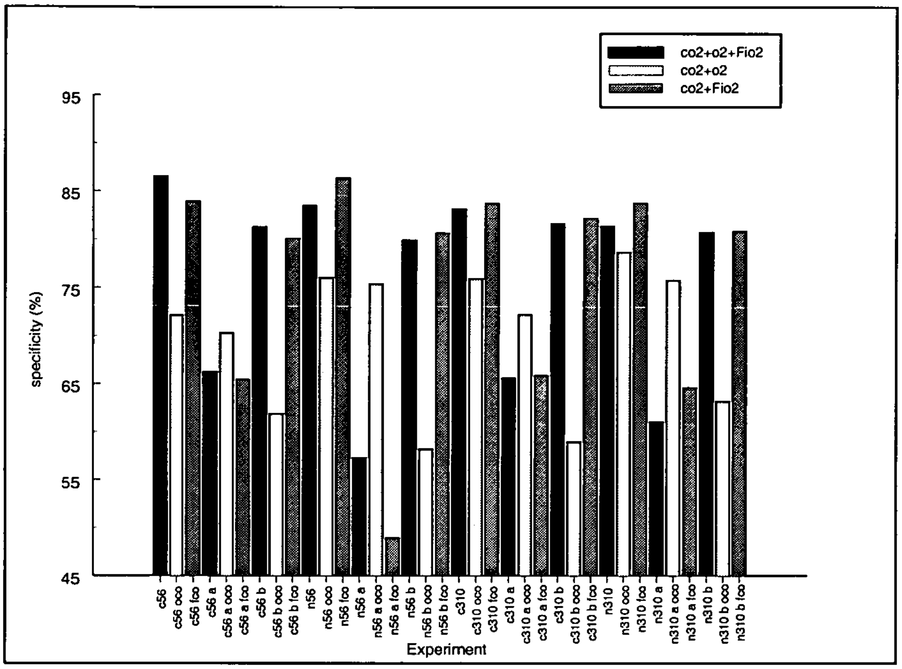


Figure 6–6: Comparison of specificity rates when different physiological signals are used

For each feature extraction technique investigated, tests were run using the same data set for the generation of training and test sets. Results of the best performing network architectures in each of these are summarised in Table 6–3. In Figures 6–9, 6–10, 6–11 and 6–12 a direct comparison is made between the performance of the classifier when using contiguously extracted features, and the performance using the nested features.

Examination of Table 6–3 suggests that there are several conclusions which can be drawn as to which type of feature extraction process should be used. There is no uniform improvement in all performance measures when a particular feature extraction process is used. However, if Table 6–3 is examined with reference to the effect of using nested features rather than contiguous features, only one conclusion can be drawn: Selectivity and specificity are reduced. This means that the false alarm rate increases and may cause problems in terms of increasing noise and stress levels in a unit.

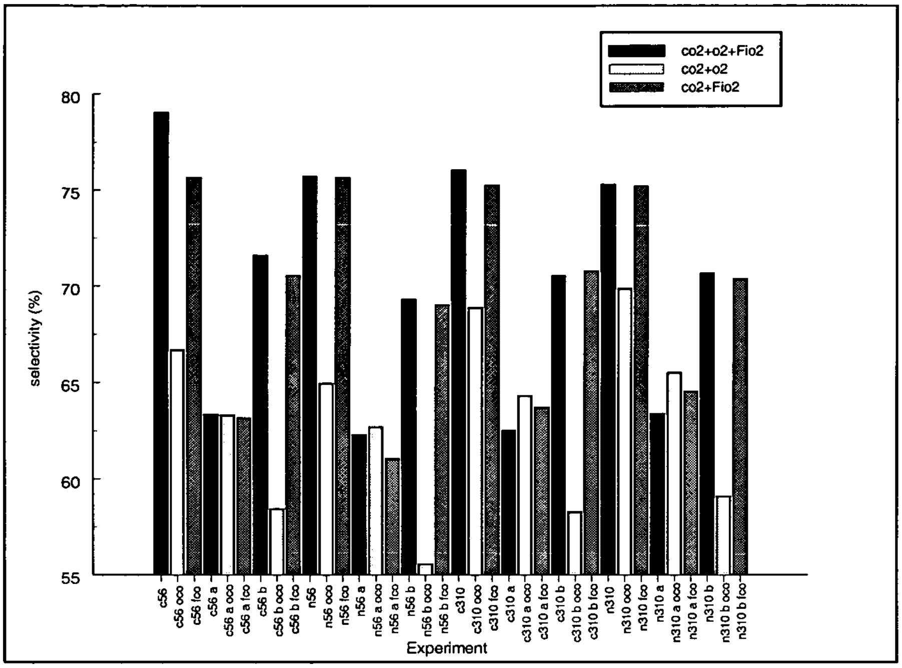


Figure 6–7: Comparison of selectivity rates when different physiological signals are used

Feature extraction	accuracy	sensitivity	specificity	selectivity
c 5 6	68.18	50.49	86.52	79.04
n 5 6	67.32	51.26	83.43	75.70
c 3 10	67.86	53.49	83.12	76.04
n 3 10	69.04	56.86	81.31	75.28

Table 6–3. Summary of performance using different feature extraction processes (%)

KEY: n signifies nested feature extraction and c contiguous feature extraction. The numbers shown denote the number of features (the first value) and what window length is used (the second value). For example n 5 6 = nested extraction, 5 features taken 6 minutes apart.

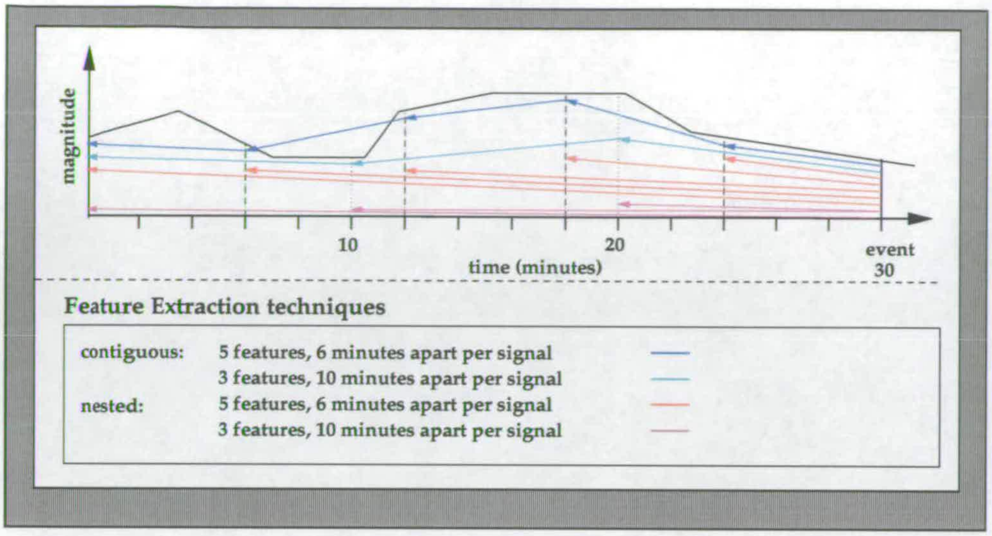


Figure 6–8: Different feature extraction methods

6.5.1 Discussion

Table 6–4 shows details of the percentage improvements in the performance parameters of the system when nested features are used in preference to contiguous features. The values in Table 6–4 have been calculated using equation 6.3 and have involved using the results from the optimal architecture found for the system under these circumstances. It can been seen that altering the method by which temporal information is included has an effect on the performance of the system. Where improvement in the system accuracy is made (for example, in the case where 3 features have been used) this improvement has been achieved by increasing the sensitivity of the system (i.e. its true alarm rate). Where 5 features have been used there is a decrease in accuracy observed when nested features are used. This decrease occurs as the sensitivity of the classifier drops, and hence its true positive rating decreases. Overall, the values shown in Table 6–4 imply that the performance characteristics of the system are not only dependent on how the features are calculated, but also on the number of features used and their relevant time interval.

$$\textit{Percentage change when nested features are used} = \frac{(nested) - (contiguous)}{(contiguous)} \text{ x } 100\% \tag{6.3}$$

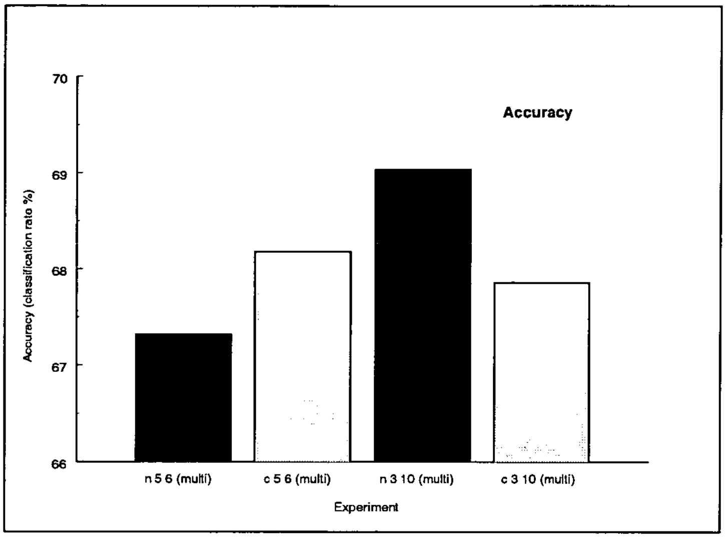


Figure 6–9: Comparison of accuracy rates when features extraction processes are used

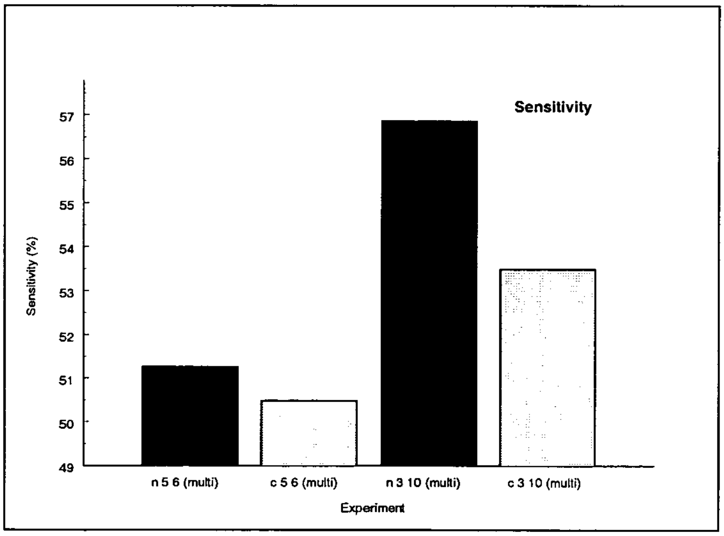


Figure 6–10: Comparison of sensitivity rates when features extraction processes are used

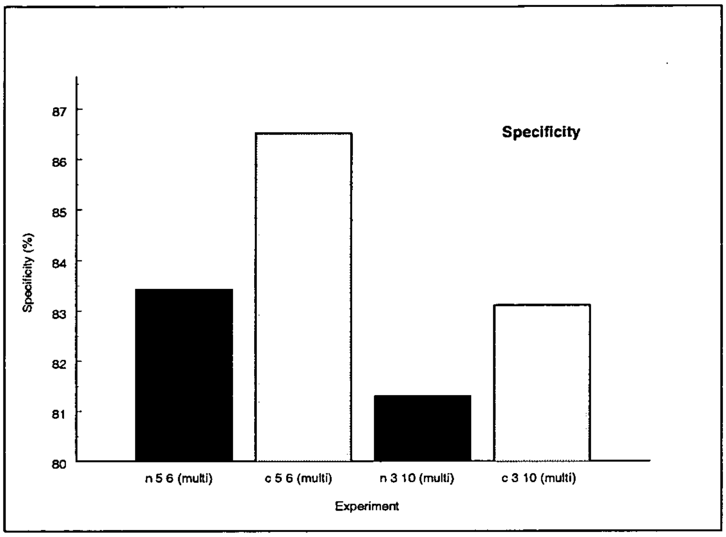


Figure 6–11: Comparison of specificity rates when features extraction processes are used

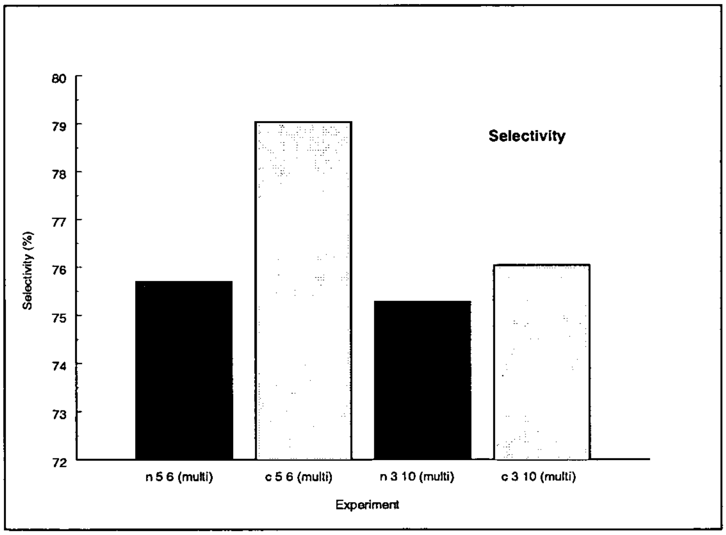


Figure 6–12: Comparison of selectivity rates when features extraction processes are used

Feature extraction	accuracy	sensitivity	specificity	selectivity
5 6	-1.26	1.53	-3.57	-4.23
3 10	1.74	6.3	-2.18	-1.00

Table 6–4. Percentage improvements in performance characteristics when nested extraction is used

6.5.2 Numbers of features and time intervals used

It is now possible to investigate the different time intervals used to extract the features. Table 6–5 and Figures 6–13, 6–14, 6–15 and 6–16 show the performance characteristics of the classifier when direct comparisons are made between the results obtained using the two different types of time interval. The values shown in Table 6–5 are calculated using equation 6.4. This table confirms the idea proposed in the previous section that the efficiency of the feature extraction method is directly related to the number of features extracted and how the temporal information is included. From this conclusion the suggestion can be made that the performance of the system can be optimised (given the techniques used) either by using a nested extraction method, in which case there should be three features extracted ten minutes apart, or by the use of a contiguous method, in which case more features are required (five, six minutes apart).

feature type	accuracy	sensitivity	specificity	selectivity
nested	-2.49	-9.85	2.61	0.56
contiguous	0.47	-5.61	4.09	3.95

Table 6–5. Comparison of percentage improvements in performance characteristics when 5 features are extracted 6 minutes apart versus 3 features 10 minutes apart

$$\textit{Percentage change when 5,6 features are used} = \frac{(5\ 6) - (3\ 10)}{(3\ 10)} \text{ x } 100\% \tag{6.4}$$

6.5.3 Discussion

From these results it is possible to draw some tentative conclusions and to suggest the optimal methods of extracting features for this particular application. The results suggest that there is little to be gained from either using contiguous or nested feature extraction or from increasing the number of features (and their associated time interval) used. However, they also suggest that the choice of the method of including temporal information is linked to the choice of the

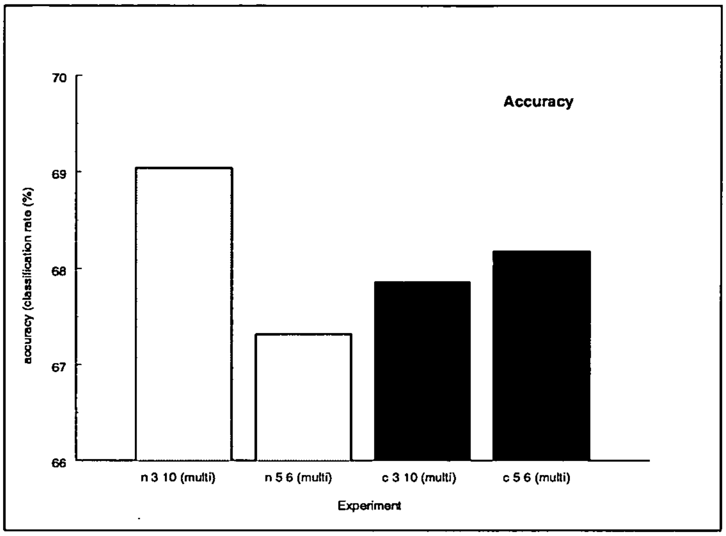


Figure 6–13: Comparison of accuracy rates when different time intervals are used

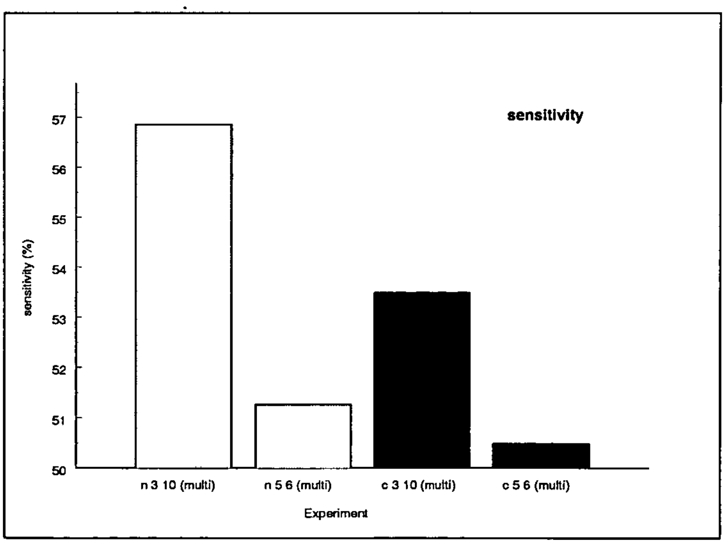


Figure 6–14: Comparison of sensitivity rates when different time intervals are used

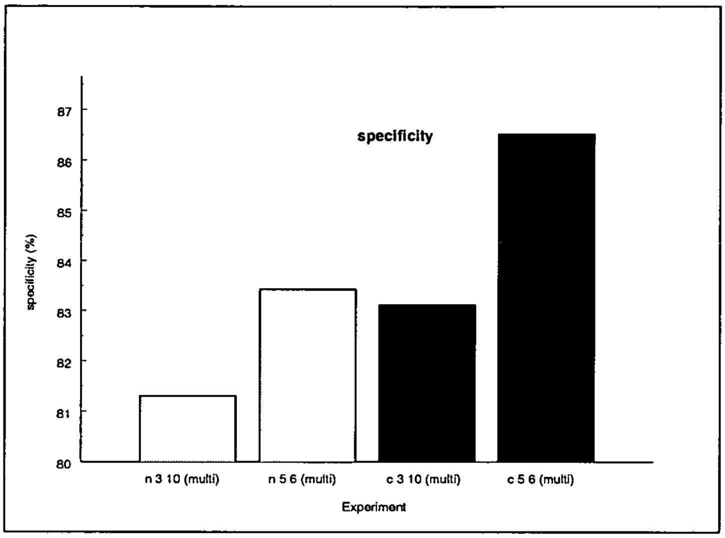


Figure 6–15: Comparison of specificity rates when different time intervals are used

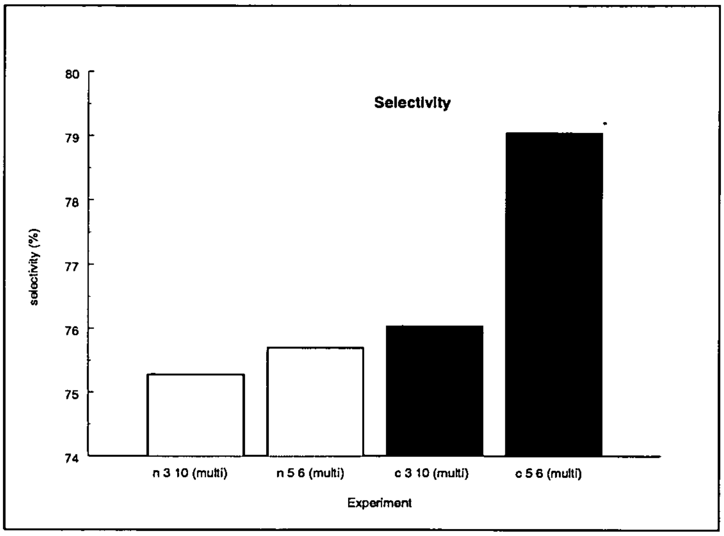


Figure 6–16: Comparison of selectivity rates when different time intervals are used

number of features (and associated time interval) and therefore one should be used to determine the other. Thus:

- If nested features are to be used then three features should be extracted at ten minute intervals within the time window
- If contiguous features are to be used then five features should be extracted at six minute intervals within the time window

Why does this link occur? A possible explanation of this may be that the two feature extraction processes used are designed to maximise different information in the signal. In the case of the nested features the information captured is that of the long-term trend in the signal, and therefore the longer time interval performs better in this instance. Contiguous feature extraction attempts to capture shorter trends in the signal and to describe the complete signal rather than its long term trend, and, therefore the shorter time interval is more effective. It may be helpful to think of the longer time interval as under-sampling the signal and, in the nested case, the shorter time interval as over complicating the required information.

This result is interesting in that it shows that two linear signal processing techniques, which effectively perform the same operation (see equation 6.5), yield different results. This may be because the extraction process and its associated number of features maximises different types of information stored within the signal. For example if a nested approach is used it extracts long term trends and therefore fewer features are required if they are taken over a sufficiently long period. These results therefore detail experiments which were carried out using a selection of feature extraction processes while varying the inputs applied to the system.

$$\begin{aligned} F_{n1} = f(T-t), F_{n2} = f(T-2t), F_{n3} = (T-3t) \\ F_{c1} = f(T-t), F_{c2} = f(T-t) - f(T-2t), F_{c3} = f(T-2t) - f(T-3t) \\ F_{c1} = F_{n1}, F_{c2} = F_{n1} - F_{n2}, F_{c3} = F_{n2} - F_{n3} \end{aligned} \tag{6.5}$$

where F_{n1} 1st nested feature, F_{c1} 1st contiguous feature.

6.6 Investigation of the use of single or multiple patient data

The type of problem under investigation is analogous to that of speaker recognition, in that it is necessary to determine whether a system which has been trained on a single patient than one which has been trained on multiple patients is more accurate. In medical applications there is an added problem in that there are plentiful data describing the "normal" behaviour of patients as he/she spend most of their time in this condition. There are, however, few data describing abnormal behaviour, as this is usually associated with crisis events.

For this particular application a study has been made of the effect that training on the normal behaviour of a single patient has in combination with the abnormal behaviour of a collection of patients and comparing this with the accuracy of the system when data from a collection of patients is used to describe the entire system.

Figure 6–17 shows that in the first instance the classifier would learn normal behaviour of a specific patient and a generalised case for abnormal behaviour derived from the examples taken from multiple patients. It would therefore be expected that the output of the classifier would yield a much higher specificity (true negative) rate than sensitivity (true positive) rate. When a generalised case has been used to describe both "normal" and abnormal behaviour it would be expected that the two performance values would be more evenly matched. Therefore, a series of experiments was carried out which would evaluate the performance of the classifier, given the method used to generate its training and test sets. These methods were as follows;

- to train and test the system on the normal behaviour of a single patient combined with the abnormal behaviour of multiple patients

- to train and test the system on the normal **and** abnormal behaviour of multiple patients

In each case the training and tests sets included equal numbers of examples of normals and abnormals. Therefore, there were approximately three thousand examples of each type of behaviour included in the data set.

Table 6–6 shows the results from the tests comparing these two training approaches. It can be seen from this that using the multiple patient approach consistently yields better classification results than procedures where the system has been trained on the normal activity of a single patient. Figures 6–18, 6–19, 6–20 and 6–21 show these results in graphical form. It can be seen from these graphs that the only conclusive results are those of accuracy and selectivity. These imply that the performance of the system is better when examples from multiple patients are used to generate the model of normal behaviour. This improvement is achieved by increasing the number of alarms being triggered. The actual percentage changes found when moving from a single patient to a multiple patient system are shown in Table 6–7. The values in this table have been calculated using equation 6.6. It can be seen from Table 6–7 that there is a decrease in almost all performance characteristics when moving between a training data set which has been derived from a number of patients and that which has had its normal set derived from a single patient.

$$change = \frac{\textit{(single patient value-multiple patient value)}}{\textit{(multiple patient value)}} \text{x}100\% \tag{6.6}$$

A single patient value is the performance produced where the normal behaviour of a single patient has been used in the model and where the multiple patient value is the result produced when multiple patients have been used to generate both normal and abnormals behaviour values.

Test (patient)	accuracy	sensitivity	specificity	selectivity
c 5 6 (multiple patient)	68.18	50.49	86.52	79.04
c 5 6 (patient a)	62.01	57.98	66.23	63.35
c 5 6 (patient b)	63.66	46.71	81.26	71.58
n 5 6 (multiple patient)	67.32	51.26	83.43	75.70
n 5 6 (patient a)	64.26	71.16	57.3	62.30
n 5 6 (patient b)	62.52	45.28	79.87	69.29
c 3 10 (multiple patient)	67.86	53.49	83.12	76.04
c 3 10 (patient a)	61.33	57.46	65.58	62.52
c 3 10 (patient b)	61.81	43.62	81.61	70.53
n 3 10 (multiple patient)	69.04	56.86	81.31	75.28
n 3 10 (patient a)	64.19	67.30	61.01	63.39
n 3 10 (patient b)	63.09	46.17	80.68	70.68

Table 6–6. Table comparing performance under single or multiple patient training (%)

6.6.1 Discussion

Conditions for the generation of training and data sets can now be proposed and these include a number of criteria.

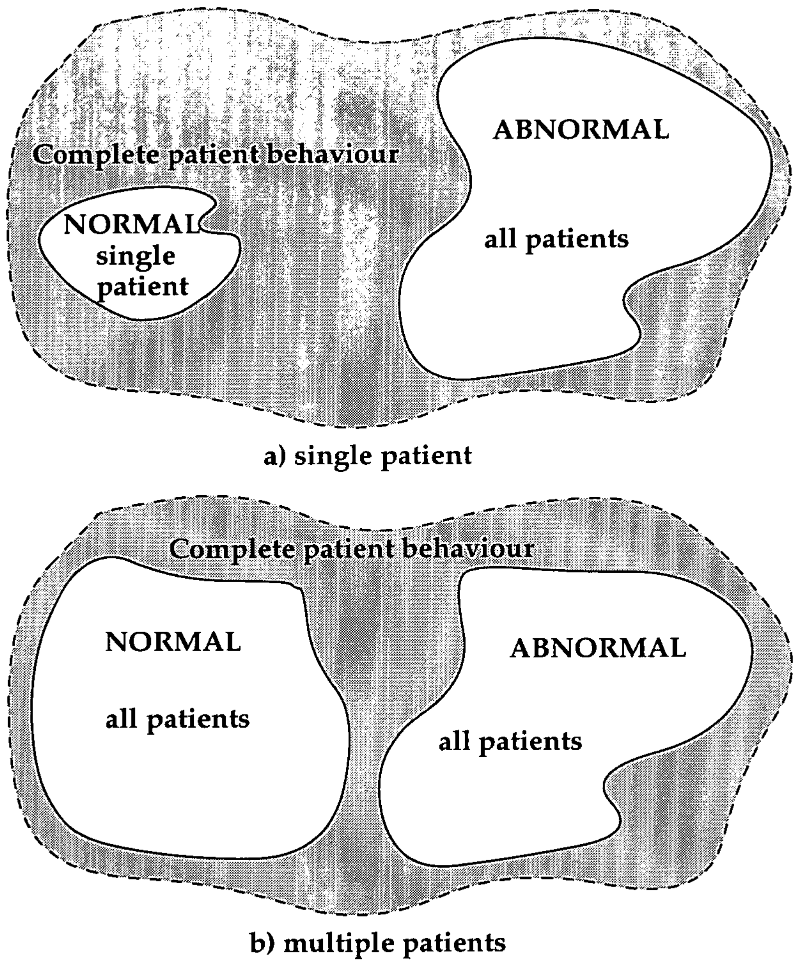


Figure 6–17: Set diagram of patient descriptions used

Feature extraction	Measure	multiple > (a)	multiple > (b)
c 5 6	Accuracy	-9.05	-6.63
c 5 6	Sensitivity	14.83	-7.49
c 5 6	Specificity	-23.45	-6.08
c 5 6	Selectivity	-19.85	-9.44
n 5 6	Accuracy	-4.55	-7.13
n 5 6	Sensitivity	38.82	-11.62
n 5 6	Specificity	-31.32	-4.28
n 5 6	Selectivity	-17.70	-8.47
c 3 10	Accuracy	-9.62	-8.92
c 3 10	Sensitivity	7.42	-18.80
c 3 10	Specificity	-21.10	-1.82
c 3 10	Selectivity	-17.78	-7.25
n 3 10	Accuracy	-7.02	-8.62
n 3 10	Sensitivity	18.36	-18.8
n 3 10	Specificity	-24.97	-0.77
n 3 10	Selectivity	-15.79	-6.11

Table 6–7. Percentage differences between single and multiple patient data for accuracy, sensitivity, specificity and selectivity.

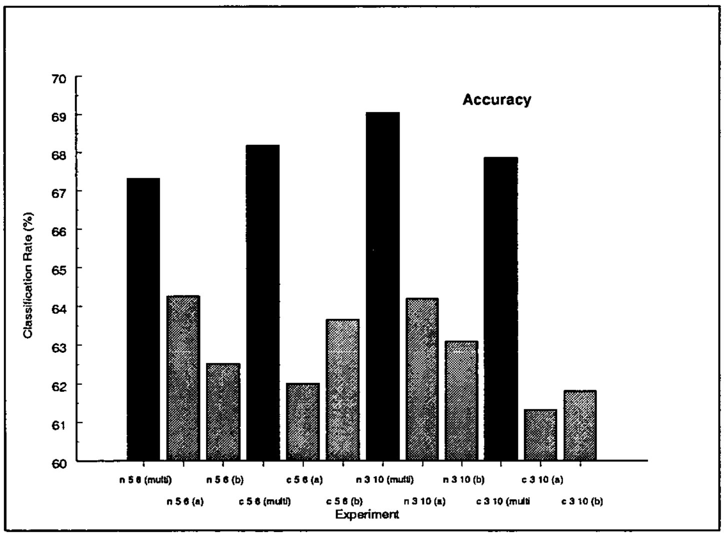


Figure 6–18: Comparison of accuracy results for single and multiple patient training sets

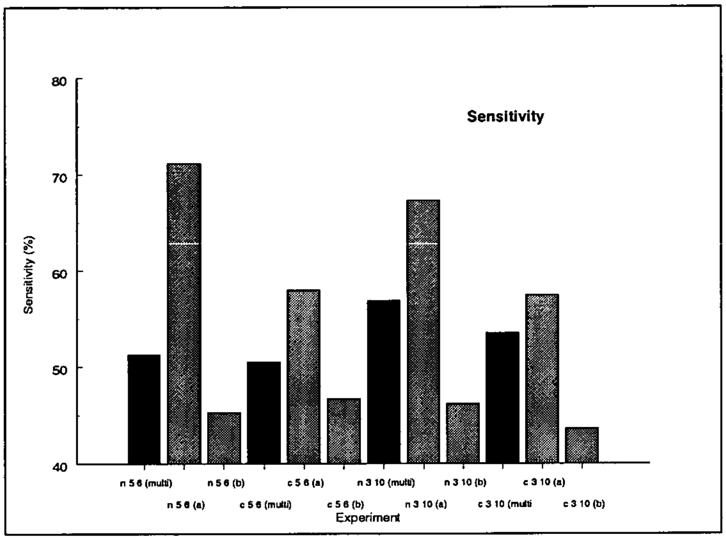


Figure 6–19: Comparison of sensitivity results for single and multiple patient training sets

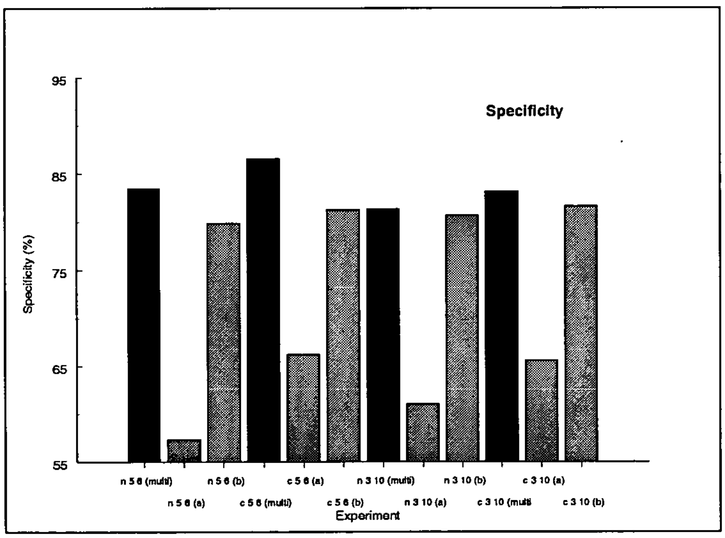


Figure 6–20: Comparison of specificity results for single and multiple patient training sets

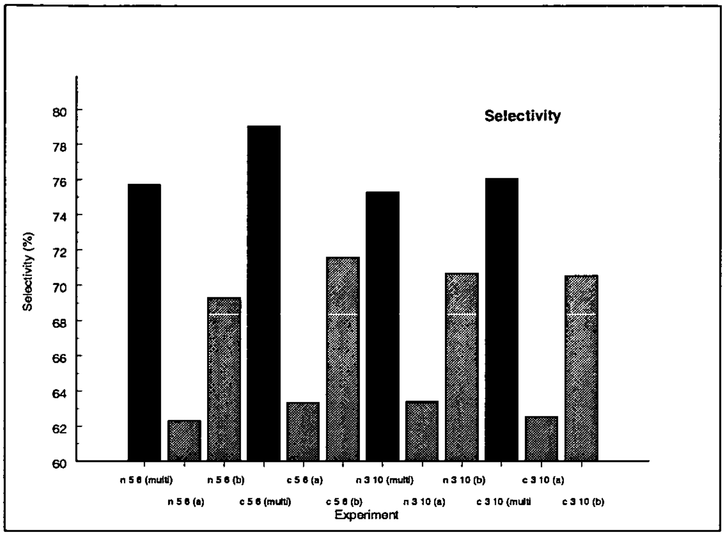


Figure 6–21: Comparison of selectivity results for single and multiple patient training sets

- Exemplars of events must be taken from as large a selection of patients as possible. This means that the system which the ANN models will be as generalised as possible

- The exemplars must be taken from areas in patient data files where it is known that an event occurred, or where clinicians expressed or felt no concern about the patient. Given the current monitoring system, this is not always possible in the latter case as clinicians are not expected to enter that the behaviour of a patient who evokes no concern.

- Exemplars of events and normal behaviour should be taken from distinctly different areas of physiological data as, if an area which is assumed to be normal represents the initial stage of developing RD, the discriminant ability of the classifier will be reduced.

- As many exemplars of both event and no-concern should be included, the model will be as general as possible.

The experiments which were run to investigate the effect of learning the behaviour of a specific patient were heavily constrained by the data available. Of the data which had been collected only two patient's records contained sufficient numbers of "no-concern" examples to permit a balanced data set of training and test examples to be formed. These are patients a and b in the examples.

It should be noted that the data from patient (a) used to generate a training set of normals used thirty-one days of physiological data, whereas that from patient (b) used twenty-five. Some of these days may also have contained exemplars of events. The set of "no-concern" examples derived from multiple patients was generated from twenty-one patients on fifty-one different days, (i.e. no two examples came from the same patient on the same day). This discrepancy between the ideal case and the actual contents of the training and test sets may have affected the results. This is particularly true in the case of the single patient data as it is in these cases that two exemplars may be found on the same day. This may slightly affect some of the performance characteristics of the classifier, in particular the specificity rating may not be as high as it might have been as the area of stable behaviour will be less distinct.

Table 6–7 shows that there is a reduction in system performance when only a single patient is used to generate the training and test set regions of normality. This test is not complete as, ideally, training and test sets would be generated from regions of normal and unusual behaviour of a single patient but this is not possible given the fact that patients will not spend equal amounts of time experiencing Respiratory Disorder and behaviour which is typical of normal behaviour. The accuracy results from these tests can be explained by referring to Figures 6–17a) and 6–17b), the improvement in accuracy is achieved by producing a more generalised model of "normal" behaviour for a set of patients. This result is interesting as it suggests that a generic system can be designed which can use approximations of both "normal" and "abnormal" behaviour as its template for training. This removes the need for training time for the system to "adapt" itself to the particular patient on which it is to be used.

Given the circumstances under which this system has been developed which are:

- No patient experiences RD for 50% of the time and normal behaviour for the other 50%
- Other physiological conditions are ignored which may affect the physiological data being measured
- No data from a single patient will ever contain enough examples of abnormal behaviour for a single patient model to be developed.

It should however be noted that the results produced may be due largely to the fact that it was not possible to generate complete training and test sets from a single patient. Given the application area, and how different the ideal case for generating training and test data is from the actual case, this will never be possible. The conclusion can therefore be drawn that it is not necessary to attempt to model the behaviour of a single patient as the model can never be complete. It is more sensible to generate a model of the general behaviour of multiple patients. This means that a generic system can be produced for the application area of neonatal respiratory monitoring.

To summarise, the results presented so far suggest that a system can be built which will be able to classify approximately 70% of the cases presented to it and that the data should be presented to it, in a certain way.

- The training and test sets should include areas of normal and abnormal behaviour from a selection of patients
- There are two possible combinations of feature extraction and temporal information inclusion possible
  - Nested features taken (three of them) ten minutes apart
  - Contiguous features taken (five of them) six minutes apart
- The data used should include FiO_2

However, before conclusions about the predictive quality and overall performance of the system can be drawn, the behaviour of the system must be evaluated when it is tested on sections of data which have not been preselected as demonstrating a particular event or behaviour type.

6.7 Testing of classifier on real-time, continuous events

Having now looked at training based on isolated intervals of patient history real-time operation must be discussed. The application area being dealt with in this instance is that of real-time physiological data taken from a ventilation-assisted neonate. For the training and testing of the classifier, sections of this physiological dataset have been selected as exemplars of patient behaviour at certain times. These exemplars do not completely describe the system that is the neonate (see Figure 6–22) and therefore the system must also be tested on complete days of physiological data and the output of the classifier compared with the annotation added by clinicians to the archived data. This section will examine the output of the classifier, that has been suggested as being the most accurate, on complete days of data.

The classifier has been trained as before, where exemplars of events have been used to generate the test and training sets. When training of the classifier was complete a number of complete days of data were applied (twelve). Although the results and conclusions are drawn from twelve patients over twelve different days, four have been chosen for the purposes of illustrating the behaviour of the classifier under these circumstances. They are broadly representative of the results obtained from all the days which were tested. Two of the chosen days (patients a and

b) contain known physiological events and areas which, for the purposes of developing the training and test sets, were taken to be exemplars of areas where no problem was anticipated. The third and fourth days (patients c and d) are where no event occurred and no anticipated problems were entered in the annotations. Initially results from patients **a**, **b** and **c** will be used to illustrate the behaviour of the classifier and patient **d** will be discussed later on for different reasons.

The conditions which have been imposed on the classifier are:

- The training and test sets are taken from multiple patients
- The feature extraction process has used either
  - nested features (three at ten minute intervals), or
  - contiguous features (five at six minute intervals)
- All possible physiological signals are used, i.e. pCO_2, pO_2 and FiO_2

It should be noted that the first section (approximately thirty minutes) of the output of the classifier should be ignored. This is because errors are introduced as part of the feature extraction process. Unless data from the day examined is concatenated with that from the previous day the feature extraction process has no data on which to operate. This is true until thirty minutes into the data file.

Figures 6–23, 6–24 show the outputs of the two classifiers when the physiological data from the different days are processed. The data sets from other patients/days also produced similar results. Figure 6–26 shows the annotations entered on the selected days. In Figures 6–23 and 6–24 the x-axis has been annotated (with for example "R" which corresponds to a reintubation event) and each "tick" corresponds to an annotation which has been included in the chart. It can be seen that in both cases there is a significant peak in the classifier output when the patient was reintubated and therefore both feature extraction techniques must be extracting information which is of importance in the diagnosis of these problems.

It should also be noted that the peaks which occur start to develop at least thirty minutes (denoted by shading) before the annotation is added to the record. This means that the system is giving an early warning of at least thirty minutes of the need for reintubation or for greater attention to be paid to the respiratory function of this patient.

Between the two systems investigated there seems to be little difference in the physical location of the peaks which appear. However, it seems that the output from the classifier which has been trained using contiguous features is less stable, there is greater variation in the output, than that of the nested features classifier. This corresponds to the fact that the nested feature classifier maximises the long term trend information and would therefore be expected to be slightly more stable.

There are other peaks included in the output of the classifier. In particular in patients **a** and **c**. In patient **a** there are three significant peaks in the output of the classifier, the first of these corresponds to the patient being reintubated. The second and third are less easy to explain. In the first instance the peak correlates to the patient being treated to all care. This process can sometimes include endotracheal suction and therefore the system may be classifying signs which are typical of a blocking tube that the clinicians have not detected. The third peak in the trace of this patient follows a heel stab. This procedure is relatively stressful and may force the patient to change his or her behaviour. In patient **b** the only significant peak corresponds directly to a respiratory event which occurred **and** which culminated in the patient being reintubated. Patient **c** is an example where no significant event has been entered on the patient records. It would be expected here that there would be no significant peaks in the output of the classifier. However, this is obviously not the case. The peaks which are generated are of short duration and cannot easily be explained. At this point it should be noted that annotations entered are not complete and are often entered later than the actual diagnosis. In some cases clinicians attempt to combat this by entering their estimated time of diagnosis but this procedure is obviously open to error.

It may be that not all of the information needed which can completely describe the precursors to respiratory disorder are extracted by the processes used here. It may be that the feature extraction processes manage to approximate this information and therefore other sections of the signals may show similar properties to those areas which are of interest. If the two graphs for the classifier output are superimposed (see Figure 6–25) it can be seen that there is little variation between the two on days where the patient has experienced difficulties (patients a and b). However, in the final case (patient c) it can be seen that there is greater variation between the two traces and therefore if both networks were used (for example the outputs "anded" together), it may be possible to remove some of the instability from this final trace. The final trace also has greater short time scale fluctuation in it than the previous two and it may be possible to combat this by including greater temporal information in the feature extraction and classification processes, for example, in the use of Hidden Markov Models on the output of the classifier (see Appendix B). It may also be possible to threshold the output in an attempt to remove some of the noise which appears around the zero line. In particular, this would clean up the trace produced for patient c. Patient c has no annotation in its record indicating concern as to its respiratory function, and this suggests that the output of the classifier should be stable. However, on examining the raw physiological data for patient c (see Figure 6–27) it can be

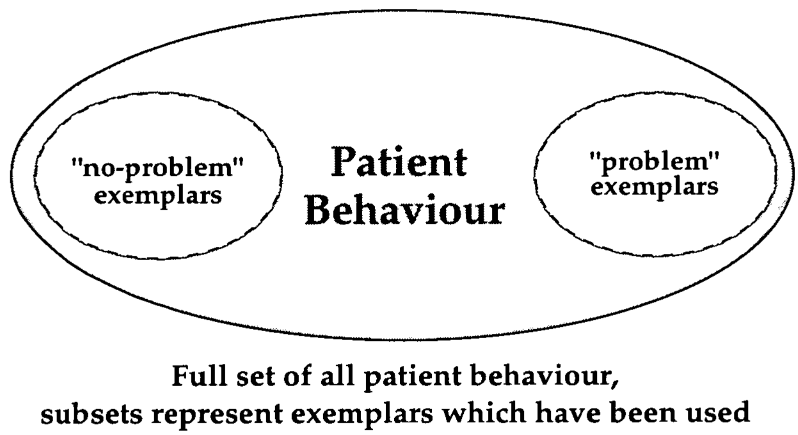


Figure 6–22: The 'patient' system description

interpreted that clinicians and staff were concerned about the behaviour of the patient as the fraction of inspired oxygen in the air mixture has been increased significantly over the twenty-four hour period. This demonstrates the need for accurate annotations and patient records to be kept, and partially explains the instability in the output of the classifier when real-time continuous data from patient c are applied. A more accurate demonstration of the behaviour of the classifier on data which it is believed are typical of normal behaviour for a patient is shown in Figure 6–28 with associated annotation. In this instance, the classifier output shows two significant peaks, one of these corresponds to the patient undergoing All Care, which may include endotracheal suction. This result suggests that without complete patient records the annotations which are used to generate the exemplars of both normal and abnormal behaviour, but in particular those used to generate the model of normal behaviour, cannot be fully relied upon.

From Figure 6–25 it can be seen that the nested approach yields the classifier output with the greatest range. This is coupled with the fact that the nested approach requires less computation at the classification stage as fewer features are used and hence fewer parameters in the classifier (see Appendix A for more details). It is therefore suggested that for these reasons, if one classification process is to be used for this application, the nested approach using the long time interval yields the best results with the lowest computational complexity.

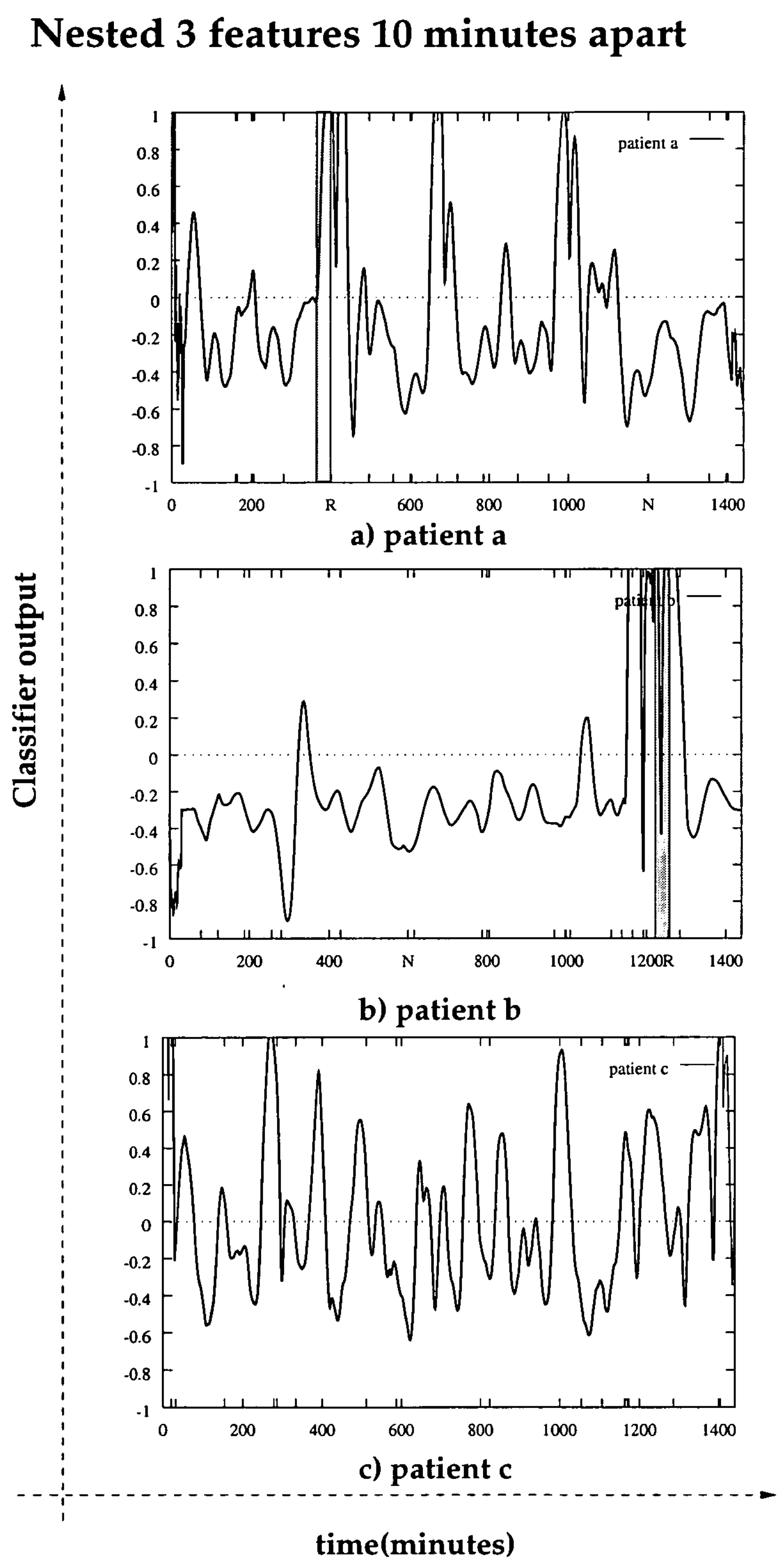


Figure 6–23: Example of classifier output for three discrete days when three features have been extracted from each physiological signal

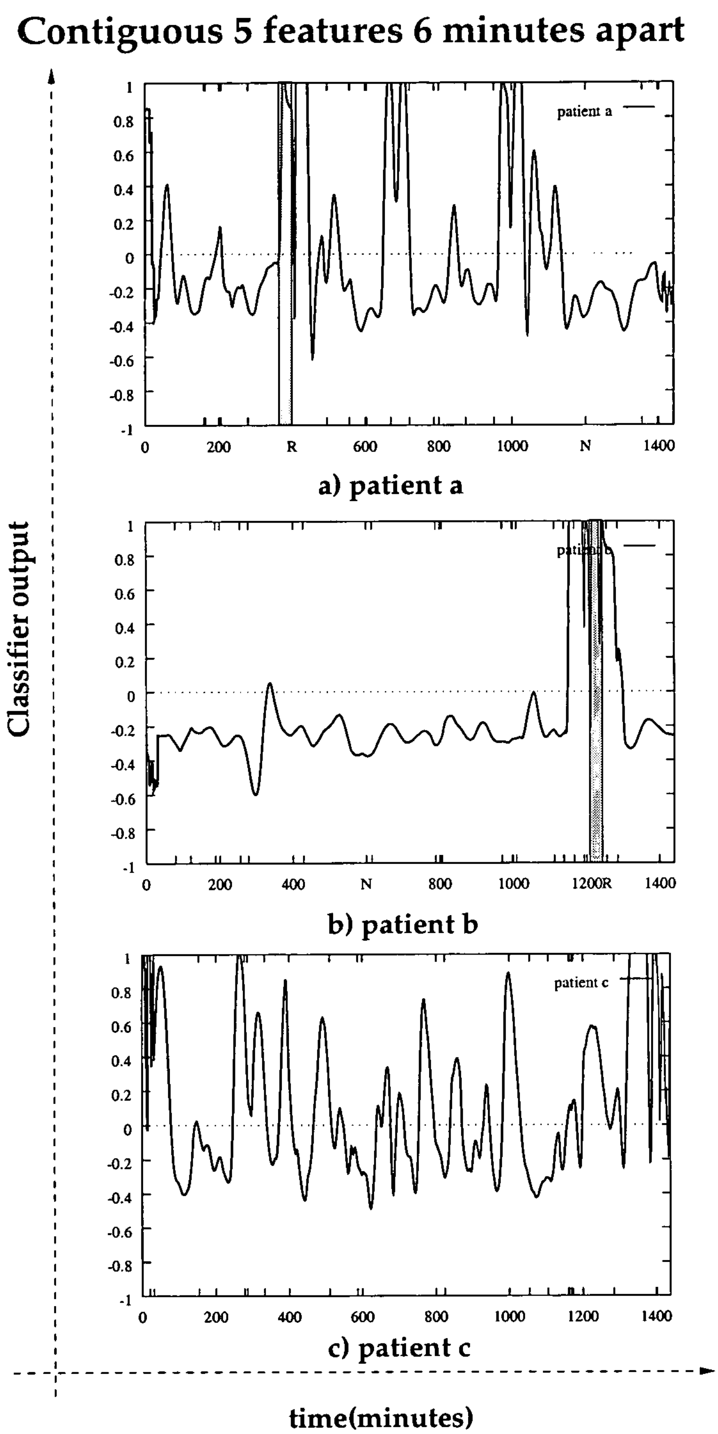


Figure 6–24: Example of classifier output for three discrete days when five features have been extracted from each physiological signal

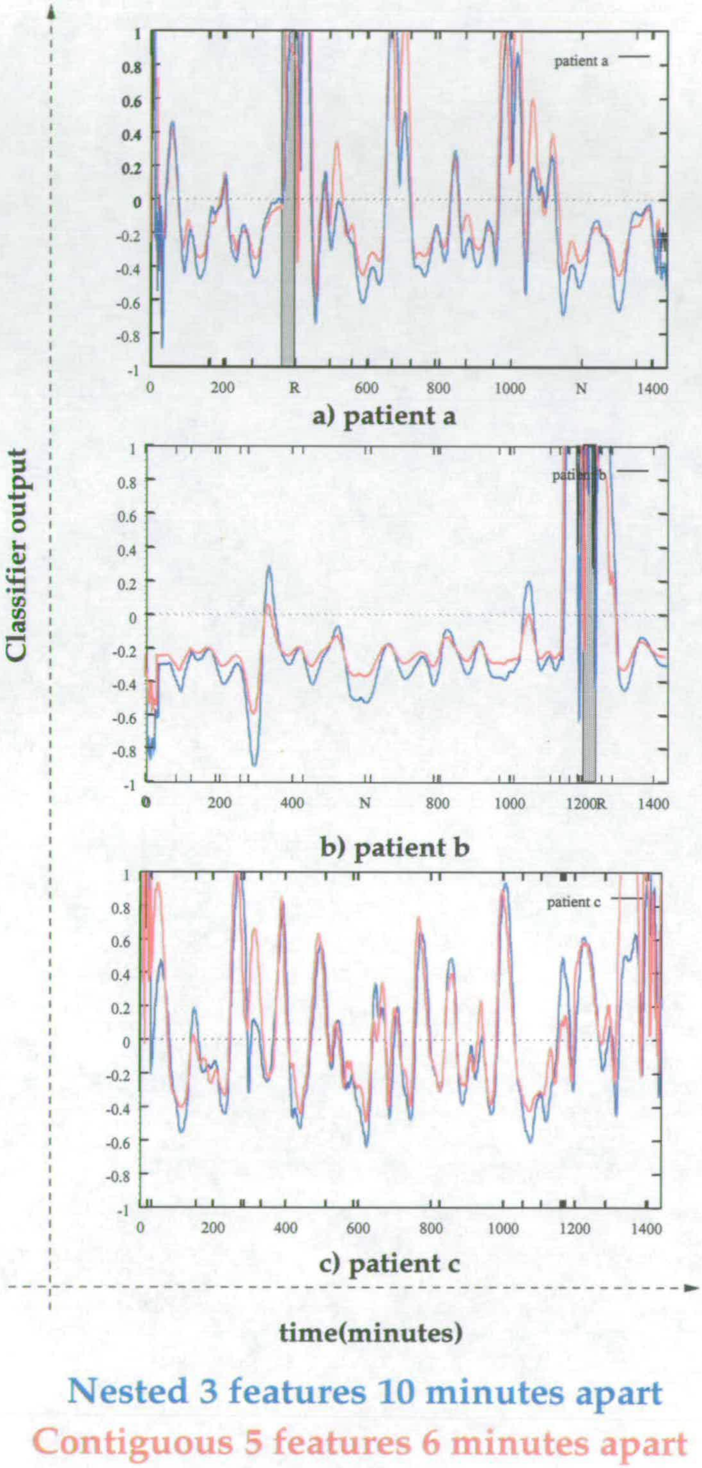


Figure 6–25: Two superimposed classifier outputs

Patient a

160	Blood for gases
164	Bolus N-G feeds
203	Physiotherapy
207	ENDOTRACHEAL SUCTION
280	Change PO2 probe
400	REINTUBATED
496	All Care
557	X RAY, Change PO2 probe
604	Bolus N-G feeds
669	All Care
720	Bolus N-G feeds
787	Bolus N-G feeds
874	Bolus N-G feeds
951	Heel stab
1354	Change PO2 probe
1396	ENDOTRACHEAL SUCTION, Physiotherapy

Patient b

79	Panc given iv
121	T2 probe resited
188	Pharyngeal suction
256	Top up transfusion started gtn up to 1ml
279	Blood for gases
429	Bed changed
615	All care
788	ENDOTRACHEAL SUCTION
806	Change PO2 probe
970	ECG leads new art line started
997	Resiting art line
1010	Failed IA line
1113	Cardiac unltrasound
1139	ENDOTRACHEAL SUCTION
1166	ACUTE EPISODE PO2 DROPPED BAGGED
1193	FURTHER EPISODE BAGGED XRAY
1256	REINTUBATED
1285	Nasal Suction

Patient c

37	Physiotherapy ENDOTRACHEAL SUCTION
38	All care, Bolus N-G feed
154	Bolus N-G feeds
278	Weighing
333	Change PO2 probe
334	All care physiotherapy
551	All care
587	in open cot
824	Bolus N-G feeds
1055	Bolus N-G feeds
1107	Bolus N-G feeds
1165	Change nappy, glycerin chip given
1165	Physiotherapy
1174	Change PO2 probe
1174	Bolus N-G feeds
1286	Change nappy Change PO2 probe
1419	Physiotherapy ENDOTRACHEAL SUCTION

Figure 6–26: Annotations for selected example days

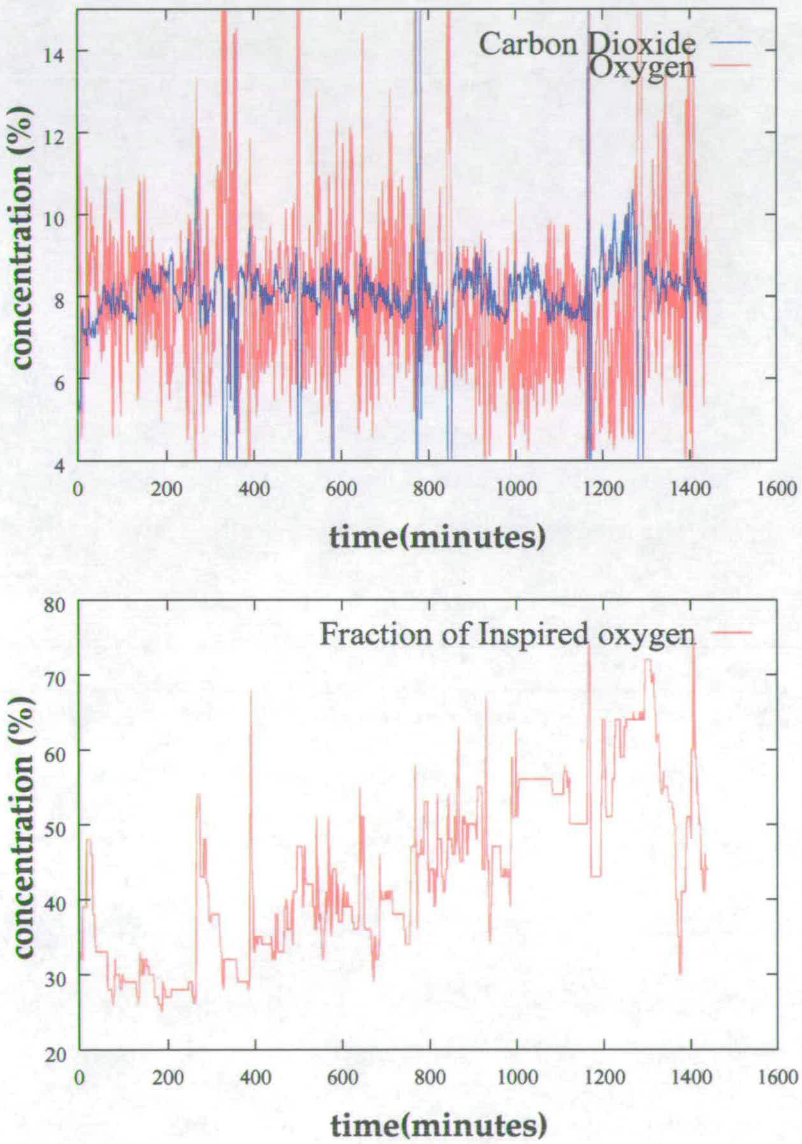
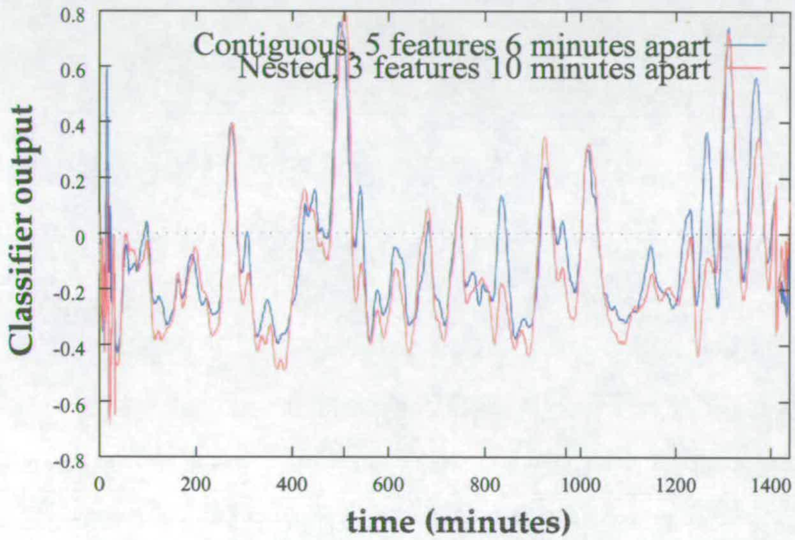


Figure 6–27: Raw physiological data taken from patient c, note the long term increase in FiO_2



	Patient d
508	**All Care**
969	**Change PO2 probe**

Figure 6–28: Classifier results on day containing no event

6.8 Discussion and conclusions

This chapter has detailed the results of a number of experiments carried out using a multi-layer perceptron neural network to classify the condition of a patient at a particular time. These tests included the determination of the optimum feature extraction technique, the method by which training and test sets should be generated, and how the classifier performs on complete sets of retrospective physiological data. It has been determined that the MLP outperforms a linear classifier in every case and that features should be extracted using a nested approach. It has been found that procedures using data from multiple patients to generate test and training sets outperform any where "normal" behaviour has been modelled for a single patient.

When the system was tested on complete days of physiological data it was possible to correlate peaks appearing in the output of the classifier with events recorded on patient records. However, there are also peaks which have no corollary. In some cases these "false alarms" may be attributable to the occurrence of a physiological problem not entered on patient records, or to other intervention to the patient which may trigger similar effects to those of a blocking tube or pneumothorax. It is also possible that the neural network is classifying another, as yet unknown condition, which has the same precursors as those under investigation.

It should be noted that the work described here has a number of drawbacks and potential problems. For example, although a feature extraction technique has been selected as one which is most efficient and accurate under the circumstances, the information which it provides to the network may only be an approximation to that which is required to fully describe the system. By examining the output of the classifier relative to complete days of data it is possible to see that the inclusion of temporal information in the feature extraction process permits appreciable peaks to be formed at the output of the classifier which remain in a steady state for significant periods of time (approximately thirty minutes). It can therefore be stated that the system has predictive ability, as the classifier produces a high output before an event has been entered onto the clinical record of the patient. In the case where no event has been entered the peaks which the classifier produces tend to be of smaller magnitude and duration and may, in part, be due to the approximation problems discussed earlier.

To summarise, the results presented in this chapter suggest that this is an application area in which a multi-layer perceptron shows promise for use a predictive diagnostic aid. There are, however, a number of limitations to this conclusion which the next chapter will address and where suggestions for the how the work might be extended will be made.

Chapter 7

Discussion and Conclusions

7.1 Overview

This chapter will discuss a number of topics. These include the results and their implications. It will suggest work which could be used to extend and improve the system which has been developed and finally draw conclusions concerning the efficacy and applicability of the system developed.

7.2 Discussion

The area which this study set out to investigate was that of the development of a diagnostic aid for use in a neonatal intensive care unit (NICU). In particular, it was decided that any system developed should be able to predict certain common physiological conditions. Often these are diagnosed later than clinicians would like. It was also to use a common classification technique to determine its applicability to the area of interest. The work described here was therefore based on a number of questions;

- Can a diagnostic aid be produced for an NICU?
- Can an Artificial Neural Network be used as the heart of such a diagnostic process?
- Can the diagnostic aid have predictive ability?
- Is expert knowledge of particular relevance in this application area?

The results presented in chapter 6 provide an answer to each of these questions. However, each carries important qualifications and these will be dealt with in turn.

7.2.1 Production of diagnostic aid for NICU

From the results presented in chapter 6 it can be seen that it is possible to produce a diagnostic aid for use in a NICU. The results generated by the system developed are not of a sufficiently high quality to be used in earnest as yet.

Possible reasons for this can be postulated. These predominantly relate to the method currently used to collect data and the assumptions made about the characteristics of the data.

Data collection

At present in the NICU at Edinburgh Royal Infirmary, physiological data is collected automatically and stored as part of the cot-side monitoring system (referred to as "Mary"). It is possible through "Mary" to annotate the physiological data files with details of treatment which has been carried out. Therefore, both the physiological data and the actual treatment record are time indexed. Despite the obvious advantages of this system that patient data can be retrospectively examined, there are a number of problems. The most significant of these is that the accuracy of the treatment record relies on the clinicians and carers in the unit. There are great demands placed on their time, particularly in moments of crisis, and it is not always possible for them to enter treatment records at the time of treatment or to enter them at all. This has meant that when the system was being developed a number of assumptions relating to the data had to be made.

Assumptions

The assumptions made about the system relate in particular to the generation of exemplars for use both in training the system and in preparing test sets for the classification process. There were two assumptions made. Firstly, when an event of interest was entered into the treatment record it was assumed that it was entered at the time of diagnosis of the event. In many cases this was not the case as treatment often occurs before the event is entered into the record. The second assumption made was that when no treatment was entered on to the record for a period of at least two hours it could be assumed that the patient was in a stable condition and therefore the central section of this period could be used as an exemplar of normal (no-concern) behaviour. The problems associated with this assumption are that treatment records are often incomplete and common treatment regimes may not be entered because staff are busy with another patient.

7.2.2 Can an ANN be used as the classifier?

It was decided that an ANN would be tested for use as the classifier in the developed system. This decision was made because a) the system which had to be modelled was non-linear (results show that the ANN consistently outperforms a linear classification technique, see chapter 6), and b) ANNs have previously been used in medical signal monitoring and condition monitoring applications, for example [78,12,101]. ANNs were also chosen as they do not require the formal ruling which some other, predominantly artificial intelligence, techniques require.

Results show that the ANN which was selected to be used (a multi-layer perceptron (MLP)) can be used as the classifier in this process. However, it may not be the best network or classification technique for use in the application. For example in other medical applications other types of ANNs have been used, e.g. Radial Basis Function Networks [51], recurrent networks [89], hierarchical networks [102] and self-organising maps [68]. It is therefore obvious that the MLP may not be the only method applicable in this application and that other techniques, both ANN and other standard classification techniques should also be investigated.

7.2.3 Can the diagnostic aid have predictive ability?

It is obvious that for a diagnostic aid to be of any use it is necessary that it must be capable of outperforming clinicians in some aspect of their work. This can be achieved in a number of ways. These range from reducing the possibility of clinicians and carers becoming habituated, by removing tedious monitoring tasks from their duties, to detecting patterns within signals which might otherwise be missed because of the nature of the application area. The system developed for this application area was required to be capable of achieving both of these aims as diagnosis in the NICU occurs on a second by second basis and patterns which may presage certain conditions are often missed or diagnosis takes place only when further invasive therapy is required.

By using expert knowledge as to how the conditions of interest develop it was possible to build predictive ability into the system developed. Temporal information was included in the pre-processing stage of the system and also at the diagnosis stage, when exemplars were being classified. Using the knowledge that the conditions for concern often develop over a few hours, it was assumed that the hour before diagnosis was made should be characterised by a detectable problem. In the feature extraction process, information regarding the long-term trends of the signal was maximised as expert knowledge told us that patterns which were thought to pertain to respiratory problems often occurred at this level.

However, as it is known that conditions develop over a period of time there may have been problems with this type of classification therefore another method of classifying the data might be examined, for example by using intermediate states or a non-step classification function.

7.2.4 Is expert knowledge of particular relevance in this application?

Expert knowledge relating to this problem was initially sparse because little is known about how the respiratory disorder develops and affects the physiological signals being monitored. However, it was possible to determine that clinicians felt that certain signals which were monitored may be of greater diagnostic importance than others. This reduced the volume of data which had to be handled and permitted an investigation of the diagnostic relevance of other signals not taken as standard.

Expert knowledge also made clear that conditions often take a significant period of time to develop and that with hindsight it is often possible to identify trends within isolated signals which might have been related to the development of the condition.

7.2.5 Summary

At this stage it is possible to state that the results achieved are promising but that further work must be carried out to maximise the effectiveness of the system. A number of questions have been raised which merit further investigation.

7.3 Future work

This section will detail the work which further merits investigation.

7.3.1 Prospective data collection

The previous discussion suggests that the treatment records which relate to the physiological data stored on "Mary" are often incomplete or inaccurate. The only method which can be used to correct this problem is to prospectively gather data which have been carefully annotated. If necessary, every single event pertaining to the particular patient should be entered as it may be of some relevance either to the development of the respiratory disorder or for partial explanation of the extra signals which currently occur in the output of the classifier.

If the data are carefully annotated it should also be possible to develop the system further, to detect the onset of other common conditions found in NICUs or ICUs. However, even with the most perfectly annotated physiological data there will still be limitations to the system in its current form and therefore other adaptations and extensions to the proposed system must be made. Clinical knowledge must also be used in the generation of the training and test sets. This would reduce the problems associated with the assumption that "no comment" corresponds to no concern.

More data must also be used in the study as the larger the number of patients included, the more complete the description of "normal" and abnormal behaviour will be.

7.3.2 Feature extraction

It is in the feature-extraction and pre-processing stages of the current system where the expert knowledge has been primarily utilised, both in the selection of the signals of interest and decisions as to how they are to be processed. If this stage does not maximise the information of greatest relevance to these conditions the classifier will never achieve better results. It is therefore suggested consideration should be given to modifying the feature extraction process. For example in other condition monitoring applications [53] ARX modelling has been used to describe the time series data [84,22] this performs a conversion of the single data point into a series of parameters which describe the signal at that point and from these parameters the original signal can, if necessary, be regenerated. This may present the relevant information to a classifier in such a way that the classifier finds it easier to distinguish between the classes.

It may also be that the inclusion of temporal information at this stage is unnecessary if the classification stage either already includes temporal information or is combined with other techniques which include temporal information.

7.3.3 Classifiers

The classifier which was used for the series of experiments detailed here was a multi-layer perceptron (MLP) neural network. There were a number of reasons why this particular architecture was chosen, the predominant one being that MLPs have previously been used in this type of domain and applied to these types of signals. Despite this the MLP may not be the most efficient method of dealing with these signals and application area. It may be that as a result of the nature of the problem, and given the number of assumptions that are made as to the development of respiratory disorder in premature neonates, it is necessary to investigate other methods.

Assumptions have been made about this area to permit training and to allow test sets of data to be formed for the classifier. These sets do not accurately represent the real situation in the NICU for a number of reasons; a) they are unbiased and in real life the patient spends the majority of his/her time in a stable condition and b) they are incomplete. There is obviously no feasible solution to the second problem without generating an infinite data set. However, it may be necessary that the generation of the training and test sets should accurately represent real life in that the data set will contain few examples of abnormal behaviour. In this case the sets would be heavily biased and the task becomes that of novelty detection rather than classification as it has been treated here [103]. There is another issue involved in the choice of the classifier. At present, all temporal information in the system is included in the feature extraction stage. This may not necessarily be the most reliable method. Classifiers exist which are capable of including temporal information. Temporal neural network classifiers for example feed prior outputs back to the inputs after a period of time [104]. These are known as for recurrent neural networks [105,106]. It is also possible to combine a traditional classification technique with another signal processing technique to capture some of the temporal element of the process being modelled. An example of this is in the use of Hidden Markov Models (HMMs).

Hidden Markov Models

Hidden Markov Models (HMMs) have been used for a number of years as methods of capturing the temporal element present in speech signals (see Appendix B). Elements of the HMMs have been used to model phonemes [31]. More recently, they have been combined with other classification techniques, for example ANNs in both speech processing [63,85,64,65] and other areas [53,20]. The work of Smyth et al. [53,107,54,55] is of particular interest as although it does not use medical signals there are a number of similarities between the application areas. For example, both the application area of this thesis and the application area of Smyth et al. involve the use of time-series data which have a large temporal element and are heavily biased towards a stable state. It is therefore suggested that developing a system similar to that of Smyth et al. for this application area may yield interesting results and point the way toward developing the system further, both in terms of the conditions it can identify and its efficiency.

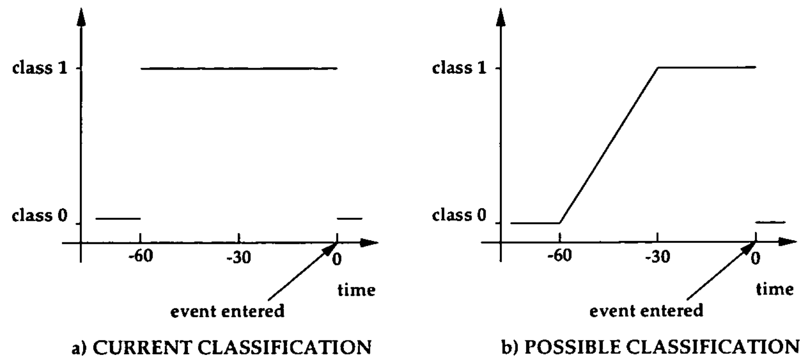


Figure 7–1: An alternative classification function

Method of Classification

Another avenue for further work may lie in the use of different classification functions. At present the decision of the network is a classification one involving a two class problem; no-concern and concern. This classification could be subdivided into three classes; no-concern, some-concern and concern or even subdivided into a completely different type of function; a linearly or exponentially increasing function (see Figure 7–1).

7.4 Conclusions

This work described in this thesis was originally designed to perform a number of tasks which included the determination of whether it was possible to produce a diagnostic aid for use in a neonatal intensive care unit.

An aid in this type of environment is designed to;

- Increase the level of patient care
- Prevent the development of certain conditions
- Prevent deterioration of the condition of a patient
- Automate access to patient records
- Reduce the risk of habituation by staff

- Reduce the number of false alarms in the NICU
- Automate the control of life support systems

After investigation of the application area and its associated problems it was decided that a subset of these aims would be investigated and these were;

- Increase the level of patient care
- Reduce the risk of habituation by staff
- Prevent the development of certain conditions
- Prevent deterioration of the condition of a patient

These were chosen as there were a number of constraints involved in the development of a diagnostic aid at this time. These included firstly lack of data. Some physiological signals are measured as standard and are not always the most applicable to the diagnosis of a condition. Secondly little was known about the development of certain conditions. A common problem occurring within the NICU is Respiratory Disorder and it was chosen as there were a number of documented examples available in archived form. Thirdly some problems commonly occurred and were thought to develop over long periods. It was felt that early diagnosis of this type would improve patient care and reduce the risk of invasive therapy.

The work therefore became an investigation of the development of respiratory disorder (RD) in ventilation assisted neonates. Multi-channel physiological data was available on which to develop a diagnostic aid. The work included an investigation into the applicability of a diagnostic aid for this type of area. It aimed to determine if any physiological signals were of greater diagnostic relevance than others, and if there was an appropriate information extraction technique which could help in the early diagnosis of RD. RD was also chosen as it occurs in patients of all ages and therefore any system developed would be capable of being transferred to any other critical care environment with similar monitoring techniques.

The results which were produced by the system suggest that a system can be produced which will predict the development of respiratory disorder thirty minutes ahead of the current diagnosis time. It uses linear signal processing techniques to extract information thought, by clinicians, to be of relevance in this application and combines this with an MLP to produce a judgement of the risk of a particular patient developing RD. Results show that at present a crucial signal for the diagnosis of RD is often omitted at the monitoring stage. This is the fraction of inspired oxygen in the air mixture and it is omitted despite the fact that its inclusion would not involve the use of invasive or stressful techniques.

7.4.1 Summary

The work carried out therefore suggests that it is possible to produce a diagnostic aid for use in an NICU. As the condition selected is a common one and also occurs in other critical care environments the system can be transferred. The techniques which have been used at the pre-processing and feature extraction stages are also applicable to any application area where time signals are thought to vary in a similar way. The results which the system produced are promising but in every case the techniques which have been applied merit further investigation so that the system can be maximised in terms of its diagnostic ability and reliability.

At present, the system is not run in real-time but this could be achieved at all stages with little alteration. Raw data could be run through the alpha filter only in the forward direction. The feature extraction process, if provided with a constant thirty minute supply of retrospective data can run real time and the classifier can be adapted for this use. Therefore, with few alterations the system would be capable of producing diagnostic predictions after thirty minutes (this value is controlled by the feature extraction process).

Clinical diagnosis is an extremely complex area in which great reliance is placed on the skills of clinicians to adapt both to changes in patients' conditions and to inter-patient variability. At present it is not possible to design a system which can achieve all that a skilled clinician can in terms of diagnosis. However, as this work shows, it is possible to develop a system which, by the monitoring of various physiological signals from a patient, can also advise clinicians on the condition of a patient. In its current form the system which has been developed shows promise, but a large amount of work needs to be carried out before a similar aid could be in place in a NICU.

Appendix A

Multi-Layer Perceptrons

A.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) is the collective name given to different models which were originally designed to capture some of the functionality of the human brain.

Computers are adept at performing numerous calculations which would take a human years to complete, however they are particularly bad at reasoning and extrapolation. If a system is required which uses rules in the form "if A is greater than B then the result is C" a computer can perform these calculations easily. However, it does not have the ability to extrapolate when presented with information which does not conform to its rules. Unlike computers, humans can extrapolate, from a set of known rules, answers to inputs which resemble previously studied information.

ANNs were designed to bridge the divide between the extrapolation abilities of humans and the rule-based approach of computers. They are "trained" to model underlying non-linear processes in a system and when previously unseen information is presented they can generalise (extrapolate) where in the model the information lies.

A.2 Neuron models

The basis of an ANN is an architecture of artificial neurons or perceptrons. These are designed to be a generalised model of a biological neuron.

A.2.1 Biological neuron

There are over 10^{10} neurons within the human brain and each of those is connected to at least 10^4 others. They consist of a soma, an axon, synaptic junctions and dendrites. A simplified diagram of a biological neuron is shown in figure A–1.

The functionality of the biological neuron is as follows; messages or impulses are passed along the dendrites where they eventually arrive at the soma. Here an output is produced if the collective sum of all the dendrite impulses exceeds a threshold level. However, each dendrite can pass its signals on with varying degrees of success. This can be due to the strength of the synaptic junctions, coupling with other neurons' outputs, and to how efficient the dendrite is at passing on its information. These can be thought of like resistors in that some materials are more efficient at conducting electricity therefore some dendrites have more influence (higher voltage contribution) at the soma. It was this simple idea of sum, different communication efficiency and threshold which the artificial neuron was designed to capture.

A.2.2 Artificial neuron

An artificial model of the biological neuron was first introduced in 1943 by McCulloch and Pitts [98]. They proposed a model which attempted to capture the functionality of the biological neuron without taking timing and propagation delays into account.

The functionality can be captured very simply by using a model similar to that shown in Figure A–2. This artificial neuron accepts inputs which have been multiplied by a connective weight (this models the dendrite and its associated efficiency) these inputs are then summed at the node (or neuron), which takes the place of the soma, and if the sum exceeds some arbitrary threshold level the neuron fires and an output of one is produced. The threshold function controls the operation of the neuron in that if the function is a stepwise Heaviside function the neuron can only behave as a linear discriminator. This means that it can only perform accurate classification on problems which are linearly separable. To solve linearly inseparable problems

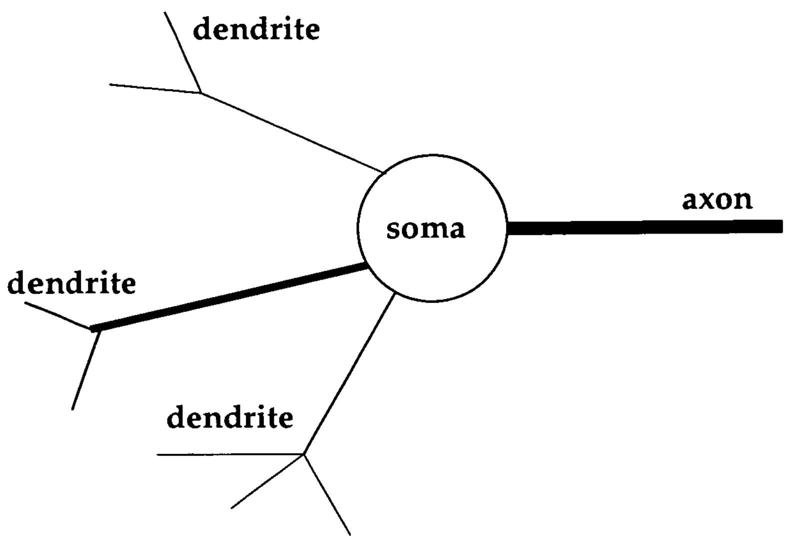


Figure A–1: Diagram of biological neuron

a non-stepwise function is required. A common choice with ANN researchers is that of the sigmoid function (see equation A.1). In this case the value a determines the slope (or steepness) of the linear section of the function, i.e. how squashed the function is.

Using a series of layers of these artificial neurons it is possible to solve linearly inseparable problems. The most common type of these layered networks is that known as the multi-layer perceptron (MLP) network.

$$F(v) = \frac{1}{1 + exp(-av)} \tag{A.1}$$

A.3 Multi-Layer Perceptrons

These networks commonly consist of three layers of neurons or perceptrons which perform the a recognition or classification task. There are input, output and hidden layers (see Figure A–3). The input and output layers, as their name suggests connect the network to the outside world. The hidden layer has no such connections and therefore its calculations remain "hidden" from outside observation.

The multi-layer perceptron operates as follows; values applied at the input layer are fed to the hidden layer through parallel multiplicative connections (these multiplicative connections are

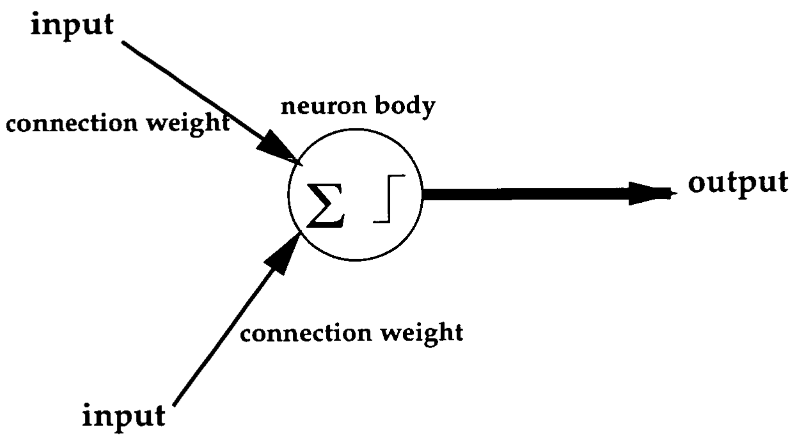


Figure A–2: Single perceptron

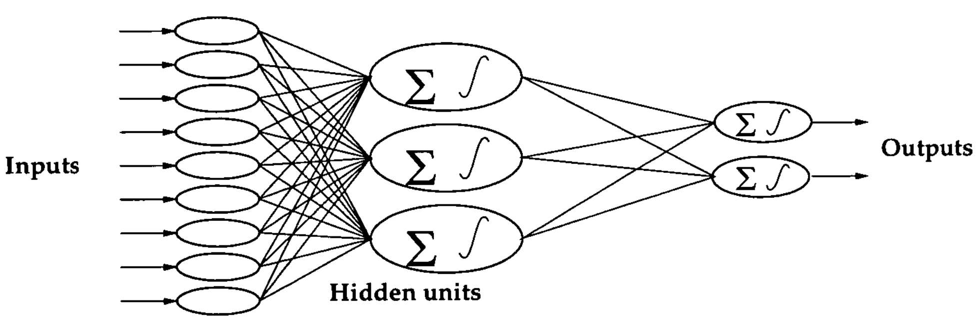


Figure A–3: Multi-layer perceptron ANN

Inputs

V_i

Hidden layer

$$x_j = \sum_i T_{ji} V_i - \theta_j \qquad V_j = f(x_j, Th_j)$$

Output layer

$$x_k = \sum_j T_{kj} V_j - \theta_k \qquad V_k = f(x_k, Th_k)$$

x_n input to nth node
V_n output from nth node
T_{no} weight from node in previous layer to node in current layer
Th_k threshold of transfer function on output node
i - input layer, j - hidden layer, k - output layer
f activation/transfer function
θ bias term

Figure A–4: Equations for MLP

called "weights"), the results of these multiplications are summed at each hidden node and if the sum exceeds a threshold level (i.e. it is passed through a transfer function) the hidden node fires. The bias term determines the activation/threshold level of the neuron. The output of the hidden node is passed on to the output layer through another series of multiplicative weights. Again these are summed at each of the output nodes and the resultant is passed through a transfer function to generate the output. This process can be expressed mathematically as shown in Figure A–4.

By adjusting the weights or multiplicative connections between the different layers it is possible to alter the values at the output nodes. This is the heart of the MLP's learning process in that during the training phase inputs are presented to the network and their corresponding outputs examined, the weights are then altered in a way which is designed to reduce the error between the generated output and that which was expected. This is called supervised learning. This learning phase continues until the error is minimised on a set of previously unseen input data. A description of the training and testing phases are shown in Figure A–5.

During this training or learning phase the MLP progressively minimises the error term associated with the validation set (see Figure A–6) and improves its model of the system underlying the input data. Once training is complete it is possible to apply new data to the network and to have a measure of the reliability of the network's output classification. The choice of error term and activation/transfer function depends on what is required at the output of the network. In

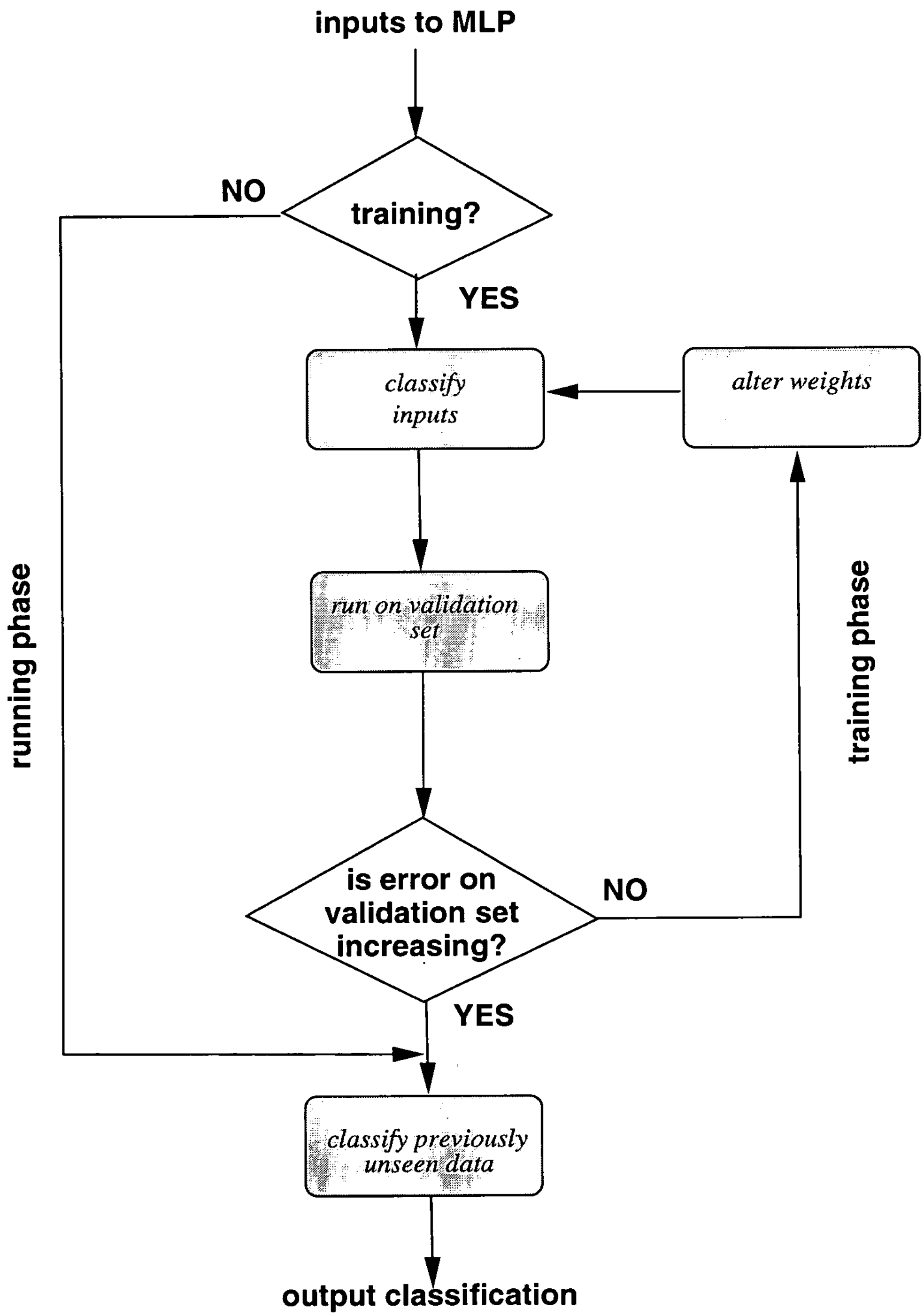


Figure A–5: Operation of an MLP

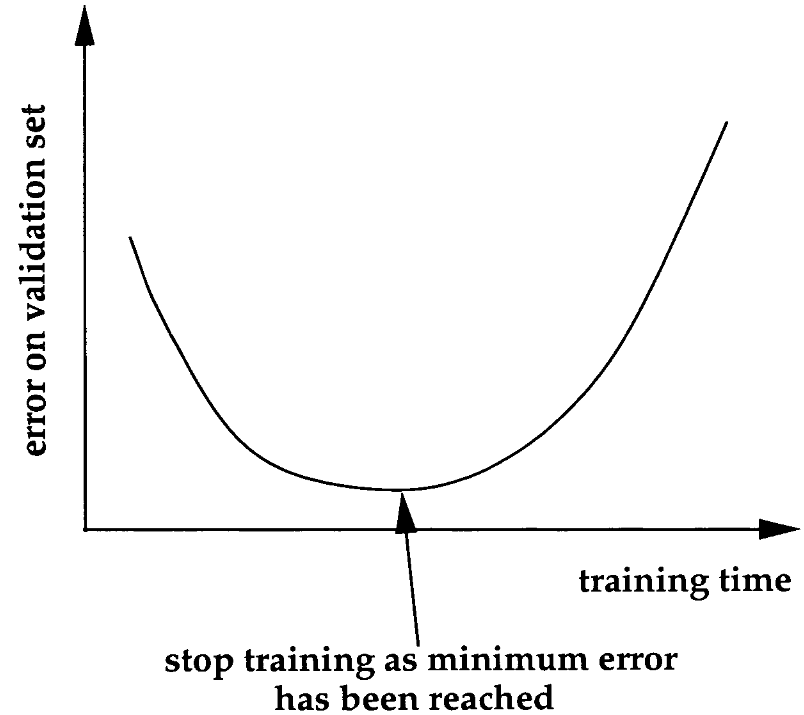


Figure A–6: Error minimisation

this case the sum-of-squares error has been used, which when combined with a simple sigmoid activation function produces an approximation of the conditional distribution of the data in terms of a Gaussian distribution. If however, conditional probabilities are required at the output of the MLP a softmax activation function is required [108].

Different training methods can also be used. These vary in weight-update methods and hence can have an effect on training time and how the network copes with local minima on the error surface. The weight-update methods try to alter the weights in a way which moves the outputs of the network closer to a global solution.

A.3.1 Architecture

If an MLP consists only of an input and output layer it behaves as a linear discriminant classifier drawing linear boundaries or planes through multi-dimensional data. If however, a hidden layer is introduced this means that the MLP can be used as a non-linear modelling technique. The greater the number of hidden units used the more flexible the boundary drawn between classes can become. There is, however, a balance to be drawn between generating too many degrees of flexibility (too many hidden units) and hence over-fitting (following the boundaries in the training data set too closely) and under-fitting (not including enough degrees of freedom or hidden units in the model). At present the formation of the optimal architecture of a network cannot be directly calculated but requires testing of different architectures until the optimal is found in terms of the network's performance characteristics. One of the variables which can constrain the size of the network is associated with the number of examples which are available to be used as training data. An ad-hoc measure of how many training examples are required to permit the network to learn adequately is given in equation A.2. Where the number of parameters is equal to the number of connections and biases in the network.

$$Examples = 10 \text{ x } parameters \tag{A.2}$$

A.3.2 Training Multi-Layer Perceptrons

To permit training to be carried out data must be collected in which it is felt adequate examples of system behaviour exist. This collection of data must be divided into a number of distinct data sets; train, validation and test sets. As their name suggests each set serves a different purpose. The training set is usually the largest of the three sets and it is used in the training phase of the process, the validation set is used to monitor how well the network is training and the test set is used in the final phase to determine the network's operating parameters (see Figure A–5) and its ability to generalise (categorise previously unseen data).

To enable conclusions about a particular network's operation to be drawn a particular network architecture must be tested a number of times. For each test the network architecture must be started using a different set of connection weights and the data sets must be randomised. This means that if any particular data set contains a majority of examples of a particular class the effect will be minimised when the network is trained again on different data. It also means that conclusions about the ability of the MLP to generalise can also be drawn as this largely depends on the network architecture.

A.3.3 Uses of Multi-Layer Perceptrons

Since their conception MLPs have been used for a number of tasks which include classification and prediction. They have also been used in a number of application areas which range from image processing [109,80] through speech processing and condition monitoring [78,76,60, 110] to medicine [71,99,72]. They are particularly useful where a system cannot be explicitly modelled using other techniques and where the MLP can be treated as a flexible model for estimating the underlying processes. They can also be used in combination with other signal processing techniques, for example Hidden Markov Models (see Appendix B) to provide a more accurate model of a system's behaviour.

A.3.4 Summary

Multi-layer perceptrons are a form of artificial neural network which are in use in a variety of different application areas. In general, their architecture consists of three layers: input, output and hidden. The hidden layer performs the system modelling via a series of interconnected weights and sigmoidal non-linearities. The input and output layers are used to communicate with the outside world. They use supervised learning to produce a model of a particular system about which they can then generalise. They have been used extensively for classification problems where there are a number of discrete classes.

A.4 Conclusions

This appendix has introduced artificial neural networks. They were originally designed to emulate the operation of the human brain in its ability to learn and generalise. In particular the appendix has concentrated on one form of artificial neural network: the multi-layer perceptron. It discusses issues surrounding both the operation and design of the network and the method of training an MLP.

Appendix B
Hidden Markov Models

B.1 Introduction

This appendix will give a brief description of the operation of Hidden Markov Models (HMMs). In particular it will concentrate on the methods of combining HMMs with Artificial Neural Networks (ANNs) as this hybridisation is a technique which could be used as an extension to the work covered in this thesis. Hidden Markov Models are an extension of Markov Models and therefore any description of HMMs must begin with a description of the Markov model.

B.2 Markov Models

Markov models or Markov chains are used to describe processes in which a number of discrete states exist. The entire system is modelled in terms of both its states (component parts) and the probability of moving between those states (*P(transition)*). The time taken to move from state to state is assumed to be zero; therefore at any time the system is in one of its pre-defined states. Markov chains can be used to describe a huge number of processes which range from biological to engineering.

The simplest form of Markov chain is one where the system has explicit start and end states and where the transitions between states are one-way transitions, i.e. states cannot recur. An example of this type of system can be seen in Figure B–1 where the process of wall papering a room is described.

However, this simplest type of chain cannot always be used as most systems contains states in which the process remains for significant periods (i.e. there is a finite time permitted per state and some states recur or repeat). An example of this type of Markov chain would be the rolling of a die (Figure B–2 (a)). In this case the value on the face of the die is the state and if the

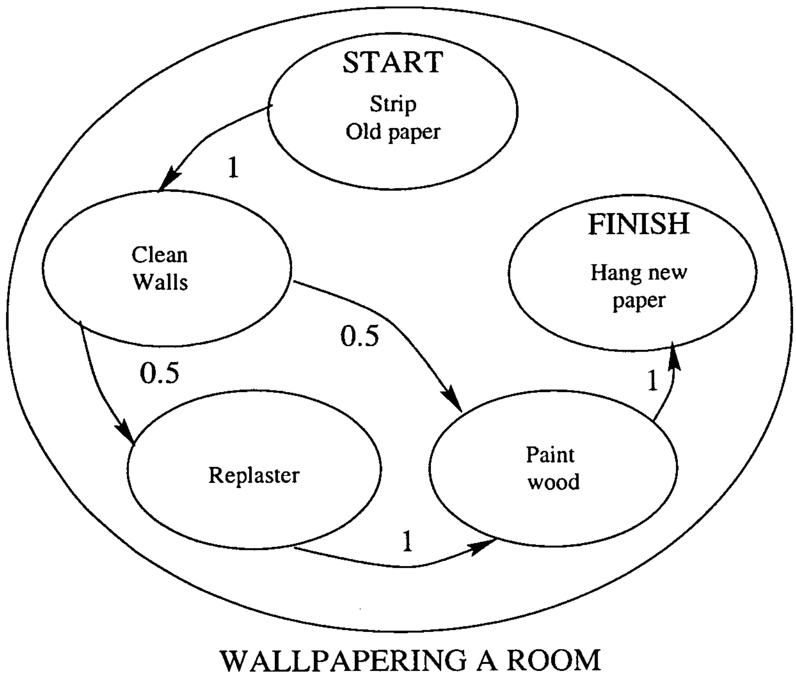


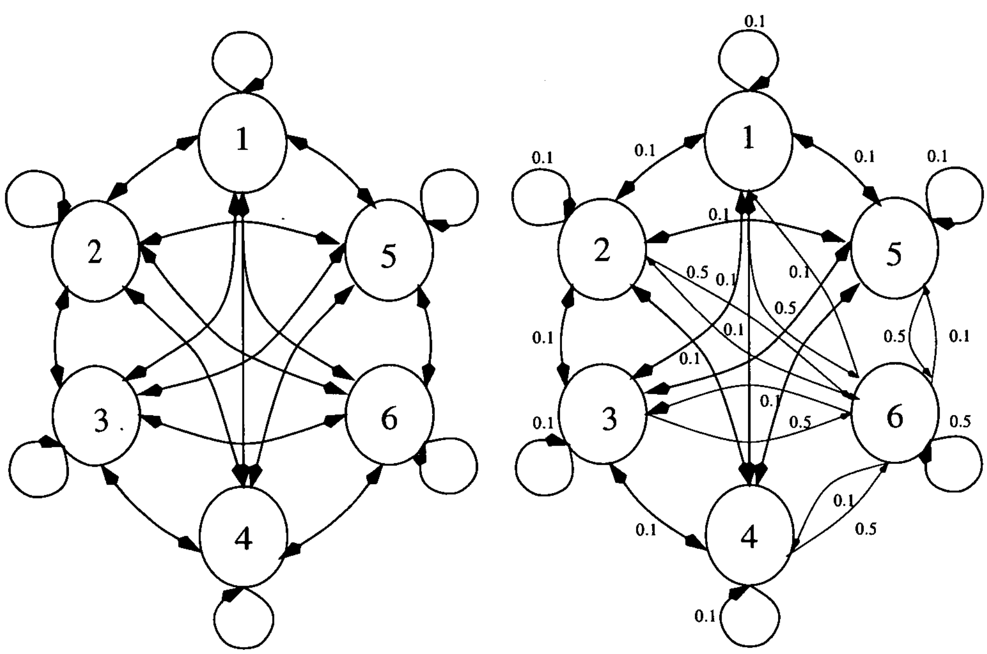
Figure B–1: A simplified Markov model of the process of wall papering a room

die is unweighted the probability of moving between one state and the next i.e. rolling a new number is equal for all numbers. This example also incorporates the information that the time spent moving from state to state is zero as the time spent rolling the die is negligible compared to the time spent on a face. If the die is weighted the Markov chain for the system changes as *P(transition)* varies from state to state. (see Figure B–2 (b)). This means that these transitional probabilities now play an important role in determining the system's behaviour.

In practice most processes are continuous and therefore cannot be described in terms of a Markov chain. However, the assumption can be made that over small time intervals the system is stationary and therefore the discrete time Markov chain can be used to describe it. In formal terms a system can be described using Markov models in terms of the transition probabilities between the states *P(transition)* where the Markov chain satisfies the following relationship B.1;

$$\begin{array}{rl} Prob & \{X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, ..., X_n = x_n\} \\ = & Prob\{X_{n+1} = x_{n+1} | X_n = x_n\}. \\ \text{where} & X_n \text{is the observation at time } n. \end{array} \tag{B.1}$$

P(transition) is defined as being the probability that the system will move from state x_n to



a) Unweighted die

The probability of moving from any state to another is equal

b) Weighted die

The probability of moving from any state to another varies

Figure B–2: Markov models for the throwing of a single die a) unweighted b) weighted) state x_{n+1} when the time parameter increases from n to $n+1$. Transition Probabilities can be denoted by the equation shown in B.2 which is independent for all n [111].

$$p_{ij} = Prob\{X_{n+1} = j | X_n = i\}. \tag{B.2}$$

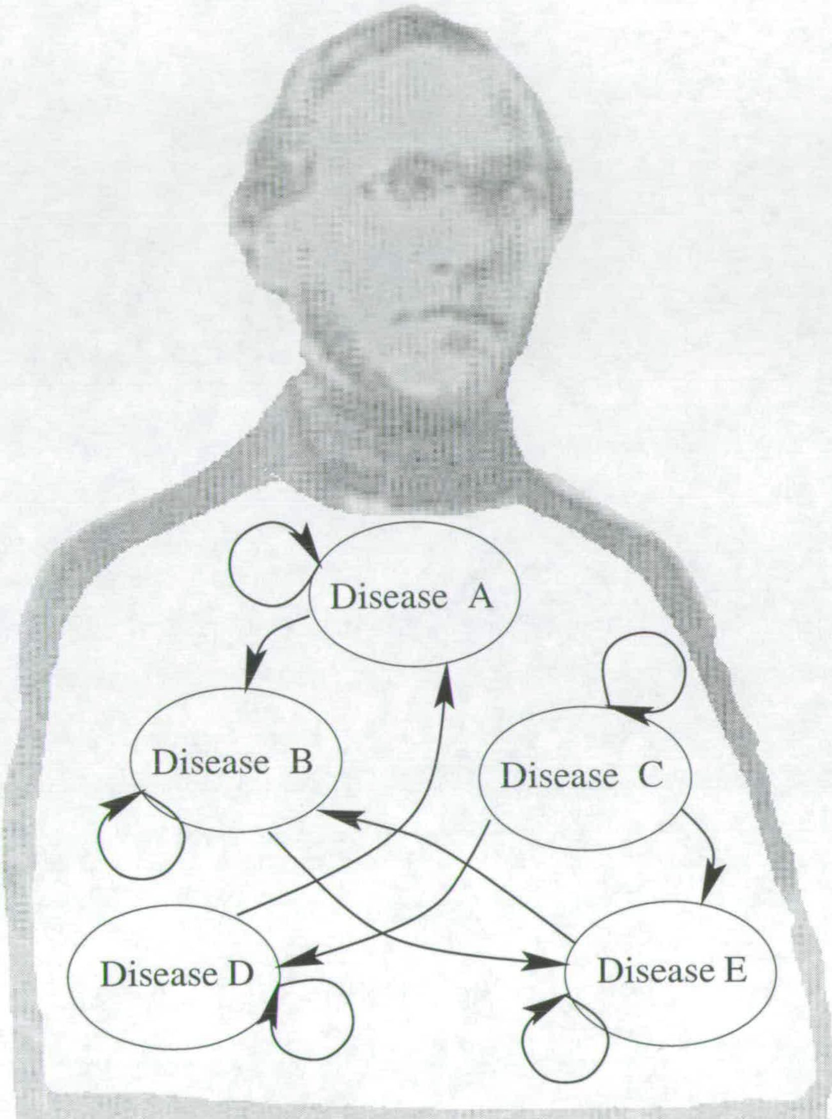
A system can therefore be described in terms of p_{ij}, its transition probabilities, and it is possible to determine minimum times to certain states if this knowledge is known. This is equivalent to finding the minimum propagation delay through an asynchronous electrical circuit. If however p_{ij} cannot be defined assumptions can be made that p_{ij} is equal for all states and a system model can still be built.

An extension of this type of model is where the system generates outputs (or observations), the system is known to move from one state to another (the transitions) and there is no known link between the outputs and the states. This type of model is called a Hidden Markov Model (HMM).

B.3 Hidden Markov Models

A complete description of HMMs can be found in Rabiner [86,112] and therefore only a brief description will be given here. Hidden Markov Models (HMMs) differ from Markov chains in that there is no direct mapping between the observations of the system and its states. They are essentially processes embedded within another process; states of a system are known but there is no knowledge of the underlying production of the outputs associated with them. An example of this type of system is that in the medical domain a patient can display a number of symptoms (observations in the HMM). These symptoms are indicative of a number of diseases (states in the HMM). The clinician uses knowledge of the likelihood that the patient will present these symptoms under the circumstances and the probability that the patient will develop another disease (transition probabilities), see Figure B–3. Explicit information about the disease, such as its cause, is not needed all that is required is prior knowledge of the cycle through which a patient's condition can move.

As in the case of the Markov model the HMM can be formally described using prior knowledge in terms of the state transition probabilities if prior knowledge of the state probabilities is available. HMMs can therefore be described by the diagram shown in Figure B–4 and can be defined in medical terms in the relationship shown in B.3 or more mathematically as shown in B.4 [113];



Disease A	Disease B		Disease C
P(symptom1)	P(symptom1)	- - - - - - -	P(symptom1)
P(symptom2)	P(symptom2)		P(symptom2)
P(symptom3)	P(symptom3)		P(symptom3)

Figure B–3: Medical example of Hidden Markov Model

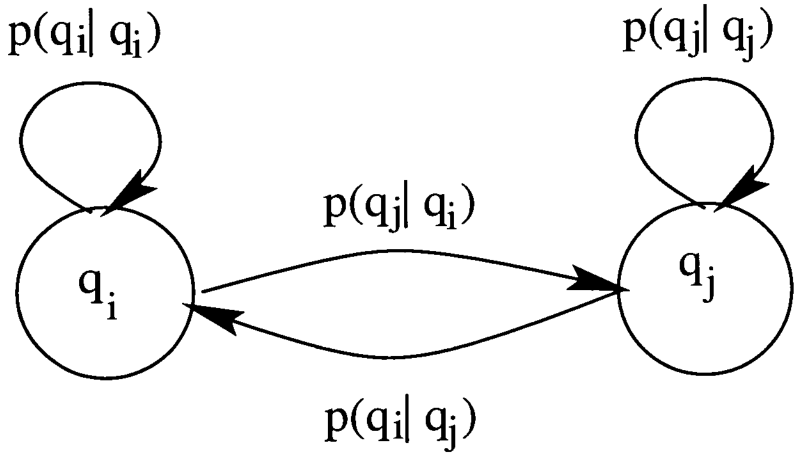


Figure B–4: Two state Hidden Markov Model

$$Prob\{Measles|spots\} = \frac{Prob\{spots|Measles\}Prob\{Measles\}}{Prob\{spots\}} \tag{B.3}$$

$$Prob\{M|X\} = \frac{Prob\{X|M\}Prob\{M\}}{Prob\{X\}} \tag{B.4}$$
where M is the model and X is the set of observations

i.e. the posterior probability $Prob\{M|X\}$ that the sequence of outputs/observations X has been generated by the model M can be calculated in terms of the conditional probability $Prob\{X|M\}$ and the prior probabilities $Prob\{X\}$ (probability of a particular sequence of outputs or observations being generated) and $Prob\{M\}$. If the prior and conditional probabilities are known posterior probabilities can be calculated and the system behaviour completely described. One method of calculating the conditional probability is to use an artificial neural network in combination with the HMM.

B.4 Hidden Markov Model Hybrids

Hidden Markov Models are an efficient method with which to describe discrete systems such as speech where the states of the HMMs can correspond to individual phonetic units [113,64]. However, they can also be used in combination with ANNs [63,85,53,65,114] to capture the temporal component ($Prob\{X\}$) of a system which is often ignored in the use of an ANN.

B.4.1 Temporal information capture

ANNs are poor at capturing the temporal aspect of a given time dependent signal such as is found in medical physiological signals. It is known that few medical conditions occur instantly and those whose symptoms occur rapidly often have warning signs, for example rising blood pressure. If these types of systems which cannot be completely described are to be studied HMMs can be combined with ANNs in an effort to capture both the underlying model of the system and its temporal component.

B.4.2 HMM-ANN combination

The combination of the two types of modelling tool predominantly occur in one particular way.

- ANN-HMM

This method which the author considers to be applicable for medical applications uses the ANN to classify the patient's state given their symptoms or the measured physiological signals. The HMM then takes the system's transition probabilities and the system's previous state into account and produces a decision as to the current state. This can be thought of as the HMM validating the output of the ANN classifier. Mathematically speaking in the ANN-HMM combination the ANN is used to produce the conditional probabilities ($Prob\{X|M\}$) that the system is in a particular state given the observations or occurrences which have been applied at its inputs. The HMM then calculates the posterior probabilities ($Prob\{M|X\}$) taking into account the transitional probabilities between states (see Figure B–5) and the prior probability

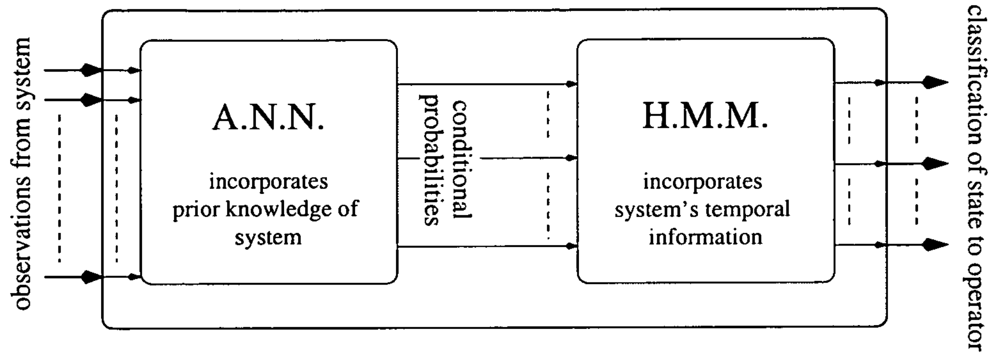


Figure B–5: A 'black-box' hybrid ANN-HMM system

that a particular state sequence will occur.

Using a hybrid ANN-HMM network means that a complete description of a system is not necessary. The ANN can be trained to model the system without an explicit mathematical system description being required. The author feels that this modelling approach is ideal for the application area studied in this thesis as the generation of respiratory disorder is difficult to model. Using the hybrid approach would also mean that the data used as inputs to the ANN would not need to be as heavily processed. It need only be filtered and have the outliers removed. Further processing can be carried out on this raw data but it would now longer be necessary to include a temporal component as this would be captured by the HMM.

The author feels that the work of Smyth et al. [53,107,54,55] may suggest a method which would be applicable in the medical application area. The application areas differ as Smyth's includes antenna pointing systems for satellite communications i.e. it is applied to mechanical problems. There are however, a number of similarities in that there is a large temporal aspect to the signal, few examples of problems exist and knowledge is insufficient to allow it to be modelled. Smyth et al. use a multi-layer perceptron ANN as pre-processing to an HMM as previously described. This combination produces much cleaner results than the single classifier approach and helps to eliminate outliers caused in the signal by similar inputs generated at different times in the operating cycle. An example of Smyth's results can be seen in Figures B–6 and B–7 [107].

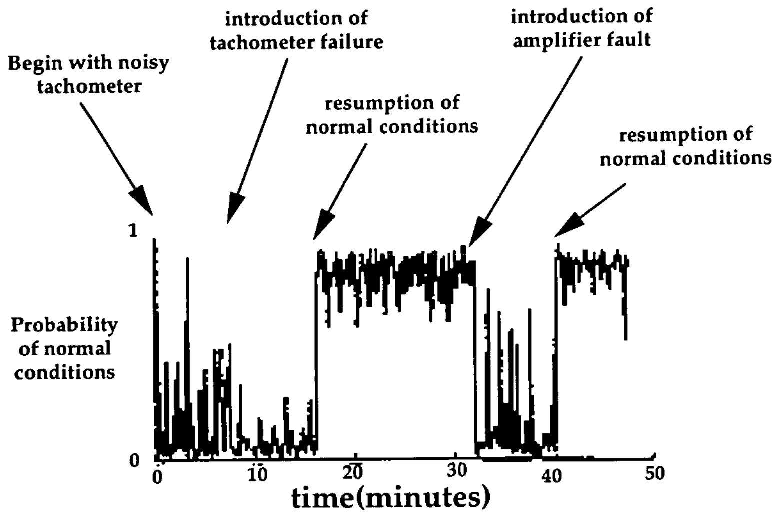


Figure B–6: Smyth's results when MLP classifier is used

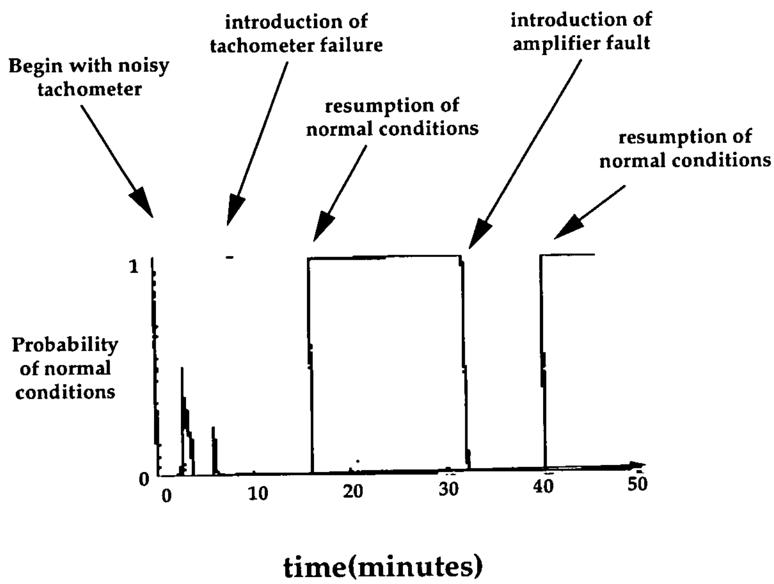


Figure B–7: Smyth's results when MLP-HMM combination is used

B.5 Summary

Hidden Markov models are an efficient method of modelling complex processes where it can be assumed that a system consists of a number of states. These may be working conditions or some system, for example condition monitoring systems; [115] or speech recognition applications where the words of speech are modelled by the HMM states [87,86] (i.e. the HMM models a certain number of words), the observations are the phonetic elements of the speech and the hidden element of the process encapsulates the knowledge that speech is highly inter-phoneme dependent. HMMs can also be combined with other signal processing methods, in particular artificial neural networks. These ANN-HMM hybrid systems use the ANN to incorporate the prior knowledge of the system states. The ANN output is the probability that the system is in a particular condition given the observations of the system which have been applied as the ANN inputs. The Hidden Markov Model can then incorporate the transition probabilities that the system is in that state given the previous state.

This type of hybrid approach is now in common use in speech recognition tasks [66,31] condition monitoring applications [88,53] and some medical signal monitoring applications [21,20]. It is felt that this method may be applicable to the domain investigated in this thesis and therefore it is suggested as a possible area for further investigation.

Appendix C
Publications

E Braithwaite, J Dripps and A F Murray "Neural Network Decision Support Device for Neonatal Intensive Care ", In *Proceedings of IPEMB* ,Leeds 1996 ,

E Braithwaite, J Dripps, A Lyon and A F Murray. "Classification of the onset of Respiratory Difficulties in ventilation assisted neonates", *IWANN*, 1997.

E Braithwaite, J Dripps and A F Murray. "Prediction of the onset of Respiratory Disorder in neonates", In *Proceedings of the International Conference on Neural Networks.* IEEE Computer Society Press, 1997.

E Braithwaite, J Dripps, A F Murray and A Lyon. "Neural Networks for prediction of Respiratory Disorder in ventilation-assisted Neonates", *Accepted for publication in International Conference on Artificial Neural Networks, Sweden 1998.*

Note: Only the three most recent publications are included in this appendix.

E Braithwaite, J Dripps and A F Murray "Classification of the onset of Respiratory Difficulties in ventilation assisted neonates ", *IWANN 1997*, In print, 1996.

Classification of the onset of Respiratory difficulties in ventilation assisted neonates

Emma Braithwaite, Jimmy Dripps, Andrew Lyon* and Alan F. Murray

Department of Electrical Engineering,
University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JL
Phone: +44 131 650 5665 (Fax: +44 131 650 6554) Emma.Braithwaite@ee.ed.ac.uk

Abstract. Intensive care units are designed to sustain the lives of the people being treated within them. However, often treatment decisions are made on a second by second basis and warning signs for other developing problems are missed. This paper describes a system which uses a multi-layer perceptron network to detect the deterioration in respiratory function of neonates who require artificial ventilation. It presents some results from the system and discusses the implications of using such a system in its current form.

1 Introduction

Intensive Care Units (ICUs) often use artificial respirators (ventilators) to assist patients to breathe. The neonatal ICU is no exception. Here, however the reason for ventilation is often associated with underdeveloped lungs (as a result of premature birth) as opposed to brain-damage. The problems associated with assisted breathing often remain the same regardless of the reason for ventilation and they will be collectively termed as respiratory disorders (RDs) for the purpose of this paper.
The paper describes a system which is being developed to detect the onset of RD in ventilated patients. The data which is being used has come from an neonatal ICU, however, the techniques developed here could easily be applied in any other ICU where artificial ventilation is carried out. The system has been designed to classify a developing problem early before more serious reprocussions occur and to allow clinicians an early warning to permit treatment.

1.1 Respiratory Disorder

Artificial respiration is designed to assist patients to breathe, either because of reasons of incapacitation (through accident) or incapability (underdeveloped lungs). It has some associated problems which must be diagnosed promptly for a number of reasons:

- If lungs are damaged for whatever reason they take a long time to heal, and in the case of the neonate (new-born baby), they may never heal properly if too much damage occurs[1]

* Dept of Child Life and Health, Edinburgh Royal Infirmary

- Any ventilation therapy plan must try to minimise potential damage while balancing the patient's need for oxygen as deprivation of O_2 for significant periods can result in brain damage and even death.
- Conversely, too much oxygen can result in damage to underdeveloped eyes [2].

All these problems can be caused by the ventilation therapy and are largely preventable as they develop gradually over a number of hours and often have simple solutions e.g. turning the oxygen concentration up or down.
During ventilation, air mixture is applied to the lungs through a tube which is passed down the trachea (throat). This endotracheal tube (ETT) requires periodic suctioning to remove a build-up of mucus from the lungs. If this is not done the mucus will continue to collect until the ET tube becomes blocked and the patient is starved of oxygen; this is one form of ventilator-induced RD. Another type of RD is that of a pneumothorax: this can form when too great an air pressure is applied to the lungs. This excess pressure can tear a small hole in the lining of the lung and a pneumothorax forms [3]. This again can deprive the patients of oxygen and its symptoms in terms of the physiological signals being measured are *very* similar to those of the blocked ETT.

1.2 Current system

Monitoring techniques used in the neonatal ICU are very similar to those used within any ICU. Every patient has his or her own dedicated medical staff who monitor progress both clinically and electronically using a number of purpose-built devices. In the Edinburgh neonatal ICU these monitors are linked to a PC (personal computer) based display. This system (called "Mary") [4] is capable of displaying a maximum of five physiological signals on the screen at any one time. The data is displayed on graphs of time (in seconds) against the magnitude of the signal. This means that conditions which develop over long periods are often missed as the signals associated with them are either not displayed or the changes take place too gradually to be noticed.
Currently diagnosis of a problem is carried out on a second-by-second basis by using a single screen of physiological signals and on examination of the patient. This means that conditions which develop over a long period (for example RD) are often missed in their early stages or diagnosed too late to prevent further damage to the neonate's lungs.

2 RD monitor

The development of this system has involved combining:

- Expert medical knowledge
- Intelligent data pre-processing
- An MLP classifier

2.1 Expert knowledge

After consultation with clinicians it was felt to be unnecessary to monitor all the available signals but only to use those in which the clinicians could, in hindsight, spot developments before the onset of RD. Table 1 gives a description of these developments in the isolated signals CO_2 and O_2. Initially blood pressure was also isolated but it was felt that as the change was of such short duration it did not merit inclusion at this time. Expert knowledge also implies that RD can develop over different periods of time; anything from one to four hours. It was decided that any feature extraction (isolating the important information in the signal) technique must take this into account. It should also be noted that treatment can take place without typical symptoms being present.

CO_2 concentration	Gradually increasing
O_2 concentration	Gradually decreasing
Blood Pressure	Initial Increase

Table 1. Typical signal trends before onset of RD

2.2 Intelligent data pre-processing

The RD monitor which is being developed is designed to generate a prediction of whether or not a patient is beginning to develop respiratory difficulties. The design of the predictor has included a number of stages (Fig 1):

- Pre-processing
- Feature Extraction
- Classification of the data

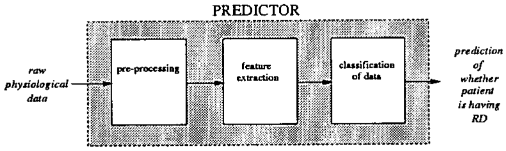


Fig. 1. Current prototype system

The objective of data pre-processing is to transform the data from its raw state into a form that will maximise a predictor's ability to discriminate between classes (problem developing, patient is stable). This implies that the new form of data representation must take account of as much information as is already known about the data as possible. In this case it was required that the pre-processing could describe trends within the data as this is what clinicians look for as indications of the onset of RD. As these trends appeared over long periods of time, anything from one to four hours, it was felt unnecessary to monitor the patient's progress second by second, and minute averages of the currently stored data (one second) were used, this removed some of the artifacts from the raw data and it also allowed the information carrying waveform to be observed. However, some outliers (signal spikes that were not physiologically possible) were still present in the resultant waveform. A filter was used to separate the noise and outlier carrying high frequency element of the signal from the low frequency trend information.

Filtering and outlier removal

A simple technique was used to extract this trend from the signal; a first order recursive mean estimator. It operates by running through the data in two directions, first forward then reverse. On the forward pass it estimates the current mean of the data, and, if the data lies outwith predefined bounds the current mean value is held (or "frozen"). In this way gross outliers representing physiologically and even pathologically impossible data are removed. On the reverse pass the calculations are checked and the delay introduced in the forward pass is cancelled. The resulting output is therefore an estimate of the mean of the data signal with zero time shift. An example of the filtered data is shown in Fig 2. This type of forward-backward filtering can only be used on historical data (ie in the generation of the training set), therefore the final system must use a different type of filtering for the raw physiological data so that it can be used in real-time.

Feature extraction

In this application some form of describing the trends hidden within the signal must be used to allow for both data compression and redundant data removal. Trends are required as that is what the clinicians use for their diagnostic processes. Gradient measures are calculated for both waveforms. Gradients are calculated at different points in time on the waveforms within a (fixed length) time window (see Fig 2). An input vector is therefore made up of gradients generated at a particular point in time for both the CO_2 and the O_2 traces. The starting point for these gradient calculations is taken as being the time where a comment pertaining to RD was entered onto the system. The difference was also taken between the CO_2 levels and the O_2 levels at the point of interest. These values were used as extra features in some of the tests. All these input vectors are then gathered together to form training and test sets for a Multi-Layer perceptron classifier.

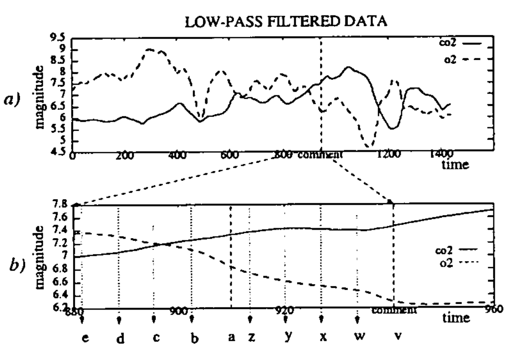
2.3 MLP Classifier

Multi-layer Perceptrons (MLPs) which are the most popular subset of Artificial Neural Networks provide a flexible method of parameterising a fairly general non-linear set of discriminant functions. In this instance the MLP is being used to classify the likelihood that a patient is suffering from some form of RD.

Data selection

To classify data correctly the MLP must be trained using a representative set of examples of the different types of feature vectors that it may encounter during operation. If this set is sufficiently complete and unbiased towards any class the MLP will be able, using its current "knowledge", to make a correct classification of previously unseen data. The training set must satisfy three important criteria:

- Large, there must be as many examples as possible



c)

Feature 1	Feature 2	Feature 3	Feature 4
a-b	b-c	c-d	d-e
v-w	w-x	x-y	y-z

Fig. 2. Method of gradient calculation a) example of filtered data b) expansion of filtered data to time interval of approx. 60 minutes c) calculation of features

- Representative, the set must contain examples of all the different types of problems the the final system might face
- Balanced, there must be approximately even numbers of all the classes which are to be discriminated amongst

The formation of the training set causes problems within medical (and most control) environments as patients spend the majority of their time in a "normal" condition and hence if all data was taken from one patient the training set would be heavily biased towards normality even to the point of being unable to classify abnormal patterns. The method used here takes examples of "normality" from one patient and an equal number of "RD developed" from a selection of other patients plus our patient to be studied. This allows generalisation for both normal and abnormal behaviour and it should produce a more complete description of both these possible states [5].

2.4 Current structure

In summary the current structure of the prototype system consists of a low pass filter, to isolate the trends in the two waveforms (CO_2 and O_2). These then have features extracted and combined. This feature vector is used as an input vector to an MLP which predicts the probability that the patient being monitored will develop RD in the next thirty minutes.

3 Experimental Method and Results

For this series of experiments training sets of data were generated using data taken from patients where this type of RD had occurred. This was determined by examining patient records stored as part of the "Mary" data-logging system. One patient was chosen for examination and where no action was taken by the staff in the treatment of the patient this period was classified as a period of "normality" for the patient. Where a comment, pertaining to the RD, was entered the period before this was assumed to be indicative of RD development.
A training set of example vectors was formed. This included a selection of input vectors which were generated at points where a comment was made and thirty minutes before. The training set contained equal numbers of feature vectors from periods of both "normality" and RD. As there are fewer numbers of RD in any patient file that periods of "normality" RD examples were used from other patients to balance the set. The features contained in this set were obtained both at isolated points were RD were diagnosed and thirty minutes before. Initially the resulting feature vector contained only gradient measures, however, later on this was extended to include the difference measure as well. Once training was complete the network was tested using a complete day's data. This means that for every point in time on a certain day gradients were calculated to describe its behaviour. The results for the initial tests including isolated gradient measures and then including difference measures are shown in Figs 4a and 4b respectively.
Upon examination of these results it was felt that the training set was too small to allow for any conclusions to be drawn, therefore it was extended. This was achieved by using sixty minute windows of data extending back from the comment. Gradients were calculated and the input vectors were assigned a classification of RD in the range between 0 and 1. A value of 1 means that the patient is suffering from RD, 0 means that there was no diagnosed RD. Where a diagnosis of RD was made a function for classification of RD was used see Fig 3. A 1 is entered from thirty minutes before the comment to where the diagnosis was made and from sixty minutes to thirty minutes before diagnosis the expected output was set to be a linearly increasing function in the range 0 to 1.
A window length of sixty minutes was chosen as it is the shortest period that RD usually develops over and as it has been decided that a thirty minute warning is required the output function was chosen to be this shape. The new training set therefore consisted of input vectors generated from these windows of data either those where at the end a comment was entered or where no action occured. Again initially the training set contained purely the gradient measures and latterly it contained difference measures included in the input vectors, see Fig 5a) and b) respectively when the network was tested on an entire day's worth of data.

3.1 Discussion

Both Fig 4 and Fig 5 show that the ANN's prediction of the patient's state correlates with the patient's condition as it was entered into the records. As an illustration, when

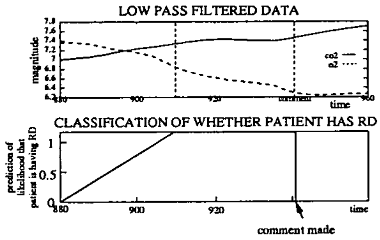


Fig. 3. Filtered data with example classification function

the patient's ET tube was suctioned a comment was entered and there should be some sign that the patient was beginning to suffer from a blocked tube by the probability of there being a problem increasing towards 1.0. As it can been seen from the figures this happens in most cases. There are large differences between the network outputs obtained from the four different systems. For example, in Fig 4 the output of the classifier is significantly smoothed and less eratic after the inclusion of the difference measure, this may be in part due to extra information being included in the input vectors. A similar change occurs in Fig 5 except in this case the output of the network is not only smoothed slightly but the range is increased as if it had become easier to separate the different categories being a looked for. However, it should be noted that the clinicians have stated that they do not feel that the use of the difference measure will enhance the system as they do not use it in their diagnostic processes. They also feel that the important information contained within it is already present in the gradient measures being used. It will however continue to be used as an extra feature for information purposes in the short term.

It should be noted however, that clinicians sometimes apply suctioning as a matter of course and none of the standard symptoms need be present. It should also be noted that the classifier is likely to indicate the likelihood of a problem for some time after a comment has been entered as it takes time for the patient's system to return to normal. The clinicians have also stated that the comments entered onto the system are often incomplete because of the treatment taking place at the time. This means that some treatments are never entered onto the system and this may explain the false positives which are being produced. This source of apparent errors will be eliminated by using data which is currently being collected in as complete a form as possible. The training data has also been applied to a single layer, linear network where the output classification on the test days had little correlation with patient behaviour at the time.

It is also clear that the output of the classifier is increasing to a maximum a significant amount of time (approx 40 minutes) before the diagnostic comment was entered, this suggested that the current type of classification function can be used as an early warning for the onset of RD.

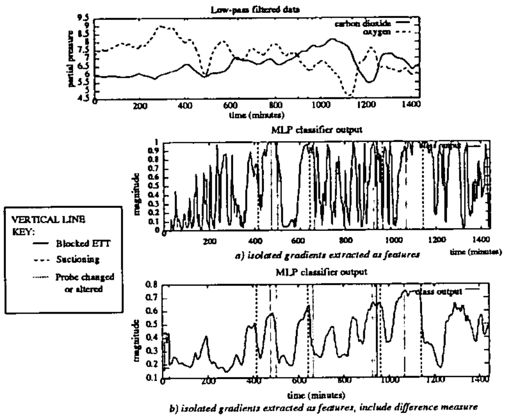


Fig. 4. Results from network trained on points of interest

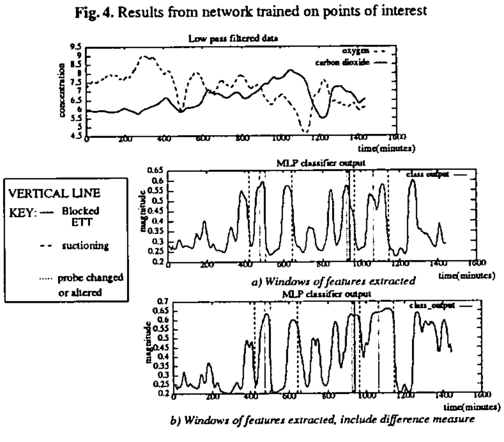


Fig. 5. Results from network trained on windows of interest

4 Further Work

This section will describe some of the adaptations and extensions to this work that are currently being investigated.

4.1 Prototype adaptations

The gradient measure metric used here has been based on a number of perhaps unjustified assumptions; for example that the minimum time taken to develop a blocked ETT is one hour, this may not always be the case and the final system must take this into account. The feature extraction technique will be re-examined and other methods tested, for example ARMA, ARX [6], difference measures [7] and overlapping gradient measures which have previously been used as features in optimised classification systems [8] to parameterise and model the raw waveforms.

4.2 Extensions to work

We plan to extend this work by adapting the prototype to take greater account of the temporal information involved in the system. This would "clean" up the system output by reducing the number of short-duration spikes in the output of the classifier. It is envisaged that this will will be achieved by using an MLP/HMM (Hidden Markov Model) hybrid classifier. The optimised MLP will operate exactly as before and its outputs will be used as estimates of prior probabilities in an HMM. The HMM will examine both the current output of the MLP and also its previous output. This will incorporate into the system the expert knowledge that a patient is unlikely to one minute be normal one minute and the next minute to be suffering from RD. The output of the HMM will be a new prediction of the patient's likelihood of developing RD and this is what will be used by the clinicians as an alarm. The HMM will "smooth" the MLP's output by ignoring obvious outliers and hence reduce the number of false alarms [9] [10].

5 Conclusions

This paper has detailed work that has been carried out within the Electrical Engineering department of Edinburgh University into the development of a prototype early warning system for the onset of respiratory disorder in neonates in ICUs. However, many limitations of the current system have been noted and further work will include improvements in the feature extraction techniques used and in the classification function. Despite the current system's success in its current form it could not be used as a diagnostic aid, however, with the enhancements which have been suggested being made it is hoped that the number of false positives will be reduced and that the value at the classifier's output can ultimately be used as a guide for the likelihood that the patient is starting to develop RD.

6 Acknowledgments

This work was funded by the EPSRC and was carried out in collaboration with the neonatal intensive care unit at Edinburgh Royal Infirmary, in particular the authors would like to thank Dr Andy Lyon, Prof Neil MacIntosh and Peter Badger for their help and advice. Emma Braithwaite is also grateful to the IEE Leslie H Paddle scholarship for its financial support.

References

1. I.R.Merkatz M.Hack, "Editorial: Preterm Delivery and Low Birth Weight– A Dire Legacy", *New England Journal of Medicine*, vol. 333, no. 26, December 1995.
2. A.Guyton, *Textbook in Medical Physiology*, chapter 42: Respiratory Insufficiency- Pathophysiology, Diagnosis, Oxygen Therapy, pp. 454–462, W.B.Saunders, 8th edition, 1956.
3. W.F.Ganong, *Review of Medical Physiology*, chapter 37, p. 541, Lange, 10th edition, 1981.
4. Hilary Kerr and Andrew Symon, "The Mary computer system- an introductory guide for unit staff", SMMP, 1993.
5. J.S.Taur and S.Y.Kung, "Prediction-Based Networks with ECG Application", in *International conference on Neural Networks*. IEEE, 1993, vol. 3, pp. 1920–1925.
6. M.H.Perrott and R.J.Cohen, "An Efficient Approach to ARMA modelling of Biological Systems with Multiple Input and Delays ", *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 1, September 1996.
7. T.Masters, *Neural, Novel & Hybrid Algorithms for Time Series Prediction*, John Wiley & Sons Inc, 1995.
8. P.Smyth, "Hidden Markov Models for Fault Detection in Dynamic Systems", *Pattern Recognition*, vol. 27, no. 1, pp. 149–164, 1994.
9. H.Bourlard and N.Morgan, "A Continuous Speech Recognition System Embedding MLP into HMM", in *Neural Information and Processing Systems*, David S. Touretzky, Ed., 1989, vol. 1, pp. 186–193.
10. P.Smyth and J.Mellstrom, "Fault diagnosis of antenna pointing systems using hybrid neural networks and signal processing techniques", in *Advances in Neural Information Processing Systems*. 1994, vol. 4, pp. 667–674, Morgan Kaufman Publishers inc.

E Braithwaite, J Dripps, A F Murray. "Prediction of the onset of Respiratory Disorder in neonates", In *ICNN*, 1997.

Prediction of onset of Respiratory Disorder in neonates

E. Braithwaite, J.Dripps and A.F.Murray
Department of Electrical Engineering,
University of Edinburgh
Mayfield Road, Edinburgh, EH9 3JL
Phone: +44 131 650 5665
(Fax: +44 131 650 6554)
Emma.Braithwaite@ee.ed.ac.uk

Abstract

Premature extremely sick babies are currently monitored by skilled medical staff using numerous dedicated non-invasive sensors and associated monitoring equipment. This paper describes a method of "fusing" a number of the physiological signals and, by examining them simultaneously and continously in time, produces an early warning for the onset of respiratory disorder (RD). The method uses a multi-layer perceptron neural network to produce probabilities that the patient is going to suffer from RD at some point within the next thirty minutes. Initial results from this classification system are shown and suggestions for further work are given.

1 Introduction

Babies which are born extremely premature are cared for within neonatal intensive care units (ICUs). ICUs are present in many large maternity hospitals and their purpose is to bring these premature infants to "term" without further mishap. Each of the babies treated in the ICU has its own set of problems, some of which are associated with its incomplete development. Patients are placed under twenty-four hour observation and clinicians treat immediate problems which occur, sometimes after they have been developing unnoticed for some period of time.

This paper describes an early warning system which is designed to detect the onset of respiratory disorder (RD), one of those conditions which develops over a long period. The system generates a prediction of RD development thirty minutes before it would normally be diagnosed and treated.

1.1 Respiratory Disorder

This problem is one of the most commonly occurring within any ICU. Ventilators are often used in ICUs to assist the patient to breathe and the neonatal ICU is no different. Here the problem being treated is that of underdeveloped and fragile lungs. If lungs are damaged for whatever reason they take a long time to heal, and in the case of the neonate (new-born baby), they may never heal properly if too much damage occurs. Any ventilation therapy plan must try to minimise potential damage while balancing the patient's need for oxygen as deprivation of O_2 for significant periods can result in brain damage and even death. Conversely, too much oxygen can result in damage to underdeveloped eyes [1]. These conditions can be caused by the ventilation therapy and are largely preventable as they develop gradually over a number of hours and often have simple solutions e.g. turning the oxygen concentration up or down. An early-warning system for this disorder would allow the current care procedures to be optimised and hence it would be of benefit to both medical staff and patients.

During ventilation, air mixture is applied to the lungs through a tube which is passed down the trachea (throat). This endotracheal tube (ETT) requires periodic suctioning to remove a build-up of mucus from the lungs. If this is not done the mucus will continue to collect until the ET tube becomes blocked and the patient is starved of oxygen; this is one form of ventilator-induced RD. Another type of RD is that of a pneumothorax: this can form when too great an air pressure is applied to the lungs. This excess pressure can tear a small hole in the lining of the lung and a pneumothorax forms [9]. This again can deprive the patients of oxygen and its symptoms in terms of the physiological signals being measured are *very* similar to those of the blocked ETT.

1.2 Current system

Monitoring techniques used in the neonatal ICU are very similar to those used within any ICU. Every patient has their own dedicated medical staff who monitor progress both

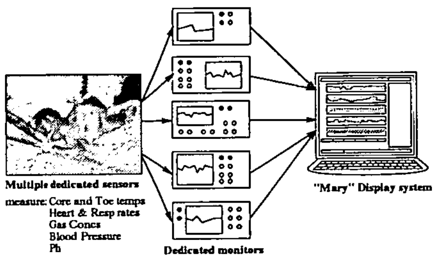


Figure 1. Single cot monitoring system

clinically and electronically using a number of purpose-built devices. As Fig 1 shows, in the Edinburgh neonatal ICU these monitors are linked to a PC (personal computer) based display. This system (called "Mary") [4] is capable of displaying a maximum of five physiological signals on the screen at any one time. The data is displayed on graphs of time (in seconds) against the magnitude of the signal. This means that conditions which develop over long periods are often missed as the signals associated with them are either not displayed or the changes take place too gradually to be noticed.

1.3 Current diagnostic techniques

Clinicians, when examining the patient, rarely take much notice of the onscreen "Mary" display. This may be due to the problems of time scale and the lack of complete information. It may also be due to the fact that many problems occurring within the ICU appear very quickly, for example, bradycardia and tachycardia (abnormal heartrates). Obviously these must be treated immediately to avoid serious repercussions. It is felt that improvement could be made in these diagnostic techniques if a system were designed which could monitor patients and, by taking into account their previous behaviour, predict their current or near-future condition.

2 RD monitor

The development of this system has involved combining:

- Expert medical knowledge
- Intelligent data pre-processing
- An MLP Classifier

2.1 Expert knowledge

After consultation with clinicians it was felt to be unnecessary to monitor all the available signals but only to use those within which the clinicians could, in hindsight, spot developments before the onset of RD. Table 1 gives a description of these developments in the isolated signals CO_2 and O_2. Initially blood pressure was also isolated but it was felt that as the change was of such short duration it did not merit inclusion at this time. Expert knowledge also implies that RD can develop over different periods of time-anything from one to four hours. It was decided that any feature extraction technique must take this into account. It should also be noted that treatment can take place without standard symptoms being present.

CO_2 concentration	Gradually increasing
O_2 concentration	Gradually decreasing
Blood Pressure	Initial Increase

Table 1. Typical signal trends before onset of RD

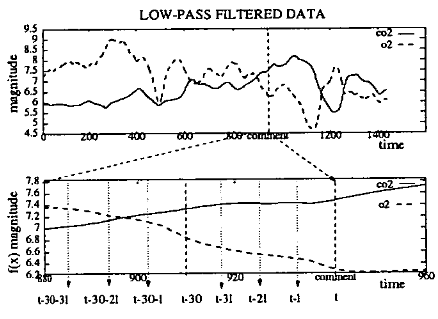
2.2 Pre-processing and Feature Extraction

The objective of data preprocessing is to transform the data from its raw state into a form that will maximise a classifier's ability to discriminate between classes. This implies that the new form of data representation must take account of as much information as is already known about the data as possible. In this instance it was required that the preprocessing could describe developmental trends within the data. As these trends appeared over long periods of time, anything from one to four hours, it was felt unnecessary to monitor the patient's progress second by second, and minute averages of the currently stored data (one second) were used.

Filtering and outlier removal

A simple technique was used to extract the low frequency trend from the signal; a first order recursive mean estimator. It operates by running through the data in two directions, first forward then reverse. On the forward pass it estimates the current mean of the data, and, if the data lies outwith predefined bounds the current mean value is held (or "frozen"). In this way gross outliers representing physiologically and even pathologically impossible data are removed. On the reverse pass the calculations are checked and the delay introduced in the forward pass is cancelled. The resulting output is therefore a estimate of the mean of the data signal with zero time shift. An example of the filtered data is shown in Fig 2.

Feature extraction



Feature 1	Feature 2	Feature 3
f(t)-f(t-l)	f(t-l)-f(t-2l)	f(t-2l)-f(t-3l)
f(t-30)-f(t-30-l)	f(t-30-l)-f(t-30-2l)	f(t-30-2l)-f(t-30-3l)

t is the time the diagnosis was made
l is a fixed time interval

Figure 2. Method of gradient calculation

In this application the trends indicative of RD must be extracted and hence gradient measures are calculated for both waveforms. Gradients are calculated at different points in time on the waveform within a (fixed length) time window (see Fig 2). An input vector is therefore made up of gradients generated at a particular point in time for both the CO_2 and the O_2 traces. The starting point for these gradient calculations is taken as being the time where a comment pertaining to RD was entered onto the system. It was decided that an early warning thirty minutes before the present action time would be an improvement on the current situation within the unit and hence input vectors were generated at this point as well. As the shortest development time for RD is sixty minutes, the minimum length of data the system might be given to classify is thirty minutes, hence the window over which classification must be achieved (and the gradients calculated) is thirty minutes. Input vectors were also generated one minute before a comment was entered as comments are always logged on the system *after* the event and this, if action is required can be at least a minute after diagnosis. All these input vectors were then gathered together to form training and test sets for a Multi-Layer perceptron classifier.

2.3 MLP Classifier

Artificial Neural Networks (ANNs), of which multi-layer perceptrons (MLPs) are a subset, accept an input vector, change its dimensionality through a "hidden" layer mapping and then further change its dimensionality at the output. This is a little like forcing the data through a bottleneck in order to select the features or parameters which are of most interest to the investigator. In this instance an MLP was used to classify the input vectors into one class: the probability that the baby was going to suffer from some form of RD within the next half an hour. The choice of an MLP allowed us to test the feasibility of using an ANN in this type of monitoring application.

Data selection

To classify data correctly the MLP must be trained using a representative set of examples of the different types of feature vectors that it may encounter during operation. If this set is sufficiently complete and unbiased towards any class, the MLP will be able, using its current "knowledge", to make a correct classification of previously unseen data. The formation of the training set causes problems within medical (and most control) environments as patients spend the majority of their time in a "normal" condition and hence if all data was taken from one patient the training set would be heavily biased towards normality even to the point of being unable to classify abnormal patterns. The method used here takes examples of "normality" from one patient and an equal number of "RD developed" from a selection of other patients plus our patient to be studied. This allows generalisation for both normal and abnormal behaviour and it should produce a more complete description of both these possible states [3].

2.4 Current structure

In summary the current structure of the prototype system is shown in Fig 3 and consists of a low pass filter, to isolate the trends in the two waveforms (CO_2 and O_2). These then have features extracted and combined. This feature vector is used as an input vector to an MLP which predicts the probability that the patient being monitored will develop RD in the next thirty minutes.

3 Experimental Method and Results

For this series of experiments training sets of data were generated using data taken from patients where this type of RD had occurred. This was determined by examining patient records stored as part of the "Mary" data-logging system. One patient was chosen for examination and where no action was taken by the staff in the treatment of the patient this period was classified as a period of "normality" for the patient. Where a comment, pertaining to the RD, was entered the period before this was assumed to be indicative of RD development.

The network was first trained using a selection of feature vectors from periods of both "normality" and RD from the

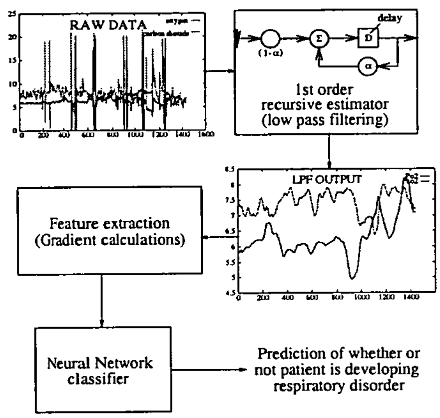


Figure 3. Current prototype system

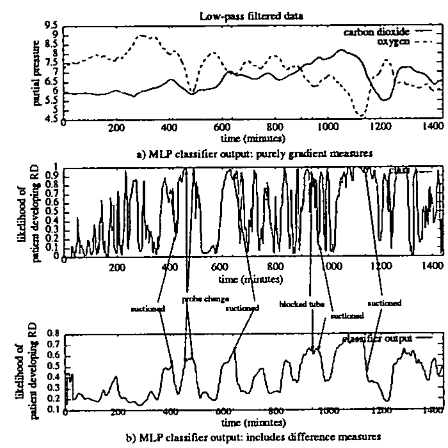


Figure 4. Results

different patients; there were equal numbers of each. Features were obtained both at the point were RD was diagnosed and thirty minutes before. The test set of data initially included a selection of features that had been obtained from similar regions on other days in the patient's stay (they were not included in the training set). The training set originally included a selection of input vectors which were generated at points where a comment was made and thirty minutes before. The test set was generated in the same way. It is extremely difficult to explicitly state how well the network performed on both training and test sets. This is partly due to the system being used which thresholded the output of the network to one or zero to determine this score. What is more interesting to examine is the output of the network and compare it to the actions that occured to the patient at a particular time on a day's worth of data as ultimately the test set included this. Results from this are shown in Fig 4. The training set was then extended and included input vectors from complete sixty-minute windows of data taken before a comment was entered. Another feature was also extracted, that of the difference between the two waveforms (CO_2-O_2). The system was then tested on a full day's data and the results obtained are shown in Fig 4.

3.1 Discussion

Fig 4 shows that the ANN's prediction of the patient's state correlates with the patient's condition as it was entered into the records. As an illustration, when the patient's ET tube was suctioned a comment was entered and there should be some sign that the patient was beginning to suffer from a blocked tube by the probability of there being a problem increasing towards 1.0. As it can been seen from the figure this happens in most cases. It should be noted however, that clinicians sometimes apply suctioning as a matter of course and none of the standard symptoms need be present. It should also be noted that the classifier is likely to indicate the likelihood of a problem for some time after a comment as been entered as it takes time for the patient's system to return to normal. In summary, these preliminary results shown are promising and warrant further investigation. It should be noted that at present it is impossible to rate the false alarm level of the system as the network has been run on retrospective data. However, this would be feasible if a clinician or diagnositician noted everytime they were concerned about a patient and not just when an action was taken. The network's output could then be compared to this. This can only be done prospectively and it is hoped that ultimately this can be achieved.

4 Further Work

This section will describe some of the adaptations and extensions to this work that are currently being investigated.

4.1 Prototype adaptations

The gradient measure metric used here has been based on a number of perhaps unjustified assumptions; for example that the minimum time taken to develop a blocked ETT is one hour, this may not always be the case and the final system must take this into account. The feature extraction technique will be re-examined and other methods tested, for example ARMA, ARX [5] and difference measures [8] which have previously been used as features in optimised classification systems [6] to parametise and model the raw waveforms.

4.2 Extensions to work

We plan to extend this work by adapting the prototype to take greater account of the temporal information involved in the system. This would "clean" up the system output by reducing the number of short-duration spikes in the output of the classifier. It is envisaged that this will be achieved by using an MLP/HMM (Hidden Markov Model) hybrid classifier. The optimised MLP will operate exactly as before and its outputs will be used as estimates of prior probabilities in an HMM. The HMM will examine both the current output of the MLP and also its previous output. This will incorporate into the system the expert knowledge that a patient is unlikely to be normal one minute and the next minute to be suffering from RD. The output of the HMM will be a new prediction of the patient's likelihood of developing RD and this is what will be used by the clinicians as an alarm. The HMM will "smooth" the MLP's output by ignoring obvious outliers and hence reduce the number of false alarms [2] [7].

5 Conclusions

This paper has detailed work that has been carried out within the Electrical Engineering department of Edinburgh University into the development of an early warning system for the onset of respiratory disorder in neonates in ICUs. However, many limitations of the current system have been noted and further work will include improvements in the feature extraction techniques used and in the classification function. The classifier will be extended to include temporal information which has hitherto been ignored. Despite the limitations of the prototype system the initial results show promise.

6 Acknowledgments

This work was funded by the EPSRC and was carried out in collaboration with the neonatal intensive care unit at Edinburgh Royal Infirmary, in particular the authors would like to thank Dr Andy Lyon, Prof Neil MacIntosh and Peter Badger for their help and advice. Emma Braithwaite is also grateful to the IEE Leslie H Paddle scholarship for its financial support.

References

[1] A.Guyton. *Textbook in Medical Physiology*, chapter 42: Respiratory Insufficiency- Pathophysiology, Diagnosis, Oxygen Therapy, pages 454–462. W.B.Saunders, 8th edition, 1956.

[2] H.Bourlard and N.Morgan. A Continuous Speech Recognition System Embedding MLP into HMM. In D. S. Touretzky, editor, *Neural Information and Processing Systems*, volume 1, pages 186–193, 1989.

[3] J.S.Taur and S.Y.Kung. Prediction-Based Networks with ECG Application. In *International conference on Neural Networks*, volume 3, pages 1920–1925. IEEE, 1993.

[4] H. Kerr and A. Symon. The Mary computer system- an introductory guide for unit staff. SMMP, 1993.

[5] M.H.Perrott and R.J.Cohen. An Efficient Approach to ARMA modelling of Biological Systems with Multiple Input and Delays . *IEEE Transactions on Biomedical Engineering*, 43(1), September 1996.

[6] P.Smyth. Hidden Markov Models for Fault Detection in Dynamic Systems. *Pattern Recognition*, 27(1):149–164, 1994.

[7] P.Smyth and J.Mellstrom. Fault diagnosis of antenna pointing systems using hybrid neural networks and signal processing techniques. In *Advances in Neural Information Processing Systems*, volume 4, pages 667–674. Morgan Kaufman Publishers inc, 1994.

[8] T.Masters. *Neural, Novel & Hybrid Algorithms for Time Series Prediction*. John Wiley & Sons Inc, 1995.

[9] W.F.Ganong. *Review of Medical Physiology*, chapter 37, page 541. Lange, 10th edition, 1981.

E Braithwaite, J Dripps, A F Murray and A.Lyon. "Neural Networks for prediction of Respiratory Disorder in ventilation-assisted neonates", Accepted for publication by *ICANN*, 1998.

Neural Networks for prediction of Respiratory Disorder in ventilation-assisted Neonates

Emma Braithwaite, Jimmy Dripps and Alan Murray
Department of Electronics and Electrical Engineering
University of Edinburgh
Edinburgh, U.K
tel: +44 131 650 5665, fax: +44 131 650 6554, email: eab@ee.ed.ac.uk

Andy Lyon
Department of Child Life and Health
University of Edinburgh
Edinburgh, U.K

Abstract

Medical Decision Support is an area of increasing research interest. It can underlie support for diagnosticians in everyday decision-making and can also alleviate habituation and fatigue - both of which can lead to errors, even in the most well-disciplined working environments. The neonatal intensive care unit (NICU) is a prime example.

Babies who are born extremely premature suffer from a number of conditions, in particular pulmonary (lung) function is not fully developed. These patients are placed on a variety of ventilators to assist them to breathe. Due to the extremely small size of the patient, problems which can occur to any ventilated patient are more common in neonates, for example blocked endotracheal tube and pneumothoraces. This paper describes a system which has been developed to detect the onset of respiratory difficulties in these neonates. The system uses a combination of signal processing, neural networks and expert knowledge to produce an evaluation of a patient's condition given their prior behaviour.

Results show that it is possible to diagnose these conditions earlier than currently. However, they also emphasise that this type of system is heavily reliant on both expert knowledge and the quality of data which is used.

1 Introduction

Babies born extremely premature require support to enable them to survive in the world *ex-utero*. One the most common forms of support which is required is that of maintaining the premature neonate's respiratory function by assisting him/her to breathe by using an artificial respirator or ventilator. The use of these artificial respirators often causes problems for clinicians in that the tubes supplying air to the patient can block easily (the tube is often less than 10mm in diameter) or too high a pressure applied to the supply of air may cause a tear in the fragile fabric of the neonate's lungs.

This paper describes a system which has been developed to detect the development of these two main problems of blocking tube and tearing lung. The system has been developed at Edinburgh and it has been designed to be used in combination with the current monitoring techniques in use in the NICU there.

2 Background

The current monitoring system at Edinburgh involves using a number of dedicated monitors which are linked to a cotside PC which logs physiological data every second. Data logging occurs automatically and physiological data can be annotated with clinical data to enable complete patient records to be kept. This data is stored for three days after which time, due to storage limitations, it is averaged every minute and archived. At Edinburgh there is currently archived minute-averaged data from over 2000 previous patients in the neonatal intensive care unit (NICU). Each of the incubator PCs are networked so that historical data on any patient can be accessed from anywhere in the unit [1].

The use of *Mary* (the PC logging system) although it improves the quality of record kept of a patient's time in the unit does not improve the diagnostic skills of the clinicians and carers in the unit. Due to the nature of the unit and the instability of the patients within it conditions are treated on a second-by-second or as-they-occur basis and little thought is given to problems which can develop over significant periods, for example a blocking air-tube in an artificially ventilated patient. Clinicians feel that a significant addition to the current monitoring process would be a system which can detect the development of these conditions while using the physiological data which is already taken as standard.

Expert knowledge about these respiratory problems exists in the clinicians can often, in hindsight, detect the onset of problems and can trace the problem's development. This development is most often seen in blood gas levels. Both Carbon Dioxide (CO_2) and Oxygen (O_2)are measured by a single probe, however in ventilated patients Oxygen levels can be artificially maintained by altering the fraction of inspired Oxygen (FiO_2) in the air mixture therefore any monitoring system for these conditions must include the use of all three of these physiological signals [2].

3 Design methodology

The methodology used for the development of the system was to incorporate as much expert knowledge as possible into its processing. The initial stages of this was to isolate the signals which were going to be used and using expert knowledge of their behaviour under the circumstances which were being investigated produce a system which will maximise this information. The system was therefore designed as follows, signals were isolated, raw data were filtered to remove artifact and outliers, filtered data were processed to maximise relevant information content, these data were then classified into two classes; 1) no concern/no problem developing 2) concern/suspected problem.

3.1 Signal isolation

Using expert knowledge it was decided that the three physiological signals which clinicians felt contained the most information relating to the development of respiratory problems would be used. Expert knowledge and experience suggested that these signals exhibited particular trends when a respiratory problem was occurring [3]. These trends are shown in Table 1. However, the raw physiological data signals often contained elements which were of little interest in this application and therefore it was required that the raw data be processed in some way to remove these artifacts.

Gas Type	Precursor
Carbon Dioxide PCO_2	increasing value
Oxygen PO_2	decreasing value
Fraction of Inspired Oxygen FiO_2	either stable or increasing

Table 1: Specific trends exhibited in gas concentrations

3.2 Filtering

It must be remembered that the data which are recorded by *Mary* are annotated with clinical commentary and this is time indexed. The time correspondence between the physiological data and the commentary must be maintained throughout the signal processing stages of the system otherwise the system's results cannot be fully evaluated. The filtering system which was developed involved a two-stage process, the first phase ran a low pass filter through the data, removing large outliers on the way. This processed introduced a time shift of thirty minutes in the data. This was eliminated when the data was run in the reverse direction through a variant of the original filter, the only difference being that the threshold limit for determining outliers was narrowed. The results of this process was a smooth waveform which approximated the long term trend in the physiological data while maintaining a time correspondence to clinical records. These data were now processed to further extract and enhance the trend information from the signal.

3.3 Feature extraction

The process of enhancing relevant information from a signal can be known as feature extraction. In this case it was decided that the information which was maximised must include temporal information about the patient's prior behaviour as well as its immediate trend. This was achieved using the technique shown in Figure 1. Historical information was included in the feature extraction process by including overall trends as well as shorter trends. The time period over which these calculations were made was chosen to be thirty minutes as diagnosis thirty minutes in advance of the current time would reduce the need for invasive treatment and prevent developing problems.
A data set was formed which contained exemplars of the two types of patient behaviour. These data sets were formed by analysis of the clinical entries to find records of diagnosis of the conditions under investigation. These time periods were assumed to be typical of the behaviour which indicated developing respiratory problems. It was more difficult to form a set of no-concern exemplars. Ultimately it was assumed that if no data entry was made for three hours in any patient record the patient's behaviour was causing no concern and therefore the central section of this period could be using as an exemplar of typical behaviour for a patient when behaving normally.
Using expert knowledge it was possible to state that as respiratory problems have significant development periods the period before a diagnosis is made must also be typical of behaviour which should evoke concern. In this way for every diagnosis of respiratory difficulty a series of exemplars, from the preceding sixty minutes, which should have caused a diagnosis to be made was also formed. Ultimately the data-set of exemplars which was formed contained over 6000 examples of patient behaviour which could be applied to the classification process.

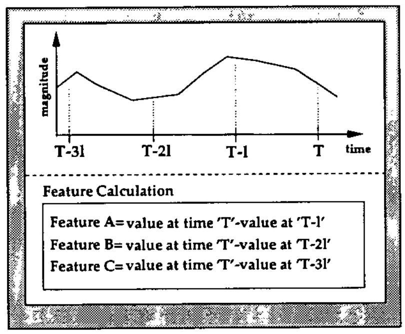


Figure 1: Feature extraction

3.4 Classification

Given that in a number of other medical signal monitoring applications multi-layer perceptrons (MLPs) have been used to classify sections of the signal [4],[5] it was decided that an MLP would be used in order to determine the feasibility of using this type of classifier for this application area.
An MLP neural network is a multi-layer system of computational nodes (or neurons) which model a system's behaviour through a series of weighted connections to a "hidden" layer. In this application the inputs applied to the MLP were the results of the feature extraction process. This means that the input vector was of nine dimensions (each physiological signal generated a three dimensional descriptor) and the output of the classification process was set to be one of two classes: no-concern or potential problem developing. The input data was also classified using a simple linear discriminant classifier to evaluate the MLPs performance.

4 Results

As it can be seen in Table 2 the MLP outperformed the linear classifier when trained on data from a number of patients. When the performance characteristics of the MLP classifier are examined (see Table 3) it can be seen that the system generalises better on the no-concern examples (high sensitivity rating). However in medical applications one of the most important measures is selectivity as this provides a measure of the number of false alarms the system would generate, a high value indicates few false alarms. For this application it can be seen that the MLP provides a relatively low false alarm rating. However, the MLP has been trained on exemplars of patient behaviour and a better indicator of how well it performs is to test it on an entire day of data. The results of these tests are shown in Figure 2.

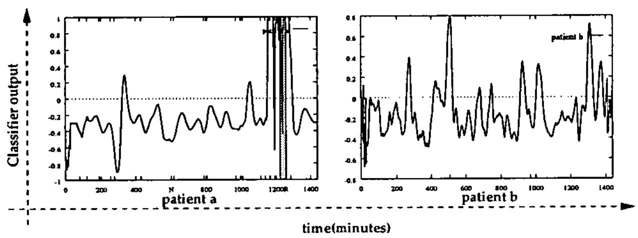


Figure 2: MLP Classifier output when tested on real-time, continuous data

The results show significant peaks in the output of the classifier. The clinical record of patient a) tells us that the patient was re-intubated where the "R" is entered on the x-axis and a peak occurs in the output of the classifier approximately thirty minutes prior to this entry. This suggests that the classifier is predicting the development of respiratory difficulties. In patient b) although there are peaks in the classifier's output there is no clinical record of respiratory problems. However, the initial peak coincides with a data entry of "All Care" and this may sometimes include a suctioning of the air tube being used. This may mean that a respiratory problem was developing and was undetected but prevented at this stage. The second peak cannot be explained in this manner as no record has been entered at all. However, it does highlight that clinical records are often incomplete, as routine treatments are often not entered, and therefore complete reliance cannot be placed on the methods used to generate the training and test sets.

Linear Classifier	MLP Classifier
64.76 %	69.04 %

Table 2: Comparison of Linear Classifier with MLP Classifier

Classification Rate	Sensitivity	Specificity	Selectivity
69.04 %	56.86 %	81.31 %	75.28 %

Table 3: Performance characteristics for MLP Classifier

5 Discussion

The results suggest that a system of this type is possible. However the results also raise a number of issues relating to data collection and hence to the categorisation of the data

used to train the classifier. In its current form the classifier relies heavily the clinical entries in a patient's records. Using these as a guide training and test sets are formed from a collection of exemplars of events and non-events. Often clinical processes are not entered onto a patient's records and therefore an exemplar which may be indicating developing respiratory problems might be used as an exemplar of non-concern.
It is therefore suggested that either clinical entries must be more accurate before the system can be further developed or if this is not possible another method of classification or data-analysis must be found to enable to accuracy of the results to be increased. These methods could include using unsupervised neural networks. These issues are currently under investigation at Edinburgh.

6 Conclusions

This paper has detailed work which has been carried out at Edinburgh University into the automatic diagnosis of respiratory problems in ventilation-assisted neonates. The work has involved using an MLP neural network in combination with other signal processing techniques to detect, before diagnosis is currently made, the development of common respiratory problems. Results are promising and show that such a system is possible. However it is noted that given the current monitoring processes and the lack of accuracy in the entry of clinical information the system could not be relied upon until the issues relating to the data collection are addressed.

7 Acknowledgements

This work was funded by the EPSRC and was carried out in collaboration with the neonatal intensive care unit at Edinburgh Royal Infirmary.

References

[1] Hilary Kerr and Andrew Symon, "The Mary computer system- an introductory guide for unit staff", SMMP, 1993.

[2] Emma Annette Braithwaite, *Neural Networks for medical condition prediction: an investigation of neonatal respiratory disorder*, PhD thesis, The University of Edinburgh, 1998, unpublished.

[3] Andy Lyon and Neil MacIntosh, ", Personal Communications, 1996.

[4] G.B.Moody S.G.Artis, R.G.Mark, "Detection of Atrial Fibrillation Using Artificial Neural Networks", *Computers in Cardiology*, pp. 173–176, 1992.

[5] B.H.Jansen, "Artificial Neural Nets for K-Complex Detection", *IEE Engineering in Medicine and Biology*, vol. 9, no. 9, pp. 50–52, September 1990.

References

[1] Louis-Giles Durand, Michel Blanchard, Guy Cloutier, Hani N.Sabbah, and Paul D.Stein, "Comparison of pattern recognition methods for computer assisted classification of spectra of heart sounds in patients with a porcine bioprosthetic valve implanted in the mitral position", *IEE Transactions on Biomedical Engineering*, vol. 37, no. 12, pp. 1121–1129, Dec 1990.

[2] E.Boersma, E.W.M.Grijseels, A.C.P.Maas, J.W.Deckers, J.A.M.Hartman and M.L.Simoons, "Improved Electrocardiographic Criteria for Confirmation of Acute Myocardial Infarction with Application in Pre-hospital Thrombolysis", *IEEE Computers in Cardiology*, pp. 725–728, 1995.

[3] M.R.S.Reddy, L.Edenbrandt, J.Svensson, W.K.Haisty, and O.Pahlm, "Neural network versus electrocardiographer and conventional computer criteria in diagnosing anterior infarction from the ECG", *IEE Proc. of computers in cardiology*, pp. 667–670, 1992.

[4] C.S.Goldschmidt A.Farnsworth, F.M.Chambers, "Evaluation of the PAPNET system in a general pathology service", *Medical Journal of Australia*, vol. 165, pp. 429–431, October 1996.

[5] R.Schmidt, B.Pollwein, L.Filipovici, and L.Gierl, "Adaptation and Abstraction as steps towards Case-based reasoning in the real Medical World: Case-based selection strategies for antibiotics therapy", In Greenes et al, editor, *MEDINFO 95*, pages 947–951, 1995.

[6] S.Snowden, K.G.Brownlee, and P.R.F.Dear, "An expert System to assist neonatal intensive care", *Journal of Medical Engineering & Technology*, vol. 21, no. 2, pp. 67–73, 1997.

[7] S.Cagoni, G.Coppini, R.Livi, R.Poli, P.T.Scarpelli, and G.Valli, "A neural network expert system for computer assisted analysis of blood pressure data", *Proceedings. Computers in Cardiology*, pp. 473–476, 1991.

[8] D.L.Hudson, M.E.Cohen, and M.F.Anderson, "Computer-assisted differential diagnosis and management", presented at *24th Annual Hawaii international conference on system sciences*, vol. 3, pages 218–226. IEEE, 1991.

[9] S.Cunningham, S.Deere, R.A.Elton, and N.McIntosh, "Neonatal physiological trend monitoring by computer", *International Journal of Clinical Monitoring and Computer*, vol. 9, pp. 221–227, 1992.

[10] D.P.de Bono. *Macleod's Clinical Examination*, chapter 5.The cardiovascular system, pages 81–124. Churchill Livingstone, Longman Group, 8th ed., 1991.

[11] S.Roberts and L.Tarassenko. *Applications of Neural Networks (ed: A.F.Murray)*, chapter Automated sleep EEG analysis using an RBF network, pages 305–322. Kluwer Academic Publishers, 1995.

[12] L.Kennedy R.F.Harrison, "From the laboratory to casualty: Neural Networks in the management of chest pain", presented at *Neural Networks- producing dependable systems conference*, pages 2.2–2.2.7. ERA Technology, November 1995.

[13] Anthony V. Sebald, "Use of neural networks for detection of artifacts in arterial pressure waveforms", presented at *11th International conference*, pages 2034–2035. IEEE Eng in Medicine and Biology Society, 1989.

[14] E.E.Kerre, "Outline of a fuzzy expert system for eeg diagnosis using fuzzy sets", *Artificial Intelligence in Medicine*, vol. 1, pp. 139–144, 1989.

[15] K.Becker, H.Kasmacher, G.Rau, G.Kalff, and H.J.Zimmermann, "A fuzzy logic approach to intelligent alarms in cardioanesthesia", presented at *Proceedings 3rd annual conference on Fuzzy Systems*, vol. 3, pages 2072–2076. IEEE, 1994.

[16] D.L.Hudson and M.E.Cohen, "Fuzzy logic in Medical Expert systems", *IEEE Engineering in Medicine and Biology*, pp. 693–698, November/December 1994.

[17] J.S.Taur and S.Y.Kung, "Prediction-Based Networks with ECG Application", presented at *International conference on Neural Networks*, vol. 3, pages 1920–1925. IEEE, 1993.

[18] NATO Advanced Reseacrh Workshop on Comparitive Time Series Analysis, *Time Series Prediction: Forecasting the Future and Understanding the past.* Addison-Wesley, 1994.

[19] D.G.Coast, "Segmentation of High Resolution ECGs using Hidden Markov Models", presented at *ICASSP*, vol. 1, pages 67–70. IEEE, 1993.

[20] D.Voudoris, M.G.Strintzis, N.Maglaveras, and C.Pappas, "Use of Hidden Marcov Models for the Analysis of Multiple Lead ECG Recordings", *Computers in Cardiology*, pp. 899–902, 1993.

[21] D.Coast, R.Stern, G.Cano, and S.Briller, "An Approach to Cardiac Arrythmia Analysis Using Hidden Markov Models", *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 826–836, September 1990.

[22] S.Cerutti, G.Chiarenza, D. Liberati, P. Mascellini, and G.Pavesi, "A Parametric Method of Identification of Single-Trial Event-Related Potentials in the Brain", *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 9, pp. 701–711, September 1988.

[23] V. Moret-Bonillo, A.Alonso-Betanzos, É.Garcia Martin, M. Cabrero Canosa, and B. Guijarro Berdinas, "The PATRICIA project", *IEEE Engineering in Medicine and Biology*, vol. 12, no. 4, pp. 59–68, December 1993.

[24] J.H.Frenster, "Expert Systems and Open Systems in Medical Artificial Intelligence", presented at *Proc Am Assoc Medical Systems and Informatics*, vol. 7, pages 118–120, Congress 89.

[25] B.Hayes-Roth, R.Washington, D.Ash, R.Hewett, A.Collinot, A.Vina, and A.Seirer, "Guardian: A prototype intelligent agent for intensive care monitoring", *Artificial Intelligence in Medicine*, vol. 4, pp. 164–185, 1992.

[26] S.W.Smye S.Snowden, K.G.Brownlee and P.R.F.Dear, "An advisory system for artificial ventilation of the newborn utilizing a neural network", *Medical Informatics*, vol. 18, no. 4, pp. 367–376, 1993.

[27] Tommi Sukuvaara, Aki Makivirta, Aarno Kavi, and Erkki Koski, "An intelligent intensive care alarming system", *Proceedings of computers in cardiology*, pp. 225–228, 1989.

[28] S.C.Lee, "Using a Translation-Invariant Neural Network To Diagnose Heart Arrythmia", In David Touretzky, editor, *Neural Information and Processing Systems*, vol. 1/2, pages 240–247, 1989.

[29] R.Bedi and J.T.E.McDonnell. Acoustic analysis of aortic bioprosthetic heart valves. unpublished, 1994.

[30] A.Casaleggio, M.Morando, and S.Ridella, "Neural network for automatic anomalous QRS complex detection", *IEEE Computers in Cardiology*, pp. 553–556, 1991.

[31] N.Morgan and H.Bourlard, "Continuous Speech Recognition", *IEEE Signal Processing Magazine*, vol. 12, no. 3, pp. 25–42, May 1995.

[32] F.T.Ulaby, F.Kouyate, B.Brisco, and T.H. Lee Williams, "Textural Information in SAR Images", *IEEE Transactions in SAR Images*, vol. GE-24, no. 2, pp. 235–245, March 1986.

[33] R.W.Conners, C.W.McMillin, K.Lin, and R.E.Vasquez-Espinosa, "Identifying and Locating Surface Defects in Wood: Part of an Automated Lumber Processing System", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 6, pp. 573–583, November 1983.

[34] M.Gorzalczany and M.McLeish, "Combination of neural networks and fuzzy sets as a basis for Medical Expert systems", presented at *IEEE Symposium on Computer Based Medical Systems*, vol. 5, pages 412–420. IEEE, 1992.

[35] Evangelia Micheli-Tzanakou, "Biomedical applications of parallel processing and artificial neural networks", *Proceedings of the 1992 International biomedical engineering days*, 1992.

[36] M.Nayagam O.A.Nasseem Husain, E.B.Butler, L.Mango, and A.Alonzo, "An analysis of the variation of human interpretation: $Papnet^{TM}$ a mini-challenge", *Analytical Cellular Pathology*, vol. 6, no. 2, pp. 157–163, 1994.

[37] Beryl Corner, *Prematurity: The Diagnosis, Care and Disorders of the Premature Infant*. Cassell:London, 1st ed., 1960.

[38] Scottish Home and Scottish Health Service Advisory Council Health Department, *Neonatal Care in Scotland*. HMSO, 1990.

[39] Cropp, L.Woods, Raney, and D.Bredle, "Name that tone : The proliferation of alarms in the Intensive Care Unit", *CHEST*, vol. 105, no. 4, pp. 1217–1220, April 1994.

[40] F.Benini, V.Magnavita, P.Lago, E.Arslan, and P.Pisan, "Evaluation of Noise in the Neonatal Intensive Care Unit", *American Journal of Perinatology*, vol. 13, no. 1, pp. 37–41, January 1996.

[41] M.Topf and E.Dillon, "Noise-induced stress as a predictor of burnout in critical care nurses", *Heart and Lung*, vol. 17, no. 5, pp. 567–574, 1988.

[42] G.Grumet, "Sounding Board:Pandemonium in the Modern Hospital", *New England Journal of Medicine*, vol. 328, no. 6, pp. 433–473, Feb 1993.

[43] A.Meyerfalcke, R.Rack, and F.Eichwede, "How noisy are anaesthesia and intensive care medicine? Quantification of the patients' stress", *European Journal of Anaesthesiology*, vol. 11, no. 5, pp. 407–411, 1994.

[44] P.Rolfe, "Neonatal critical care monitoring", *Journal of Medical Engineering & Technology*, vol. 10, no. 3, pp. 115–120, May/June 1986.

[45] A.Guyton. *Textbook in Medical Physiology*, chapter 83: Fetal and Neonatal Physiology, pages 929–938. W.B.Saunders, 8th ed., 1956.

[46] L.I.G.Worthley. *Synopsis of Intensive Care Medicine*, chapter 32: Adult respiratory distress syndrome, pages 331–338. Churchill/Livingstone, 1994.

[47] L.I.G.Worthley. *Synopsis of Intensive Care Medicine*, chapter 90:Paediatric respiratory failure, pages 895–910. Churchill/Livingstone, 1994.

[48] A.Guyton. *Textbook in Medical Physiology*, chapter 42: Respiratory Insufficiency-Pathophysiology, Diagnosis, Oxygen Therapy, pages 454–462. W.B.Saunders, 8th ed., 1956.

[49] S.D.Ghazanshahi and M.C.K.Khoo, "Optimal Application of High-Frequency Ventilation in Infants: A Theoretical Study", *IEEE transactions on Biomedical Engineering*, vol. 40, no. 8, pp. 788–796, August 1993.

[50] D.W.Hill and A.M.Dolan, *Intensive Care Instrumentation*. Medical Physics Series. Academic Press, 1982.

[51] S.Roberts and L.Tarassenko, "New method of automated sleep quantification", *Medical and Biomedical Engineering and Computing*, vol. 30, pp. 509–517, 1992.

[52] L.Dodd, N.Dodd, and M.Kenrick, "Neural Network alarm monitoring in intensive care wards", presented at *Neural Networks- producing dependable systems conference*, pages 2.1.1–2.1.5. ERA Technology, November 1995.

[53] P.Smyth, "Hidden Markov Models for Fault Detection in Dynamic Systems", *Pattern Recognition*, vol. 27, no. 1, pp. 149–164, 1994.

[54] P.Smyth, "Detecting novel fault conditions with Hidden Markov Models", presented at *Int. Conf. Pattern recognition in Practice IV*, pages 525–536. Elsevier, June 1994.

[55] P.Smyth and J.Mellstrom, "Fault diagnosis of antenna pointing systems using hybrid neural networks and signal processing techniques", presented at *Advances in Neural Information Processing Systems*, vol. 4, pages 667–674. Morgan Kaufman Publishers inc, 1994.

[56] L.F.Lollar S.C.Lee, "Development of a component centered fault monitoring and diagnosis knowledge based system for Space Power system", presented at *International Energy Conservation Engineering Conference*, pages 377–382, 1988.

[57] C.Lee G.B.Jasmon, L.H.Callistus, "Prediction of Voltage Collapse in Power Systems using a reduced system model", presented at *International Conference on Control*, vol. 1, pages 32–36. IEEE, 1991.

[58] D.L.Morrison and D.M.Upton, "Fault Diagnosis and Computer Integrated Manufacturing Systems", *IEEE Transactions on Engineering Management*, vol. 41, no. 1, pp. 69–83, January 1994.

[59] A.Kundu and G.C.Ghen, "An Integrated hybrid neural network and hidden Markov model classifier for sonar signal classification", presented at *ICASSP*, vol. 5, pages IV/3587-3509. IEEE, 1995.

[60] Sigeru Omatu and Tomoji Yoshida, "Pattern classification for remote sensing using a neural network", *IGARRS: International Geoscience and remote sensing symposium*, vol. 2, pp. 899–901, 1993.

[61] L.Devillers and C.Dugast, "Hybrid system combining expert-TDNNS and HMMs for Continuous Speech Recognition", presented at *ICASSP*, vol. 2, pages II/165–168. IEEE, 1994.

[62] H-P. Hutter, "Comparison of a new hybrid connectionist-SCHMM approach with other hybrid approaches for speech recognition", presented at *ICASSP*, vol. 5, pages IV/3311–3314. IEEE, 1995.

[63] W.Reichl and G.Ruske, "A Hybrid RBF-HMM system for continuous speech recognition", presented at *ICASSP*, vol. 5, pages IV/3335–3338. IEEE, 1995.

[64] J.Dai, "Hybrid approach to speech recognition using hidden Markov models and Markov chains", *IEE Proc.-Vis.Image Signal Processing*, vol. 141, no. 5, pp. 273–279, 1994.

[65] W.Y.Huang and R.P.Lippman, "HMM Speech Recognition with Neural Net Discrimination", In David Touretzky, editor, *Neural Information and Processing Systems*, vol. 1/2, pages 194–202, 1989.

[66] G.Zavaliagkos, Y.Zhao, R.Schwartz, and J.Makhoul, "A Hybrid Segmental Neural Net/Hidden Markov Model System for Continuous Speech Recognition", *IEE Transaction on Speech and Audio Processing*, vol. 2, no. 1, pp. 151–160, January 1994.

[67] D.S.Hickey, D.J.Prendergast, and H.J.Roberts, "Metapattern recognition in medical diagnosis", presented at *Colloqium on AI in medical decison making*, vol. 62, pages 1–2. IEEE, 1990.

[68] C.N.Schizas, C.S.Pattichis, and L.T.Middleton, "A new approach to medical diagnosis", *Proceedings of the 1992 International Biomedical Engineering Days*, pp. 207–212, 1992.

[69] M.Chudleigh, C.Berridge, R.May, J.Butler, and I.Poole, "The SADLI project: Safety critical software research in the medical diagnostic domain", *IEE Computing and Control*, vol. 6, no. 5, pp. 211–215, October 1995.

[70] L.Tarassenko, "Diagnostic systems application 1", presented at *5th Annual Cambridge Neural Networks Summer School*, vol. 2, 1995.

[71] Metin Akay, Yasemin M.Akay, and Walter Welkowitz, "Neural Networks for the diagnosis of coronary heart disease", *IJCNN International Joint Conference on Neural Networks*, vol. 2, pp. 419–424, 1992.

[72] S.G.Artis, R.G.Mark, and G.B.Moody, "Detection of Atrial Fibrillation Using Artificial Neural Networks", *Computers in Cardiology*, pp. 173–176, 1991.

[73] Abdalla S.A.Mohamed and Hazem M.Raafat, "Recognition of heart sounds and murmurs for cardiac diagnosis", *9th International Conference on Pattern Recognition*, vol. 2, pp. 1009–1011, 1988.

[74] W.H.Nailon, S.McLaughlin, T.Spencer, N.W.W.McDicken, and K.A.A.Fox, "Can statistical texture analysis of unprocessed intravascular ultrasound (IVUS) signal dicriminate red and white thrombi acid plasma", *Circulation*, vol. 94, no. 8 SS, pp. 3587, 1996.

[75] Teresia Ramdén. On Condition Monitoring of Fluid Power Pumps and Systems. Master's thesis, Linköping University, Sweden, 1995. Licentiate Thesis.

[76] E.Ayanoglu, "Robust and fast failure detection and prediction for fault tolerant communication links", *Electronics Letters*, vol. 28, no. 10, pp. 940–941, 1992.

[77] H.Z.Liu, H.M.Wang, and X.Q.Ni, "Fault Diagnostics on Jet Engine Starting", presented at *ICIASF '91*, pages 392–396, 91.

[78] J. Palmberg T.Ramdén, P.Krus, "Fault Diagnosis of Complex Fluid Power Systems Using Neural Networks", presented at *4th Scandinavian Conference on Fluid Power*, Tampere, Finland, 1995.

[79] E.M.Strand and W.T.Jones, "Tracking Ventricular Arrythmias with an Artificial Neural Network", *Proceeding of Computers in Cardiology*, pp. 291–294, 1990.

[80] M. Hintz-Madsen, L.K. Hansen, J.Larsen, E.Olesen, and K.T.Drzewiecki, "Detection of malignant melanoma using neural classifiers", presented at *Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN)*, June 1996.

[81] H.Bourlard, "Neural Networks: Theory and applications to Speech recognition", presented at *5th annual Neural Networks summer school*, pages 1–65. International Computer Science Institute, 1995.

[82] H.Bourlard, "Applications to speech recognition", presented at *5th Annual Cambridge Neural Networks Summer School*, vol. 2, 1995.

[83] H.E.Bassil, J.H.Dripps, and P.M.Grant, "Detection and correction of outliers in foetal heart rate time series", *Electronics Letters*, vol. 28, no. 4, pp. 382–383, Feb 1992.

[84] M.H.Perrott and R.J.Cohen, "An Efficient Approach to ARMA modelling of Biological Systems with Multiple Input and Delays ", *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 1, pp. 1–14, September 1996.

[85] H.Bourlard and N.Morgan, "A Continuous Speech Recognition System Embedding MLP into HMM", In David S. Touretzky, editor, *Neural Information and Processing Systems*, vol. 1/2, pages 186–193, 1989.

[86] L.Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–285, February 1989.

[87] C.J.Leggetter and P.C.Woodland. Speaker Adaptation of HMMS using Linear Regression. Technical Report CUED/F-INFENG/TR.181, Cambridge University Engineering Department, June 1994.

[88] L.P.Heck and J.H.McClellan, "Mechanical system monitoring using Hidden Markov models", presented at *Proceedings IEEE Internation conference on Acoustics, Speech and Signal Processing*, pages 1697–1700. IEEE, 1991.

[89] V.Tresp, J.Moody, and W.R.Delong. Neural Network modelling of physiological processes. In S.J.Hanson:T.Petsche:M.Kearns:R.Rivest, editor, *Computational Learning Theory and Natural Learning Systems*, vol. II: Intersections between Theory and Experiment, chapter 21, pages 363–378. MIT Press, 1994.

[90] W.F.Ganong. *review of Medical Physiology*, chapter 34-Pulmonary Function, pages 507–519. LANGE, 10 ed., 1981.

[91] Rosanna Degani and Giovanni Bortolan, "Fuzzy decision making in electrocardiography", *AI in medicine*, vol. 1, pp. 87–91, 1989.

[92] C.C.Lee, "Fuzzy logic in control systems: Fuzzy logic controller- part i", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 2, pp. 404–418, March/April 1990.

[93] C.C.Lee, "Fuzzy logic in control systems: Fuzzy logic controller- part ii", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 2, pp. 419–435, March/April 1990.

[94] L.Tarassenko, A.Nairac, and N.Townsend, "Novelty Detection", presented at *Neural Computing -The Key Questions*, Bath, September 1997. NCAF.

[95] J.Dai, "Isolated word recognition using Markov chain models", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, pp. 458–463, 1995.

[96] C.A.Cruz, *Understanding Neural Networks : a primer*. Arlington press, 1991.

[97] Simon Haykin, *Neural networks: A comprehensive foundation*. MacMillan College Publishing Company, 1st ed., 1994.

[98] R.Beale and T.Jackson, *Neural Computing : an introduction*. Institute of Physics publishing, 1994.

[99] B.H.Jansen, "Artificial Neural Nets for K-Complex Detection", *IEE Engineering in Medicine and Biology*, vol. 9, no. 9, pp. 50–52, September 1990.

[100] E.Hatzipantelis, A.Murray, and J.Penman, "Comparing Hidden Markov Models with artificial neural network architectures for condition monitoring applications", presented at *Artificial Neural Networks*, pages 369–374. 4th Int Conf on Artificial Neural Networks, 1995.

[101] G.A.Rovithakis and M.A.Christodoulou, "Adaptive Control of Unknown Plants Using Dynamical Neural Networks", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 3, pp. 400–412, March 1994.

[102] L. Ohno-Machado and M.A.Musen, "Hierarchical Neural Networks for Partial Diagnosis in Medicine", presented at *World Congress in Neural Networks*, vol. I, pages 291–296, 1994.

[103] S.Roberts and L.Tarassenko, "A probabilistic resource allocating network for novelty detection", *Neural Computation*, vol. 6, no. 2, pp. 270–284, 1994.

[104] K.Swingler and L.S.Smith, "Producing a Neural Network for monitoring driver alertness from steering actions", *Neural Computing and Appliications*, vol. 4, no. 2, pp. 96–104, 1996.

[105] Z.Zeng, R.Goodman, and P.Smyth, "Self-clustering recurrent networks", presented at *International conference on Neural Networks*, vol. 1, pages 33–35. IEEE, 1993.

[106] JM.Kuo, J.C.Principe, and B.deVries, "Prediction of Chaotic Time Series using recurrent neural networks", presented at *Neural Networks for Signal Processing II*. IEEE Workshop on Signal Processing, 1992.

[107] P.Smyth, "Markov monitoring with unknown states", *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 9, pp. 1600–1612, December 1994.

[108] C.M.Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1st ed., 1996.

[109] A.D.Kulkarni. *Artificial Neural Networks for Image Understanding*, chapter 5: Supervised Classifiers, pages 93–108. Van Nostrand Rienhold, 1994.

[110] J.R.Chen and P.Mars, "The Feasibility of Using MLP Neural Networks for System Identification", presented at *IEE Colloquim on Neural Networks for systems: Principles and Applications (Digest No 19)*, vol. 3, pages 1–3. IEE, 1991.

[111] W.J.Stewart, *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.

[112] L.Rabiner, "An Introduction to Hidden Markov Models", *IEEE ASSP Magazine*, pp. 4–16, January 1986.

[113] H.A.Bourlard and N.Morgan, *Connectionist Speech Recognition-a Hybrid Approach*. Kluwer Academic Press, 1994.

[114] Y.Chauvin P.Baldi, "Hybrid Modeling, HMM/NN Architectures, and Protein Applications", *Neural Computation*, vol. 8, pp. 1541–1565, 1996.

[115] E.Hatzipantelis and J.Penman, "The use of Hidden Markov models for condition monitoring electrical machines", presented at *6th Int. Conf. on Electrical machines and Drives*, pages 91–96. IEE, 1995.