Reservoir Computing approaches to EEG-based Detection of Microsleeps

Sudhanshu Ayyagari

Department of Electrical and Computer Engineering

A thesis presented for the degree of

Doctor of Philosophy

University of Canterbury

Christchurch, New Zealand

February, 2017

"Facts are stubborn, but statistics are more pliable"

- Mark Twain

Abstract

Long-haul truck drivers, train drivers, and commercial airline pilots routinely experience monotonous and extended driving periods in a sedentary position, which has been associated with drowsiness, microsleeps, and serious accidents. Consequently, the detection and preferably prediction of the microsleeps in subjects working in these high-risk occupations is very important to workplace safety. Therefore, the aim of this project was EEG-based characterization and early detection of microsleeps during a sustained attention task. The overall approach was to identify reliable physiological cues of lapses of sustained attention and microsleeps, to develop a microsleep system which could be used to detect, or better yet, predict the onset of microsleeps in real time and trigger an alert to rouse the user from an impending microsleep. The main motivation of this project was to develop a state-of-the-art lapse detection system by employing novel classifier schemes based on reservoir computing (RC), specifically echo state networks (ESNs) with cascaded-leaky-integrator-neurons and liquid state machines (LSM) to increase current benchmark performances on microsleep detection.

This is the first project and study to have implemented and evaluated EEG-based microsleep detection using RC models for the characterization and detection of microsleeps from the underlying EEG. Moreover, the novelty of the ESN-based cascaded-leaky-integrator neuron approach is in its simplicity (as networks with only 8 or less neurons could achieve optimal performance) and its superior microsleep detection performance.

In this research, previously collected behavioral EEG data from fifteen healthy male, non-sleep-deprived volunteers performing a 1D-visuomotor tracking task for 1 hour, was used to form classifier models capable of detecting microsleeps with second-scale resolution. The performance of the microsleep detector was measured both in terms of its ability to detect the lapses-of-responsiveness states and microsleep states (in 1-s epochs). The previous lapse detection benchmark performance on this data, used a simple linear discriminant analysis (LDA)-based classifier, fitted with a meta-learner model. This LDA-based system reported the best performance in terms of its mean phi correlation (φ) = 0.39± 0.06, receiver operator characteristics. An epoch length of 2 s and an overlap window of 1 s (50%) between successive epochs were used in the analysis (AUC-ROC) = 0.86 ± 0.03, and precision recall (AUC-PR) = 0.43± 0.09.

Models based on EEG power spectra, and power in the traditional bands, were used to detect the changes in the EEG during microsleeps. Normalized EEG epochs with z-scores > 30 were excluded from analysis, resulting in rejection of 8.3% of the epochs. This process was referred

to as data pruning. Reduced features from 6 independent feature reduction schemes including, principal components analysis (PCA), kernel PCA (KPCA), probabilistic PCA (PPCA), symmetric neighbourhood embedding (SNE), Nearest neighbour embedding (NNE), and stochastic proximity embedding (SPE) were passed as an input to the classifier models. Classifier models evaluated included the RC-based models including the ESNs with sigmoidal neurons, cascaded ESNs with leaky-integrator neurons and LSMs. The RC-based models were compared to other standard classifier models, such as, support vector machines with polynomial kernel (SVMP), LDA, spiky neural networks (SNN), and k-nearest neighbour (KNN) classifiers.

Best microsleep detection was achieved using cascaded ESNs with cascaded-leaky-integrator neurons and 50-60 PCs from PCA of the overall 544 power spectral features. This configuration resulted in $\varphi = 0.51 \pm 0.07$ (mean ± SE), AUC-ROC = 0.88 ± 0.03, and AUC-PR = 0.44 ± 0.09. LSM-based detectors had a lower performance of $\varphi = 0.42 \pm 0.06$, AUC-ROC = 0.83 ± 0.03, and AUC-PR = 0.43±0.06, compared to the cascaded-leaky-ESN approach. The PCA-based feature reduction modules showed the highest overall performances of the 6 feature-reduction schemes evaluated. This high performance of PCA-modules was found on all classifier schemes. PPCA-based methods followed the PCA schemes in terms of the best microsleep detection performances. Analysis also showed that creating multiple microsleep detection models (ensemble learning) and combining them to form an overall detector resulted in an improvement in performance over a single classifier model. Microsleep detection was also found to have higher accuracy than the other metrics of flatspots, video microsleeps and definite microsleeps.

To study the effect of pruning the data, performances were determined for the classifiers when presented with unpruned data in its entirety for training. Performance was compared with a previous study which used a long short-term memory (LSTM) recurrent neural network (RNN) for which $\varphi = 0.38 \pm 0.07$, AUC-ROC = 0.84 ± 0.02, and AUC-PR = 0.41 ± 0.08). Similar to the pruned datasets, ESNs with cascaded-leaky-integrator neuron models outperformed all the other classifier schemes and set a new benchmark for EEG-based microsleep detection of $\varphi = 0.44 \pm 0.06$, AUC-ROC = 0.88 ± 0.04, and AUC-PR = 0.45 ± 0.09. This performance, albeit lower than for the pruned datasets, is deemed the best overall performance for microsleep detection as it was for the full behavioural dataset.

In summary, the cascaded-leaky-integrator-ESN approach has provided a new benchmark performance for microsleep detection, which is significantly higher ($p = 0.012$) than by all previous approaches. Notwithstanding, the performance of these EEG-based microsleep detection systems is still considered to be modest. Further research is needed to identify

additional cues in the EEG leading to devices capable of more accurate detection and prediction of microsleeps.

"Science is not about building a body of known 'facts'.
It is a method for asking awkward questions and subjecting them to a reality-check,
thus avoiding the human tendency to believe whatever makes us feel good."
- Terry Pratchett

Acknowledgements

When I began this PhD I was devoid of all confidence, background and good sense, and accordingly gave myself even odds on completing it; that it ever arrived at the binders is very much a reflection on the people I have had around me.

Accordingly, over the last four years I feel privileged to have 'cut my teeth' in a good portion of very exciting science and feel satisfied with the challenges that I have overcome. I am glad that I have undertaken this ginormous task as it has allowed me to rebuild several aspects of my mental self and hence accomplish the goals I had set out for myself when I started in this programme. None of it would however be possible if I was not lucky enough to be surrounded by some very special and talented people. I describe a few of them here, who have helped me considerably through this journey and also made it memorable and adventurous.

I am enormously grateful to my advisors Dr Steve Weddell and Prof Richard Jones for their time, advice and guidance. Over the past four years, Steve has served at various times as a good boss, a mentor, a collaborator, a middleman and a friend. Four years is a long time for anything, and I do not think I would have made it this far without Steve, who kept on believing in me long before I have. In addition to our academic collaboration, I greatly value the close personal rapport that Steve and I have forged over the years. I have great difficulty imagining a kinder or more encouraging individual.

Richard with his leadership, guidance, meticulous attention to detail and friendly encouraging attitude, has allowed me to work hard and take pride in my work. Richard balanced his careful, thorough scepticism of every word, figure, and thought I presented him with a strong undercurrent of encouragement and support, which improved me greatly as an independent researcher and a young professional. I would also like to thank both of you for tolerating, and sometimes even supporting, the distracting side projects I tend to succumb to, and always letting me find my own way back to research productivity.

While on the subject of kind and thoughtful advisors, I have been fortunate enough to have had a host of wonderful mentors:

- Prof Phil Bones for his constant encouragement, support and advice throughout this PhD programme. Phil's insights, both practical and philosophical, have greatly influenced me as a young researcher. Needless to say I am extremely grateful to Phil for organizing and chairing my oral confirmation proceeding.

- Prof Mark Billinghurst and the Human Interface Technology Laboratory at the University of Canterbury for offering me the opportunity to work as a research assistant on the 'Cosense' project. The valuable research insights I have picked up from Mark during our collaboration were extremely helpful during the latter phases of my PhD programme. It still amazes me that I had such good luck in stumbling upon someone whose style and approach suited me so perfectly. I would also like to thank the SAMSUNG Thinktank group for funding this project.

I would also like to express my gratitude to all the people who have enriched my stay at the Electrical Engineering department and helped me in so many ways. In particular, I would like to thank my friends and colleagues, Mr. Reza Shoorangiz, Ms. Nellie Sibaeva, Mr Paul Agger and Mr David Healy for their infectious energy, friendship and constant encouragement.

Finally, I am eternally indebted to my parents for their never-ending support that I have received. I am also grateful for them to have put up with me more than any parent should. Hence, I take this opportunity to dedicate this work to my Mom and Dad; I hope I have done you proud.

I am also forever indebted to the thousands of people that have contributed to the millions of lines of open-source software that this project builds upon. I feel as if I have stood on the shoulders of giants.

I will end by thanking those who did not make this list but would like to be on it. I'm deeply indebted to ---------- for more things than I can possibly list. (Please write your name in the space provided).

Contents

"If the brain were so simple we could understand it, we would be so simple we couldn't."

- Lyall Watson

Preface

This thesis is submitted for the degree of Doctor of Philosophy in Electrical and Electronic Engineering at the University of Canterbury. The research described herein was conducted under the supervision of Doctor Steve Weddell and Professor Richard Jones in the Electrical and Computer Engineering Department, University of Canterbury. The research presented is part of the Lapse Research Programme within the Christchurch Neurotechnology Research Programme.

Results of this research have been presented at a number of international conferences including 36th Annual International Conference of the IEEE Engineering in Medicine and Biology (Chicago, USA, August 2014), Annual Conference of the Australasian College of Physical Scientists and Engineers in Medicine (ACPSEM) (Christchurch, NZ, November 2014), NZ Physicists & Engineers in Medicine Conference (Christchurch, NZ, November 2014), 13th International Conference on Neuro-Computing and Evolving Intelligence (Auckland, NZ, February 2015) and 37th Annual International Conference of the IEEE Engineering in Medicine and Biology (Milan, Italy, August 2015).

The following publications were generated during this PhD research:

Ayyagari, **S. D. P. S.**, Jones, R. D. and Weddell, S. J. EEG-Based Event detection Using Optimized Echo State Networks with Leaky Integrator Neurons, *Proceedings of the 36th International Conference of the IEEE Engineering in Medicine and Biology Society,* Chicago, USA 2014, 36: **5856-5859**.

Ayyagari, **S. D. P. S.**, Jones, R. D. and Weddell, S. J. EEG-Based microsleep detection using supervised learning", *Australasian Physical & Engineering Sciences in Medicine, 2015*, **38: 194**.

Ayyagari, **S. D. P. S.**, Jones, R. D. and Weddell, S. J., EEG based event detection using optimized echo state networks with leaky integrator neuron models, *Australasian Physical & Engineering Sciences in Medicine*, **2015, 38: 193**.

Ayyagari, **S. D. P. S.**, Jones, R. D. and Weddell, S. J. Event based microsleep detection using echo state networks with leaky integrator neurons. *Proceedings of the 13th International Conference on Neuro Computing and Evolving Intelligence*, Auckland, 2015**; 13:26-7.**

Ayyagari, **S. D. P. S.**, Jones, R. D. and Weddell, S. J. EEG-Based Microsleep detection Using Optimized Echo State Networks with Leaky Integrator Neurons, *Proceedings of the 36th International Conference of the IEEE Engineering in Medicine and Biology Society*, Milan, Italy, 2015, 37: **3775-3778.**

"Ironically, work on this sleep research, was an outcome of numerous sleepless nights."

- Sudhanshu Ayyagari

CHAPTER 1

Introduction

1.1 Research Overview

Maintaining responsiveness is necessary for safe and successful performance of most human activities. The role of a human operator becomes more supervisory, as the level of automation increases in many systems. Tiredness and fatigue can cause an individual to lose concentration/ motivation and become less effective as a controller, especially while engaged in some active task. Instances of severely diminished alertness in high-risk occupations such as truck drivers, pilots, air-traffic controllers, and surgeons can have serious consequences, including multiple fatalities. Cases of tired drivers falling asleep while driving are not uncommon and there is considerable anecdotal evidence of brief *lapses of consciousness* – or *'micro-sleeps'* – during tasks such as driving and attending a talk (Bittner *et al.,* 2000).

Lapses are transient/phasic disruptions in sensory-motor/cognitive performance, from ~0.5–15 s, during an active task. Lapses usually range from brief pauses to microsleeps, which are involuntary events of lapses in attention or responsiveness, during which a person has a brief suspension of performance and appears to fall asleep momentarily (Peiris *et al*., 2006). Although, microsleeps are considered to be distinctly different from sustained attention lapses, both physiologically and behaviourally (Poudel *et al*., 2010), they are far from being disassociated constructs. Microsleeps are thought to emanate from a homeostatic drive for sleep and a complex interaction between the brain's arousal and attention system (Buckley, 2013).

Lapses in responsiveness can manifest themselves from the complex interaction of a number of factors such as the influence of circadian rhythms (Freund *et al.,* 1995), physical and mental exhaustion, lack of sleep or reduced quality of sleep, and boredom. Sleep-deprivation is particularly serious because of its detrimental effect on neurocognitive functions. Lapses can be observed even in healthy non-sleep-deprived subjects performing a range of vigilance

and sustained attention tasks, without a preceding phase of subjectively experienced drowsiness (Makeig & Inlow, 1993; Sagberg 1999; Doran et al., 2001; Peiris *et al.,* 2006; Poudel et al., 2006; Innes *et al*., 2010; Jones *et al*., 2010).

Microsleeps are more common in individuals suffering from obstructive sleep apnoea (OSA), narcolepsy, and other sleep disorders (Colten *et al.,* 2006). According to a report from a Committee on Sleep Medicine and Research (Colten *et al.,* 2006), more than 70% of such individuals suffering with sleep disorders are believed to go undiagnosed. As is evident, these lapses are of particular concern in occupations in which public safety depends on extended unimpaired performance. People on sedative medications, night-shift workers, and chronically sleep-deprived individuals are also at a higher risk of traffic and work accidents due to fatigue.

A transition from wakefulness to the sleep state is often believed to occur "when excitatory and inhibitory post-synaptic potentials become activated along with increasing simultaneity, ultimately resulting in the appearance of an EEG containing more limited variety of frequencies and, hence, increasing their amplitude" (Ogilvie, 2001).

Compounding this situation, in many cases, even well-rested individual, often not sleep-deprived, may also experience "sleep-related states", without a preceding phase of subjectively experienced drowsiness (Sagberg, 1999). The neural dynamics of sleep transition as seen in both the EEG spectral power and task performance measures in drowsy individuals performing an active task, such as driving, usually vary rapidly between periods of wakefulness and sleep (Makeig *et al.,* 2000). Despite the prevalence of these lapses in everyday life, we have yet to develop a system-wide understanding of the brain mechanisms underlying lapses.

Because of the serious real-world consequences that lapses pose on a day-to-day basis, elucidating and understanding the relationships between lapses and microsleeps with respect to task type is an important aspect of research.

Consequently, the aim of this research has been to recognise and reduce the risks to safety arising from tiredness and mental fatigue, and thereupon, develop a state-of-the-art *lapse prediction system*, in which an individual's state of responsiveness is monitored continuously, which could ultimately be used to trigger an alert to rouse them from a predicted microsleep or an attention lapse, potentially avoiding a catastrophe.

1.2 Significance of lapses

Momentary lapses in responsiveness often impact on a goal-directed behaviour, sometimes causing serious consequences such as accidents and errors. Thus, the capability of sustaining attention has long interested researchers in consciousness, in both basic and applied fields (Parasuraman & Davies, 1984). Loss of operator alertness is almost always preceded by psycho-physiological and/or performance changes (Wierwille *et al.*, 1992; Dingus *et al.*, 1987; Vallet *et al.*, 1988; Knipling and Wierwille, 1993)

In many occupational groups, such as the aviation, transportation, and process and control industries, it is essential that individuals remain highly attentive during their job at all times. For example, long-haul truck drivers, train drivers, and commercial airline pilots routinely experience monotonous and extended driving periods in a sedentary position, which has been associated with drowsiness, microsleeps, and, consequently, serious accidents. Mental fatigue affects task performance, a reduction in alertness, longer reaction times, poorer psychometric coordination, and memory lapses. Tiredness and fatigue can lead to brief instances of people falling asleep while engaged in an active task such as driving a motor vehicle. A study on fatigue by the General Association of German Insurance Industries (Byrne 2002), identified microsleeps as the principal cause of 24% of fatal motorway accidents.

Microsleeps are involuntary and, hence, other than awareness of feeling drowsy, don't come with a prior warning. A microsleep that occurs while driving can have fatal results due to the speed of the vehicle and the distance travelled while out of control of the driver. For example, if an individual is driving at a speed of 100 km/h and has a microsleep lasting four 4 s, the vehicle will have travelled 111 metres while completely out of the control of the driver. Unfortunately, however, Itoi *et al.* (1993) reported that in many situations, drivers themselves are often unaware of their own deteriorating condition or, even when they are aware, are often motivated to keep driving.

Furthermore, a *'Sleep in America poll'* conducted by the National Sleep Foundation in 2005, reported that around 60% of adult drivers in America – about 168 million people, had experienced a sleep-related incident while driving, and more than one-third (37% or 103 million people) have actually fallen asleep at the wheel! Furthermore, of the 103 million people who had nodded off while driving, 13% (~13 million people) say they have done so at least once a month. Moreover, 4 % (~11 million drivers) said they had had an accident or near accident due dozing off or were too tired to drive (National sleep Foundation, 2005). A study for the AAA Foundation for traffic Safety perceives driver drowsiness to be the second major reason for car accidents with the involvement of alcohol being the first (Tefft *et al.*,

2014). It has also been emphasized that sleepiness, like alcohol, impairs functioning at the level of the central nervous system (CNS) (Arnedt *et al.*, 2001).

Driver drowsiness is a major, though elusive, cause of traffic crashes. Drowsy driving is the dangerous combination of driving and sleepiness or fatigue. The danger, risks, and often tragic results of drowsy driving are alarming. Several studies on road safety and sleep education have identified the relationships between fatigue, lapses, and sleepiness and quantified the impact of the drowsy driving on the public safety in terms of the proportion of all the accidents that are attributed to lapsing and sleepiness.

According to the National Highway Traffic Safety Administration (NHTSA), has estimated approximately 1.4% of all motor vehicle crashes in the United States, 2.2% of those that resulted in injuries, and 2.5% of all fatal crashes in years 2005 – 2009, involved a drowsy driver, and those crashes resulted in a total of 5,021 deaths over those years (NHTSA, 2011). Another report from NHTSA estimated that drowsy driving was responsible for 72,000 crashes, 44,000 injuries, and 800 deaths and $12.5 billion in monetary losses in 2013 alone (NHTSA, 2013). Furthermore, an independent report estimated fatigue as a most likely cause in 4 – 7% of civil aviation mishaps (Kirsch, 1996). Data from the U.S. Army Safety Center suggests that fatigue is involved in 4% of Army accidents (Luna, 2003), and statistics from the Air Force Safety Center blame fatigue, at least in part, for 7.8% of U.S. Air Force Class A mishaps (FAA, 2007). NHTSA also reports that 6-10% of self-reported car accidents are the consequence of drivers falling asleep (NHTSA, 2008).

It is also critically important to understand and recognize that the possibility of lapsing is a major problem across many occupations and is not confined only to individuals in the transport industry; for example, occupations which involve night shifts and require prolonged periods of minimal physical activity. Complete loss of attention even for a few seconds, can have disastrous consequences in many industries requiring continuous vigilance by a human operator.

1.3 Motivation for project

Despite substantial research in this area over many years, there have been no reports of any commercial device or technique capable of providing a reliable indicator of every level and every stage of awake-sleep transition with high temporal resolution. Current technologies also appear to have difficulty coping with EEG data corresponding to different drowsiness levels in different subjects. Even more importantly, no study has successfully demonstrated a technique able to predict microsleeps prior to their occurrence.

The detection of microsleeps, especially in subjects working in high-risk occupations, is very important to workplace safety. Ultimately, a real-time lapse warning device, calibrated using several wireless, dry EEG scalp electrodes is required, where an individual's state of responsiveness is monitored continuously and could be used to trigger an alert to rouse a user from a predicted microsleep or an attention lapse, potentially avoiding a catastrophe.

A lapse prevention system such as the aforementioned, which uses advanced techniques in signal processing, in addition to the superior machine learning algorithms such as reservoir computing (RC) for detecting/ predicting the imminent transient events in the brain to a substantially higher level and dimension, will be valuable across many high-risk occupations, such as commercial truck drivers, pilots, air traffic controllers, and medical staff.

1.4 Risk factors and counter measures against sleepiness and microsleeps

Several risk factors such as sleep-deprivation, time-of-day, boredom, circadian rhythms, and physical and mental fatigue, and ultimately, sleepiness and microsleep can result in poor performance (Poudel, 2009). Sleepiness impairs certain elements of human performance that are critical for better physical and mental functioning (Dinges and Kribbs, 1991). Relevant impairments identified in a laboratory setting and in-vehicle studies include:

- *Slower reaction time:* Even minor reductions in reaction time can have a serious effect on crash risk, particularly at high speeds. Sleepiness reduces optimum reaction times, and moderately sleepy people can have a performance-impairing increase in reaction time that will hinder stopping in time to avoid a collision (Dinges *et al.*, 1997).

- *Deficit in information processing:* Processing information takes time and integrating the processed information takes much longer, as the accuracy of short-term memory decreases, the performance drops (Dinges, *et al.*, 1998).

- *Reduced vigilance:* Performance on attention-based tasks drops with sleepiness, including increased periods of non-responding or delayed responding (Haraldsson *et al.*, 1990; Dinges and Kribbs, 1991).

Often, many individuals use physical activity and stimulants to cope with sleep deprivation, masking their level of sleepiness. However, when they sit still, perform repetitive tasks (such as driving long distances or attending long lectures), get bored, or let down their coping

defenses, sleep comes quickly (Mitler *et al.,* 1982; National Transportation Safety Board, 1995).

Knipling and Wang (1995) suggest that drivers younger than 30 years of age accounted for almost two-thirds of drowsy-driving crashes. Several reports also indicate that shift workers are another group of people who are more prone to lapsing. Dinges *et al.* (1987), Williams *et al.* (1959), and Hamilton *et al. (*1972) found that circadian phase disruptions caused by rotating shift work are often associated with lapses of attention, increased reaction time, and decreased performance. Research also indicates that individuals with untreated sleep disorders such as untreated sleep apnea syndrome and narcolepsy are more likely to lapse (Aldrich, 1989).

A wide range of countermeasures can be taken to minimize the occurrence of lapses both the on road and in the workplace. A report from the NHTSA expert panel on driver fatigue and sleepiness suggest few categories of countermeasures including, the behavioural and medical (NHTSA, 2008). In addition to acquiring sufficient sleep, some of the behavioural interventions include scheduled naps. Napping for brief periods of time has been shown to improve subsequent performance, particularly in sleep-deprived people (Dinges *et al.,* 1987; Philip *et al.,* 2001). Consuming caffeine, even in low doses, substantially improves alertness in drowsy individuals (Griffiths *et al.,* 1990; Lorist *et al.,* 1994). Horne and Reyner, (1995) demonstrated with a driving simulator test that consuming caffeine in sleep-deprived drivers reduced lane deviations and sleepiness, thus avoiding potential crashes. Alertness-enhancing drugs such as amphetamines and methylphenidate have also been suggested in the past for sleep-deprived individuals to sustain attention and wakefulness (Philip *et al.,* 2001).

In jobs with extended hours, the scheduling of work and rest periods to conform to circadian rhythms promotes better sleep and performance (Stampi, 1994). McCartt *et al.* (2000) also proposed adequate rest breaks at the work place as another countermeasure for lapsing. Listening to the radio and setting periodic alarms to check drowsiness levels was proposed by Reyner and Horne, (1998). Ultimately, practice of good sleep hygiene is recommended to avoid unwanted lapses and drowsiness at the workplace.

1.5 Research objectives, focus and goals

The overall goal of this project was EEG-based characterization and early detection of lapses of responsiveness states and microsleeps, during a sustained attention task. The overall expectations from this project were identifying different reliable physiological cues, indicative of lapses of sustained attention and microsleeps, to develop a real-time microsleep

system which can be used to detect, or better yet, predict the onset of lapses and microsleeps in real time and trigger an alert to rouse the user from an impending microsleep. Thus allowing for a precautionary measure to maintain safety.

The key goals within this research were to:

1. Conduct a review of the literature to explore and evaluate the existing approaches to the lapse and microsleep detection, based upon the electrophysiological, behavioural traits.

2. Investigate the propensity for lapses and microsleeps using noisy EEG and video recordings.

3. Evaluate subject-independent features within EEG, which could provide reliable indication of lapses and microsleeps.

4. Examine the efficient pre-processing techniques for EEG signals and extract the most informative 'features' indicative of microsleeps.

5. Employ supervised feature selection/reduction models to improve the class separability (microsleep vs. non- microsleep).

6. Assess the effectiveness of novel machine learning algorithms including reservoir computing (i.e., echo state networks and liquid state machines) to maximally acquire new cues for detecting microsleeps from deep neuroelectric activity in the brain (Buteneers *et al.*, 2013, Jonmohamadi *et al.*, 2014) and changes in neural connectivity (Toppi *et al.*, 2012).

7. Apply reservoir computing models to determine the distribution of each class.

8. Develop an improvised, automated EEG-based microsleep detection and increase the current performance traits such as, sensitivity, specificity, and selectivity and reduce system latency.

9. Determine the minimum/optimal number of EEG channels/electrodes, and placement of electrodes to achieve satisfactory performance.

10. Investigate and evaluate the characteristics of lapses/ microsleeps across independent subjects.

These notwithstanding, the main focus of this dissertation, as will be highlighted in the following chapters, was on employing complex architectures within reservoir computing

(RC) models such as the echo state networks (ESN) with leaky-integrator-and-fire-neuron model, and liquid state machines (LSM) towards EEG-based microsleep detection.

1.6 Original contributions

The work demonstrated in this thesis incorporates substantial original contributions. The reservoir-computing approach, previously used for epileptic seizure detection [Buteneers, et al., 2011], provided a basis for new research in EEG-based microsleep detection. Consequently, a novel cascaded ESN architecture, based on the leaky-integrator and fire-neuron model, was developed for the detection of microsleeps. The advantage of this cascaded-leaky-integrator-based ESN approach lies in its simplicity, as networks configured with only 8-leaky neurons are able to achieve optimal microsleep detection. This cascaded-leaky-integrator-based ESN approach also substantially reduced the training time required to train an enhanced neural network while attaining state-of-the-art performance.

As will be presented in the following chapters, the cascaded-leaky integrator-based ESNs have a global control parameter, called 'the leakage factor', which needs to be optimized in addition to input and output feedback scaling and the spectral radius of the reservoir weight matrix to attain optimal performances. The simplified training requirements, supported by the application of a cascaded-leaky-integrator-based ESN classifier, provided the basis for the best microsleep detection performance achieved to date on our Study A datasets.

A novel cascaded architecture within the leaky-integrator-based ESNs (as depicted in §7.3), consisting of two parts – a reservoir that maps the input to a higher dimensional space and a linear readout which is used for training and to produce an output – was developed. In this cascaded approach, each non-zero input sample to a reservoir excites the state dynamics of the system and pushes the system to a new state. Essentially, this is achieved by mapping the input data, which consists of only 1 feature vector, to a higher dimensional space. In this higher dimensional space the probability increases that the microsleep and non-microsleep samples are linearly separable (Cover, 1965).

A new variant for the creation of the state matrix for this cascaded ESN approach is explored in §7.1. This approach includes the partitioning of each state sequence into a number of sub-sequences of equal length. This way of partitioning creates an individual reservoir state. For this set-up, a reservoir of 8 neurons was used; this value was chosen by extensive optimization of the state dynamics. Observation, with only 8 neurons, showed higher accuracies for microsleep detection. Furthermore, these networks had

markedly different dynamic properties across their random instantiations, which was not the case for larger ESNs where inter-network differences became insignificant with growing network size. It was hypothesized that each reservoir created through this approach could be considered a detector for a random, 8-dimensional, nonlinear, dynamic feature of the input signals.

Because of the intrinsic nature of microsleeps, which are rare events, the datasets used in this work often span several hours (training using 7 EEG sessions, each 2 hours in duration). If each feature of the data is mapped to one data point, this corresponds to millions of data points which is very large from a machine learning perspective. Therefore, to be able to process the datasets of this size, this work proposes a different learning algorithm in RC that has a training time in the order of hours as opposed to days. These RC-based algorithms have been designed to deal with the special characteristics of microsleep datasets, such as the rareness of microsleeps and the variability of EEG.

This is the first study to have implemented and evaluated EEG-based microsleep detection using RC models for the characterization and detection of microsleeps from the underlying EEG. It is also the first study to have compared RC-based models (both cascaded-leaky-integrator ESN and LSMs) to other classifiers such as linear discriminants analysis, support vector machines, and spiking neural networks for microsleep detection.

Integration of the leaky-integrator-based ESN approach, with the advanced feature-reduction techniques discussed in Chapter 6, also forms a substantial portion of new work. Moreover, this is the first study to have compared microsleep detection performance on different gold-standards (lapses, video microsleeps, flatspots, and definite microsleeps and, flatspots) provided with the Study A datasets. Performance analysis of the cascaded-leaky-integrator-based ESN classifier modules on an independent EEG-based study (BCI competition dataset IIa), as highlighted in Chapter 5, also formed a substantial portion of original contributions to this work.

Finally, as discussed in §1.5, the main focus of this research was on EEG-based detection of microsleeps, using several physiological measures such as EEG, video recordings, and performance data which had been recorded in a previous NeuroTech study (Peiris et al., 2006b; Davidson et al., 2007; Peiris et al, 2011). Several key research questions and corresponding hypotheses relating to this project were formulated and evaluated.

1.7 Hypotheses

Hypothesis 1 – Detection of microsleeps using non-linear classification techniques

Gap in the literature: There is little evidence on whether non-linear classifiers can enhance the detection of microsleeps over that obtainable from linear classifiers.

Hypothesis: Non-linear classification techniques will be able to detect microsleeps with greater accuracy than a linear classifier, such as linear discriminant analysis.

Rationale: Non-linear classifiers have non-linear and, possibly discontinuous decision boundaries (Bradley and Mangasarian, 2000). In highly imbalanced data, it is likely that the boundary between two classes is complex, with the classes often overlapping and with nonlinear boundaries. It is well established that a non-linear classifier can represent the data into a higher dimensional space and form better classifier models (Bradley and Mangasarian, 2000). For example, non-linear classifiers have been shown capable of predicting the onset of epileptic seizures several minutes in advance, despite no visible changes in the power spectra or visually in the raw EEG (Le Van Quyen *et al.*, 1999; Le Van Quyen *et al.*, 2001). Therefore, it is reasoned that the application of such techniques may be appropriate to the microsleep detection problem to find 'hidden' changes in the EEG as the brain state changes from alert to non-responsive. Notwithstanding, Peiris *et al.* (2011) found no improvement in detection of microsleeps with three non-linear features (fractal dimensions, approximate entropy, and Lempel-Ziv complexity) over that of spectral features alone.

Significance: Subtle changes in the EEG during microsleeps may potentially be classified using non-linear classification methods. These changes may provide cues that could be incorporated into a reliable microsleep detection system.

Hypothesis 2 – Detection of microsleeps via biologically-inspired neural network paradigms in reservoir computing.

Gap in the literature: It is not known whether or not biologically-inspired neural networks in RC can enhance the detection of microsleeps over that obtainable from other neural network approaches using linear or non-linear classification techniques.

Hypothesis: A reservoir-computing paradigm, referred to as a cascaded-leaky-integrator ESN, will be able to detect microsleeps with greater accuracy than other

linear and non-linear classifiers, such as linear discriminant analysis and polynomial-based support vector machines.

Rationale: Reservoir computing is an innovative approach to the design, training, and analysis of RNNs (Jaeger, 2001). RNNs are mostly used for modelling dynamical systems because of their inherent memory states. The RC approach is based on the observation that if a random RNN possesses certain algebraic properties, training only a linear readout is often sufficient to achieve superior performance in many practical applications (Jaeger, 2001; Jaeger, 2007; Yildiz, 2012). RC methods have been extensively applied towards several machine learning applications with positive results (Buteneers 2012; Jaeger, 2007; Maas 2002). Therefore, it is reasoned that an RC-based classifier could provide maximal detection performance, in terms of accuracy and minimum latency.

Several variants of artificial neural networks (ANN) are touted to be neuro-biologically inspired paradigms which attempt to emulate the functioning of a human brain in a variety of fundamental ways. The human brain is a highly structured entity with localized regions of neurons specialized in performing specific tasks (Azam, 2000). The non-explicit structure and monolithic nature of current mainstream ANNs results in a lack of capability, in terms of systematic incorporation of functional or task-specific a priori knowledge in the artificial neural network design process. It has been established that, using models, such as leaky neuron and spiky neural networks, the temporal characteristics of a learning task can be exploited (Jaeger, 2007; Yildiz 2012; Kasabov, 2010).

Significance: Reservoir computing essentially uses the path of iterative updating on recurrent networks as an indication of the input. Since a microsleep can occur at any time, it is, in principle, a superior method for spatial-temporal pattern recognition. Additionally, ESNs with leaky neuron configurations contain additional 'global control parameters' (Jaeger *et al.*, 2007; Jaeger, 2012), such as the spectral radius of the reservoir weight matrix and a leaking rate (leakage factor), that can be optimized particularly for low frequency input signals which microsleeps are expected to be. This in turn may be reflected in the performance of classifier systems built on these models.

In conjunction with the distinctions of RC-based approaches, RC techniques have not been applied previously to microsleep detection and, hence, will demonstrate the originality of this research.

1.8 Thesis organization

This dissertation is organized into 8 chapters. The key objectives of the research, motivation and brief statistics on lapse problem affecting individuals in an operational environment are highlighted in the current chapter. Chapter 1 concludes by providing the key research objectives in addition to proposing the hypotheses. Chapter 2, in particular, provides the reader with the necessary understanding of the terminologies and physiological techniques used throughout this thesis. Chapter 3 provides an overview of relevant literature on devices and techniques for detecting microsleeps and measuring drowsiness, as well as for other cognitive monitoring applications. It also provides insights into different studies conducted in the Christchurch NeuroTechnology Research Programme and the current state of its research into microsleep detection. Chapter 3 concludes with an introduction to the continuous tracking task study (Study A), and the generation of gold standard of the Study A datasets used in this research.

Chapter 4 begins with an introduction to learning systems and provides an insight into various classification techniques and different classification schemes developed in this research. Chapter 5 provides an insight into the development of a novel cascaded-leaky-integrator ESN for microsleep detection. In addition, Chapter 5 validates the novel cascaded leaky-integrator ESN classifier models on another independent EEG-based study, the BCI competition IV dataset IIa.

Chapter 6 describes several pre-processing techniques applied to clean the experimental data used in this study and highlights the analysis and key findings from several supervised and supervised feature reduction methods implemented on the Study A datasets. Chapter 7 contains the analysis and key experimental findings on the Study A datasets, and concludes with a brief discussion on the major results and findings from this thesis. Finally, Chapter 8 contains the key conclusions and findings of the project, a critique of the research, and thoughts for future research in this topic.

The original contributions in this project are presented in Chapters 5, 6 and 7.

In summary, a new cascaded-leaky-integrator ESN-based method for microsleep detection is proposed. Validation of this new cascaded-leaky-integrator-ESN architecture on another independent EEG-based study, where the information of the gold standard event was precisely known is outlined in Appendix A. Different optimization parameters considered for the feature reduction models presented in Chapter 6 and their trustworthiness scores are presented in Appendix B.

CHAPTER 2

Physiological measures for microsleep detection

This chapter provides an overview of the key concepts and the background of physiological measures and principles used in the development of this work. The objective is to outline the current state of this research, biosignals associated with attention and arousal and their use in the identification of lapses and microsleeps. Following this, a review on wakefulness, sleep, and the transitional phase between the two is provided.

2.1 The electroencephalogram

The processes of wakefulness and sleep are regulated by complex neurobiological mechanisms controlled by networks of neurons within the brain. Subtle shifts from the wakefulness to sleep can be identified by quantifying the changes in the electrical activity of the brain (Makeig and Inlow, 1993). As the electrical activity is generated directly by the neurons, the electroencephalogram (EEG) is the most direct non-invasive measure we have of brain activity. A summary of the key aspects of EEG and different types of electrical activity in the brain are described below.

2.1.1 Neurophysiology

The human brain consists of approximately 100 billion nerve cells or neurons interconnected in a very intricate network within which the information is transmitted by electro-chemical impulses (Carlson, 2004). Each neuron consists of a cell body, or soma, with several receiving processes, or dendrites, which prolongs to a nerve fibre, or axons which carry electrical

impulses from the cell body to target sites such as muscles, glands, or other neurons (Bronzino *et al.*, 2000). The structure of a neuron is depicted in Figure 2-1.

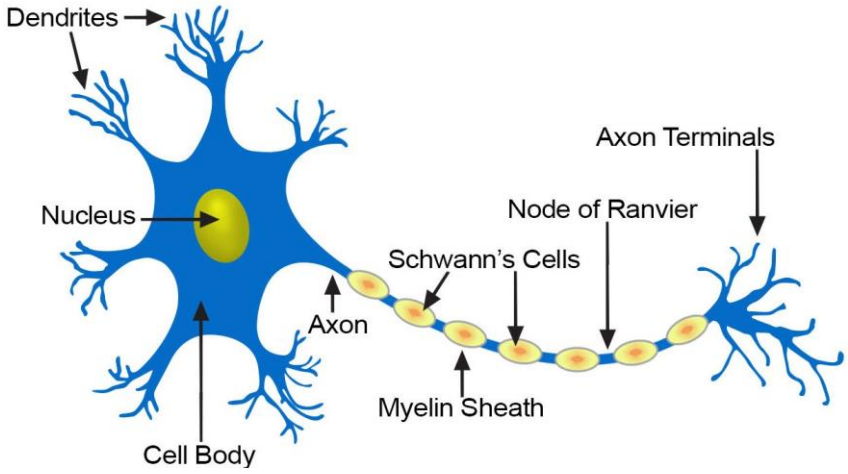


Figure 2-1. A schematic diagram of a neuron depicting its cell body, nucleus, synapse, axon and dendrite (adapted from Duffy *et al.* 1989).

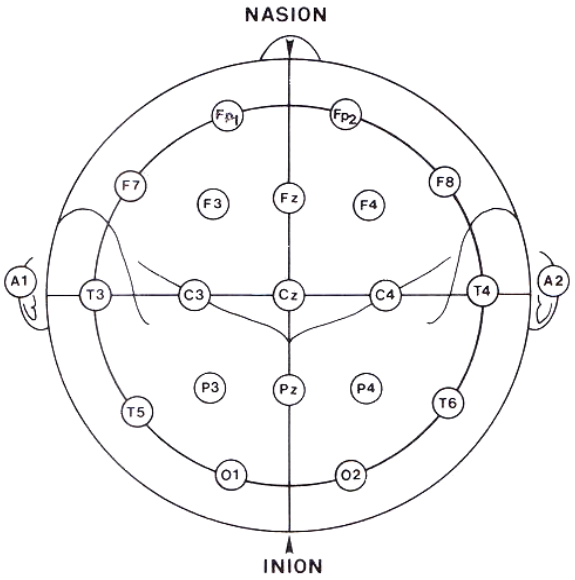
Electroencephalography is the measurement of continuous brain-wave patterns, or electrical activity of the brain, as recorded with the placement of small metal-disc electrodes positioned in a standardized pattern on the scalp. The resulting tracing reflects the summation of the activity of millions of individual neurons. EEG has been used in sleep research for over 75 years[1]. Matousek and Petersen (1983) were the first to develop a method to estimate the attention levels using the EEG frequency patterns. A shift from desynchronized to a more synchronized patterns in EEG can be described as the transition from wakefulness to sleep (Ogilvie, 2001). EEG has also been mainly used in many cognitive sciences, medical research (neurosciences), psychology and many more brain research applications (Valley *et al.*, 1983; Pomfrett and Pearson, 1996; Jung *et al.*, 2000; Peiris *et al.*, 2011).

The EEG is recorded by amplifying the potential differences between two electrodes located on the scalp. The signals recorded at the scalp primarily reflect the cortical activity. An electrode is a liquid-metal junction used to make the connection between the conducting fluid of the tissue in which the electrical activity is generated and the input circuit of the amplifier (Binnie & Osselton, 1995).

[1] The first recording of the electrical activity of the brain was made by Caton in 1875 using rabbits, monkeys and other small animals (Caton, 1875), the first non-invasive EEG measurement in humans was described in 1929 by Hans Berger.

EEG is widely regarded as the physiological "gold standard" for the assessment of alertness, even though numerous physiological parameters, including cardiovascular indices, pupil diameter, and eye closures have been employed (Torsvall *et al.*, 1988; Akerstedt *et al.*, 1995; Berka *et al.*, 2004). It has also been suggested that a variety of factors including biochemical, metabolic, circulatory, hormonal, neuroelectric, and behavioural factors affect the cortical EEG patterns (Bronzino *et al.*, 2000).

Modern EEG recording systems are capable of recording 100 or more channels simultaneously from electrodes placed on the scalp and can provide comprehensive representation of the spatial distribution of the electric potentials across the entire scalp (Kamp and Lopes da Silva, 1999).The most commonly used system for the placement of the electrodes is the so-called "10-20 International System of Electrode Placement" which relies on standard landmarks on the skull (Duffy *et al.*, 1989). Figure 2-2 depicts the 10-20 International System of Electrode Placement.



$Fp_{1,2}$ – pre-frontal	$F_{7,8}$ – anterior
$F_{3,4}$ – frontal	F_Z, – frontal mid-line
$C_{3,4}$ – central	C_Z – Central vertex
$P_{3,4}$ – parietal	P_Z – parietal mid-line
$O_{1,2}$ – occipital	$A_{1,2}$ – mastoid
$T_{3,4}$ – mid-temporal	$T_{5,6}$ – posterior temporal

Figure 2-2. The 10-20 International System of electrode placement depicting different scalp locations (Duffy *et al.*, 1989).

An orderly arrangement of a group of EEG derivations displayed simultaneously is termed a *montage.* When all the channels are referenced to the same electrode (usually mastoid processes A1 for the right side of the scalp, and A2 for the left side of the scalp, or a common site located at the nose or at the chin) the montage is called *referential*. If all the channels represent the difference potential between two consecutive electrodes on the scalp, the montage is said to be *bipolar* (Sharbrough, 1990). Both the referential and bipolar montages have pros and cons associated with their use (Duffy *et al.*, 1989).

The bipolar montage considers both electrodes to be active, and the varying voltage difference between the two electrodes is recorded. This can sometimes lead to the cancellation of the two signals at the amplifier, depending on their magnitude and phase (Duffy *et al.*, 1989). On the other hand, the referential montage, uses a common reference electrode to compare the electrical activity at different electrode sites. Usually, the reference electrodes are placed near the earlobes and, ideally, the common reference electrode should not be affected by any bioelectric activity, but no such ideal reference exists as the earlobes tend to pick up temporal cerebral activity which presents problems with localization (Duffy *et al.*, 1989).

In addition, the scalp electrodes can also pick up other signals whose sources are not in the brain, but are near or strong enough to interfere with its electrical activity. These signals, referred to as the extra-cerebral potentials or artefacts, can totally obscure the EEG, making the recording uninterpretable. They can subtly mimic normal EEG activity or distort normal activity, leading to misinterpretation. There are multiple sources of artefacts within the EEG; these include, other bio-electrical signals which do not originate from the brain, such as, signals from eye movements, blinks, cardiac signals, muscle noise, line noise, and extraneous environmental sources (e.g., 50 Hz power), which are orders of magnitude larger than brain-generated electrical potentials and are one of the main sources of artefacts in EEG data.

That notwithstanding, EEG data recorded continuously can be used as an important indicator of an individual's brain state. EEG has also been used clinically for diagnosis of epilepsy, detecting seizures, or classifying sleep stages (Duffy *et al.*, 1989). The electrical activity associated with specific perceptual and cognitive processes can often be investigated using EEG signal, time-locked to some stimulus, averaged over many trials.

2.1.2 Types of EEG activity

Cortical activity as measured by the EEG is traditionally categorized into different frequency bands and is identified using the Greek letters alpha, beta, theta, and delta. These waves are specified in terms of their individual amplitude and frequencies. The characteristics of different frequency bands in EEG are summarized in Table 2-1.

The corresponding EEG patterns vary substantially in each individual, and even between sleep/drowsiness episodes of the same individual, making it very difficult to rely on EEG alone to determine the level of drowsiness (Chee *et al.*, 2008).

The main focus of this research was to understand the transitional phase between the awake and sleep states and most of the research in the awake-sleep transition, which involved correlating the electrophysiological signals (EOG, fMRI and EOG). Although our focus is to

use EEG for sleep research, other methods can be adapted from the field of neuroscience to enhance our investigations.

Table 2-1 Characteristics of the main EEG waveforms adapted from (Duffy *et al.*, 1989; Lal and Craig (2001), Fell *et al.* (2002); Horne, 1988; Niedermeyer, 1999).

Activity type	Frequency	Amplitude	Amplitude/frequency/spatial characteristics
Delta	0.5 – 4 Hz	High amplitude (over 100 μV)	• Increases with deepening sleep. • Present during transition to drowsiness and during sleep.
Theta	4 – 8 Hz	25 – 30 μV	• Associated with lower alertness. • Does not occur in the normal awake adult. • Occurs prominently during drowsiness and sleep.
Alpha	8 – 13 Hz	50 μV or less	• Rhythmical activity . • Occurs in an awake, relaxed adult. Best seen when the eyes are closed. Opening the eye results in the attenuation of the alpha rhythm. • Also termed as *"paradoxical alpha".* • Amplitude and abundance of alpha rhythms may increase with awake state or eye opening (instead of decreasing) following a period of drowsiness..
Beta	13 – 32 Hz	Low amplitude (< 10 μV)	• EEG potentials associated with increased alertness, arousal, and excitement. • Masked by alpha activity, beta waves are mostly seen in anterior and posterior regions. • Beta waves occur in the cerebrum when alert or anxious.
Gamma	> 32 Hz		• EEG potentials associated with perceptual and cognitive processing. • High frequency waves.

2.2 The electrooculogram

Both physiological and behavioural measurements are used in microsleep research. Physiological measurements are those that directly depend upon neurophysiological events, including EEG, functional magnetic resonance imaging fMRI, and electrooculogram (EOG)

(Golz *et al.*, 2005). Behavioural measurements do not depend directly on a specific physiological parameter, and include performance on behavioural tests and video recordings of eye movements (Krajewski *et al.*, 2008). Lowered arousal, which can result in drowsiness and arousal-related lapses, is often associated with overt behavioural signs such as eye-closure, loss of facial tone, and head-droop. Some of these signs can be determined from vertical-electrooculography (VEOG) and/or eye-video.

In the EOG, the eye forms an electric dipole, the orientation of which can be detected by placing electrodes near the eyes. A change in potential difference across the two electrodes can be detected when the eyelids close, which is recorded as the VEOG (Lins *et al.*, 1993). EOG has been found as a possible measure method for all different types of eye movements except for intra-ocular movements (Stern *et al.*, 1994; Stern and Ray 2001). In addition to measuring eye closure and opening, other eye measurements recorded by the EOG include saccades, involuntary slow horizontal eye movements, and smooth pursuit (vertical and horizontal). An advantage of EOG over other physiological methods is its correspondence with the saccadic eye movements, which are the conjugate fast eye movements from one fixation point to the other. Additionally, another distinguishing feature of the EOG method is that one can measure eye movements even when the eyes are closed (as in REM sleep). That said, however, the application of EOG to drowsiness/lapse detection is mostly limited to research in the psychophysiological laboratories because it requires electrodes to be attached to an individual's face causing discomfort over extended periods.

2.2.1 Eye closure and eye movements

Although primarily used in ophthalmological diagnosis, using EOG it may be possible to detect drowsiness through measurement of slow eye movements (SEMs) (De Gennaro *et al.*, 2001; Porte, 2003). Measures based on eye-video recording including percent eye-closure (PERCLOS) (Wierwille and Ellsworth, 1994), duration of eye-closure, and frequency of eye-closure which are important markers used in vigilance research (Peiris *et al.*, 2006; Stern *et al.*, 1994). Mallis *et al. (*1999) reported that, the percentage of time in a minute that the eyes were more than 80% closed (PERCLOS) was found to be a good indicator of lapse frequency and duration. A series of studies by Åkerstedt and his colleagues found that SEMs during task performance with eyes open were the most sensitive measure for differentiating subjective sleepiness from alertness (Åkerstedt and Gillberg, 1990; Torsvall and Åkerstedt, 1987; Torsvall and Åkerstedt, 1988).

2.2.2 Facial video

Non-electrophysiological measurements such as behavioural data and video recording of eyes have been used for microsleep detection (Peiris *et al.*, 2006b; Bergasa *et al.*, 2006; Golz & Sommer, 2010). Video can detect both the timing and duration of eye closure events. Eyelid movements were detected on video by measuring the percentage of eye closure (PERCLOS) (Bergasa *et al.*, 2006); Knopp, (2015). Behavioural test performance offers another avenue for lapse detection (Doran *et al.*, 2001).

Video systems are often considered to be more practical in real-world applications because of their non-intrusive nature (Bergasa *et al.*, 2006); BMW Connected Drive[2]. Despite this, such technology requires tremendous improvement before making it suitable for use in a commercial product because, just by using the video data alone it is very difficult to determine any stage of sleep; video systems are more useful for the estimation of drowsiness states (Peiris *et al.*, 2011). Video systems integrated with other physiological signals, such as EEG, could be used for the detection of microsleeps (Peiris *et al.*, 2005b).

A video camera directed towards capturing an individual's eyelid closure, rate of blinking and degree of closure face, can provide valuable information regarding the level of arousal. Furthermore, the video recordings can also provide information on additional features such as, eye gaze, yawning, head droop/nodding, and facial muscle tone, which can ultimately be used to estimate the level of arousal of a subject.

2.3 The electromyogram

It has been suggested in the past that wakefulness is characterized by a high tonic electromyogram (EMG) in the facial muscles (Carskadon and Rechtschaffen, 2000). EMG is usually recorded from three electrodes placed on the muscles beneath the chin, overlying the mentalis/submentalis muscles after careful skin preparation. Although it seems reasonable to assume that the sleep onset, the moment when a person falls asleep, is related to certain muscle activities, Erwin *et al.* (1973) report that muscle activity measurements alone cannot offer reliable predictive information on sleep onset.

To overcome this problem, and accurately identify the transition between alert and sleep states, several other studies have used a combination of physiological measures including, but not limited to, EEG, EOG and EMG, in addition to behavioural data (Krajewski *et al.*,

[2] http://www.bmw.connecteddrive.info/en

2008; Carskadon and Rechtschaffen, 2000). Figure 2-3 depicts the most common electrophysiological measures used in the lapse detection.

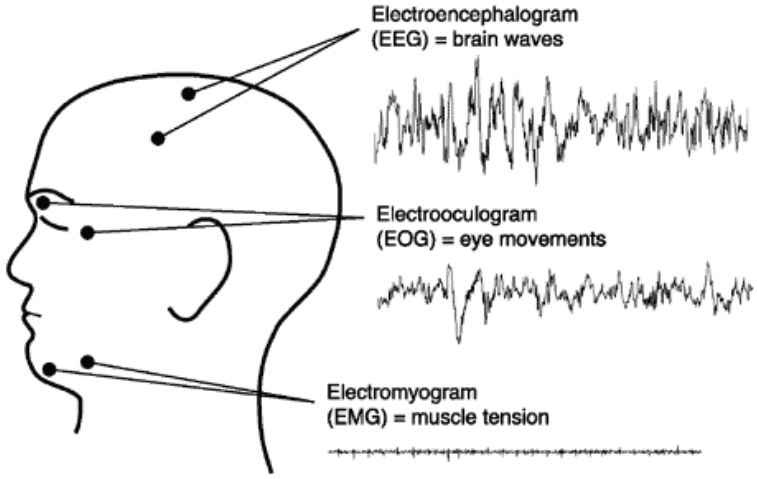


Figure 2-3. Examples of the common bio-electrical signals used for monitoring sleep adapted from (Carskadon and Rechtschaffen, 2000).

2.4 Functional magnetic resonance imaging

The magnetic resonance (MR) signal is generated using the nuclear magnetic resonance (NMR) phenomenon observed in the quantum-mechanical magnetic properties of an atomic nucleus. In the presence of an externally applied magnetic field, blood oxygenation level dependent (BOLD) changes in the intensity of the MR signal and provides a mechanism for imaging the functional human brain. One of the main advantages of the fMRI is its higher temporal resolution, which is used to reveal functional changes in the brain with a greater temporal accuracy.

To explore this further, several studies have investigated the interaction between attention and arousal using simultaneous fMRI and EEG (Foucher *et al.*, 2004; Laufs *et al.*, 2006; Poudel, 2010). In particular, these studies have predominantly investigated brain networks modulated by changes in alpha rhythms in the EEG. Alpha rhythm usually increases during relaxed eyes-closed state and can decrease either due to an increase in alertness (such as endogenous alertness or eyes-opening) or due to a decrease in alertness (such as wake-sleep transition) (Laufs *et al.*, 2006).

It has been suggested that, by using novel fMRI techniques to track changes in functional connectivity of the brain across time, we may be able to better isolate brain areas that work in tandem to compensate for decreased arousal during increasing drowsiness.

2.5 Measures of task performance

In tasks that require timely responses to discretely presented stimuli, lapses manifest as delayed or slowed responses. There is considerable anecdotal evidence to suggest that low levels of cortical arousal in an individual can lead to impairment in reasoning abilities, perceptual skills, better decision-making skills, and judgement (Dinges and Kribbs, 1991). The term *lapse* was arbitrarily defined by Williams *et al.* (1959) as having reaction times greater than twice the subject's baseline mean during a performance task.

Performance tasks are primarily used to assess the fluctuations in the cognitive processing and can be divided into two broad categories – psychomotor tasks and cognitive tests. Psychomotor tasks mostly include auditory and visual reaction time (simple or multiple choice), tracking, and tapping tests; whereas cognitive tasks, on the other hand, include tests on attention, memory, and logical reasoning. The following sections highlight a few key auditory and visual psychomotor task tests that have been used in vigilance research in the past.

2.5.1 Psychomotor tasks – auditory/visual

One of the most common and widely used tests to measure visual psychomotor performance in drowsiness research is the psychomotor vigilance task[3] (PVT). During the PVT, a subject responds to cues at 2-10 s over 10 min. The PVT was developed to track the time-course of dynamic changes induced by the interaction of the homeostatic drive for sleep and the endogenous circadian pacemaker (Dorrian *et al.*, 2005). The PVT has been shown to be sensitive to sleep deprivation and is generally considered to be a validated, reliable, and sensitive test of vigilance and simple visual reaction time (Dinges *et al.*, 1998). PVT involves a combination of prefrontal cortex executive attention and traditional stimulus-response testing (Dorrian *et al.*, 2005).

Failure to respond in a timely manner to the stimuli manifests as a prolonged reaction time or response error. Particularly in tasks such as the psychomotor vigilance task (PVT), responses longer than 500 ms are considered to be lapses (Dinges and Powell, 1985; Dorrian *et al.*, 2005), while in continuous tasks, such as visuomotor tracking, lapses are complete

[3] The psychomotor vigilance task (PVT) is commonly used for detecting the effects of drowsiness/fatigue/sleepiness. The subject must respond to a series of visual stimuli presented at random intervals of 2–10 s. A lapse is defined as a response time of more than 500 ms [Drummond et al. 2005].

failures to respond resulting in loss of tracking (Peiris *et al.,* 2006a). The term 'lapses of responsiveness' (LOR) was first used by Peiris *et al. (*2006a) to refer to loss of tracking or incoherent tracking during a visuomotor tracking task. They used a continuous tracking task (CTT), where a subject must move a marker towards a moving target. Task performance was combined with physiological signals to detect lapses of responsiveness. Figure 2-4 depicts the LOR during a continuous tracking task.

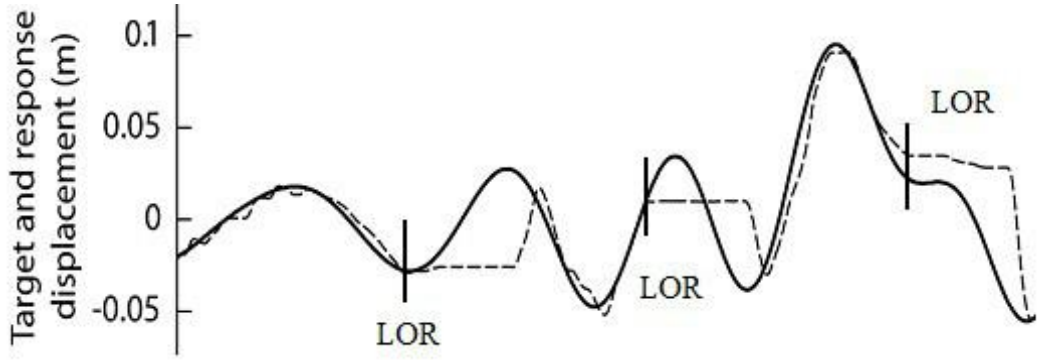


Figure 2-4. Lapses of responsiveness (LOR) during visuomotor tracking tasks manifest as flat response (from Peiris et al., 2006).

Other visual and auditory tasks commonly used for determining vigilance that have been mentioned exhaustively in the literature and include the *Reaction Time Test (RTT)* (Conradt *et al.,* 1999), *Four-Choice Reaction Test* (Wilkinson and Houghton, 1975), Simulated Assembly Line Task (Walsh *et al.,* 1992), sustained attention to response task (SART), Simple Reaction Time Test (Glenville *et al.,* 1978), and the Continuous Auditory Performance Task (Arruda *et al.,* 1999).

2.5.2 Driving simulators

For reasons of safety and reproducibility, laboratory-based driving simulators have been used in the past to investigate the effects of sleep deprivation (Fairclough and Graham, 1999; Lenne *et al.,* 1998) and drowsiness (Lin *et al.,* 2006; Lin *et al.,* 2005). Other advantages of driving simulators include their ease of controllability (e.g., vehicle control - steering and pedal control, position tracking, and speed deviation), reproducibility, and standardization. Simulators are often designed to replicate actual driving experiences through all types of scenarios and, hence, can provide the researchers with valuable objective information concerning the driver's state of mind during a diminished arousal state.

2.6 Wake-sleep transition – Evaluation by electrophysiology and behaviour

Several psychological constructs have an influence over task performance. The ones that play a crucial role in the ability to perform most tasks are attention, arousal, and sleep–wake processes. The following subsections provide a review of the fundamental concepts of wakefulness, sleep, and the transitional phase between these two fundamental states.

2.6.1 Wakefulness

Contemporary models of the wake-sleep regulatory system are based on the seminal research conducted by von Economo *et al. (*1930). The process of wakefulness is mediated by an ascending arousal system beginning in the brainstem, which remains active following midbrain interruption of the classical sensory pathways. Consequently, wakefulness-promoting systems cause low-voltage, fast activity in the EEG accompanied by a high muscle tone in the EMG (Brown *et al.,* 2012). Posteriorly-dominant alpha rhythm is one of the most prominent features of the EEG during wakefulness (Duffy *et al.,* 1989).

In the awake-EEG, alpha rhythms appear in posterior cortical recordings whilst the theta rhythms increase in frontal cortical regions (Makeig *et al.,* 2000). This phenomenon is most visible during relaxed wakefulness with eyes closed and attenuates with attention and when the eyes are opened (Carskadon and Rechtschaffen, 2000). The EEG of a highly attentive person whose eyes are open contains a predominance of EEG activity in the beta range (15–25 Hz). It has also been suggested that during quiet or drowsy wakefulness, the slower EEG frequencies become more prevalent (Brown *et al.,* 2012). The two most common types of eye movements as seen in the EEG are during the awake state are *blinks* and *mini-blinks* (Santamaria and Chiappa, 1987).

2.6.2 Sleep

Sleep is a widespread biological phenomenon, which can be defined as a state of immobility with greatly reduced responsiveness (Siegel, 2005), and approximately a third of human life is spent sleeping (Sejnowski and Destexhe, 2000). An orchestrated neurochemical process involving sleep promoting and arousal centers in the brain, a sleep state is typically marked by lessened consciousness, lessened movement of the skeletal muscles and a slowed-down metabolism, and has an essential restorative function and an important role in memory consolidation (Zisapel, 2005).

The process of falling asleep can be optimally measured by considering a mixture of behavioural, EEG, physiological, and subjective information (Ogilvie, 2001). The amplitude of the changes in brain metabolism and neuronal activity that occurs during sleep exceeds those which occur during most waking periods (Sejnowski and Destexhe, 2000). It has also been suggested that the sleep propensity depends on the amount of sleep deprivation and on the circadian rhythms (Zisapel, 2005).

2.7 Wake-sleep transition

Sleep is associated with characteristic changes in central nervous activity as measured by polysomnography (EEG, EMG, and EOG). The standard criteria used to define sleep onset and subsequent sleep stages depend on the presence of specified patterns of physiological activity. The R&K criteria (Rechtschaffen and Kales, 1968) use spectral and morphological changes in EEG to separate sleep into different stages and is one of the most commonly used criteria in terms of sleep scoring (Table 2-2).

2.7.1 Sleep stages and physiological parameters

Sleep structure usually alternates between rapid-eye-movement (REM) sleep and non-REM sleep during extended periods of sleep (Rechtschaffen and Kales, 1968). Non-REM sleep, REM sleep, and the wakefulness state are reflected in the cortical EEG activity. During the awake period, differences in the timing of cognitive, perceptual, and motor processes, result in the cortical EEG typically containing desynchronized beta waves. These beta waves correspond to high-frequency, low amplitude waves in the 14-30 Hz range. However, when at rest with eyes closed but still awake, the EEG activity is more synchronized and manifests predominantly as alpha waves in the 8-12 Hz range (Rechtschaffen and Kales, 1968).

According to the R&K criteria, the Non-REMsleep state is further divided into four different sub-stages. Stage 1 refers to the transition of the brain from alpha waves (common in wake state) to theta waves. Stage 2 is characterized by low amplitude slow-wave background with two morphologically distinct waveforms: K-complex and sleep-spindle (Silber *et al.*, 2007). The K-complex is a biphasic or triphasic sharp wave complex which occurs at a frequency of 1–1.7 per min with an average duration of 0.63 s (range 0.5–1.0 s), and peak-to-peak voltage of 100–400 μV (Bremer *et al.*, 1970; Roth *et al.*, 1956). Sleep spindles or a run of alpha rhythm can immediately follow K-complexes. Stages 3 and 4 are characterized by the appearance of delta waves (1-3 Hz) in the EEG, sometimes referred to as slow-wave activity.

REM sleep is characterized by high frequency, low amplitude, desynchronized EEG activity similar to that of stage 1 sleep and wakefulness and, is typically accompanied by episodic

bursts of rapid eye movements in contrast to wakefulness and Non-REM. REM sleep is also characterized by loss of muscle tone**.** It is understood that the sleep usually begins with a brief period of stage 1 Non-REM sleep followed by periods of stage 2, 3, and 4 sleep before cycling back through stage 3 and 2 to enter a period of REM sleep (Buysse *et al.,* 1994).

Table 2-2 Rechtschaffen and Kales sleep-staging scale (modified version) adapted from (Carskadon and Rechtschaffen, 2000; Empson, 1986 and Silber *et al.*, 2007).

Stage/state	EEG	EOG	Scoring Criteria
Relaxed wakefulness	*Eyes closed:* rhythmic alpha; attenuates with attention *Eyes open:*, mixed frequencies and relatively low voltage signals	Voluntary control; saccadic movements; blinks.	>50% of the page (epoch) consists of alpha (8-13 Hz) activity or low voltage, mixed (2-7 Hz) frequency activity.
Stage 1	Relatively low voltage, mixed frequency;	Involuntary, slow, rolling eye movements.	50% of the epoch consists of relatively low voltage mixed (2-7 Hz) activity, and <50% of the epoch contains alpha activity. Slow rolling eye movements lasting several seconds often seen in early stage 1.
Stage 2	*Background:* relatively low voltage, mixed frequency *Sleep spindles:* waxing, waning, 12–14 Hz *K-complex:* negative sharp wave followed immediately by slower positive component;	Occasional slow eye movements near sleep onset.	Appearance of sleep spindles and/or K complexes and <20% of the epoch may contain high voltage (>75 μV, <2 Hz) activity. Sleep spindles and K complexes each must last >0.5 seconds.
Stage 3	High amplitude (≥75 μV), slow frequency (≤2 Hz) waves; frontally maximal.	None, picks up frontal EEG.	20%-50% of the epoch consists of high voltage (>75 μV, <2 Hz) activity.
Stage 4	high amplitude, slow frequency waves.	None, picks up frontal EEG.	>50% of the epoch consists of high voltage (>75 μV, <2 Hz) activity.
REM	Relatively low voltage, mixed frequency; saw-tooth waves; theta activity; slow alpha.	Phasic rapid eye movements.	Relatively low voltage mixed (2-7 Hz) frequency EEG with episodic rapid eye movements and absent or reduced chin EMG activity.

Davis *et al. (*1935); (1938) proposed that characteristics of the wake-to-sleep transition are determined by changes in alpha waves, including its fragmentation and interruption by theta followed by its complete disappearance. The twelve-stage wake-sleep continuum spans

the sleep-onset period from wakefulness to the onset of standard REM sleep and is depicted in Figure 2-5.

During REM sleep, the high-amplitude low-frequency rhythms are replaced with low-amplitude high-frequency activity in the neocortex, characteristic of the wakeful state (Steriade *et al.*, 1993). In sleep, the cortex alternates between periods of slow-wave sleep in the 2–4 Hz range and periods of rapid eye movement (REM) sleep, characterized by sharp waves of activity in the pons, thalamus, and occipital cortex.

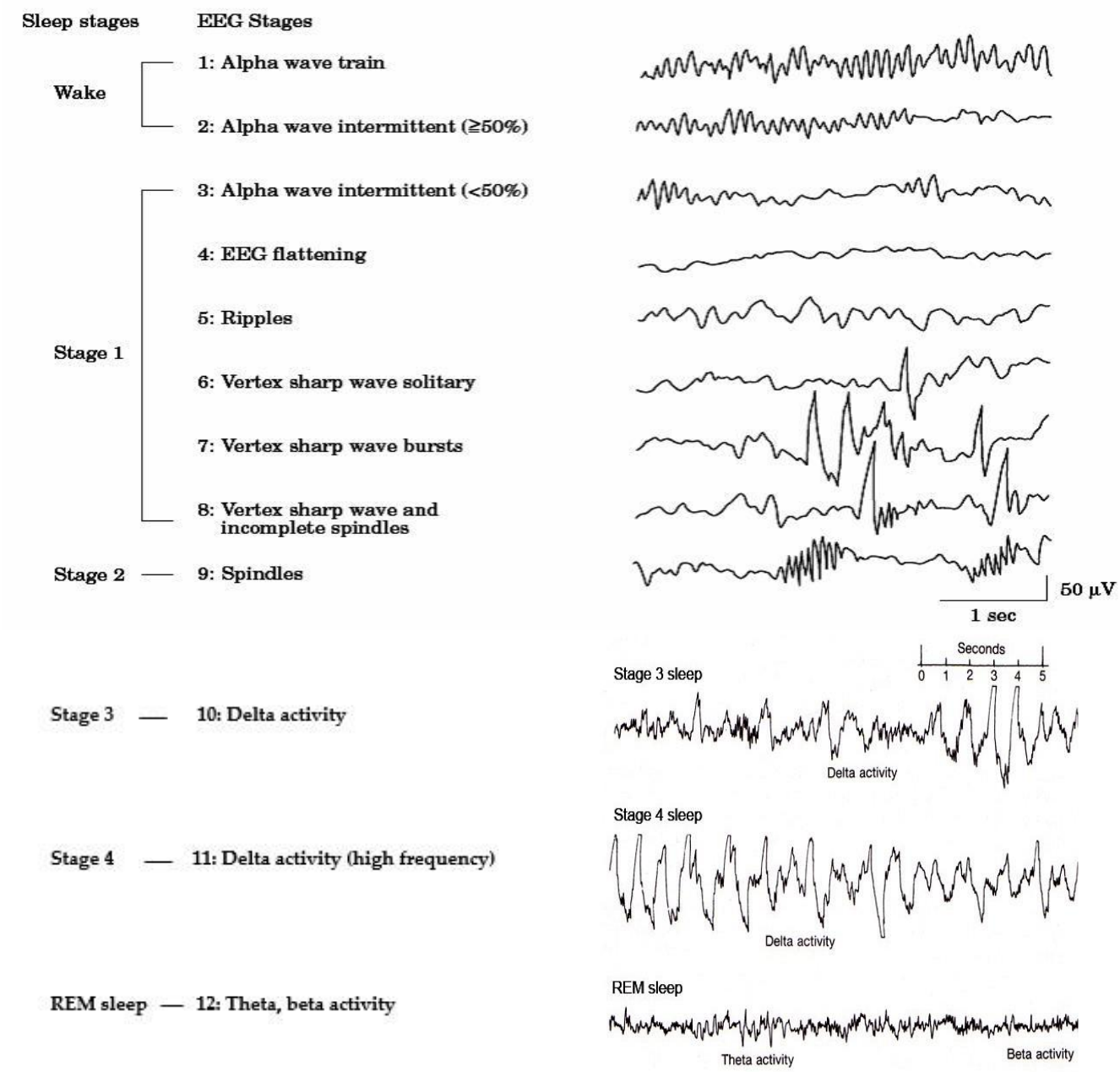


Figure 2-5. Wake-sleep continuum modified (Hori *et al.*, 1994). From top to bottom, the tracings show typical EEG features for each of the stages from wakefulness to REM sleep (Horne, 1988).

2.7.2 Sleep deprivation: physiological and behavioural effects

Sleep is a vitally important biological process in all creatures, as it maintains several aspects of like mental, physical, and emotional wellbeing. Sleep plays a critical role in a variety of functions, including restoration of endocrine and metabolic processes (Spiegel *et al.,* 1999), energy conservation (Moorcroft, 2013), memory consolidation (Diekelmann and Born, 2010), and recovery of cortical functioning (Drummond et al., 1999).

Attempts to determine the impact of sleep deprivation on psychomotor and cognitive performance date back to the 1930s. Warren and Clark (1937) reported that when the subjects were sleep deprived for 65 h the frequency of slow responses increased markedly. Consistent with this finding, Williams *et al,.* (1959) reported that 78 h of sleep deprivation increased the number of slowed responses (lapses) by 18-fold, and lapses were most likely to occur when EEG showed sleep-like characteristics.

Harrison and Horne (1996) utilized a marketing simulation game, to report that 36 hours of total sleep deprivation led to stereotyped decisions and failure to integrate previous feedback, resulting in large financial losses and production errors. In operational settings, similarly global outcome measures are reported (e.g., Friedl *et al.,* 2004; Weinger and Ancoli-Israel, 2002). A meta-analysis of 60 sleep deprived subjects found that "clinical outcomes", the culmination of many decisions, were negatively impacted by physician sleep loss (Philibert, 2005). The increase in frequency of lapses after sleep-deprivation led to the conclusion that sleep deprivation leads to intrusion of performance by brief moments of low arousal during which subjects are unable to respond to the task at hand ('lapse hypothesis') (Wilkinson and Houghton, 1961; Williams, 1959).

Evidence from several behavioural-neuroscience studies indicates that biological processes are altered following a brief period of sleep deprivation. Drummond *et al.* (2000) studied behavioural and neural outcomes in free recall memory tasks and highlight that though the behavioural outcomes showed no significant change, neural responses following sleep deprivation were significantly different. Several other studies have reported a small increase in brain activation and intact performance during sleep deprivation on a variety of tasks (Drummond *et al.* 2001; Portas *et al.* 1998; Chee & Choo. 2004; Dickinson *et al.,* 2008). Another study by Stricker *et al. (*2006) reported changes in neural networks that perform a given task after sleep deprivation

Sleep deprivation has been implicated in several major historical disasters, including the Space Shuttle Challenger explosion, the Exxon Valdez oil spill, and the Chernobyl Nuclear plant explosion (Coren, 1996). The impact of sleep deprivation in the workplace and on society as a whole, while difficult to measure precisely, is massive.

2.8 Relevant Terminology

As mentioned at the start of this chapter, several psychological constructs affect task performance and among these are attention, arousal, and sleep–wake processes which play a crucial role in the ability to perform most tasks. Although researchers have been analysing the wake-sleep transition for over 7 decades, the lapse/drowsiness research area is riddled with different interpretations of concepts and terms and still lacks a set of accepted standard definitions. The following sub-sections provide a clarification of the definitions used throughout this thesis.

2.8.1 Alertness

Alertness refers to a general state of readiness to receive an input. It might also be defined as an ability to sustain attentiveness on a particular task for a certain period of time or the ability to respond to rare and infrequent events (Parasuraman *et al.*, 1998). Loss of alertness is defined as "an operator's internal state, during which time the processing of input information fails to reach the conscious cognitive threshold necessary for the operator to provide an appropriate response" (Freund *et al.*, 1995). There is a definitional overlap in the terms *alertness* and *arousal*, although, alertness is distinguished by its focus on cognitive processing (Oken *et al.*, 2006).

2.8.2 Arousal

Arousal refers to non-specific activation of the cerebral cortex in relation to sleep–wake states. It can also be defined as "a physiological state involving the activation of the reticular activating system in the brainstem and the autonomic nervous system" (Oken *et al.*, 2006). An individual's alertness can be determined by the level of arousal (Kahnemann, 1973). It is a low level process which controls the consciousness, attention, and information processing in the human brain (Segalowitz *et al.*, 1994). To understand the physiology behind the lapses it is essential to first understand the state of arousal.

2.8.3 Attention

Attention is a term that slightly overlaps arousal but more specifically includes some cognitive processing (Nebes and Brady 1989; Posner and Petersen, 1990). It refers to the allocation of processing resources or a more focused activation of cerebral cortex that enhances information processing. It can be divided into several other components, such as sustained attention, focused attention, divided attention, and switching attention.

Sustained attention requires motivation and cognitive processing and can be of particular interest in lapse research as it is directly influenced by sleep deprivation and tiredness. In simple terms, attention can be defined as a measure of the focus of a person on a specific task, ruling out any distractions that may occur during the task.

2.8.4 Drowsiness

Drowsiness can be defined as "The transitional state between wakefulness and sleep associated with a number of subjective feelings and symptoms, sometimes referred to as 'subjective sleepiness" (Shen *et al.*, 2006). Considered as 'stage 1 sleep', it is a transitory state between fully awake state and sleep (Rechtschaffen and Kales, 1968). Both eye movements and EEG are used for identification of drowsiness.

2.8.5 Fatigue

There are many different definitions of fatigue. Crummy *et al. (*2008) gave a loose definition, calling it a broader term than sleepiness, possibly including physical and psychological effects as well as effects from sleepiness. Stedman's medical dictionary defines fatigue as "the decreased capacity or complete inability of an organism, an organ, or a part to function normally because of excessive stimulation or prolonged exertion". The term fatigue has also been used interchangeably with the term "sleepiness" in the literature.

Shen *et al. (*2006) made distinctions between acute and chronic fatigue, physiological and psychological fatigue, and central and peripheral fatigue. Their overarching definition, however, is "an overwhelming sense of tiredness, lack of energy and a feeling of exhaustion, associated with impaired physical and/or cognitive functioning". This definition describes only the effect, making no claims about the cause.

Vanlaar *et al. (*2008) defined fatigue as "a disinclination to continue performing the task at hand, caused by physical labour or repetitive and monotonous activities, such as monitoring a display screen or driving long distances". By this definition, fatigue is not necessarily associated with feeling tired.

May and Baldwin (2009) defined sleep-related fatigue, as an active and passive task-related fatigue. Sleep-related fatigue matches the definition of sleepiness given above. Active task-related fatigue is caused by mental or physical overload due to high-demand conditions, such as busy traffic or poor visibility. Passive task-related fatigue is caused by mental under-load due to low-demand conditions, such as driving on monotonous roads with little traffic.

2.8.6 Sleepiness

Sleepiness may be defined as "the propensity to fall asleep and can only be cured or remedied by falling asleep" (Shen *et al.,* 2006). It is a complex heterogeneous phenomenon (Lal and Craig, 2002). Sleepiness and fatigue often tend to coexist as a result of sleep deprivation.

2.8.7 Lapses

Lapses of responsiveness ('lapses') are brief episodes in which an individual unintentionally stops responding to the task they are performing (Peiris *et al.,* 2006a). Many factors can induce a lapse, including: boredom, mental-fatigue, monotonous environment, circadian rhythm, and sleep deprivation (Lal & Craig, 2001; Ogilvie, 2001; Oken *et al.,* 2006). Lapses might occur due to several factors within the body and the nervous system.

Lapses can be further categorized as EEG microsleeps, behavioural microsleeps, sustained attention lapses and diverted attention lapses.

Sustained attention lapses

Sustained attention in a task may mean the ability to respond to frequent neutral signals and withhold response to rare critical signals. A sustained attention lapse may occur if an individual is highly fatigued or drowsy.

Diverted attention lapses

Schacter and Dodson (2001) define a diverted attention lapse as "an unwarranted distraction of attention from the object of focus by irrelevant thoughts or environmental events". In simple terms it can be defined as the diversion of attention from a given task.

EEG microsleep

An EEG microsleep is an instance of EEG-rated sleep which lasts for a minimum duration of 0.5–10 s and to a maximum of 14–30 s. Harrison *et al. (*1996) define EEG microsleep as a brief period of sleep identified by EEG dominated by a theta activity (4–8 Hz) and an absence of alpha activity (8–13 Hz). EEG characteristics in a microsleep are very close to that of early REM stage 1 or stage 2 sleep. EEG microsleeps have been detected while a participant continued to perform a task, and thus, have been observed without any behavioural signs of momentary sleep (Peiris *et al.,* 2006a).

Microsleep

A microsleep is defined as a brief episode of behavioural pattern of sleep and cessation of response to the moving target while performing a visuomotor tracking task (Peiris *et al.,* 2006a). They are characterized by significant signs such as closure of the eyelids, nodding of the head, drowsy behaviour, and absent task responsiveness. In some cases, it has been reported that microsleeps occur even in well-rested individuals undertaking an extended monotonous task during the early afternoon (Peiris *et al.,* 2006b).

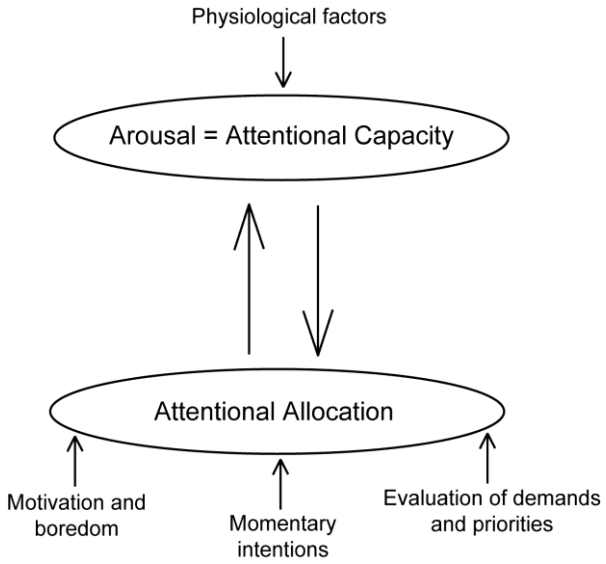


Figure 2-6 Schematic of the attention-arousal model outlined by Kahnemann (1973).

2.9 Summary

This chapter provided a review of various aspects related to lapses of responsiveness including the underlying psychological constructs, classification, measurement, and factors that influence their occurrence. Three important psychological constructs including attention, arousal, and the sleep-wake spectrum were discussed. Multiple factors that influence the occurrence of lapses were also presented. Sleep-deprivation was presented as the most important factor for the occurrence of lapsing. Other contributing factors such as, circadian rhythm and task monotony were also identified as they tend to influence the propensity of lapsing during task-performance due to their interaction with the awake-sleep continuum.

CHAPTER 3

Microsleep detection and drowsiness estimation: A review

Although the dangers of microsleeps are well recognized, research into how to prevent, predict, and detect them has been going on for several decades. Chapter 2 presented a review of the important psychological constructs used to investigate microsleeps and arousal, and an overview of the sleep-wake spectrum. This chapter outlines the current state of this research, in addition to providing a review of knowledge and research in the areas of drowsiness estimation and microlseep detection. A brief overview of research into other cognitive monitoring applications is also presented.

As discussed in Chapter 1, the focus of this research project is on the transitional phase between the awake and sleep states. Ogilvie suggested that a transition from wakefulness to the sleep state is believed to occur "when excitatory and inhibitory post-synaptic potentials become activated along with increasing simultaneity, ultimately resulting in the appearance of an EEG containing more limited variety of frequencies" and, hence, "increasing their amplitude" (Ogilvie, 2001).

Most research into the awake-sleep transition has involved correlating the electrophysiological signals (EEG and EOG) with the performance level of a subject to identify the best indicators of a subject's level of vigilance. However, due to the variety of studies and approaches, inconsistencies may have risen regarding the relationship between physiological events and performance. Non-electrophysiological measurements, such as behavioural data and video recording of eyes, can provide crucial information for microsleep detection (Peiris *et al.*, 2006b; Bergasa *et al.*, 2006; Golz & Sommer, 2010). Video systems can detect the timing and duration of eye closure events. Eyelid movements can be detected on video by measuring the percentage of eye closure (PERCLOS) (Bergasa *et al.*, 2006; Knopp *et*

al., 2012). Behavioural test performance offers another avenue for microsleep detection (Doran *et al.,* 2001). Therefore, several studies in the past have used the combination of both physiological and behavioural metrics to address these concerns (Davidson *et al.,* 2007; Krajewski *et al.,* 2008; Poudel *et al.,* 2010).

3.1 Drowsiness estimation

This section provides a summary of the key EEG-based, eye-closure-based, video-based and head-position based approaches used for the drowsiness estimation.

3.1.1 EEG-based approaches for drowsiness detection

It is generally accepted that sleep and drowsiness have more complex and variable patterns than the alert EEG pattern (Santamaria and Chiappa, 1987). Consequently, EEG has been utilized for decades in sleep research, although the methods and standards differ widely. Ogilvie observed a shift from desynchronized to more synchronized patterns in EEG during the transition from wakefulness to sleep (Ogilvie, 2001). A shift in the awake, posterior-predominant EEG activity from (9-10 Hz) to a low amplitude activity at (3-6 Hz) was reported by Santamaria and Chiappa (1987). While several EEG indicators are clearly correlated with reduced levels of vigilance and performance, none have proven to be highly reliable, at least over short periods of time. Additionally, there are no completely reliable and generally accepted automatic methods for analysing the EEG during studies of vigilance.

Bjerner (1949), observed a reduction in alpha rhythm of EEG which correlated with slowed responses during a reaction time task after sleep-deprivation. Williams *et al. (*1961) found that response failure in an auditory vigilance task is not only associated with reduced alpha but also with increased theta activity. Furthermore, Torsvall and Akerstedt (1984) classified arousal-related lapses according to the quantity of alpha bursts and appearance of theta waves. Townsend and Johnson (1979) showed that errors of omission in sleep-deprived subjects correlate with changes in the EEG spectrum in alpha and theta bands, although the structure of this relationship varies across the subjects. In an auditory detection task, Jung *et al. (*1997) found strong evidence for a monotonic relationship between fluctuations in EEG power and performance in two relatively narrow bands, near 3.7 Hz and 14.7 Hz. Jung *et al. (*1997) also found that alpha activity, generally considered a reliable indicator of drowsiness, is of little use in performance estimation in subjects whose eyes are open.

The relative spectral amplitudes in the alpha and theta bands and the mean frequency of the EEG spectrum were found to be the best combination for predicting auditory alertness level (Huang *et al.,* 2001). The authors later used advanced EEG signal processing techniques to

identify occipital, somatomotor, and central medial components in the EEG that were modulated in different ways by tonic and phasic fluctuation in task performance (Huang *et al.,* 2008). Although the criteria used often overlaps with drowsiness estimation, EEG is also used to identify episodes of EEG-microsleeps (Harrison *et al.,* 1996). EEG-microsleep is defined as a short period (5–14 s) of sleep identified by an EEG dominated by theta activity (4–7 Hz) and an absence of alpha activity (8–11 Hz) (Harrison *et al.,* 1996). EEG-microsleeps usually consist of short episodes, as opposed to drowsiness, which can be longer in duration. Several other researchers have also used similar EEG characteristics to define EEG-microsleeps (Boyle *et al.,* 2008). Tirunahari *et al.* (2003) suggest that EEG-microsleeps are commonly associated with excessive daytime sleepiness and their presence may be a more sensitive indicator of sleepiness than the multiple sleep latency test.

Matousek and Petersen (1983) conducted an EEG study on 41 subjects to develop a technique for automatic evaluation of the vigilance level using the EEG frequency pattern. Their study included 477 epochs of 5-s duration from routine EEGs. Epochs considered to represent either fully alert or stage 1 sleep were selected by visual inspection. The *'gold standard'* for the selected epochs was marked either '0' (alert) or '1' (sleep). Test data used for validation consisted of 265 epochs which belonged to another set of EEGs. Twenty two variables were calculated based on the EEG spectra using 8 EEG channels. The spectra were divided into 6 bands (delta, theta, alpha 1, alpha 2, beta 1, and beta 2). EEG data was related to the assigned level of vigilance using multiple regression. The authors found that the automatic estimates were in good agreement with the individual visual findings in the test data with a correlation coefficient of 0.86. The correlation between the automatic and visual findings of alertness was 0.89 in the training set and 0.86 in the test set.

Makeig and Jung (1995) recruited 15 participants in an eyes-open dual-task (auditory and visual) paradigm experiment. All subjects performed 5 sessions of 30 min each whilst the EEG was recorded at central and parietal/occipital channels referenced to the right mastoid at a sampling rate of 312.5 Hz (pass-band 0.1–100 Hz). The experiment was conducted in a small, warm, and dimly-lit experimental chamber. Normalized log spectra were calculated and 40 frequency bins between 0.61 and 24.4 Hz The discontinuous, irregularly-spaced auditory detection performance index was smoothed using the 95-s exponential window to obtain a continuous local error rate. Two sessions each from 10 (of the 15) subjects were selected for further analysis. A "correlation spectrum" was calculated to find the correlation between smoothed log spectral power and local error rate for all 40 frequency bins for each of the 20 sessions. The mean correlation spectrum was calculated by averaging over the 20 sessions. A 'grand correlation spectrum' was also calculated using data from all 20 sessions with the normalized data described above. They also found that found that a single principal component of EEG spectral variance is linearly related to minute-scale changes in detection

performance. Makeig and Jung (1995) found that mean and grand correlation spectra were nearly identical (r = 0.995, RMS difference = 0.023) and the correlations between EEG power and performance were significant in four frequency bands.

Makeig and Jung (1996) extended their approach further through an auditory detection task and observed that the mean activity levels in the frequency bands, delta (< 4 Hz) and theta (4–6 Hz) and "sleep spindle frequency (14 Hz)" were higher during periods of lowered performance of the 15 subjects from their previous study (Makeig and Jung, 1995). They also observed "lapses" when the 4–6 Hz theta activity increased approximately 10 s prior to an undetected auditory target. The authors also noted that alpha activity was not phase-locked to the subjects' auditory performance fluctuations. A system exclusively based on decomposing the EEG data into different frequency bands and extracting a number of statistical features from the sub bands using the concept of wavelet transforms (WT) was proposed by Subasi *et al.* (2005). They used the extracted features from the sub-bands to represent the distribution of the wavelet coefficients. The obtained features were then used to train an ANN using an error back-propagation algorithm to analyse different levels of alertness.

Different types of ANN schemes, including linear network trained by Widrow-Hoff rule, feed-forward network trained with the Levenberg-Marquardt learning rule, and learning vector quantization (LVQ), were applied to EEG data collected from 17 normal, non-sleep-deprived subjects during normal working hours (Vuckovic *et al*., 2002). They used a training set consisting of 20 epochs of drowsiness and 20 epochs of alertness from each subject and the validation set consisting of 40 epochs of alertness and 40 epochs of drowsiness from each subject. The results from this experiment indicated that the LVQ network had the best classification performance of the implemented ANNs. Vuckovic *et al*. (2002) reported that, when trained on 5 subjects and validated using the remaining 12 subjects, the LVQ ANN showed an agreement of 94 ± 2% between its output and the expert rating.

Jung *et al.* (1997) used nonlinear adaptability of multilayer perceptrons to estimate alertness from the EEG power spectrum and reported an improvement in the performance estimation over linear regression techniques and a reduction in the RMS error. The best performance in terms of RMS was obtained for a network with only 1 hidden layer with 3 nodes. A combination of linear and nonlinear methods for feature fusion, which were obtained by a modified periodogram and delay vector variance (DVV) using the decision process from support vector machines (SVM) and learning vector quantization (LVQ) neural networks, was introduced. The performance of this methodology was illustrated on the "detection of sudden and non-anticipated lapses of attention in car drivers due to drowsiness" (Martin *et al*., 2007).

Jung *et al.* (1997) reported that SVM, utilizing a Gaussian kernel function, outperforms other classifiers such as LVQ and DVV. Also, to get an estimate of whether a signal had undergone a modality change or not during the microsleep event under analysis, DVV was utilized in order to study the common estimation of power spectral densities (Martin *et al.*, 2008). It was reported that SVMs lead to high classification accuracies. However, though prediction was feasible it had much higher errors.

Kiymik *et al.* (2004) proposed a neural network trained on spectral coefficients of a Discrete Wavelet Transform (DWT) for the prediction of the level of drowsiness. Gamma, theta, alpha and beta sub-frequencies of the EEG signals were extracted by WT and the wavelet spectra of extracted EEG signals was used as input to an ANN that could be used to discriminate between alert, drowsy, and sleep states. Pal *et al.* (2008) also developed an unsupervised adaptive neural network approach, primarily to avoid the need for training a classifier using a training set. Kim and Rosen (2010) proposed training free autoregressive (AR) model parameters for the prediction of drowsiness states. Pre-recorded EEG data was used to estimate the internal parameters of an AR model prior, and following the drowsiness states; principal component analysis (PCA) was used. In order to guarantee an efficient training free signal processing, a weighted least square estimation (WLSE) model was utilized.

An EEG-based driver alertness estimation system using both linear and non-linear analysis methods was evaluated in Papadelis *et al*. (2007). This study comprised 21 subjects (20 M, 1 F), sleep-deprived for 24 hours, performing an on-road driving task, driving approximately 200 km on a monotonous highway. This experiment was conducted using a modified car consisting of double support pedals accessible to the instructor, the advanced driver assistance system and sensors, such as a Lane Detection System (LDS) (used for estimating the orientation and position of the test vehicle by identifying the lane borders), and infra-red eye-lid sensors to detect eye blink duration and per minute averaged blink duration (Papadelis *et al*., 2007). The EEG signals, in addition to the horizontal and vertical EOG, ECG, and EMG, were recorded during the driving task. The authors reported that cross-approximate entropy was found to be the most sensitive measure of detecting "sleepiness onset" and that per minute averaged blink rate increased in the minute preceding a driving error.

Mardi *et al.* (2014) tried to demonstrate the separability of the sleepiness and alertness signals with an appropriate margin by extracting suitable features. To this end, they used the EEG data from 10 volunteers, whilst the subjects were engaged in a virtual driving game for 45-min. Subjects were asked to refrain from sleep 20 hours prior to the experiment. The data was labeled either drowsy or alert by using times associated with pass times of the barriers or crash times to them, and chaotic features (including Higuchi's fractal dimension and

Petrosian's fractal dimension) extracted. A two-tailed t-test was applied to measure significant difference ($p<0.05$) between drowsiness and alertness features across all of the EEG channels. They reported that the accuracy of classifying each feature using an artificial neural network was 83.3%.

Muhammad *et al. (*2014) used a driving simulator based study to observe the significant changes that occur in the EEG power spectrum during monotonous driving, in nine healthy volunteers. They calculated the absolute band power of the EEG signal by taking the FFT of the time series signal and then the power spectral density by using Welch method. They concluded that alpha and theta band powers increase substantially ($p<0.05$) when a subject moves from alert state to drowsy state. They also reported that these changes are more dominant in the occipital and parietal regions. Agustina *et al. (*2013) aimed to develop an automatic method to detect the drowsiness stage in EEG records using time, spectral and wavelet analysis. They used a total of 19 features from only one EEG channel to differentiate the alertness and drowsiness stages. After a selection process based on lambda of Wilks criterion, they chose 7 parameters to feed into a neural network classifier. Their method resulted in an accuracy of 87.4% and 83.6% of alertness and drowsiness respectively. The results obtained indicate that the parameters can differentiate both stages.

Goovaerts *et al. (*2014) proposed a method to detect drowsiness based on features of ocular artifacts in EEG signals. They used the ocular artifacts derived from the EEG signals by using Canonical Correlation Analysis (CCA) and Wavelet transforms to automatically select components containing eye blinks. Sixteen features are then calculated from eye blinks and used for drowsiness detection. Their first method is based on linear regression, the second on fuzzy detection. For the first method, they reported that the drowsiness level is correctly detected in 72% of the epochs. Whist, the second method uses fuzzy detection and detects the drowsiness correctly in 65% of the epochs. The best results are obtained when using one single eye blink feature.

3.1.2 Eye-closure and eye-movement based drowsiness systems

Much like EEG, the eyes provide several valuable cues regarding alertness, sleep/wake state. The eyes can be monitored with a video camera, mounted either remotely or on the subject's head, or by electrooculography. With EOG, it may be possible to detect drowsiness through measurements of slow eye movements as they provide the most sensitive measure for differentiating subjective sleepiness from alertness (De Gennaro *et al.*, 2000; Leong *et al.*, 2007). Golz *et al. (*2007) used EOG as part of an automated multi-input system for detecting microsleep events and found vertical EOG to be particularly useful. Whereas, Peiris *et al. (*2008) found that EOG was of little or no value to expert raters identifying lapses during

continuous tracking task. Whether useful for detecting lapses or not, the EOG requires electrodes to be placed around the eyes, causing subject discomfort for a user over extended periods which limits its suitability for a commercial device (Alba *et al.,* 2010). However, another researcher has found that using the independent component analysis (ICA) on the frontal EEG channels can extract much of the EOG signal without the need for electrodes on the face (Alba *et al.,* 2010).

Computerized, non-intrusive, behavioural approaches have been widely used for determining the drowsiness level of drivers by measuring their abnormal behaviours (Lew *et al.,* 2007). Wierwille and Ellsworth (1994) developed PERCLOS (a measure of percentage eye closure over a minute) — a video based scoring of slow eye closures by trained observers — in order to better capture behaviours such as changes in visual perception due to eye-lid closure and associated changes in performance during drowsiness. Several researchers in the past have found PERCLOS to be a good indicator of lapse frequency and duration in the vigilance tasks (Dinges and Grace, 1998; Mallis *et al.,* 1999). However, although PERCLOS captures the average time eyes are closed during drowsiness, it does not provide instantaneous measures of eye-closure that may be associated with individual arousal-related lapses.

The PERCLOS measurement has been found to be a reliable measure to predict drowsiness (Dinges and Grace, 1998) and has been used in commercial products such as Seeing Machines[4] and Lexus[5]. Some researchers have also used multiple facial actions, including inner brow rise, outer brow rise, lip stretch, jaw drop and eye blink, to detect drowsiness (Yin *et al.,* 2009; Lew *et al.,* 2008). However, research on using other behavioural measures, such as yawning (Smith *et al.,* 2003) and head or eye position orientation (Murphy and Trivedi 2010; Xue *et al.,* 2009), to determine the level of drowsiness, is ongoing.

The main limitation of using a computer vision-based approach for the drowsiness detection is the lighting. Normal cameras do not perform well at night. In order to overcome this limitation, Bergasa *et al. (*2006) have used active illumination utilizing an infrared Light Emitting Diode (LED). However, although these work fairly well at night, LEDs are considered less robust during the day. In addition, most of these methods have been tested on data obtained from drivers simulating drowsy behaviour rather than on real video data in which the driver gets naturally drowsy. Mostly, images are acquired using a simple web

[4] http://www.seeingmachines.com/product/dss/

[5] http://www.lexus.eu/range/ls/key-features/safety/safety-driver-monitoring-system.aspx

camera during day (Shen *et al.,* 2012) and an IR camera during night (Flores *et al.,* 2010) at around 30 fps.

A video-based drowsy-driver detection system for heavy vehicles using the PERCLOS measure was proposed by Grace *et al. (*1998). This system consists of two infra-red light sources at different frequencies and two charged-couple-device cameras situated at a 90-degree angle to each other. This system also used retinal reflection differences caused by the two infra-red sources to calculate the percentage of eye closure. The authors also reported that the system can cope adequately with users wearing glasses. Swirski *et al.* (2012) developed a pupil localization algorithm inspired by *"starburst"* which takes into account the difficulties arising from using video taken at a very oblique angle to the eye. They used features from Haar classifier to identify a region of interest around the pupil. Then, a k-means clustering algorithm is implemented on the image histogram to segment the dark pupil from the lighter iris. A Canny edge detector is then applied to locate the edges of the segmented region and a modified version of RANSAC is used to fit an ellipse to the edge points. The authors claim that this algorithm was able to locate the pupil within 5 px of a manually annotated gold standard in 87% of the 600 frames in their test data set. By contrast, *"starburst"* achieved 15% on the same data set.

Malla *et al. (*2010), developed a computer vision technique to automatically measures eye closure in real-time. Their algorithm uses a cascade of detectors to narrow down the search area for more efficient computation. It also uses a Haar classifier for face detection (Viola and Jones, 2004; Lienhart and Maydt, 2002) and smoothens the result using a Kalman filter. Typical anthropometric proportions are used to place regions of interest around the eyes, and then weighted template matching is applied in these regions to locate the centres of the eyes. Malla *et al. (*2010) acknowledged that this approach had limited success since the cascaded nature of the algorithm compounds the errors in each stage. Consequently, their algorithm was only able to discriminate between three degrees of eye closure, and unreliably at that.

Golz *et al.* (2010) identified another limitation of behavioural measures which was highlighted through their experiment. They evaluated various drowsiness monitoring commercial products, and observed that driver state cannot be correlated to driving performance and vehicle status based on behavioural measures alone.

3.2 EEG-based microsleep detection

Sommer *et al. (*2001) used EEG to develop a microsleep detection system, in which clear microsleep and non-microsleep events were selected and small segments of those signals

were analysed and the spectral densities of these segments were classified using three methods of Learning Vector Quantization (LVQ). They reported classification results of up to 91% accuracy with inclusion of all used EEG and EOG channels. Previously investigated pattern recognition techniques for microsleep detection or epileptic seizure detection include neural networks with back propagation (Vuckovic *et al.,* 2002), fuzzy-logic-based classifiers (Coufal, 2009), self-organizing maps (Golz *et al.,* 2001; James *et al.,* 1999), support vector machines (SVM) (Golz and Sommer, 2010), long short-term memory recurrent neural networks (Davidson *et al.,* 2007), and linear discriminant analysis (Peiris *et al.,* 2008).

Golz *et al.* (2007) compared the periodogram, delayed vector variance (DVV), and a combination of the two for extracting features from EEG. The periodogram is an estimator of the true power spectral density (PSD) and assumes linearity and stationarity of the signal. DVV, on the other hand, estimates to what extent the signal has non-linear or stochastic components. The features obtained by these techniques were fed into neural networks for classification. They found that the periodogram was more effective than DVV for extracting features, and even adding DVV to the periodogram provided only minor improvement. Therefore, one can form a hypothesis based on the results from the periodogram approach along with the results of Peiris et al. (2011) that traditional linear techniques could be sufficient for extracting features from EEG suitable for classification.

Wavelet transform has also been used in the past for the detection of epileptic spikes (Goelz et al, 1999 & 2000). Saxena *et al,* (2002) demonstrated that the superior performance of the wavelet method for feature generation. They attribute the superior performance to the specific feature of the WT which characterizes the local regularity of EEG signals. This local regularity feature was then used to distinguish EEG waveforms from noise, artefacts, and baseline drift.

To avoid the need to train a classifier on either individuals or large populations, Pal *et al*. (2008) developed an unsupervised adaptive approach. Their system establishes a baseline alert model by measuring several minutes of EEG when the subject starts driving. It is assumed that they remain alert during this period. When testing subjects in a driving simulator, they observed that deviations in EEG power from the alert baseline measurement corresponded to changes in driving performance. This relationship can be used to detect changes in the subject's alertness. Note that this approach does not actually estimate the level of drowsiness (decreased alertness), but rather changes in alertness. For instance, a change could be detected if the subject has an excited conversation with a passenger.

Peiris *et al. (*2011) developed a detector based on spectral power features from EEG and aimed at detecting microsleeps with second-scale resolution. This method has comparable

performance to the long short-term memory (LSTM) recurrent neural network implementation from their earlier paper (Davidson *et al.* 2007) indicating that non-linear techniques may not provide any more useful information than simpler linear ones. It should be noted, however, that their method in its current form, cannot be applied in real time since the entire EEG feature vector was required for a normalisation step. It may be possible to remove this limitation by using mean values from previous sessions but this has not been tested.

3.3 Other approaches to lapse and microsleep detection

A common behavioural test in drowsiness research is the psychomotor vigilance task (PVT). During a PVT, a subject responds to cues often 2-10 s over 10 min. The ability to respond to various signals and paying attention can be captured by the PVT (Dorrian *et al.*, 2005).The reaction time test (RTT) (Conradt *et al.*, 1999) is another behavioural test developed in which the response latency of the subjects is measured discretely.

The vigilance task is another task which requires continual attention from the subject. A type of vigilance test is the continuous tracking task (CTT), in which a subject must move a marker towards a shifting target (Peiris *et al.*, 2006a). Task performance could be combined with physiological signals to detect microsleep (Peiris *et al.*, 2008). Neural networks with varying architectures, such as a tapped delay line (TDL) linear perceptron and LSTM were used to classify the data based on a sliding feature window (Davidson et al., 2007).

Video systems are often considered to be more practical in real-world applications because of their non-intrusive nature (Bergasa *et al.*, 2006) such as BMW Connected Drive[6]. Despite this, this technology requires tremendous improvement before making it suitable for usage in a commercial product because, just by using the video data alone it is very difficult to determine the stage of sleep; video systems are more useful for the estimation of drowsiness states (Peiris *et al*, 2011). Video systems integrated with other physiological signals such as EEG could be used for the detection of microsleeps (Peiris *et al.*, 2005b).

Instead of using a video camera to observe the eyes, Johns et al. (2007) employed an infrared reflectance oculography technique. This device, the Optalert™ Drowsiness Measurement System[7], consists of a glass frame with an IR LED and phototransistor mounted below the eyes. This device samples at 500 Hz, which is much faster than a video camera, possibly

[6] http://www.bmw.com/com/en/insights/technology/connecteddrive/2013/

[7] http://www.optalert.com/

because it only measures a single value; a one pixel image, as it were. When tested in a driving simulator, the Optalert device has been shown to predict the vehicle completely leaving the road within next 15 min, with sensitivity 83.3% and specificity 60.9%. No figures are given for non-sleep-deprived subjects, however, so we cannot know the false positives under normal conditions or how often non-sleep-deprived people make mistakes. Their drowsiness metric was based on the ratio of the amplitude to the velocity of eyelid movements during eye blinks and eye closure.

Tsuchida *et al,* (2009) used ECG to measure heart rate variability as an indicator of autonomic activity. Their ECG data was fed into a multi-class classifier along with eyelid closure with which they were able to achieve a drowsiness level classification accuracy of 89% with a resolution of 10 s. No indication of the heart rate variability contributing to the classifier's accuracy was presented, and hence it is difficult to say if it was any improvement over eye closure alone.

The eyes provide several useful cues regarding sleep/wake state. Malla *et al.* (2010) developed a computer vision technique which can automatically measure the eye closure in real-time. They begin with a Haar classifier for face detection (Lienhart and Maydt, 2002; Viola and Jones 2004) and smooth the result with a Kalman filter. Typical anthropometric proportions are used to place regions of interest around the eyes, and then weighted template matching is applied in these regions to locate the centres of the eyes. The local extrema of the vertical gradient in these regions mark the upper and lower eyelids, from which the eye closure can be measured relative to a known eyes-open condition. This approach had limited success since the cascaded nature of the algorithm compounds the errors in each stage.

Few groups have used methods based on artificial neural networks (ANN) for extracting features from ECG (Ghongade and Gatol, 2008; Fatemian and Hatzinokas, 2009). However, it is commonly understood that these techniques using ANNs, require exhaustive training settings and estimation of model parameters and, hence, are computationally complex and time consuming. To overcome these limitations, another technique using WT has been used by Li et al. (1995) for feature generation, which reports a high detection rate for the ECG characteristic point detection by using quadratic-spline WTs.

3.4 Alerting techniques and cognitive monitoring applications

A device capable of detecting lapses in real-time and alerting the user in advance would be extremely useful in many occupational sectors. Consequently, that device should provide clear alerts and intervene during a performance lapse. Several methods of providing the alerts have been proposed in the past (Dinges *et al.*, 1998; Grace 2001), including: a beep, buzz, recorded voice messages, a visual gauge of drowsiness, an audible tone, a peppermint scent, temperature changes and vibrations. A device is capable of detecting performance lapses in real-time is of little use unless something critical is done with that information. The device's output needs to either trigger an alert to rouse the subject or intervene in what they are controlling in some way.

The effect of auditory lane-departure warnings on subjects in a driving simulator was studied by Jung *et al. (*2010). The simulated vehicle was subjected to random drifts away from the centre of the lane which the participant was required to correct. In 50% of the cases in which a participant's reaction time to these events was much longer than an alert baseline response time, their system sounded an audible alert (1750 Hz tone, ~68.5 dB). The authors reported that the reaction time for events immediately following those in which an alert was given were significantly shorter than those for which no alert was given. Additionally, the EEG power in the alpha and theta bands was found to be closer to the baseline for the events following an alert, indicating that audible alerts are effective at improving responsiveness.

Mental workload involves various processes, where neurophysiologic, perceptual, and cognitive processes are included (Baldwin & Coyne, 2003), and can be defined as the proportion of information processing capability used to perform a task (Brookhuis & De Waard, 1993, 2000; De Waard, 1996; Kahneman 1973). It has been suggested that the individual capabilities and characteristics (e.g., age, driving experience), motivation to perform the task, strategies applied on task performance, as well as physical and emotional state affect the workload experienced (e.g., Verwey, 2000). A review of research into the effect of mental workload, drowsiness and fatigue on the EEG-based task performance of airline pilots and car drivers was presented in Borghini *et al*. (2014). They report that the current accuracy of the best techniques is around 90%. However, most of the techniques reviewed, operate offline and "no device or convincing algorithm has been published or practically applied for a robust online recognition of such mental states to date" (Borghini *et al*., 2014).

3.5 Limitations of various approaches to microsleep detection and drowsiness estimation

Despite an immense amount of research in this area over many years, there have been no reports of any commercial device or technique capable of providing a reliable indicator of every level and every stage of awake-sleep transition with high temporal resolution. Current technologies also appear to have difficulty coping with the EEG data corresponding to different drowsiness levels in different subjects.

Many drowsiness estimation approaches have been reported in the past, but only a very few have addressed the detection of microsleeps employing artificial neural networks trained on the EEG data. As mentioned in the previous section, Sommer *et al. (*2001) and Golz *et al. (*2007) made promising advances towards detecting microsleeps (albeit offline) using multiple signal sources such as EEG and EOG signals, in which clear microsleep and non-microsleep events were selected and small segments of those signals analyzed and spectral densities of these segments classified using three methods of Learning Vector Quantization (LVQ). The authors reported discrimination performance of 0.90 ± 1.4 between a microsleep and a non-microsleep. However, their system was not capable of prediction of microsleep events.

A combination of linear and nonlinear methods for feature fusion, which were obtained by a modified periodogram and delay vector variance (DVV) using the decision process from support vector machines (SVM) and learning vector quantization (LVQ) neural networks, was introduced by Martin *et al. (*2007). The performance of this methodology was illustrated on the "detection of sudden and non-anticipated lapses of attention in car drivers due to drowsiness." The authors reported that SVM utilizing a Gaussian kernel function outperformed the other classifiers and led to high classification accuracies. However, although prediction was feasible, performance in terms of accuracy was compromised.

From the related field of sleep studies, Álvarez-Estévez *et al. (*2009) developed a fuzzy-logic approach to detect sleep stages from polysomnograms (EEG + ECG + EOG + EMG). The usual method of quantifying sleep stages uses six discrete stages ranging from awake, through sleep 1– sleep 4, to REM sleep. To calculate a more useful measure than discrete stages they also used fuzzy logic to determine the degree of membership of each stage and so obtain a continuous measure of sleep stage. While their work is not directly applicable to identifying microsleeps, the general approach of using fuzzy rules to classify states may be.

Lin *et al.* (2005) tried to confirm a close relationship between "minute scale" changes in the driving performance and the EEG power spectrum. However, due to the variability in

different subjects, an inconsistency in the mean correlation coefficient was found between actual driving performance and the model estimate. It is also not known whether lapses in performances correspond to an increase or decrease in the spectral asymmetry between different brain sites. Usually, changes in the spectral symmetries contribute to some additional cues which might be helpful in enhancing a lapse detection system. Golz & Sommer (2010) used a combination of biometric signals (EEG and EOG) alongside self-reported sleepiness. However, they stated that their study only had reasonable success.

Performance tests have also been used to detect the onset of microsleeps, in which performance drops during drowsy periods (Poudel *et al.,* 2010). Behavioural and cognitive tasks require a user to be constantly performing them, limiting their applicability in the field. There have also been reports of normal non-sleep-deprived subjects having multiple lapses and behavioral microsleeps for several seconds while performing sustained attention tasks (Peiris *et al.,* 2006b; Innes *et al.,* 2013; Poudel *et al*; 2014).

An innate limitation in EEG-based microsleep detection is the relevant brain-states responsible are highly variable and speculative. As a result, ratings based on both video and behavioural recordings can be used to compensate for this (Peiris *et al.,* 2011). The quality of the EEG electrode connections and impedances can vary between sessions and even over the course of the same session, limiting efforts to quantify the relevant brain-states (Othman *et al.,* 2009). While EEG signal quality may be "improved" through filtering, baseline removal, artefact rejection, and similar methods, doing so could distort relevant information.

As such, ICA and several filtering techniques have been used on the EEG to remove ocular artefacts and noise (Davidson *et al.,* 2007; Peiris *et al.,* 2005a; Peiris *et al.,* 2006a; Peiris *et al.,* 2011; Mullen *et al.,* 2013). In spite of the advanced EEG processing, the exact brain-state of a given microsleep is still very speculative, given that the EEG recorded during a microsleep can be highly variable in quality, and that any features derived from that EEG segment may also be similarly variable in quality. In addition, EEG-based microsleep detection tends to be challenged by very noisy and imbalanced data (Davidson *et al.,* 2007; Peiris *et al.,* 2011)

One of the major disadvantages of an EEG-based alerting system is the requirement of an individual to wear EEG electrodes continuously, which may prove be highly impractical for a person in certain environments. A video-based system, by contrast, does not depend on electrodes. Whilst the advantages and disadvantages of a video-based system can be debated, an EEG-based system is still considered viable (Peiris *et al.,* 2011), as it can provide critical information on an individual's brain state in real-time during a sustained attention task. An EEG-based system is also highly desirable in transportation sectors/occupations

(pilots, air-traffic controllers, truck drivers, etc.), and is the only means by which microsleeps can be detected soon after their onset or even prior to their occurrence.

Consequently, for an EEG-based microsleep detection to be reliable, the system must be able to perform accurately and in real time. Several feature extraction and classification algorithms to have been useful in other applications have not yet been applied to the EEG-based microsleep detection problem. The requirements of EEG-based detection mirror demands in other fields, which may assist in performance improvements.

3.6 NeuroTech's Research Programme – previous findings and studies on microsleep detection

To put this project in context, it is useful to review the prior research undertaken as part of NeuroTech's Lapse Research Programme. Over the past 10 years, NeuroTech has conducted experimental studies using several electrophysiological and behavioural tests to investigate the nature and causes of lapses of responsiveness. Using the experimental data from these studies, several state-of-the-art techniques have been, and continue to be, developed to automatically identify, characterize, detect , and predict microsleeps.

3.6.1 Study A: EEG

The first behavioural and EEG dataset to be acquired in NeuroTech's lapse research is the Study A dataset. The Study A dataset has been used extensively in prior research, establishing the current baseline for microsleep detection performance (Davidson *et al.,* 2007; Peiris *et al.,* 2011). In addition, the findings from the Study A dataset formed the main focus of the current research, resulting in several substantial original contributions.

3.6.2 Subjects

Fifteen normal healthy male subjects aged 18–36 years (mean = 26.5) were recruited in Study A. Peiris *et al. (*2011) indicated that the purpose of the age range and gender restriction was to limit the sources of variation in the data. Furthermore, the authors highlighted that none of the 15 subjects had a current or previous neurological or sleep disorder and all had visual acuities of 6/9 (= 20/30) or better in each eye. In addition, all the subjects considered that they slept normally the previous night (mean = 7.8 h, SD = 1.2 h, min = 5.1 h) and, hence were considered non-sleep-deprived. The subjects used a steering wheel (395 mm diameter, wheel-to-screen gain = 1.075 mm/deg) to control an arrow-shaped cursor located near the bottom of the screen. The eye-to-screen distance was 136 cm (Peiris, 2008)

3.6.3 Continuous tracking task (CTT)

A pseudo-random target (bandwidth 0.164 Hz, period 128 s) with an 8-s preview was provided to all the subjects that participated in the study. A preview on the screen in front of the subjects scrolled downwards at a rate of 21.8 mm/s. Subsequently, a target signal was generated by summing 21 sinusoids evenly spaced at 0.00781 Hz intervals and with random phases (Peiris, 2008). A screenshot of the target waveform is depicted in Figure 3-1. Furthermore, the subjects were instructed to keep the arrow-head of the cursor (which could only move horizontally) on the target waveform as closely as possible. This 1-D continuous visuomotor tracking task was conducted for 1 hour, in two separate sessions which were a week apart from one another.

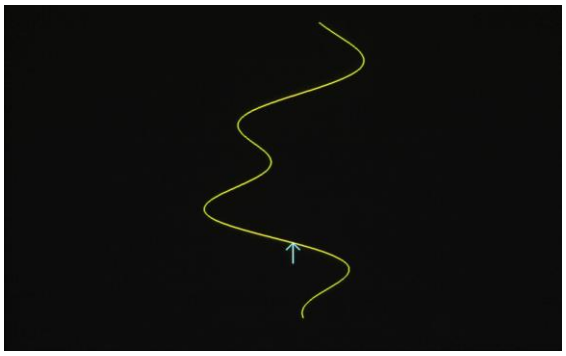


Figure 3-1. Target waveform used for the CTT: a random preview (adapted from Peiris, 2008).

3.6.4 Neurophysiological and behavioral measures

While the subjects were conducting the aforementioned task, EEG was recorded from electrodes at 16 scalp locations, band-pass filtered (0.5–100 Hz), and digitized at 256 Hz with a 16-bit A-D converter. Bipolar derivations used to calculate power spectra were: Fp1–F3, Fp1–F7, Fp2–F4, Fp2–F8, F3–C3, F4–C4, F7–T3, F8–T4, T3–T5, C3–P3, P3–O1, T5-O1, C4–P4, T4–T6, P4–O2, and T6–O2. (Peiris *et al.*, 2011). Bipolar derivations were preferred over the referential due to their ability to minimize the common mode noise.

Facial video was also recorded during the tracking sessions. Video-based microsleeps, were identified by prolonged eye-lid closure, and sometimes accompanied by rolling upward or sideways movements of the eyes, head nodding, and often terminated by waking head jerks. Transitions in the video recording had a time resolution of 1.0 s (Peiris *et al.*, 2011).

3.6.5 Generation of the gold standard

Validation of training and testing data required properly-labelled states indicating a microsleep. The presence of a microsleep was treated as a binary state, where "1" indicated

the presence of a microsleep and "0" indicated the responsive or baseline state. Data rated by human experts served as the gold standard for gauging performance of an automated classifier.

Tracking task performance and video rating were the two independent measures used to identify when subjects were in the microsleep state. Performance lapses in the tracking task were recorded when the tracking response was non-coherent with the target and when the response cursor stopped moving for an extended period while the target was still in motion. These performance lapses were referred to as *flatspots.* Video recordings, synchronized to both EEG and tracking, provided a level of alertness in each subject. EEG data from the subjects who had at least a single definite microsleep over the two sessions in both the video and the tracking response were selected for the microsleep detection system (N = 8).

In addition, this study used a 6-point scale to measure alertness: 1=alert, 2=distracted, 3=forced eye closure while alert, 4=light drowsy, 5=deep drowsy, and 6=sleep (including microsleeps). 'Alert' periods included the fast eye blinks associated with normal facial tone. 'Distracted' intervals were the ones which were associated with few momentary diversions from the task. 'Forced eye closure' was when a subject's eyes closes while remaining alert. The 'light drowsy' state was characterized by the subject's blink rate slowing and facial tone shifting. The 'deep drowsy' was characterized with fewer eye movements and partial eye closure. Finally, the 'sleep' state had prolonged eye-lid closure with head repeated nodding and some jerks (Peiris, 2008).

Flat spots with longer durations were also examined for overlaps with video microsleeps. Definite microsleeps were defined as a combinations of flat spots and video lapses. While rating was performed conservatively, some segments were ambiguous as to whether they were microsleeps or not. Only the segments which were rated 6 on the scale, in addition to the definite microsleeps, were considered as the gold standard in this study. While this gold standard was not a perfect measure of the brain-state during microsleep events, it was as close as human experts could provide.

Different gold standards were also considered, corresponding to different scenarios for defining events as discussed in prior research (Davidson *et al.,* 2007; Peiris *et al.,* 2011). The different gold standards considered for Study A included,

1) Simultaneous video microsleeps and flat spots = 1, and all other states = 0. (Definite microsleep)

2) Flat spots alone = 1, and all other states = 0.

3) Video microsleeps alone = 1, and all other states = 0.

4) Lapses (video microsleeps and/or flat spots) = 1, and all other states = 0. The presence of either a video microsleep *or* a flat spot usually provides a sound, albeit a conservative, indicator of the presence of a microsleep.

3.6.6 Current performance standards

Peiris *et al.* (2006b) reported that during microsleeps, an increase in EEG power in the delta, theta, and alpha spectral bands, and a decrease in the beta and gamma bands was observed. They also reported that 14 of the 15 subjects had at least one lapse, with an overall rate of 39.3 ± 1.9 h^{-1} (mean ± SE) and duration of 3.4 ± 0.5 s (Peiris *et al.*, 2006b).

Davidson *et al. (*2007), developed a microsleep detector using a long short-term memory (LSTM) recurrent neural network operating continuously at 1 Hz, with normalized EEG log-power spectrum inputs from two bipolar EEG derivations. Microsleeps, identified using a combination of video rating and tracking behaviour were used to train this LSTM-based detector. The authors reported that the overall performance of the microsleep detector was satisfactory, with $\varphi = 0.38 \pm 0.05$, area under the curve from receiver operating characteristic analysis of 0.84 ± 0.02 (mean ± SE) and area under the precision-recall curve of 0.41 ± 0.08. This technique was claimed to be the first application of a LSTM network to an EEG analysis problem, and also compared the LSTM-based detector to a tapped delay-line linear perceptron (TDL-LP), tapped delay-line multilayer perceptron (TDL-MLP) Davidson *et al. (*2007). They reported that the LSTM-based microsleep detector achieved performance equivalent to the best tapped delay-line multilayer perceptron (TDL-MLP) but did not require an input buffer. Peiris *et al. (*2011), reported the performance of a lapse detector, estimating the lapse/not-lapse state at 1 Hz based on power spectral features, approximate entropy, fractal dimension, and Lempel-Ziv complexity of the EEG. Their system used a linear discriminants analysis (LDA) to form detection models based on individual subject data, and a stacked generalization approach to combine the outputs of multiple classifiers to obtain the final prediction. They stated that the best lapse state estimation performance was achieved using the detector model created, and that employed power spectral features with an area under the curve from receiver operating characteristic analysis, of 0.86 ± 0.03 (mean ± SE) and an area under the precision-recall curve of 0.43 ± 0.09. They claimed that this approach was able detect microsleep states from spectral power features with moderate accuracy (sensitivity = 73.5%, selectivity = 25.5%). The authors also report that this method resulted in the highest lapse detection performance seen to date (at the time of publication) of $\varphi = 0.39 \pm 0.06$.

To further the claims from Peiris *et al. (*2011), more recently, LaRocco (2015), reported the performance of a microsleep detector system with the same configuration (LDA to form

detection models based on individual subject data, and stacked generalization approach to combine the outputs of multiple classifiers). He used input features from principal component analysis (PCA), However, his results were moderate at best, with a Pearson correlation coefficient $\varphi = 0.40 \pm 0.05$. LaRocco (2015) also implemented a supervised feature reduction algorithm, average distance between events and non-events (ADEN)-on the same data, but reported a drop in lapse detector performance with $\varphi = 0.33 \pm 0.05$.

LaRocco (2015) also used a variation of ADEN on a Gaussian-SVM classifier. This method, however, did not achieve an improvement in the performance, while in fact demonstrated a drop in phi ($\varphi = 0.32 \pm 0.04$). On the same Study A dataset, LaRocco (2015) further reported that an LDA-based classifier was able to outperform the Gaussian-based SVM classifier in terms of the phi correlation. However, LaRocco does not provide a complete overview (sensitivity, specificity, AUC-ROC, or AUC-PR) into the detection performance of the Gaussian-SVM classifier.

In summary, the overall best performance of microsleep detection on data from Study A was achieved by Peiris *et al. (*2011) using an LDA and stacked generalization based technique on EEG-power spectral features, and is considered to be the current benchmark performance for lapse detection. Figure 3-2 highlights the current performance measures of different systems on the Study A (N=8) dataset. Section §6.1 presents different findings and observations on the Study A training set, yielded from this research.

Consequently, the main focus of this project was to develop a state-of-the-art lapse/microsleep detection system by employing advanced classifier schemes, such asreservoir computing (RC). As a result, the following chapters comprehensively describe and critically analyze the novel microsleep detection system, which used RC-based models in addition to multiple supervised feature selection/ reduction techniques.

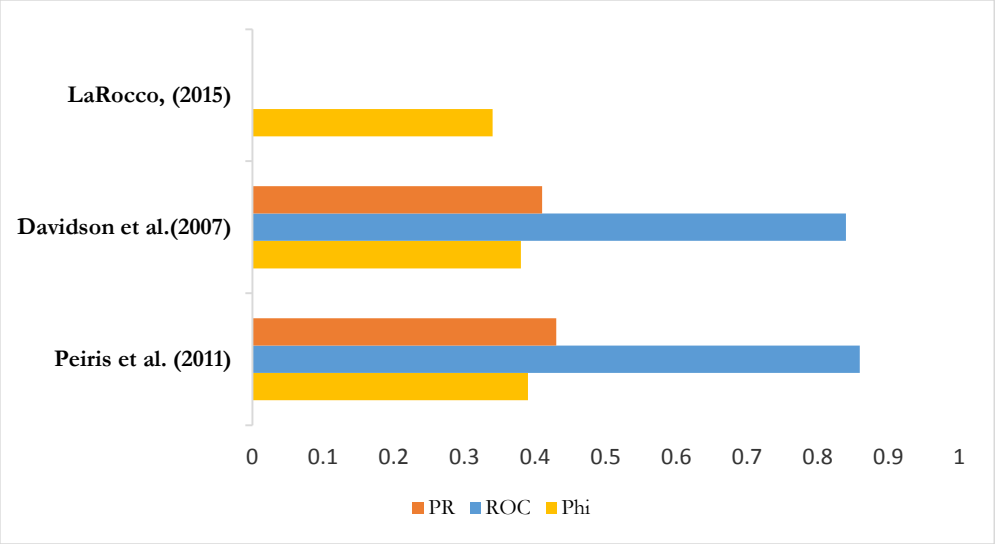


Figure 3-2 Current performance measures on the (N=8) dataset. Three independent parameters (phi correlation, receiver operator characteristics and precision recall) are depicted as percentages (scale 1: 100).

3.7 Summary

The two physiological signals used most often in prior lapse detection research are EEG and video of the eyes, while behavioural signals can be used in specific situations such as driving. This chapter provided a review of several algorithms used to detect classify lapses of responsiveness states and microsleeps. A review was also made on current state of research in NeuroTech. As stated in Chapter 2, detecting lapses of responsiveness and microsleeps is extremely difficult as there is no one signal or feature which is a reliable indicator of microsleep events, so we will have to resort to a combination of signals in an attempt to achieve an acceptable level of accuracy. Consequently, one of the aims of this research project was to explore new methods of automating microsleep detection and, in the process, increase the accuracy of detection.

CHAPTER 4

Machine learning systems

The idea of an artificially intelligent (Turing, 1950) and independent learning machine has fascinated humans for many decades (Russell and Norvig, 2003). The field of classical artificial intelligence (AI) coalesced in the 1950s drawing on an understanding of the brain from neuroscience (Brownlee, 2011). Arthur Samuel, in 1959, defined the concept of machine learning (ML) as the "ability of computers to learn to function in ways that they were not specifically programmed to do" (Samuel, 1959). This chapter provides an in-depth overview on learning from data, different types of systems and methods applied for learning from data, and various algorithms used in learning.

4.1 What is learning?

Learning is broadly defined as the ability of a computational system to adaptively improve its performance with experience accumulated from the observed data. Learning from data plays a crucial role in the fields of artificial intelligence (AI), data mining and statistics, intersecting with many areas of biomedical engineering and other disciplines. Similar to the field of statistics, the basic premise of learning from data is using a set of observations to uncover an underlying process (Abu-Mostafa *et al.,* 2012). The main distinction between the fields of statistics and machine learning is that statistics mainly relies on testing hypotheses, whereas machine learning relies on formulating the process of generalization as a search through possible hypotheses (Back *et al.,* 1999).

Learning models that continually evolve produce increasingly reliable results, reducing the need for human intervention. These adaptive models can be used to automatically produce accurate and repeatable decisions. Learning models can be applied in cases where the desired outcome is known or not known beforehand. One way of visualizing a learning

model, that distinguishes it from statistical approaches, is to visualize a search through a space of possible concept descriptions that fits the available data.

One area of ML that is rigorously applied in the fields of speech and object recognition, image segmentation, modelling language and human motion, etc., is the artificial neural network (ANN). ANNs are a group of algorithms that are loosely based on our understanding of the brain. ANNs are often depicted as computational models in the field of neuroscience and are based on biological neural networks (Bishop, 1995). They relate to the understanding of the human brain as a computer. These ANNs are often referred to as dynamical systems and allow us to understand how the dynamics can be represented as computational functions. So, how can a neural network learn from experience? The classic paradigm is that the strength of the synaptic connections is adjusted in response to experience, i.e., learning. Neural network learning processes are often considered as adaptive learning and are typically applied to function approximation and pattern recognition domains (Brownlee, 2011).

A few other commonly used algorithms in ML include: decision trees, support vector machines, local search optimization techniques (e.g., genetic algorithms.), kernel density estimation, hidden Markov models, Gaussian mixture models, Bayesian networks, learning automata, nearest-neighbour mapping, and self-organizing maps.

4.2 Types of learning

Among the different types of learning tasks, a crucial distinction is drawn between supervised, unsupervised, semi-supervised, and reinforcement learning. Most ML applications use supervised learning schemes. The main focus of this research is catered towards the supervised learning schemes, as the majority of the algorithms used in this work use these models.

4.2.1 Supervised learning

Supervised learning is the most studied and most utilized type of method in ML, and is commonly used in applications where historical data is used to predict the likelihood of a future event (Buteneers *et al.*, 2013). Supervised learning algorithms are trained using systematically labelled examples, where the training data contains explicit examples of the precise outputs for a given set of inputs. The learning algorithm begins with a set of inputs and their corresponding correct outputs. The training algorithm learns by comparing its actual output with the correct outputs to find errors, and then modifies the model accordingly. Using techniques such as regression, classification, prediction, and gradient

boosting, supervised learning uses patterns to predict the values of the label on additional unlabelled data (Abu-Mostafa *et al.,* 2012).

Supervised learning schemes can be further sub-categorized into feed-forward neural networks (FFNs) and recurrent neural networks (RNNs). Feed-forward networks are networks without internal feedback. Every unit in a layer is connected with all the units in the previous layer without any feedback. Hence, the name *feedforward* neural network. Recurrent networks on the other hand, are more complex. RNNs, unlike FFNs, deal with feedback by supporting lateral and top-down connections.

Some of the advantages of RNNs are their adaptability, input-output mapping, fault tolerance, and parallelism. However, one of the major disadvantages with these networks is their large processing time due to their complex architecture.

4.2.2 Unsupervised learning

In contrast to the supervised learning scheme, unsupervised learning is often used against data that has no historical labels. The learning system is not provided with the "correct answer." With a goal to explore the data and find some structure within, the training algorithm is left to figure out from the data presented (Cohen, 1995). Some popular techniques in this type of learning include, self-organizing maps, nearest-neighbour mapping, and singular value decomposition. These algorithms have been used in the past to segment text topics, recommend items, and identify data outliers (Dy and Brodley, 2004; Ranzato *et al*., 2007; Pan *et al*., 2013). Similarly, semi-supervised learning also makes use make use of unlabeled data (i.e., data with no historical labels) for training.

4.2.3 Reinforcement learning

Often used in the fields of robotics, gaming, and navigation, reinforcement learning (RL) is learning by interacting with an environment (Woergoetter and Porr 2008). With RL, the learning algorithm is not told what actions to take, as in most forms of machine learning, but instead the algorithm must discover which actions yield the most reward by trying them through trial and error.

4.3 Classification & types of classifiers

Pattern recognition/ classification is the assignment of a label to a given input value. It is the science of making inferences from perceptual data, using tools from statistics, probability, computational geometry, machine learning, signal processing, and algorithm design. Thus, it

is of central importance to artificial intelligence and computer vision, and has far-reaching applications in engineering, science, medicine, and business (Snapp *et al.*, 2001). Unlike most traditional statistical models, the models created by machine learning algorithms are often nonlinear and can have many thousands of rules or parameters that define the mode. For reasons that will be outlined in the following sections of this thesis, reservoir computing models will be used as the basis for a superior classifier in the EEG-based microsleep detection system.

4.4 Linear discriminant analysis

LDA is a statistical technique used to determine which continuous variables can discriminate between two or more groups (Ripley, 1996). LDA assumes that the group memberships of the initial cases (training set) are known correctly. This analysis yields information which can then be used to classify a future case with an unknown group membership, into a group (Hilbert, 2009). This method maximizes the ratio of between-class variance to the within-class variance in any particular data set, thereby guaranteeing maximal separability.

In essence, LDA calculates the within-group and between-group variances of training data and draws a boundary between them. Depending on which side of the boundary a new observation is assigned to, it is assigned a different group label. LDA was used to form classification models capable of detecting microsleeps by Peiris *et al.* (2011) and will be set as the baseline for comparison with other classifier models.

The algorithm used for training a microsleep detection system using LDA is as follows:

1. Compute the *d*-dimensional mean vectors for the different classes from the input dataset.
2. Compute the scatter matrices (in-between-class and within-class scatter matrix).
3. Compute the eigenvectors $(e_1, e_2, ..., e_d)$ and corresponding eigenvalues $(\lambda_1, \lambda_2, ..., \lambda_d)$ for the scatter matrices.
4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a $k \times d$ dimensional matrix $\mathbf{W}$ (where every column represents an eigenvector).
5. Use the obtained $k \times d$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the mathematical equation: $\mathbf{Y} = \mathbf{X} \bullet \mathbf{W}$, where $\mathbf{X}$ is a $n \times d$-dimensional matrix representing the n samples, and $\mathbf{Y}$ represents the transformed $n \times k$-dimensional samples in the new subspace.

4.5 k-Nearest neighbours classifier

Neighbour-based classification is a type of instance-based learning or non-generalizing learning. This algorithm does not attempt to construct a general internal model, but simply stores instances of the training data (Alexandr and Indyk, 2006). Classification is computed from a simple majority vote of the nearest neighbours of each point. A query point is assigned to the data class which has the most representatives within the nearest neighbours of the point. k-nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure, e.g., distance functions.

The kNN algorithm implements learning based on the k-nearest neighbours of each query point. A case is classified by a majority vote of its neighbours, with the case being assigned to the class most common amongst its k-nearest neighbours measured by a distance function based on the Euclidean distance. The basic kNN classification algorithm uses uniform weights: that is, the value assigned to a query point is computed from a simple majority vote of the nearest neighbours. Under some circumstances it is better to weight the neighbours such that nearer neighbours contribute more to the fit. A detailed implementation and the mathematical formulation of kNN classifiers can be found in Bawa *et al.* (2005).

4.6 Support vector machines

Kernel-based algorithms (such as Gaussian processes, Bayes point machines, kernel principal component analysis, and support vectors) signified a major development in machine learning algorithms. Support Vector Machines (SVM) are a set of supervised learning techniques used for regression, classification, and outlier detection based on the concept of decision planes which define decision boundaries (Ruping, 2001). SVM performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels (Hill & Lewicki, 2006). Support vectors turn high dimensionality problems into linear classification problems. SVM algorithms are also well equipped to handle multiple continuous and categorical variables.

SVMs provide a unique solution, since their optimality problem is convex (Shawe-Taylor and Cristianini, 2004). This is an advantage compared to the neural networks schemes, which have multiple solutions associated with local minima, and for this reason may not offer a robust solution over different samples. A few other key advantages and disadvantages of SVM-based approaches are discussed in Table 4.1.

Table 4-1 Advantages & disadvantages of SVM-based approaches. (Shawe-Taylor and Cristianini, 2004; Hill & Lewicki, 2006)

Advantages
High dimensional space SVM algorithms are extremely effective in high-dimensional spaces and in cases where number of dimensions is greater than the number of samples.
Kernel functions. Kernel functions provide versatility to SVM methods as different custom kernels can be specified for the decision function.
Memory efficiency – support vectors SVM uses a subset of training points in the decision function called support vectors which have direct bearing on the optimum location of the data points on the decision surface. Support vectors help in finding the optimal hyperplane which stems from the function class with the lowest "capacity".
Modelling non-linear features SVM maximizes the margin of error, so that the final model is slightly more robust (compared to linear regression methods) but more importantly, using kernels functions of SVMs even non-linear relations can be derived with in the data.
Disadvantages
Probability estimates SVM (apart from relevance vector machines) do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.
Over-fitting Similar to other well-known classifier techniques, if the number of features is much greater than the number of samples, the SVM method is likely to give poorer generalized (cross-validation) performance due to over-fitting of the training data.
Computation Quadratic programming introduces complexity as the higher order polynomials can often become computationally expensive.

Implementation

Given a training set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ and training labels $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$, SVM aims to create a hyperplane $\mathbf{w}.\mathbf{x} + b$ (where $\mathbf{w}$ is the weight vector and b is the bias) which can separate the data into two binary classes e.g., microsleep vs. non-microsleep.

$$\mathbf{w}.\mathbf{x}_i + b \geq 1 \quad \xi_i \qquad \text{if} \qquad \mathbf{y}_i = 1, \tag{4.1}$$

$$\mathbf{w}.\mathbf{x}_i + b \leq \xi_i - 1 \qquad \text{if} \qquad \mathbf{y}_i = -1, \tag{4.2}$$

$$\xi_i \geq 0 \ \forall \ \ i, b. \tag{4.3}$$

The number of correct classifications are represented with $(\xi_i = 0)$. The separating hyperplane with the maximum distance to the closest training sample is found when $(\xi_i = 0)$.

SVM uses an iterative training algorithm (quadratic programming) to construct an optimal hyperplane and minimize the error function. The final classification (or model output for a given x) is denoted by *f(x)*, and is computed as

$$f(x) = \sum_{i=1}^{r} \alpha_i y_i K(\mathbf{x}, \mathbf{x_i}) + b, \tag{4.4}$$

where α_i are Lagrange multipliers associated with the input $\mathbf{x}$ and $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function that implicitly maps the pattern vectors into a suitable feature space. The training samples for which the corresponding Lagrange multiplier are zero are removed from the training set without affecting the position of the final hyperplane (Hastie *et al.,* 2008). The training samples for which the Lagrange multipliers are non-zero are called support vectors and, hence, the name *support vector machines*.

The best separation of the decision boundary (hyperplane) can be achieved by finding the values of $\alpha_i y_i K(\mathbf{x}, \mathbf{x_i}) = \beta$ and b which minimize $||\beta||$ for all the data points $(\mathbf{x}_j, \mathbf{y}_j)$ such that $\mathbf{y}_j f(\mathbf{x}_j) \geq 1$ (Christianini and Shawe-Taylor, 2000).

The support vectors are the $\mathbf{x}_j$ on the boundary, those for which $\mathbf{y}_j f(\mathbf{x}_j) \geq 1$. The above formulation is a well understood quadratic programming problem for which the optimal solution (β, b) enables classification of a vector *z* as follows (Hill & Lewicki, 2006):

$$\text{class}(z) = \text{sign}(z'\hat{\beta} + \hat{b}) = \text{sign}(\hat{f}(z)), \tag{4.5}$$

where $\hat{f}(z)$ is the classification score and represents the distance of the vector *z* from the hyperplane. The other equally important case in SVM literature is the mathematical dual formulation which is used when the data is not linearly separable. In this case, the training involves minimization of the error function

$$e(x) = \frac{|w|^2}{2} + C \sum_i \xi_i, \tag{4.6}$$

where C is capacity constant, a parameter used to find the correct estimate to fit the training data. It should be noted that the larger the C, the narrower are the margins, and smaller the training error. However, the test error may or may not be higher. Therefore, C is always chosen with care to avoid over-fitting.

For $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i y_i = 0,$ which can be translated to the following dual problem

$$\min \sum_i \alpha_i - \frac{1}{2} \sum_i \alpha_i \alpha_j y_i y_j x_i x_j. \tag{4.7}$$

Previously, SVMs with linear and Gaussian kernel functions have been used for lapse detection (Golz *et al.*, 2007; Krajewski *et al.*, 2008; Golz and Sommer, 2010). More recently, LaRocco (2015) applied SVMs for microsleep detection on the same Study A dataset and reported mediocre performance for SVM with a polynomial kernel (Qiao et al., 2010; Xu et al., 2010) and poor performance with a Gaussian kernel.

4.7 Spiking Neural Networks

In recent years there has been a growing interest in the field of spiking neural networks (SNN), which are more biologically realistic than conventional neural network models. Biological neurons have action potentials, spikes, or pulses which are short, and sudden increases in voltage are required to send and receive information (Du Bois-Reymond 1848). A substantial amount of neural information is encoded within the firing rate of the individual neurons (Ponulak & Kasinski, 2011). Although based on recent physiological experiments in many parts of the nervous system, neural code is, however, founded on the timing of individual action potentials (Beierholm *et al.*, 2001). This section provides a basic overview on the concept of spiking neurons and spiking neural networks, with a summary on models based on synaptic plasticity and spike-based information coding.

Most current ANN models are based on highly simplified brain dynamics and have been extensively used as computational tools to solve complex problems in pattern classification and pattern recognition (Ghosh-Dasstidar and Adeli, 2009). Spiking neural networks are more biologically inspired and are more efficient than their non-spiking predecessors as they can encode temporal information in their signals (Vreeken, (2003). Information transfer in SNN models mimics the information transfer in biological neurons. SNNs make use of the precise timing of spikes or a sequence of spikes (Ghosh-Dastidar and Adeli, 2009). As SNNs offer functional similarities to biological neurons, they are more powerful in analysing elementary processes in the brain, including learning, synaptic plasticity, and neural information processing (Kasabov, 2010). Several researchers also claim that a novel understanding into the dynamics of the human brain can be achieved using SNNs by adding the temporal dimension for information encoding (Ponulak & Kasinski, 2011; Ghosh-Dastidar and Adeli, 2009). This representation of SNNs could in turn result in compact representations of large neural networks.

The ability of synaptic connections to change their strength is referred to as synaptic plasticity. Synaptic plasticity is thought to be the fundamental process underlying learning and memory in biological neural networks (Baudry 1998). Many real-world learning tasks, including speech recognition, pattern classification, event detection, motor control, and spatial navigation are associated with a significant component of temporal dynamics, and as such, learning these systems using the SNNs can be more effective than perceptron-based functions and sigmoidal activation functions (Maass 1997). Spiking neuron models share very distinct and common properties with their biological counterparts, some of which are listed below.

- The ability of SNNs in processing information coming from various inputs and producing single spiking output signals;
- The probability of producing a spike, firing information in SNNs, is increased by excitatory inputs and decreased by inhibitory inputs;
- The dynamics of SNNs are characterized by at least one state variable, i.e., when the internal variables of the spiky neuron model reach a certain state, they generate one or more spikes.

The sequence of a spike train (firing times) is mathematically described as

$$S(t) = \min \sum_{f} \delta(t - t^{f}), \tag{4.8}$$

where f is the label of the spike and $f=1, 2, \ldots, n$. $\delta(.)$ is a Dirac function with $\delta(t) \neq 0$ for $t = 0$ and $\int_{-\infty}^{+\infty} \delta(t)dt = 1$.

Spiking neural network models can be further divided into 2 sub-categories: integrator-and-fire neurons (IF), and leaky-integrator-and-fire neurons (LIF) (Vreeken, (2003). All integrate-and-fire neurons can be stimulated by external current or by synaptic input from presynaptic neurons (Gerstner and Kistler 2002). A schematic of both a real and an integrate-and-fire neuron is depicted in Figure 4-1. The basic circuit for this configuration (LIF) is portrayed inside the dotted circle in the bottom corner of the right-hand side. The basic circuit of an integrate-and-fire model consists of a membrane capacitor C in parallel with an input resistance R driven by a current I(t), and a voltage u(t) is compared to a threshold ϑ (Gerstner and Kistler 2002). The dynamics of the LIF unit are described as

$$C\frac{du}{dt}(t) = \frac{1}{R}u(t) + (i_o(t) + \sum w_j i_j(t)), \tag{4.9}$$

where u(t) represents the model state variable (neural membrane potential), $i_0(t)$ represents the external current driving the neural state, $i_j(t)$ represents the input current from the j-th synaptic input and w_j is the weight of the j-th synapse. In both the IF and LIF models, a neuron fires spikes whenever the membrane potential reaches a firing threshold. The IF model is differentiated from the above LIF model when the resistor R→ ∞.

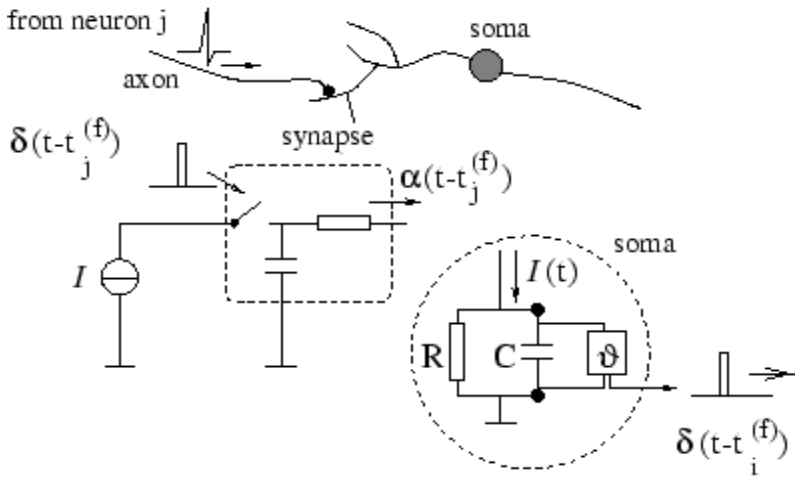


Figure 4-1: Schematic of both the real neuron (on the top) and integrate-and-fire neuron model (adapted from Gerstner and Kistler, 2002). On the left hand side is a the low-pass filter that transforms a spike to a current pulse *I*(t) that charges the capacitor C. On the right hand side, the schematic version of the soma, which generates a spike when voltage u(t) over the capacitor C crosses the threshold ϑ.

SNN-based classifiers have been extensively used in several tasks that require classification of temporal signals, including classification of spike patterns (Maass *et al.*, 2002a, Nikolic *et al.*, 2009; Ponulak & Kasinski 2010; Kasabov *et al.*, 2014), speech recognition (Hopfield & Brody 2000; Verstraeten *et al.*, 2005) and epilepsy detection (Ghosh-Dastidar & Adeli 2009).

Presence of noise

Noise is the influence of random effects that affect the quality of the training sets. The performance of neural network systems decreases with the presence of noise (Horn & Opher, 1999). In contrast, there is evidence that spiking neuron models can easily be equipped with noise-models like noisy threshold, which include parameters that contain a reset function and an integration to enhance the classification performance on noisy datasets (Maass *et al.*, 2002a). Due to these distinct advantages of SNNs, coupled with their novelty in the microsleep detection area, these techniques will be explored for the classification of microsleeps.

4.8 Reservoir computing

In machine learning, vigorous studies have been performed on processing of non-temporal problems using feed-forward structures, such as kernel methods and graphical Bayesian models (Breiman, 1996). However, many major real-world tasks such as prediction, epileptic seizure prediction, noise reduction, and vision and speech recognition, are temporal (Schrauwen *et al.,* 2007). One way of solving these temporal real-world tasks might require adding recurrent connections to the existing feed-forward structures or introducing a tapped delay line to the input stream. Artificial recurrent neural networks represent a large and varied class of computational models that are analogous with biological brain modules and support an architecture based on signal feedback (Mandic and Chambers, 2001). A recurrent network is modelled in such a way that it can develop self-sustained temporal activation dynamics along its recurrent connection pathways, even in the absence of input (Lukoševičius and Jaeger 2009). This topology may allow an RNN to act as a dynamical system, while feed-forward networks act only as functions.

From a dynamical systems perspective, there are two main classes of recurrent networks. Energy-minimizing stochastic dynamics and symmetric connections comprise the first class. Hopfield networks (Hopfield, 1982), Boltzmann Machines (Ackley *et al.,* 1985), and the recently emerging Deep Belief Networks (Hinton & Salakhutdinov, 2006) are the best known implementations of this class of the RNNs. These types of networks are typically trained in some unsupervised learning scheme and use targeted network functionalities such as associative memories, data compression, unsupervised modelling of data distributions, and static pattern classification. This model is run for multiple time steps per single input instance to reach some type of convergence or equilibrium (Taylor *et al.,* 2007).

The second class of RNN models typically incorporate features such as the deterministic update dynamics and directed connections. These models implement nonlinear filters which can transform an input time series into an output time series using their nonlinear dynamics. In these networks, the standard training mode is supervised. These systems appear to be highly promising and fascinating tools for nonlinear time series processing applications, mainly because, they can be shown that, under fairly mild and general assumptions, such RNNs are universal approximators of dynamical systems (Funahashi & Nakamura, 1983). In addition, biological brain modules almost universally exhibit recurrent neural connection pathways, hence indicating that RNNs may be potentially powerful tools for many engineering and biomedical applications.

A fundamentally new approach to RNN design and training was proposed independently by Herbert Jaeger, Wolfgang Maass and Jochen Steil under the name of e*cho state networks*

(Jaeger, 2001), l*iquid state machines* (Maass *et al.,* 2002) and the *back-propagation decorrelation learning* (Steil, 2004) respectively. All these types are referred to as reservoir computing (RC). Since its inception, RC has attracted much attention in the neural network community due to a combination of simplicity, high performance, and ease of use on a wide variety of applications and difficult benchmark tasks. While RC is rooted in the research field of neural networks, its fundamental concepts can be transposed to other excitable media and technologies (Verstraeten, 2009).

According to Verstraeten *et al.* (2007), the RC paradigm can be defined as a randomly generated non-linear dynamical system with fixed parameters (weights)". An RNN such as this is used to map inputs to high-dimensional space in which classification or linear regression can be efficiently accomplished. The states of this dynamical system are linearly combined in an adaptive-readout-output layer, which is the only trained part of the architecture. Some of the distinct advantages of RC-based approaches are described below.

Biological plausibility: RC-based principles have been widely associated to architectural and dynamical properties of mammalian brains. It has been suggested that RC-based models could provide explanations of why biological brains can carry out accurate computations with an 'inaccurate' and noisy physical substrate (Buonomano & Merzenich, 1995; Haeusler & Maass, 2007), and the way in which visual information is superimposed and processed in primary visual cortex (Stanley *et al.,* 1999; Nikolić *et al.,* 2007). RC models may also offer a functional interpretation of cerebellar circuitry (Yamazaki and Tanaka, 2007). Lukoševičius (2012) explains how a central role can be assigned to an RC circuit in a series of models explaining sequential information processing in human and primate brains, most importantly of speech signals (Dominey, 1995; Dominey *et al.,* 2003; Dominey *et al.,* 2006).

Modelling capacity and accuracy: RC-based models can be modelled universally for continuous-time and continuous-value real-time systems with bounded resources (Maass, 2003). In terms of detection accuracy, RC-models have starkly outperformed previous methods of non-linear system identification, prediction, and classification. In several applications, RC-based approaches have been shown to outperform other approaches, including, predicting chaotic dynamics (Weddell, 2008), nonlinear wireless channel equalization (Jaeger and Hass, 2004), financial forecasting (Jaeger, *et al.,* 2006), and Japanese vowel recognition (Jaeger *et al.,* 2007a).

Extensibility: Due to the presence of catastrophic interference from noise, extending previously-learned models of a neural network by new items without impairing or destroying previously learned representations is extremely difficult (French, 2003). To address this phenomenon, RC offers a simple and principled solution: In RC-based systems,

items are represented by new output units, which are appended to the previously established output units of a given reservoir. Since the output weights of different output units are independent of each other, catastrophic interference would not be as high when compared with other models (Lukoševičius, 2012).

SVM-style readout: From a machine learning perspective, a dynamical reservoir is randomly generated and sparsely connected, and functions as a temporal kernel. During simulation, reservoir states form a trajectory, which are dependent on the current external sensory input, but still contain memory traces of previous stimuli. Computation in the output layer occurs by linearly reading out instantaneous states of the reservoir. In this way, reservoir architectures can inherently process spatiotemporal patterns (Schrauwen et al., 2007).

The concept of RC can be established in the field of RNN theory but a different and equally interesting view can also be presented when reservoirs are compared to kernel methods (Scholkopf and Smola, 2002). Indeed, the key idea behind kernel methods is to pre-process the input by applying a form of transformation from input space to feature space, where the feature space is usually considered to be of higher dimension than the input. Classification or regression is then performed in this high-dimensional feature space. The power of kernel methods lies mainly in the fact that this transformation does not need to be computed explicitly, which would be either too costly or simply impossible. This transformation into a higher-dimensional space is similar to the functionality of the reservoir, but there are two major differences: first, in the case of reservoirs, the transformation is explicitly computed, and, second, kernels are not equipped to cope with temporal signals (Maass and Natschlager, 2004).

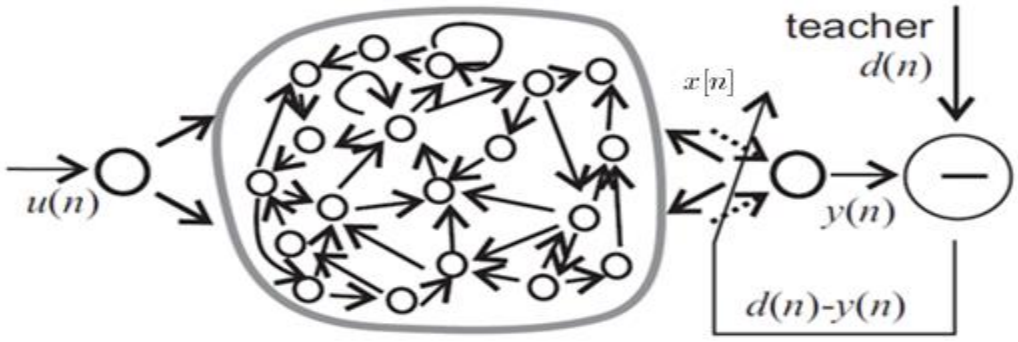
The standard linear readout (output weights) from a reservoir, acting as a temporal kernel, can be trained using the same loss functions and regularizations as in SVM and support vector regression (SVR). Different versions of this approach were investigated (Shi & Han, 2007). Schmidhuber *et al. (*2006) proposed that a standard SVM (with a kernel function) can also be used as a readout from a continuous-value reservoir. Similarly, LSM-type reservoirs can also be read out using special kernel types as discussed by Schrauwen & Campenhout, 2006.

Due to all of these distinct advantages of RC-based methods discussed above, RC methods have quickly become popular in the fields of pattern classification and recognition.

4.8.1 Echo state networks

According to Jaeger (2007), echo state networks (ESNs) are a novel engineering approach to analyse and train RNNs. They are the counterpart of another novel structure, liquid state machines, which are designed to model biological networks. The ESN approach is based on the observation that if a random RNN possesses certain algebraic properties, training only the linear readout of the ESN is often sufficient to achieve better performance in practical applications. The untrained component of an ESN is called a dynamical reservoir, and the resulting states $\mathbf{x}(n)$ are termed echoes of its input history (Jaeger, 2001).

Figure 4-2 shows the configuration of an ESN. The network consists of an input node, the dynamical reservoir that is made up of a large number of sigmoidal units, linear trainable output weights, and an output node. The output signal produced by the network $\mathbf{y}(n+1)$ is compared with the actual value of the process at each time instant and the error is used to update the output weights. The output node could simply be a linear representation that calculates the weighted sum of the networks states or it could be also be a sigmoidal unit, in which case the network output will be restricted to the interval $[-1,+1]$. An activation function $\varphi(\bullet)$ is used to ensure data remains bound and provides non-linear, output capability. $\varphi(\bullet)$ is a sigmoidal neuron activation function, typically a symmetric $\tanh(\bullet)$.



Input Node Dynamic reservoir, W_{DR} Adjustable output weights

Figure 4-2. Echo state network (Image of reservoir layer adapted from (Jaeger, 2007).

A fixed sparse matrix, $\mathbf{W}_{DR}$ is used to implement the recurrent network, and a linear readout, $\mathbf{W}_{out}$, is used for training and to produce an output. The teacher weights matrix is represented by $d(n)$. The state vector, $\mathbf{x}(n)$ maintains the relationship between the input vector, $\mathbf{u}(n)$, and output vector, $\mathbf{y}(n)$, and this is expressed as

$$\mathbf{x}(n) = (\mathbf{W}_{in}\mathbf{u}(n)^T + \mathbf{W}_{DR}\mathbf{x}(n-1)^T + \mathbf{W}_{back}\mathbf{y}(n-1)^T), \quad (4.10)$$

$$\mathbf{y}(n) = (\mathbf{W}_{out}\mathbf{x}(n)^T), \quad (4.11)$$

where $\mathbf{W}_{back}$ represents the feedback matrix; $\mathbf{W}_{DR}$ is the dynamic reservoir matrix; $\mathbf{W}_{out}$, and $\mathbf{W}_{in}$ are output and input weight matrices, and where $n \in \{1, 2, \ldots, N\}$. The output of the network, $\mathbf{y}(\bullet)$, can be expressed in terms of the input vector, $\mathbf{u}(\bullet)$, as

$$\hat{\mathbf{u}}(n+1)=\mathbf{y}(n), \tag{4.12}$$

where, $\hat{\mathbf{u}}(n+1)$ is the predicted input vector, one time-step into the future (Jaeger, 2002).

Network types other than ESN and LSM have also found applications in processing of non-stationary signals in the literature. The general idea is to split data into a number of stationary parts of the signals but this has not proven to work well. It has also been observed that prediction of non-stationary signals has attained relatively less attention in the literature. There have been few attempts in this area but those approaches have been either too general or too restricted to be useful in practical applications (Verstraeten *et al.,* 2006; Jaeger, 2002a).

Jaeger and Haass (2004) illustrated the echo state approach on two tasks: chaotic time series prediction by the Mackey-Glass system and equalization of a wireless communication channel. These tasks were important because they deal with non-stationary signals. The performance of an ESN equalizer is compared with standard channel equalization techniques, and was shown to outperform those by a significant amount (Jaeger and Haas, 2004).

Jaeger (2002) applied an ESN to a non-linear adaptive system identification task. An ESN was used to track the output of a 10th order non-linear autoregressive moving average model in an un-supervised learning fashion to train the output weights of the recursive least squares (RLS) algorithm. The performance of the network was compared to another study on the same model that employed a recurrent multilayer perceptron, where the ESN showed significant improvement, in terms of minimizing the mean squared error (MSE). ESNs have been used for reinforcement learning (Bush and Anderson, 2005). Applications using ESNs in the field of DSP have been quite successful, especially in speech recognition (Maass et al, 2002; Skowronski and Harris 2006; Verstraeten *et al.,* 2005), noise modeling (Jaeger and Haas 2004), and spatio-temporal prediction (Buteneers *et al.,* 2013).

Justification for using echo state networks for microsleep detection

The echo state network is a relatively new approach to designing, training, and analysis of RNNs. More specifically, the fundamental principle of the echo state, which distinguishes it from other types on RNNs, can be summarized as follows

1. ESNs use a large and random sparse matrix as an excitable medium - called a reservoir. When driven by input signals, each unit in the RNN creates its own nonlinear transform of the input;

2. output signals in an ESN model are read out from the excited RNN by some readout mechanism, typically a simple linear combination of the reservoir signals;

3. outputs from an ESN model can be trained using supervised learning, which map high-dimensionality space.

According to Jaeger (2002), a distinction can be been made between traditional approaches to design and application of RC. RC provides a simple but powerful framework for harnessing and using the computational capabilities of nonlinear dynamical systems (Verstraeten, *et al.* 2007). The fundamental idea underlying RC is that the reservoir performs nonlinear mapping with fading memory of the input signals into a higher dimensional space. This then enables the use of relatively simple and computationally undemanding linear classification or regression algorithms.

The structural difference between ESNs and other RNNs is in how they are trained. Generally, most RNNs are layered structures with partial or full feedback connections. Conversely, an ESN is not a layered structure. ESNs have a single hidden layer called a d*ynamical reservoir (DR),* since it stores a large variety of input signal dynamics. The DR has random connectivity. In more conventional RNNs for example, the recurrent multilayer perceptron (RMLP), with 2 fully-connected hidden layers having 10–20 neurons in each layer, is sufficient for characterisation of good results (Chundi *et al.,* 2004). In many studies (Atiya and Ji, 1997; Jaeger 2001), an RMLP network larger than a certain size has been shown to not improve results. As opposed to the relatively compact structure of an RMLP, the ESN typically contains more neurons. On average, a DR of ESN might contain 100-10,000 neurons. However, these distinctions are misleading because, even though the approach of RMLP looks more favourable due to its relatively low size, when it comes to training, ESNs are simpler to implement (Jaeger, 2005).

The reservoir is a temporal kernel (Scholkopf and Smola, 2002). The theory underlying kernel methods is founded in probabilistic statistics rather than biology or neuroscience. Cristianini and Shawe-Taylor (2004) explained kernel methods as transforming input vectors using a nonlinear map into a high-dimensional space. By projecting the input into a higher-dimensional space using the right kernel, the computational performance of these methods can be boosted considerably. Thus, there is some similarity between ESNs and SVMs (Muthuramalingam *et al.,* 2008)

A conventional ESN architecture has a set of linear weights that connect its DR to the output nodes which are the only weights that need to be trained. Also, due to the linearity of the transformation from DR to the output nodes, training an ESN is a simple linear regression which makes them easier to train compared to other RNNs using algorithms such as back propagation through time (Jaeger, 2002). A disadvantage of ESN, however, is the high variability in their performance due to a randomly connected hidden layer. In addition, similar to other classifier techniques such as SVMs, if the number of features is much greater than the number of samples, the ESN method is likely to perform poorly due to extreme over-fitting of the training data. Nevertheless, Jaeger (2005) demonstrated an efficient and more deterministic way to create connections in the hidden layer with a performance better than randomly connected hidden layers without needing to excessively iterate over the training data.

Another benefit of ESNs concerns the initial assignments of their *network states*. ESN have a state-forgetting property; an ESN algorithm will gradually forget the initial condition of its network states. Jaeger (2005) demonstrated that using ESNs for long-term prediction of several chaotic time series signals resulted in an increase in performance (i.e., accuracy) as compared with other RNN techniques.

These properties suggest the superiority of ESNs over other RNNs for nonlinear system modeling and control. In addition, there is evidence which suggests that ESNs are equipped to perform in fields of neuroscience and neuroengineering, e.g., for detection of epileptic seizures (Buteneers et al., 2013). However, despite their compelling properties, ESNs have not been used previously for microsleep detection, which highlights the novelty of, and justification for, their application to microsleep detection.

4.8.2 Liquid state machines

LSMs were developed from a computational neuroscience background to elucidate the principal computational properties of neural microcircuits in neuroscience, aiming at cognitive modelling. In LSM literature, the reservoir is often referred to as "liquid", following an intuitive metaphor of the excited states as ripples on the surface of a pool of water (Lukosevicius, 2010). Inputs to LSMs consist of spike trains in their readouts. The readout function is implemented by a population of model of neurons that receives input from all the liquid-neurons and computes the model output. During training for a particular task, only the synaptic strengths of the connections from the liquid to the readout are modified. The connection between LSM and ESN is that LSMs are continuous-time spiking networks, while ESNs are discrete-time non-spiking networks (Natschlager et al., 2002).

The liquid in LSM corresponds to a DR in the ESN, and the readout unit of LSM refers to the linear layer of the ESN. However, a complex DR with a simple readout may not constitute as an optimal machine for every task.

A LSM model typically uses a multilayer feed-forward neural network structure (of either spiking or sigmoid neurons), or linear readouts similar to the ESNs.

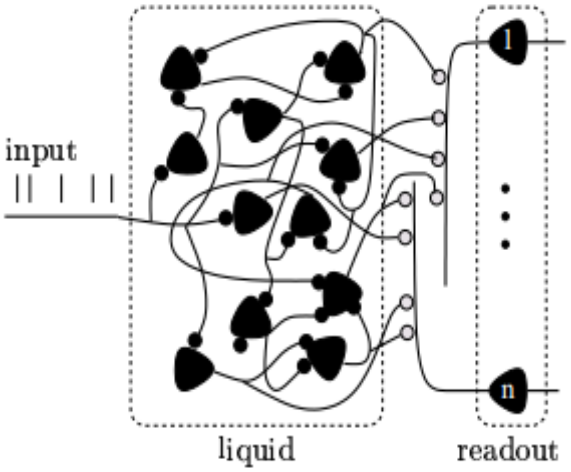


Figure 4-3. Liquid state machine (Image of reservoir layer adapted from (Maass, 2002a).

In a theoretical sense, an LSM consists of two parts

- A high-dimensional 'filter' that maps the current and past inputs onto a state vector;
- readout function (usually memory-less) that maps the state vector onto an output vector.

In order to be computationally useful, the filter should have the point-wise separation property, which states that different input signals should be mapped to different state vectors. Moreover, the readout function should have the universal approximation property. The main theoretical contribution of the LSM model to reservoir computing is their consistent analytical characterizations of computational power of such systems, in terms of measuring the reservoir activations and their supervised tuning of the dynamic reservoir. (Maass et al., 2006). By filtering the information contained in the network and feeding this to a linear readout function, the information processing capabilities of the model can be evaluated very easily. Because of this, most early descriptions of LSMs use a network of biologically plausible spiking neurons, which are excited by external inputs. Partly because of the superior computational power of spiking neurons, LSMs have also been used for

engineering applications as discussed by Verstraeten *et al.* (2005), and robotics (Joshi and Maass, 2005).

In robotics, LSMs have been extensively used in applications to control a simulated robot arm (Joshi and Maass, 2005), to model an existing robot controller (Burgsteiner, 2005), and to perform object tracking and motion prediction (Maass and Legenstein, 2002). LSMs have also been used for event detection (Jaeger, 2005) and in several applications for motor control (Ozturk and Principe, 2005; Salmen and Ploger 2005).

Formal Definition: Similar to ESNs, LSMs comprise an input layer, a large randomly interconnected unit (dynamic reservoir) which supports intermediate states transformed from input, and an output layer. LSMs use a microcircuit as a "liquid filter" which acts as an unbiased analogy of a fading memory from the current and preceding inputs to the circuit. Typically, LSMs employ RNN structures with leaky-integrate and fire neuron. The basic schematic of the LSM, as a continuous stream of 'liquid', is depicted in the Figure 4-4.

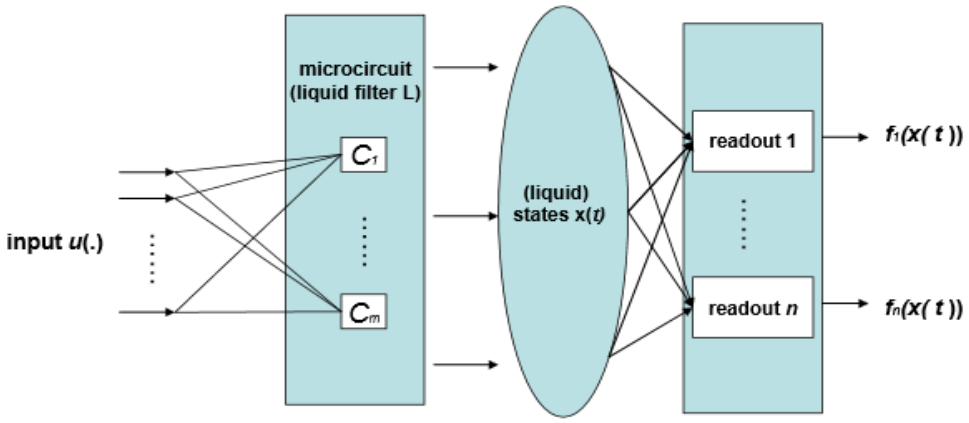


Figure 4-4. An illustration of the dynamical reservoir in the LSM as a continuous stream of liquid (adapted from Maass, 2002).

The liquid filter unit L serves like an excitable medium core to pre-process the input $\mathbf{u}(\bullet)$ and transforms the input into liquid states $\mathbf{x}(t)$

$$\mathbf{x}(t) = (\mathrm{L}\,\mathbf{u})(t). \tag{4.13}$$

The temporal features extracted are passed to the readout unit that is said to be a *memory-less* function f (Maass, 2002), which maps the input $\mathbf{u}(t)$ at time t into the output $f(t)$, with an assumption that the liquid integrates the input $\mathbf{u}(t')$ for $t' < t$ into the liquid states $\mathbf{x}(t) \forall (t)$

$$\mathbf{v}(t) = \mathrm{f}\ \mathbf{x}(t). \tag{4.14}$$

To summarize, LSM models are not trained as such but are initialized with numerous 'temporal neurons' (e.g., leaky-integrate & fire neurons), while their thresholds are selected randomly. The connections between them are also selected randomly (i.e., not each neuron has to have a common connection with each of the other neurons). To 'learn' from an input sequence $\mathbf{u}$, the occurrence of the output $f(t)$ occurred for $\mathbf{x}$ time-units, the model waits for all the $\mathbf{x}$ time-units with the leaky-integrator fire (LIF) detectors, to see which neurons fired at that specific moment. Then, it begins training using a simple classifier (e.g., feed-forward network).

LSM uses many 'temporal neurons' in the 'liquid' layer, to utilize many possible different subsets of firing neurons, hence a specific subset of firing neurons becomes almost unique (Maass, 2002). However, another distinct advantage of LSMs is that these structures not always necessarily take the temporal aspects into account for a learning task, since all temporal processing is done implicitly in the recurrent circuit. Since LSMs use more sophisticated and biologically-realistic models of spiking integrate-and-fire neurons and dynamic synaptic connection models in the reservoir, LSM models perform better than other RC-based approaches in terms of classifying physiological data (Maass, 2007).

4.8.3 Liquid state machines for microsleep detection

As discussed in §4.7, liquid state machine implementations are conceptually similar to ESNs. The distinguishing difference is that LSMs are based on a group of spiking neurons instead of the recurrent artificial neural networks used by the ESNs (Maass, 2002a). RNNs with spiking neurons represent a non-linear dynamical system with a high-dimensional internal state, which is driven by the input.

The internal state vector $\mathbf{u}(k)$ is given as the contribution of all the neurons in the LSM to the membrane potential of a readout neuron at the time k. The complete internal state of the system is determined by the present and past inputs that the network is presented with. Therefore, the history of inputs (local memory) is inherently preserved within the LSM networks (similar to the concept of LSTM networks).

When applied to the Study A dataset, the liquid layer in the LSMs consisted of 6 layers of 20 spiking neurons ($6 \times 20 = 120$) total nuerons). The final algorithm used for training a microsleep detection system using an LSM-based network is as follows.

1. Define the size of the reservoir (*liquid*).

2. Initialize the internal weights of the LSM by generating random values from a Gaussian distribution.

3. Feed the training data $\mathbf{x}(\bullet)$ and record the internal state $\mathbf{u}(k)$ of the LSM over time.

4. Use a supervised training algorithm to compute the weights of each readout neuron based on a given target vector $\mathbf{y}(k)$.

5. Generate the weights of the readout neuron.

6. Run the LSM on the test data.

Similar to the ESN approach, several of the LSM parameters also require reasonable fine tuning to attain high detection rates (Maass, 2002).

4.8.4 Leaky-integrator variants of the ESN

ESNs are mostly used for modelling dynamical systems because of their inherent memory states. Although ESNs with sigmoidal activation functions have been extensively used in the past (Bush and Anderson, 2005; Jaeger and Haas 2004; Jaeger, 2005; Verstraeten, *et al.*, 2005), one of the main disadvantages of these type of networks is that they do not have a time constant which means that their dynamics cannot be "slowed down" (for example, similar to the dynamics of a differential equation). Therefore, an alternative to this was proposed in the form of the leaky-integrator model. In leaky-integrator ESNs, the temporal characteristics of a learning task can be exploited by using the individual state dynamics of the system (Jaeger, 2001; Jaeger *et al.*, 2007). Jaeger (2001) notes that it is almost impossible with standard sigmoid networks to implement ESN generators that model very low frequency waveforms.

Lastly, the continuous-time neuron model, also known as the leaky-ESN model uses the standard sigmoidal activation neuron to perform a leaky integration activation from one of its previous time steps. Leaky-ESN models incorporate a 'time constant' and can be slowed down. Additionally, according to Jaeger (2002), to attain optimal performance, the spectral radius (largest absolute eigenvalue) $\rho(\mathbf{W}_{DR})$ must operate on the edge of the stability (approaching but not equal to unity). Jaeger (2002) also defines these parameters as the heuristics of the ESNs (e.g., $\rho(\mathbf{W}_{DR}) < 1$ and $\rho(\mathbf{W}_{DR}) = \max|\lambda(\mathbf{W})|$, which are considered to be the fundamental requirements to maintain the stability in a network.

When using the leaky-integrator properties of the ESN, the most essential criteria needed for consideration is the 'echo-state' property (Yildiz *et al.*, 2012). The echo-state property states that: a network E: $\mathbf{X} \times \mathbf{U} \rightarrow \mathbf{X}$, where $\mathbf{U}$ is the input range and $\mathbf{X}$ is the state space, and has the echo state property with respect to $\mathbf{U}$: if for any left infinite input sequence $\mathbf{u}^{-\infty} \in \mathbf{U}^{-\infty}$ and any two state vector sequences $\mathbf{x}^{-\infty}$ and, $\mathbf{y}^{-\infty} \in \boldsymbol{X}^{-\infty}$ compatible with $\mathbf{u}^{-\infty}$, it holds that, $x_0 = y_0$.

ESN with leaky neuron model:

According to Jaeger (2002), the continuous-time dynamics $x(t)$ of a leaky-integrator neuron is described by the differential equation

$$\dot{\mathbf{x}} = \frac{1}{\tau}(-a\mathbf{x} + f(\mathbf{w}.\mathbf{x})), \tag{4.15}$$

where **W** represents the weight vector of connections from all the units **x** which feed into a neuron, and *f* represents the neuron's output non-linearity function (typically a sigmoid). The term τ represents the 'time constant' (the larger the time constant, the slower the resulting dynamics of the system) and $-a\mathbf{x}$ refers to the decay potential of each neuron and is responsible for each neuron to retain its previous state. The term $'a'$ denotes the decay constant (the larger the value of *a*, the faster the decay from the previous state, and the larger the relative influence of input from other neurons). Subsequently the dynamics of the method can be represented as

$$\dot{\mathbf{x}} = \frac{1}{\tau}(\mathbf{A}\mathbf{x} + f(\mathbf{W}^{in}\mathbf{u} + \mathbf{W}\mathbf{x} + \mathbf{W}^{back}\mathbf{y})). \tag{4.16}$$

The above dynamics are observed when the entire dynamic reservoir of an ESN is made from the leaky-integrator neurons $\mathbf{x}_i$ and decay constants a_i. In this setup, the term **A** represents a diagonal matrix of the decay constants. Conversely, for discretized time functions, the dynamics change to

$$\mathbf{x}(n+1) = (I - \frac{\delta}{\tau}\mathbf{A})\mathbf{x}(n) + \frac{\delta}{\tau} f(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n)), \tag{4.17}$$

where δ is the step size and an approximate discretized version, ***I*** is the identity matrix. Jaeger, (2003) proposes that to achieve a dynamical stability for the system, the values of δ, τ and A must be chosen such that the absolute value of $I - \delta/\tau \geq 1$. Simplifying this even further by substituting $1 - a_i = r_i$ (where r_i = retainment rate), we arrive at

$$\mathbf{x}(n+1) = \mathrm{R}\mathbf{x}(n) + f(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n)). \tag{4.18}$$

In the above equation, **R** is the diagonal matrix with the retainment rates ($0 \leq r_i < 1$). Ideally, if the retainment rate is ≈ 1, then the unit dynamics are dominated by the retained previous state. If the retainment rate is ≈ 0, then the unit dynamics degenerate towards the sigmoidal units as none of the states would be retained. The ideal situation in this paradigm would be

when a neuron neither retains, nor leaks, more activation than it had earlier. In this model, choosing the values of **R** and **W** becomes more trivial when compared to the sigmoidal activation. Essentially, the leaky neuron approach works similar to a low-pass filter, as it only allows for certain spectral components in the neurons below the pre-set cut-off frequencies to be passed for the initialization of the internal reservoir (Jaeger, 2007).

In addition, ESNs with leaky neuron configurations contain an additional 'global control parameter' standard sigmoid ESNs. Such global control parameters contain an additional spectral radius of the reservoir weight matrix and the leaking rate, which can be optimized particularly for low frequency input signals such as the EEG-based microsleep feature sets.

4.8.5 Overview of reservoir networks

Reservoir networks are neural networks with feedback loops, where the feedback weights have random values and are not tuned during training. Instead, only the weights from the "reservoir" (hidden layer) to the output layer are trained. RC networks are particularly useful for time series analysis, since the numerous feedback loops act as memory in the network, i.e., when successive input signals are sent. RC networks are also referred to as the 'echo states' because of this memory property.

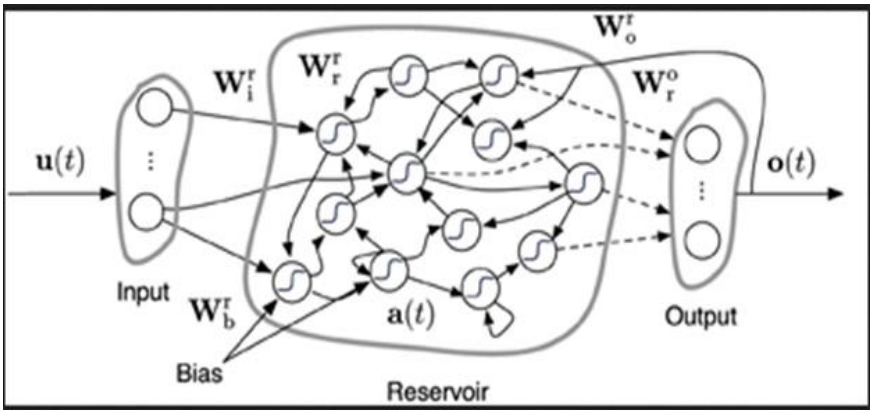


Figure 4-5. Description of an ESN. Dashed arrows are the connections which can be trained. Solid arrows are fixed. $\mathbf{W}^r$ is a weight matrix, and the letters r, i, o, b denote reservoir, input, output, and bias, respectively. $\mathbf{u}(t)$, $\mathbf{o}(t)$ and $\mathbf{a}(t)$ represent the input, output, and reservoir states, respectively (modified from Jaeger, 2002).

As discussed in §4.8, ESNs generally use neurons with a sigmoid activation function. Given this model, the output of a neuron at any given time is a function of the input of the neuron from the previous time step. A LSM, on the other hand is a reservoir network, which is instead made up of spiking neurons. That is, the neurons fire when the successive sum of inputs pass a threshold value. These neurons therefore have an intrinsic, while the ESN only have indirect memory capabilities, i.e., through the use of feedback loops.

Both the sigmoid and spiking neurons have been used extensively in the fields of artificial neural networks. The significance of the reservoir approach however, is its ability to tune the weights corresponding to the output layer. To put Figure 4-5 in perspective, an input signal (left), creates numerous, varying dynamics in the reservoir (centre) and the tuned weights create the desired output (right).

Comparison of SNN, LSM and ESN models

A Spiking neural network is a neuron model, whereas a liquid state machine is a network model. ESNs can also be classified as network models, but only if they support global feedback (i.e., connections from the output back into the reservoir as an additional input). LSMs are part of a group of network models with spiking neurons (also called graded response or analog response). Similarly, ESNs also have the same units, comparable to a typical perceptron algorithm and are thus part of the other (more popular) paradigm where neurons fire at each propagation cycle.

The basic principle behind the LSM is that the neurons are not binary (on/off) but analogous as they decode the time between spikes which is thought to be main source of information transfer between the neurons. The human brain is both analog and digital, there is evidence for both as well as the true mechanism being something completely different (Maass, 2002a). As discussed in the §4.7, LSM is a spiking neural network model. Its main properties are recurrency and that the neuron model has spiking activity (and is usually comprised of complex synaptic models). The recurrent weights in an LSM are not trained in a supervised manner (unlike the typical neural networks/RNNs). LSMs are also used in computational neuroscience (Ju *et al.*, 2013) to study functional properties of neural circuits by abstraction.

Alternatively, ESNs come from an engineering/machine learning background. Usually the neuron model is simple and continuous (like tanh activations). Similar to the LSM models, the recurrent weights ESN are usually left untouched, only the network-to-output connections are trained. ESNs can be seen as a natural extension of linear models (with nonlinear-basis, like polynomials) using time-dependent basis functions (Lukoševičius, 2012).

4.9 Summary

This chapter provided a review of the fundamentals of machine learning algorithms that were used in this research. A distinction between feed-forward and recurrent neural networks was highlighted in detail. In addition, the significance of the non-linear modelling in ML context was also critically emphasized. A number of biologically-inspired neural

network paradigms were also discussed. The two significant models (ESNs and LSMs) in RC were presented in comparison to the traditional SVM-based approaches and the significant distinctions in terms of their mathematical formulation were also offered. In summary, this chapter provides a reader with the necessary understanding of the ML algorithms that will be used for EEG-based microsleep detection in chapter 7.

CHAPTER 5

Development of cascaded-leaky-integrator-echo state network system for EEG-based detection

5.1 Introduction

This chapter is organized as follows. An overview of the development of cascaded leaky-integrator ESN model for microsleep detection on Study A dataset is provided in §5.2. Following this, a summary of mathematical preliminaries and different global control parameters of the cascaded-leaky-integrator ESNs used on the same Study A dataset is also discussed in §5.2. Lastly, an independent experiment, using data from the BCI competition to validate the cascaded-leaky-integrator ESN classifier and its global control parameters, are discussed in §5.3.

5.2 Cascaded-leaky-ESN-based approach for microsleep detection

As discussed in §4.8, a cascaded-leaky-integrator ESN-based algorithm uses a randomly created *reservoir*, which is untrained, in contrast to other traditional RNNs. A distinct advantage of the reservoir-based approach over other RNNs is that parameters, such as training time and stability, can usually be avoided without losing the desired generalization abilities (see §4.8.1) (Jaeger, 2001).

To reiterate, in the leaky-integrator ESN setup, each non-zero input sample excites the dynamical system and pushes the reservoir to a new state. In practice this can be seen as a projection of the input features to higher dimensional feature space, comparable to non-linear regression (Buteneers, *et al.*, 2013). Additionally, a significant distinction in RC-based algorithms is that information from previous inputs is inherently present in the system, which can be slowly forgotten with an exponential decay.

A traditional ESN architecture consists of an input layer, a dynamical reservoir with numerous sparsely inter-connected neurons, and an output layer as described in §4.8 (Jaeger, 2002). This ESN approach also includes a sigmoidal activation function on the simplest additive units. A schematic representation of the influence of the input and bias scaling for a sigmoidal activation function in the form of a hyperbolic tangent for is depicted in Figure 5-1. However, this section of the thesis describes the ESN structures with additional parameters within ESN architecture such as the 'cascaded-leaky-integrator-neurons', where the reservoir units are built with cascaded-leaky-integrator units as opposed to the simple additive units.

Both the standard ESN (sigmoidal neuron) and cascaded-leaky-integrator-ESN-based algorithms were implemented and evaluated for microsleep detection. In the following sections of this chapter, application of both of these methods to a microsleep detection system is explained. In §5.2.2, a more insightful approach for an ESN-based algorithm is described to show how the various individual parameters influence the performance of the training algorithm. Several optimization techniques implemented in order to achieve the best possible microsleep detection rates are also introduced.

As discussed in §4.8.1, the main advantage of this type of architecture is that classification can be enhanced by varying the various parameters of each leaky-integrator unit (Jaeger, 2007). The operation of a reservoir was shown in Figure 4-2; its mathematical formulation and optimization parameters can be further described in the aforementioned section.

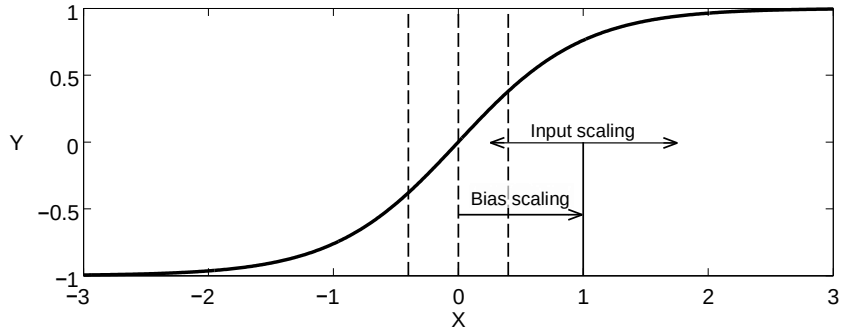


Figure 5-1 A schematic representation of the influence of the input and bias scaling for a hyperbolic tangential function for an ESN. Area between the dashed lines is considered as the linear operational area, whereas outside this area the function is considered non-linear.

The terminology used in the RC networks, as described in §4.7, includes an input $\mathbf{u}(n)$ for which the mathematical formulation of an individual neuron's *activation* values in a reservoir is represented with $\mathbf{x}(n)$ where n is the time, $\mathbf{y}(n)$ is the desired output, and $\hat{\mathbf{y}}(n)$ is the approximated output generated by the ESN.

The inputs of the neurons in the reservoir are connected with input weights, and *bias* in the system is defined as a constant input. Figure 5-1 depicts the output of all neurons in the reservoir for a hyperbolic tangential function. The individual weights of the connections between the input, bias and the dynamic reservoir are represented by the weight matrices $\mathbf{W}_{bias}$ ($N \times 1$) and $\mathbf{W}_{DR}$ ($N \times N$) respectively. Here, N denotes the number of neurons and k denotes the number of inputs. Typically in an ESN, these weight matrices are randomly initialized and are not changed during the training. The ESN initialization operation that was used for the microsleep detection is as follows:

1. Distribute the elements in $\mathbf{W}_{bias}$ i.e., the bias weight matrix, uniformly and between −1 and +1.
2. Initialize the weights of the internal reservoir matrix $\mathbf{W}_{DR}$
3. Randomly set all the input weights in the weight matrix $\mathbf{W}_{in}$ to −1 or +1.

Two ESN models, i.e., the sigmoidal neurons and a new cascaded-leaky neuron structures, were evaluated for the microsleep detection on Study A. The state update equation for both methods is depicted as a weighted sum followed by a hyperbolic tangential function (in Figure 5-1). In the case of the sigmoid neurons, this state update equation is defined as

$$\mathbf{x}[n+1] = \tanh\left(\mathbf{W}_{DR}\,\mathbf{x}[n] + \mathbf{W}_{in}\,\mathbf{u}[n] + \mathbf{W}_{bias}\right). \tag{5.1}$$

Sigmoidal neuron-based ESNs are extensively used in several machine learning applications (Jaeger, 2001; Jaeger, 2002; Jaeger *et al.*, 2007; Yildiz *et al.*, 2012). Their main disadvantage, however, is that it is extremely difficult to slow down their internal network dynamics, as the time constant within their standard sigmoid neurons cannot be constrained. As a result, these networks are more suitable, and are mainly used for, modelling discrete time systems with high frequency changes (e.g., mixed sinusoidal oscillators and noise modelling) making their behaviour quite vulnerable in some occasions where the system needs to perform on slow dynamics.

To overcome the disadvantages of these sigmoidal neurons, a new ESN architecture, referred to in the literature as leaky-integrator neurons, where the internal units supported by the novel cascaded-leaky neuron structures, was evaluated for microsleep detection. This cascaded leaky neuron model incorporates a time constant with multiple individual state

dynamics that can be exploited in various ways to adapt the network to the temporal characteristics of a learning task (Jaeger *et al.,* 2007; Lukoševičius, 2012). The state update equation for a cascaded-leaky-integrator-neuron model is given by

$$\mathbf{x}[n+1] = (1 - \mathbf{y})\mathbf{x}[n] + a\tanh(\mathbf{W}_{DR}\mathbf{x}[n] + \mathbf{W}_{in}\mathbf{u}[n] + \mathbf{W}_{bias}). \quad (5.2)$$

Here, a is the *leakage* rate or the decay potential of each neuron and it represents the leakeage factor at which the previous reservoir state is *leaked* and replaced by a following reservoir state. In essence, the leaky-integrator neuron approach can be interpreted as a sigmoid neuron followed by a first-order low-pass filter (Jaeger *et al.,* 2007). A leaky-integrator approach sets the cutoff frequency of the low-pass filter in the neurons. This extra parameter is used to tune the reservoir memory and timescales (Jaeger *et al.,* 2007; Lukoševičius, 2012).

Both ESN approaches (sigmoidal and cascaded-leaky-integrator-neuron) map the input features to a higher dimensional feature space and generate the following output from the reservoir states using a simple linear regression algorithm:

$$\hat{\mathbf{y}}[n] = \mathbf{W}_{out}\begin{pmatrix}\mathbf{x}[n]\\ \mathbf{1}\end{pmatrix}, \quad (5.3)$$

where the row vector $\mathbf{1}$ is the output bias and $\mathbf{W}_{out}$ is the output weight matrix.

Essentially, ESNs with cascaded-leaky neuron configurations contain additional *global control parameters,* compared to their standard sigmoidal counterparts. These global control parameters contain a spectral radius term of the reservoir weight matrix, in addition to the leaking rate that can be optimized, and in particular, for low frequency input signals such as EEG-based microsleeps.

5.2.1 Creating a microsleep detection model using cascaded-leaky-integrator neurons of the ESNs

As stated in §4.8, echo state networks transform the information at the input to an excited state, where the outputs represent various class hypotheses formulated from the information in state vectors (Jaeger, 2001). This is performed by connecting the output matrix to the inputs, and then the reservoir units to an output weight vector. Following this, the linear regression of the targets on the input matrix is computed. However, this approach yielded results which were unsatisfactory as the sigmoidal neurons failed to capture the low frequency shifts in EEG-based microsleeps and, hence, led to the investigation of a more enhanced architecture that utilized novel cascaded-leaky-integrator-neurons.

In this novel cascaded-leaky-integrator-ESN approach, a single large state sequence of a reservoir matrix is segregated into K state sequences, where K is a small integer, and where each segment is of equal size. This reflects a few equally spaced snapshots of the state vector development when the ESN reads a sample. The regression weights are then computed for all segregated states and the class hypotheses formulated for those weights. The next step is to use each K segregated sequence for a training sample and concatenate them into a single joined state vector, where the number of mappings (connections) from each K sequence to the final state vector is determined by m. Lastly, all output samples are then used for classification facilitated by the computation of linear regression. An illustration of this approach is depicted in Figure 5-2.

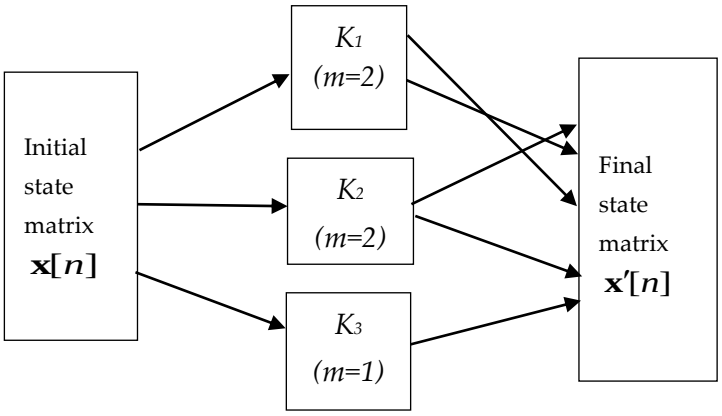
The major advantage of this innovative architecture is that even smaller networks, i.e., in the order of 5-8 cascaded-leaky-integrator-neuron structures, are capable of outperforming their counterparts that used 1000 sigmoidal neurons That said, however, one of the disadvantages of this approach is that, by increasing architectural complexity, the possibility exists of over-fitting the network. This is particularly the case if global network parameters and reservoir size are not chosen carefully. The following section includes a detailed discussion on the optimization of the global control parameters in a cascaded-leaky-integrator-ESN.

Given the nature of ESN architecture, since the weight matrices are randomly generated and fixed for the entire training cycle, this model exhibits a small degree of randomness. However, the results obtained from this approach were found to be equivocal for the given input, especially for cases in which the data were contaminated with noisy epochs.

As discussed in §4.8.4, an alternative solution was proposed to overcome the disadvantages associated with cascaded-leaky-ESNs. This included joining of several individual networks from the aforementioned design into a combined classifier model. Accordingly, the class hypotheses from each of the classifiers were combined and the mean of the individual votes was calculated for each classifier. Calculating the mean of the vote combination was performed because it averages out the vote fluctuations due to a single classifier's biases (Duin, 2002).

For the cascaded-leaky-integrator-ESN-based microsleep detection model discussed in this thesis, each randomly created leaky-integrator ESN module, comprising p neurons (where p is the number of neurons, and, in this case, 5) and K sequences ($K = 3$), was treated as an individual classifier. It was also discovered from the initial observations that, with only up to 5 neurons, the modified cascaded-leaky-integrator structures had markedly different dynamic properties across their random instantiations, which was not the case for larger

ESNs where inter-network differences became insignificant with their growing network size (Jaeger *et al.*, 2007).



Segregated sequences of the state vector.

Figure 5-2 A schematic representation of the state vector construction. Here, K_1, K_2 and K_3 represent the 3 segregated sequences of the state vector.

Therefore, each reservoir in the aforementioned cascaded-leaky-integrator ESN structure can be considered an individual detector for random, p-dimensional, nonlinear, dynamic features of the input signals. Consequently, joining all K extended state vectors creates a series of p-neuron ESNs which can transform an input sequence into a static (p + **M**) × **K**-dimensional feature. Here, **M** is the output vector and is the contribution of all reservoir state components that vary across each of these network instantiations.

The combined classifier, i.e., a cascaded-leaky-integrator ESN-based approach is presented in Figure 5-3. As the figure depicts, outputs of all individual classifiers are combined to form a single detection model.

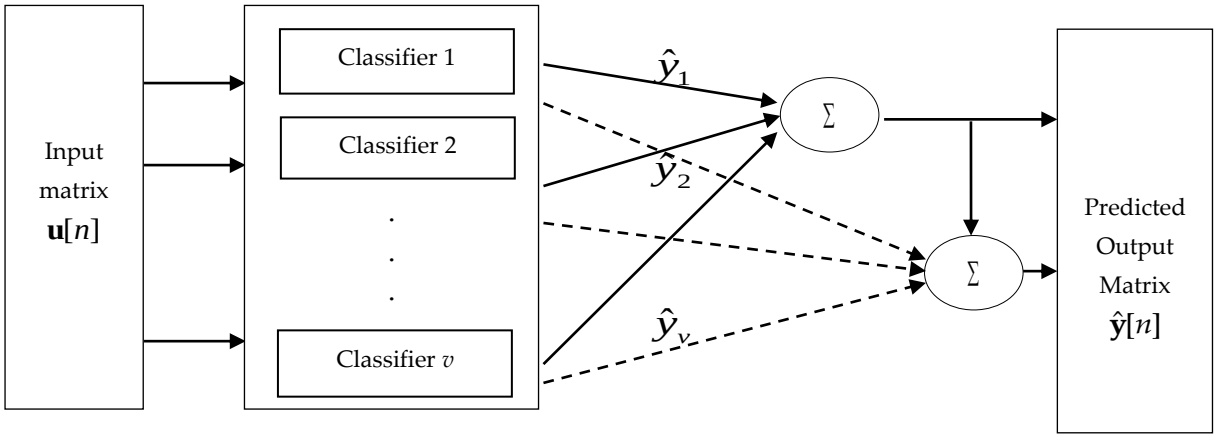


Figure 5-3 A schematic representation of combined classifier model of the cascaded-leaky-integrator ESN. Here, $\hat{y}_1$, $\hat{y}_2$ and $\hat{y}_3$ represent the outputs of each of the n classifiers.

5.2.2 Optimizing the global control parameters

The cascaded-leaky-integrator-ESN formulation depicted in the previous section uses 4 parameters, 3 of which include scaling parameters to the randomly-generated connection weights in the reservoir. Leakage factor or leaking rate is the fourth and the most crucial optimization parameter for this model. Apart from the scaling parameters and the leakage rate, reservoir size can also be considered as a fifth parameter that must be determined, which is, however, not optimized (Lukoševičius, 2012).

Scaling parameters

The three scaling parameters that need to be optimized are bias scaling, input scaling and spectral radius. Bias scaling is used to push the reservoir states closer to -1 or +1, which corresponds to the non-linear region of the hyperbolic tangent function. Therefore, for the higher values of the bias scaling, the reservoir tends to move towards non-linearity (Jaeger, 2002). Usually, scaling values set for the bias are either 0, 0.1 and 1.

Typically, input scaling is used to set the influence of the input to the reservoir states (Jaeger, 2002). The larger the input scaling, the bigger the effect of the input on the reservoir states (Jaeger, 2002). However, with a smaller input scaling the current reservoir dynamics are less likely to be disturbed by a new input. Input scaling also has an influence on the non-linearity property because of the hyperbolic tangent function, and, usually for normalized inputs, the input scaling values range from 0.001 to 0.1 and can be optimized in steps of $10^{0.2}$ on a log scale (Jaeger *et al.*, 2007).

Finally, the spectral radius is defined as a scaling parameter for the connection weights between the neurons, and is set to have a significant influence on the non-linear dynamics of a reservoir network (Jaeger, 2002). The spectral radius signifies the largest absolute eigenvalue of the connection weight matrix $\mathbf{W}_{DR}$.

A linear reservoir is said to have the *echo state property*[8] if the spectral radius is smaller than 1 (Jaeger, 2002). Linear reservoirs with a spectral radius greater than or equal to 1 can be unstable, since the reservoir states do not fade out out over time (Lukoševičius, 2012). However, in the case of a non-linear reservoir this rule is not always true. This is because the hyperbolic tangent function usually limits the reservoir states once they reach the non-linear region, i.e., in extreme cases, the reservoir states will saturate to -1 and 1. Reservoirs with a

[8] A network is said to have an echo state property when the input fades away like an 'echo'.

very small spectral radius have a shorter memory (Lukoševičius, 2007). The values for optimizing the spectral radius usually range from 0.4 to 1.0 and are optimized in steps of 0.1 (Jaeger, 2007).

Leakage rate/ leak factor

The leak factor represents the rate at which the previous reservoir state is replaced by the current reservoir state. It is typically implemented as an extra recurrent connection in the reservoir that is essentially a low-pass filter on the reservoir states (Verstraeten *et al.*, 2007). A leaking rate of 1 represents a reservoir with the low-pass filter as it realistically smoothens the reservoir states over time. Similar to the bias parameters, the leakage rate can have a substantial effect on the fading of the reservoir dynamics and can push the reservoirs with a spectral radius greater than 1 into more stable conditions. The influence of the leak rate is highly frequency dependent, as an input signal with a frequency component significantly below the cut-off frequency will not be dampened in a linear reservoir (Verstraeten *et al.*, 2007). Leakage rate is usually optimized in log steps of $10^{0.25}$ (Jaeger *et al.*, 2007). The leakage factor is highly reliant on the temporal scale of the learning task and, therefore, an ideal range within which the leak rate should be optimized cannot be determined.

Reservoir size

Reservoir size is related to the network memory and network model complexity. It is identified as one of the most important factors which can influence the overall performance of the system (Verstraeten *et al.*, 2007). Since it is directly related to the system memory, it is usually assumed that with a larger reservoir, the performance increases (if the system is properly regularized). However, this performance gain will decrease significantly with excessively large reservoir sizes. Additionally, when the network is not accurately regularized, there is a possible chance of over-fitting. Moreover, increasing reservoir size can exponentially increase the computational cost (processing time) of the network. These disadvantages can be overcome with the use of the leaky-integrator neuron ESN structures as they can scale the reservoir size quadratically. Therefore, a reservoir size is usually selected at a point where increasing the size has little to no effect on the overall performance of the network. The memory capacity of a network is strongly associated with the non-linearity of the system (Verstraeten *et al.*, 2007). The more non-linear the reservoir, the shorter the memory. Hence, the classification tasks that require a long memory usually require a very large reservoir or a linear reservoir.

5.2.3 Global control parameters for the optimal microsleep detection

The two main properties that contribute to a reservoir's processing power are its quasi-exponential decaying memory of its past inputs, and its ability to perform a non-linear transformation on the inputs (Buteneers, *et al.*, 2009). Each non-zero input sample to a reservoir excites the dynamical system and pushes the system to a new state (Lukoševičius, 2012). Moreover, it maps this input data to a higher dimensional space and in this higher dimensional space, the probability that the microsleep and non-microsleep samples are linearly/ non-linearly separable gradually increases.

When trained with the Study A dataset, a reservoir comprising 5-10 neurons was used. This value was intentionally chosen after rigorous fine-tuning of the network, which produced the most consistent and optimal performance during cross-validation using this training set. As discussed in §5.2.1, there are 4 other parameters that typically determine the dynamic properties of the reservoir. Table 7-8 (in Chapter 7) depicts the optimal cascaded-leaky-integrator-ESN global control parameters that were observed to achieve an optimal EEG-based detection on the BCI competition dataset.

High values for parameters, such as bias scaling and spectral radius, imply pushing the reservoir into a highly non-linear space (Jaeger, 2002). This phenomenon is also evident from the values depicted in the Table 7-8. Additionally, the global control parameters are extremely sensitive to any variations and the effect of varying these parameters is also presented in Table 7-8.

5.3 System validation with BCI Competition IV Dataset IIa

Before the optimal EEG-based feature sets for microsleep detection were established, the classifier models were tested on an independent dataset. Experimental data from this independent study was evaluated on the leaky-integrator ESN model as presented in §5.2.1. The intension of this experimental simulation was to validate the classifier modules and show how the methods applied on other independent datasets performed, comparable to microsleep data from Study A. The independent dataset evaluated included the BCI competition IV dataset 2a (Tangermann *et al.*, 2012).

5.3.1 Dataset description

BCI Competition IV[9] is an open competition aimed towards evaluating various signal processing and classification approaches used in brain-computer interfaces (BCI) (Brunner *et al.,* 2008). In addition, BCI competition IV also aims for a reliable/ efficient measurement of performance for each classification algorithm by comparing them on a standard data set. The BCI competition dataset IIa consisted of EEG recorded from twenty-two Ag/AgCl monopolar electrodes (inter-electrode distances 3.5 cm). Left mastoid was chosen as the reference and the right mastoid as ground (Brunner *et al.,* 2008). EEG was sampled at 250 Hz and bandpass-filtered at 0.5 and 100 Hz. An additional 50 Hz notch filter was performed to suppress the underlying line noise.

5.3.2 Experimental Paradigm

A cue-based experimental paradigm was performed on four different motor imagery tasks, namely the imagination of movement of the:

1. left hand (class 1),
2. right hand (class 2),
3. both feet (class 3),
4. and tongue (class 4).

Accordingly, the EEG data from 9 subjects was recorded in two independent sessions conducted on separate days. Both the sessions consisted of 6 runs separated by short breaks; and each run consisted of another 48 trails (12 trials for each of the motor imagery classes), yielding in a total of 288 trails per sessions (Brunner *et al.,* 2008). In each trial, the subjects were instructed to perform one of the four motor imagery tasks based on the cues provided on their screens. Overall, a single session from the motor imagery experiment consisted of 288 trials (72 trials for each motor imagery classes).

A fixation cross was shown on their computer screens at the same time a warning tone was played. Two seconds, following the fixation cross, a cue was shown instructing the subject to respond to the corresponding motor imagery task. Following this, after another 1.25 s the stimulus changed back to the fixation cross and the motor imagery task was carried on until 6 s, after which a break was taken. A time-course of each trial from the experimental paradigm is depicted in the Figure 5-4.

[9] http://www.bbci.de/competition/iv/

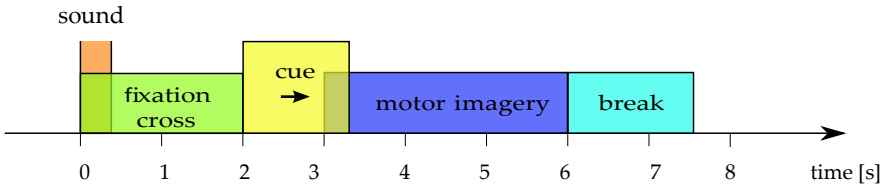


Figure 5-4 The structure of each EEG trial of the dataset IIa from the BCI Competition IV (Pszachewicz, 2009).

5.3.3 Pre-processing

Similar to the Study A, several pre-processing methods were implemented on the BCI competition datasets. A 50 Hz notch filter was implemented to remove the mains activity. Following this, ICA was performed on all the sessions to eliminate the ocular artefacts. A window size of 512 samples (2 s) and an overlap of 50% between successive windows were used to calculate the EEG band power via a 4^{th}-order Butterworth IIR filter model. EEG band powers from the frequency bands, 8–12 Hz and 14–18 Hz were calculated for all the 22 channels and the resulting 44 band power features were extracted. PCA was used to reduce the redundant features passed to the classifiers.

5.3.4 Validation of the classifier systems on the BCI competition dataset IIa

This study on the BCI competition datasets allowed for performance of the classifier modules, specifically developed for microsleep detection, be compared to another cue-based motor imagery experiment. The intuition behind this comparison was to show the validation of the classifier models on other EEG related experiments. Another purpose of this experimentation was largely to ensure classifier arrangements functioned properly. Consequently, all variants of the classifier modules were subjected to the same battery of tests. Classifier modules evaluated included LDA, SVM with a linear kernel (SVML), SVM with a polynomial kernel (SVMP), and cascaded-leaky-integrator ESN. The following sub-section discusses the results obtained on the BCI competition dataset IIa using different classifier modules.

5.3.5 Global control parameters for the detection on the BCI competition dataset

Table 5-1 depicts the optimal cascaded-leaky-integrator ESN global control parameters that were observed to achieve an optimal detection on the BCI competition dataset. Similar to the optimized control parameters for EEG-based microsleep detection (Table 7-10), the higher values for the control parameters such as, the bias scaling and spectral radius were also reported on the BCI competition datasets.

Table 5-1 Optimal cascaded-leaky-integrator ESN global control parameters for microsleep detection.

Parameter	Value	Parameter	Value
Spectral radius	0.92	Leakage rate	0.08
Input scaling	0.25	No. of neurons	6-12
Bias Scaling	0.30		

5.3.6 Results

The single-trial classification performances of the classifier modules of multi-class extensions were investigated on all the 9 subjects. The performance was evaluated in terms of the mean kappa value using 10 × 10-fold cross-validations. Similar to Pearson's phi, Cohen's Kappa (κ) is an index that measures inter-rater agreement for categorical items, and its values range between -1 and +1. A value of κ corresponding to '1' indicates perfect agreement in successfully identifying all events and non-events. The same system architectures for combined classifier modules (ensemble structures) were used similar to the simulated EEG events datasets. The results of the cross validation across all the subjects of the Dataset 2a are shown in Table 5-2.

Table 5-2 The κ values for each subject obtained using the (10 × 10) cross validation scheme on both the linear and non-linear methods for the dataset IIa. SVM polynomial provides the benchmark κ values followed by the cascaded-leaky-integrator ESN.

Method	Mean Kappa	Subjects								
		1	2	3	4	5	6	7	8	9
Linear										
LDA	0.34	0.38	0.22	0.48	0.22	0.13	0.18	0.39	0.47	0.58
SVML	0.29	0.30	0.20	0.42	0.20	0.10	0.14	0.36	0.41	0.52
Non-linear										
SVMP	0.52	0.64	0.38	0.48	0.45	0.35	0.24	0.71	0.72	0.68
Cascaded-leaky ESN	0.49	0.57	0.30	0.70	0.42	0.22	0.26	0.78	0.57	0.54

Results in the Table 5-2 reflect the inputs from 20 meta-features of the PCA. Interestingly however, it was found that increasing the number of meta-features did not have any significant impact on any of the classifier modules. The results showed that the SVM with a polynomial kernel (SVMP) yielded the best averaged mean kappa value (0.52). Whereas, the SVML-based approach yielded relatively the worst performance (κ = 0.29) in comparison to the other approaches. The LDA-based approach yielded slightly higher mean kappa value (0.34) compared to the SVML. The cascaded-leaky-ESN based approach was the second best

in terms of mean kappa (0.49) and was seen the highest in 3 subjects (3, 6, and 7). The κ values for SVMP were seen the highest in 6 subjects (1, 2, 4, 5, 8 and 9). A paired t-test was conducted on the cascaded-leaky ESN and SVMP approaches which revealed no significant difference (p = 0.88) between both the approaches.

While not presented, single classifier modules were also evaluated on the dataset IIa and were found to be weaker than the ensemble modules. However, differences in performance were not as high as in the linear classifier models, while the differences in the non-linear classifier models was substantially high.

5.3.7 Discussions and summary

From analysing the results in Table 5-2, it can be inferred that the non-linear kernel-based classifiers performed better than the linear classifiers on all subjects. The results obtained from applying all the linear and non-linear classifier methods at each level and each segment were compared to those achieved by the other authors as shown in Table 5-3. Almost all of the methods depicted in Table 5-3 used the common spatial pattern (CSP) based approach for feature reduction. The algorithm presented by Ang *et al. (*2012) used a filter-bank CSP (FBCSP) to extract features from multiple frequency bands. FBCSP finds the most discriminative features using mutual information from the underlying class labels. The FBCSP algorithm is considered the benchmark algorithm for the BCI competition IV dataset IIa and was awarded the winner of the competition. Other variants of CSP-based algorithms have also been used by Pszachewicz (2013), Song *et al. (*2012), and Alonso *et al. (*2013). In addition, it was reported that the application of CSP-based algorithms resulted in an increase in classification accuracy of 10% (Pszachewicz, 2013).

Table 5-3 Comparison of the κ values for each subject as reported by other groups.

Method	Mean Kappa	Subjects								
		1	2	3	4	5	6	7	8	9
Ang *et al. (*2012)	0.57	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61
Pszachewicz, (2013)	0.45	0.74	0.11	0.53	0.34	0.18	0.31	0.62	0.61	0.63
Song *et al. (*2012)	0.31	0.38	0.18	0.48	0.33	0.07	0.14	0.29	0.49	0.44
Kam *et al. (*2013)	0.60	0.74	0.35	0.76	0.53	0.38	0.31	0.84	0.74	0.74
Alonso *et al. (2013)*	0.63	0.77	0.39	0.85	0.60	0.43	0.31	0.79	0.77	0.72
SVMP	0.52	0.64	0.38	0.48	0.45	0.35	0.24	0.71	0.72	0.68
Cascaded-leaky-ESN	0.49	0.57	0.30	0.70	0.42	0.22	0.26	0.78	0.57	0.54

The results presented in the Tables 5-2 and 5-3 demonstrate the validity of cascaded-leaky-ESN-based structures with optimized global control parameters, specifically developed for

microsleep detection model on another independent dataset which has been extensively used in the BCI literature.

Although, the highest κ values obtained using the SVMP and cascaded-leaky-ESN modules were marginally less than the previous approaches, it has to be noted that input for both the classifiers consisted of meta-features from simple algorithms such as the PCA. Additionally, the differences in the κ values of the SVMP and cascaded-leaky-ESN classifiers do not show an appreciable difference in terms of the performance. Therefore, it can be hypothesized that evaluating advanced pre-processing methods and supervised feature reduction methods, which are more adjusted to the characteristics of the specific task, could result in an increase of performance for both network types.

5.4 Summary

This chapter provided a review of the development of the novel cascaded-leaky-integrator ESN approach developed for microsleep detection on the Study A dataset. A major advantage of this innovative architecture was also presented. This novel approach also dramatically reduced the training time required to train these recurrent neural networks while still attaining competing performance. Various global control parameters for the cascaded-leaky-ESNs, and their optimization techniques required to attain the optimal performances were also presented in §5.2.3. Finally, in order to demonstrate the validity of this novel cascaded-leaky-ESN-approach, this Chapter also presented the performance analysis of the novel cascaded-leaky-integrator ESNs on an independent BCI competition dataset.

As will be shown in the following Chapters using this novel cascaded-leaky-integrator architecture, even smaller networks, i.e., in the order of 8 cascaded-leaky integrator neuron structures, were capable of achieving better performance compared to non-leaky sigmoidal architectures that support 1000-2000 neurons.

CHAPTER 6

Microsleep detection Methods: EEG Preprocessing and Feature Reduction

6.1 Pre-processing methods assessment on Study A datasets

The proposed microsleep detection system, as depicted in Figure 6-1, involves pre-processing/conditioning, feature extraction, feature selection/reduction, and pattern classification stages. Signals from eye movements, blinks, cardiac signals, muscle noise, and, line noise can be in orders of magnitude larger than brain-generated electrical potentials and are one of the main sources of artefacts in EEG data. Therefore, in order to overcome this problem, a comprehensive set of pre-processing methods were implemented on raw EEG data. Pre-processing on the Study A dataset comprised various stages, encompassing EEG data acquisition, artefact removal, mean removal, rescaling, data filtering, and feature matrix generation.

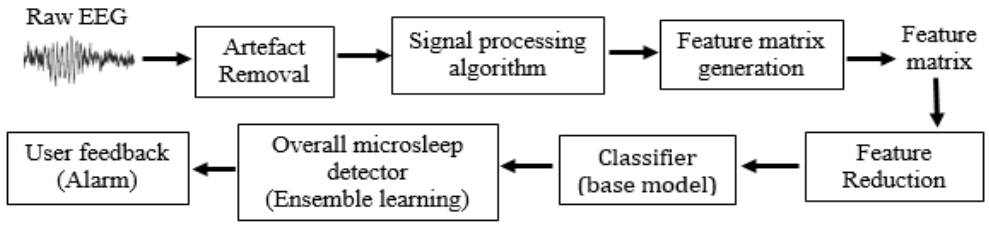


Figure 6-1 Proposed prototype microsleep detection system.

The feature extraction stage converts the processed EEG data into a set of EEG features which act as input for the classifiers. Given the high dimensionality of EEG and EEG-feature datasets, feature reduction was required so as to pass only significant features to, and not overload, the pattern classification stage. EEG feature reduction algorithms generate meta-

features from original features, using, for example, PCA, so as to minimize and optimize the number of features passed to the classifier without loss of significant information from the feature sets. A distinction is that feature selection reduces the existing data by selecting, but not altering, an optimal subset of features. Consequently, the pattern classification stage assigns class labels to a given input value based on the training algorithm.

The following sections of this chapters outline several pre-processing methods, feature selection and feature reduction methods evaluated in this research.

6.1.1 EEG Analysis

An infinite impulse response (IIR) notch filter (Jangsri, 1988) with a Q-factor of 35 was applied to the EEG data bandpass filtered between 1 and 70 Hz. A notch filter was required to remove the 50 Hz interference (mains activity) from the underlying EEG. Each EEG channel was processed by rejecting epochs contaminated with artefacts. As previous studies suggest, the removal of artefacts from the EEG improves detector performance (Peiris et al., 2011, LaRocco, 2015). Electrode noise and EEG artefacts reduce classifier effectiveness.

Eye-blink artefacts were removed manually (epoch by epoch) with the help of ICA (Delorme and Makeig, 2004). Bipolar derivations used to calculate power spectra were: Fp1–F3, Fp1–F7, Fp2–F4, Fp2–F8, F3–C3, F4–C4, F7–T3, F8–T4, T3–T5, C3–P3, P3–O1, T5-O1, C4–P4, T4–T6, P4–O2, and T6–O2. Bipolar derivations were preferred over the referential due to common mode noise rejection. A window size of 512 samples (2 s) and an overlap of 50% between successive windows were used to calculate power spectra via a 40^{th}-order autoregressive Burg model (Peiris *et al.,* 2011). Each 2-s epoch from each channel had its DC component removed, and its power spectrum estimated. Three types of power spectral features were produced:

1. Spectral power – the mean power across each of the standard frequency bands described in Table 6-1 (Peiris et al., 2011).

2. Normalized power – spectral power features divided by the mean power across all bands.

3. Power ratios – the ratio of power in two different frequency bands.

Using the power of a signal is common in EEG analysis and several researchers have used EEG spectral power (Zhang and Parhi, 2015a; Zhang and Parhi, 2015b and Zhang and Parhi, 2016) to detect changes in the level of alertness and arousal (Huang *et al.,* 2001; Jung *et al.,* 1997; Makeig and Inlow, 1993; Makeig and Jung, 1995; Makeig and Jung, 1996). The power spectrum of EEG provides an estimate of power at each frequency of the signal. This

approach can be used to estimate the composition and power of different frequency bands within the EEG. The power spectral density of the signal can be obtained by estimating the periodogram, based on the finite Fourier transform for each segment in the EEG. Peiris *et al.* (2011) used an autoregressive spectral estimation with the Burg algorithm to calculate the Least Mean Squares (LMS).

The 50% overlap window (1 s) allowed for a temporal resolution of 1 s for the spectral power in the EEG bands ranging from delta (1.0–4.5 Hz), theta (4.5–8.0 Hz), alpha (8.0–2.5 Hz), beta (12.5–25.0 Hz), gamma (25.0–45.0 Hz), and higher (45.0–100 Hz) bands. Due to the sampling rate of the tracking task ~63.99 Hz, class labels were decimated to match the number of EEG epochs generated. This was done by splitting up the gold standard across each of the 1-s epochs. Following this, if more than 50% of the epoch was labelled 'responsive' then that epoch was labelled as a non-microsleep. Similarly, if more than 50% was labelled non-responsive, the 1-s epoch was labelled a microsleep.

Table 6-1 EEG spectral features for microsleep detection calculated from each EEG derivation as depicted in (Malik, 2010).

Feature	Frequency band
Mean spectral power	
Delta (δ)	1.0 – 4.5 Hz
Theta (θ)	4.5 – 8.0 Hz
Alpha 1 (α_1)	8.0 – 10.5 Hz
Alpha 2 (α_2)	10.5 – 12.5 Hz
Alpha (α)	8.0 – 12.5 Hz
Beta 1 (β_1)	12.5 – 15.0 Hz
Beta 2 (β_2)	15.0 – 25.0 Hz
Beta (β)	12.5 – 25.0 Hz
Gamma 1 (γ_1)	25.0 – 35.0 Hz
Gamma 2 (γ_2)	35.0 – 45.0 Hz
Gamma (γ)	25.0 – 45.0 Hz
High	> 45.0 Hz
All frequencies	Included only in SP
Spectral power ratios	
θ/β , θ/α, α/β, δ/θ, α/δ, β/δ, β_1/α, β_2/α, β_1/β_2	–

In order to eliminate the electrode-pop artefacts, each derivation was normalized into z-scores. The z-scores of each epoch were computed using the mean and standard deviation of the first 2-min of each 1-hour-long record of data of the same session. To exclude high noisy data, any epochs with absolute z-scores over 30 (z-score > 30) were deleted and removed from further analysis. This process resulted in deletion of over 580 epochs across both the sessions per subject on average. The corresponding gold standard events were also excluded

from analysis at an average of 26 flat spot epochs and 34 video-lapse epochs per subject in both the sessions. Table 6-2 demonstrates this effect as the number of EEG epochs and their corresponding gold standard events are depicted pre and post processing.

Each of the four behavioural metrics (flat spots, video microsleeps, definite microsleeps, and lapses) were decimated to a resolution of 1 Hz. The presence of a microsleep was treated as a binary state, where "1" indicated the presence of a microsleep and "0" indicated the responsive or baseline state. Data rated by human experts served as the gold standard for gauging performance of an automated classifier. The thirty four spectral features per derivation (13 spectral power + 12 normalized power + 9 power ratios) were calculated for each of the 16 channels using a 2-s sliding window function, stepping at 1-s intervals, resulting in 34 × 16 = 544 spectral features over the 16 channels.

Table 6-2 Depicting the number of EEG epochs and their corresponding goldstandard events for each subject (definite microsleeps) pre-and post-processing.

Subject	804	809	810	811	814	817	819	820
Epochs								
Pre-processing	7198	7197	7197	7198	7198	7198	7198	7197
Post processing	6092	6857	6669	6960	6960	6984	6888	6669
Epochs (Definite microsleeps)								
Pre-processing	937	166	8	54	154	180	103	663
Post processing	881	166	8	54	153	179	103	650
Epochs (Lapses of responsiveness)								
Pre-processing	1897	449	64	338	262	620	468	1263
Post processing	1664	441	61	266	257	615	465	1216

Peiris *et al. (*2011) used principle component analysis (PCA) to reduce the total number of features. PCA (Pearson, 1901) has often been used to reduce the redundancy within the original features and transform feature vectors into orthogonal components to aid in the formation of classification models for over 100 years. With PCA it is possible to reduce the dimensionality of the data without any significant loss in the information. The meta-features generated from PCA are used as input to the classifiers.

The overall algorithm of the pre-processing phase used by Peiris *et al. (*2011) is described below:

1. Raw EEG (referential montage) is recorded from 16 scalp electrodes.

2. Referential EEG is converted to bipolar montage (Bipolar derivations were preferred over the referential derivations because the referential derivations could more effectively reject the common mode noise).
3. Perform 50 Hz, IIR notch filter (for the removal of mains activity).
4. Perform ICA source separation for eye blink removal.
5. Calculate the mean and standard deviation of the first 2-min of each 1-hour-long recording of data of each session.
6. Perform z-score transform to normalize each EEG derivation (to remove the electrode-pop artefacts).
7. Reject EEG epochs with absolute z-scores > 30 (data cleaning).
8. Generate EEG epochs using window size of 512 samples (2 s) and an overlap of 50% between successive windows (1.0 s).
9. Calculate power spectra via a 40^{th}-order autoregressive Burg model.
10. Generate power spectral features (34 spectral features × 16 channels = 544 features).
11. Reduce the redundant features using feature reduction algorithms.

Following these steps, the heavily processed data is presented as an input to the classifier schemes (elaborated in the next chapter) for microsleep detection. Evidence also suggests that highly processed EEG-data used to eliminate label noise may lead to higher classification performance (Chan *et al.,* 1998; Sörnmo and Laguna, 2005; Astolfi *et al.,* 2006). Subsequently, the following sections of this chapter focus on improving the quality of data presented to the classifier system.

6.1.2 Data pruning

The process of rejecting noise induced epochs and setting them aside from classification is referred to as data cleaning or data pruning. Sometimes the analysis on the quality of data from a data source is performed through manual inspection and approving/rejecting noisy labels, thereby making changes to the data. EEG is highly susceptible to various forms and sources of noise, which present significant difficulties and challenges in analysis and interpretation of EEG data. There have been a number of techniques used to deal with noise effectively, both at the time of EEG recording as well as during preprocessing of recorded EEG data. Electrode noise and EEG artifacts (ocular and motion) could be counterproductive during the performance evaluation and classifier model formation, and could well in fact reduce the overall effectiveness of the classifier. This effect is depicted in Figure 6-2, where an EEG window consisting of 10 epochs (1570 - 1580) is illustrated before and after applying

the filtering algorithms. The upper portion of the figure demonstrates the raw data from one of the subjects of Study A, contaminated with artefacts and the bottom portion signifies the signal after applying a band-pass filter.

The goal of this section is to emphasize the importance of noise removal and review methods for prevention and removal of noise in EEG recording, including elimination of noise sources, signal averaging, data rejection and noise removal, along with their key advantages and challenges. Whenever noise in recorded, data is sparse and easily recognizable, the most obvious way of dealing with this problem is to eliminate the parts of the data where the noise is present. The most straightforward procedure for rejection of noisy data is by visual inspection (Repovš, 2010). Most eye-movements, blinks and movement artefacts are relatively easily recognizable and can be marked for rejection before averaging and data analysis.

The microsleep detector performance reported by Peiris *et al.* (2011) used an artefact removal scheme, where all the epochs with z-scores >30 from the normalized EEG were excluded from analysis. This could have more than likely biased their system to enhance performance. To verify this hypothesis, similar methods used in Peiris *et al.* (2011) were implemented again on the RC and SVMP-based classifiers. The results from this analysis are presented in Chapter 7.

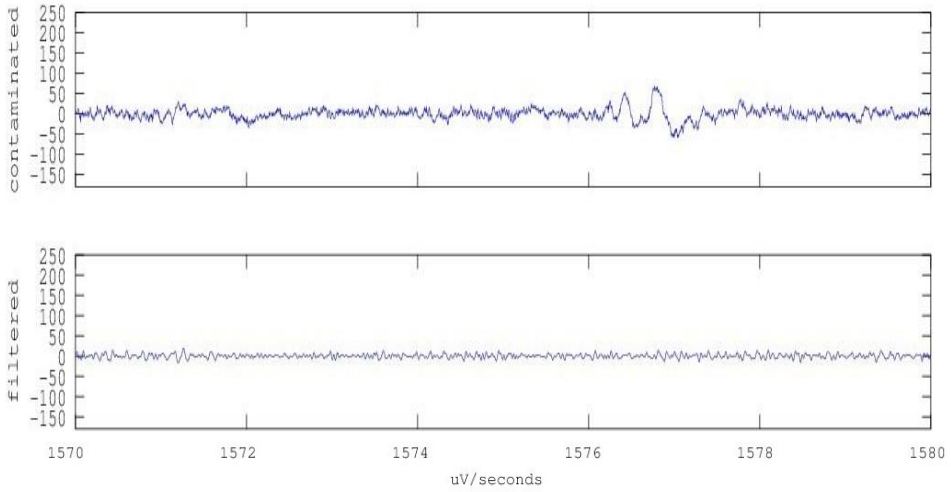


Figure 6-2 Depicting the removal of ocular artefacts using a 7–30 Hz band-pass filter. Data from session 1 subject 804 is plotted across the noisy epochs from 1570–1580. The top half of the figure portrays the channel Fz contaminated with artefacts. Whereas, the bottom half of the figure portrays the same channel, but after band-pass filtering.

A visual inspection into the data pruning effect can provide an insight into significant challenges faced by a detection system. Figures 6-3 and 6-4 depicts the scatter pattern of the

spectral features before and after pruning feature data from Subject 804 of Study A. The scatter pattern of spectral features pre-pruning and post-pruning are compared to demonstrate the impact of noisy epochs on the EEG data. Three power spectral features from α frequency band (8.0 – 12.5 Hz) are plotted on the three axes. Features containing events (microsleep states) are identified with red circles and the features without an event (alert states) are identified with blue circles (non-events).

Despite taking several precautionary steps to clean the noisy epochs before assigning them to corresponding class labels, a few artefacts will remain which cannot be cleaned during any of the processes of standard filtering, mean removal, eye-blink removal and other artefact-removal phases. This phenomenon is clearly evident in the scatter plot in Figure 6-3. The dataset subject to this discussion has gone through a series of signal processing stages but a significant amount of noise, of relatively higher power than the source signal, is evident. In some cases, this effect may cause the classifier to overfit the overall detection model. To overcome this effect, EEG was manually inspected to find anomalies that may bias the system, and any epochs that were superimposed with high amounts of noise (generally contaminated with artefacts such as electrode pop) were discarded from the analysis (after a z-score transformation, with any z-scores > 30 as discussed earlier.

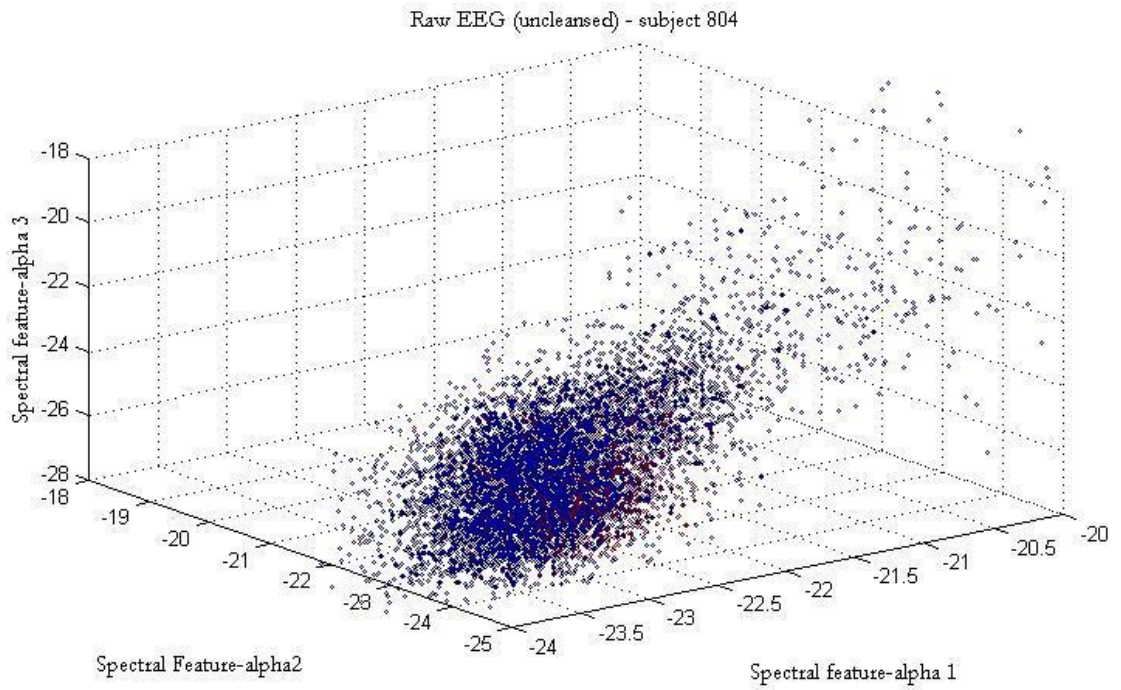


Figure 6-3 3D scatter plot depicting the class distributions on the unpruned EEG data from the subject 804. The three axes represents the power spectral features of the frequency bands α (8.0 – 12 Hz). The blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

The scatter plot in the Figure 6-4, demonstrates the new pruned data with noisy artefacts removed. As will be demonstrated in §7.5.1 and §7.8.3 of this work, the pruning effect has a

significant impact on certain types of classifier performance. Noise can present a significant challenge in EEG analysis and data interpretation, necessitating efficient strategies for noise prevention and removal, which might enhance the overall detection accuracy of the system. Additionally, a number of methods and algorithms can be employed to reject noisy data, reduce noisy signals, and thereby improve signal-to-noise ratio of the data.

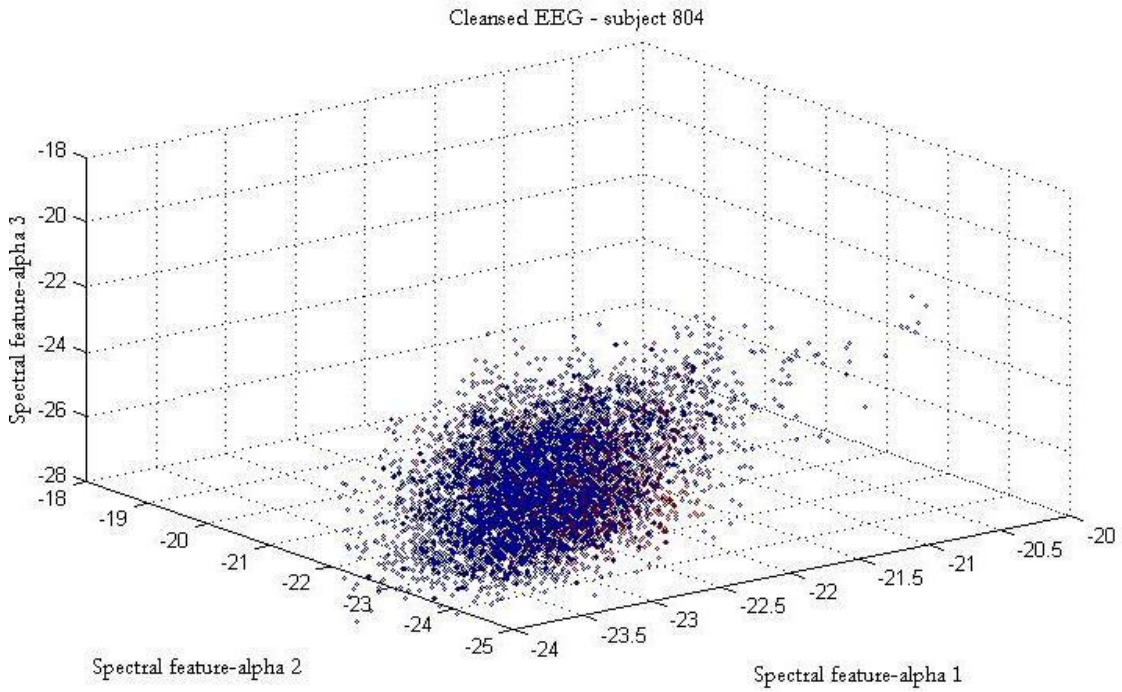


Figure 6-4 3D scatter plot depicting the class distributions on processed, pruned EEG data from the subject 804. The three axes represents the power spectral features of the frequency bands α (8.0 – 12 Hz). The blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.2 Feature reduction

Features are the functions of raw data which represent data instances to a classifier. The functions are intended to give the classifier as much useful information for distinguishing between the classes, while avoiding irrelevant information being presented. An EEG *feature* is defined as an arbitrary time series extracted from a single EEG derivation using a given signal processing algorithm (Peiris, 2008). A feature vector is a vector of all feature values for a particular data instance.

Given the high-dimensionality of the EEG datasets, forming a classifier model becomes complicated, especially for some types of classifiers (Chen, 2013; Hutapea, 2014). The exploration of new features could produce very large feature vectors which are impractical for learning, as the space of possible classifier models will become too large to search.

In fields such as image processing, speech processing, and biomedical engineering, the measured data vectors are usually high-dimensional (Jimenez and Landgrebe, 1997). Hence, there is a need to either reduce these large number of features to a smaller number, thereby, simplifying the complexity of the problem, or by passing only the significant features to form a classifier model. Such problems can be addressed by using algorithms based on feature selection or feature extraction. The difference between both methods, i.e., feature selection and feature extraction, is that feature extraction requires the generation of features from an already existing dataset, whereas feature selection means reducing the existing dataset by carefully choosing the optimized subset of features from the existing dataset.

Feature selection methods are typically partitioned into filter methods and "wrapper" methods. Typically filter methods are used as a pre-processing step whereas, the wrapper methods are usually embedded into a learning algorithm (Kohavi and George, 1997; Guyon and André, 2003). Filter methods rank features using multiple metrics which indicates how useful they are individually, such as a correlation with class labels, or average distance between events and non-events (ADEN) (LaRocco, 2015). Feature selection methods can also be used to efficiently discard large numbers of irrelevant features. The drawback of filter methods is that, in some instances, they cannot predict how effective a combination of features will be and, therefore, cannot be used to find the best feature sets by themselves (Domingos, 2012). Wrapper methods on the other hand, involve selecting a subset of features and training a classifier on that subset, allowing the evaluation of combinations of features. This is, of course, far more computationally expensive than filtering but is necessary to evaluate the performance of groups of features.

Feature reduction or dimensionality reduction can be defined as the transformation of high-dimensional data into a meaningful representation of the reduced dimensions by extracting the essential information from the dataset (Freund and Schapire, 1997). Ideally, the reduced representations should contain minimum number of parameters needed to account for the observed properties of the data [Fukunaga, 1990]. This property is called the intrinsic dimensionality of the data. Feature reduction is significant in many disciplines, since it mitigates the curse of dimensionality of the high-dimensional datasets [Jimenez and Landgrebe, 1997]. In most of the cases, high-dimensional datasets (like EEG classification) usually present many mathematical challenges as all the measured variables are not important for understanding the underlying phenomena of interest. While certain computationally expensive novel methods can construct predictive models with high accuracy from high-dimensional data, it is still of interest in many applications to reduce the dimension of the original data prior to any modelling of the data.

Feature reduction schemes can be extremely useful in tasks such as classification, visualization, and compression of high-dimensional data. Traditionally, feature reduction has been performed using standard linear techniques such as the PCA [Pearson, 1901; Hotelling, 1933] and factor analysis [Spearman, 1904]. However, it has been widely proven that these linear techniques cannot adequately handle complex nonlinear data (Burges, 2005; Venna, 2007). Consequently, in the past few decades there has been a significant increase in the number of nonlinear techniques for feature reduction that have been proposed (Saul *et al*., 2006; Venna, 2007).

Most of the real-world tasks involve datasets which are nonlinear in nature. Therefore, using a nonlinear feature reduction techniques over the linear ones may offer a distinct advantage. Additionally, numerous studies have also shown that nonlinear feature reduction techniques outperform their linear counterparts (Roweis and Saul, 2000; Tenenbaum *et al.*, 2000). Motivated by the success of the nonlinear feature reduction methods, this section presents a comprehensive comparative study of the most important linear dimensionality reduction technique (PCA), to other advanced nonlinear dimensionality reduction techniques.

Notwithstanding, the primary aims of this section are to:

- Investigate the extent to which the nonlinear-feature reduction techniques outperform the traditional linear techniques, such as the PCA on the current microsleep detection problem, and,
- determine the computational complexity of non-linear feature reduction methods relative to PCA.

Overall, eleven non-linear feature reduction methods have been evaluated as an alternative to PCA, and a 3D scatter plot analysis and other techniques, such as the calculation of trustworthiness, were performed on the reduced feature sets. Based on visual inspection of the class separation from the scatterplots, a decision was taken on whether or not to implement the algorithm in the final microsleep detection system. Performance measures from other parameters such as the trustworthiness of the feature-reduction methods (Venna, 2007) were also evaluated in addition to the 3D scatter analysis. The following subsections of this chapter provides insights into various linear and nonlinear feature reduction methods evaluated, and their algorithms are explained in greater detail. The results from the 3D scatter plots depicted in the following subsections only include the visual analysis of the best performing subject in Study A (subject 804).

Trustworthiness of the feature selection methods

One of the parameters used to decide the effectiveness of an approach in this study was to perform 'trustworthiness' analysis on the feature reduction algorithm. The trustworthiness metric was specially proposed for low-dimensional embedding. Venna and Kiski, (2006) were the first to propose the measure of trustworthiness for feature selection. Trustworthiness measures the proportion of points that are too close together in low-dimensional space (Venna and Kiski, 2006 and Venna, 2007). The trustworthiness $T_{trust}(k)$ measure is defined as

$$T_{trust}(k) = 1 - A(k) \sum_{i=1}^{N} \sum_{j \in U_k(i)} (r(i,j) - k), \tag{6.1}$$

where i, j are low-dimensional datapoints, and $r(i, j)$ represents the rank of a low-dimensional datapoint (j) to the pairwise distances between the low-dimensional datapoints. $U_{(k)}(i)$ is a variable which represents the set of points that are among the k nearest neighbours in the lower dimensional datapoint (i).

The term $A(k)$ scales the measure to be between zero and one. The error gets its maximum value when the ranks in the input and output space are reversed. The scaling term can then be found by considering that the maximum error in each data points neighbourhood is the sum of the k last ranks (minus the neighbourhood size k). Thus the scaling term becomes

$$A(k) = \frac{2}{Nk(2N - 3k - 1)} \text{ if } k < \frac{N}{2}, \tag{6.2}$$

$$A(k) = \frac{2}{N(N - k)(N - k - 1)} \text{ if } k \geq \frac{N}{2}. \tag{6.3}$$

Essentially, the trustworthiness measure is very closely related to the precision measure for the case where the objects are ranked based on their relevance (Venna, 2007). The neighbourhood size k in the trustworthiness measure defines the number of items retrieved. There is a distinct advantage over measuring the reconstruction errors when evaluating the trustworthiness and generalization errors, because a high reconstruction error does not necessarily imply that the dimensionality reduction technique performed poorly (Venna and Kiski, 2006, Van der Maarten, 2009).

6.2.1 Principal component analysis

PCA is a linear technique which has been used in a wide range of research areas as a non-parametric method of extracting relevant information from complex and often confusing datasets and is a well-known unsupervised dimensionality reduction technique (Jolliffe, 2002). To reduce the redundancy within the original features, PCA is used to transform the feature vectors into orthogonal components to aid the formation of the

classification models. The transformed feature vectors for a matrix of "principal components" (PCs), i.e., where the first component, PC 1, is the projection of the data on to the first principal axis, where principal axis represents the direction of the highest variance of the data. The direction of the second component, PC 2 represents the highest of the remaining variance orthogonal to the first component, etc.

Therefore, it is possible to reduce the dimensionality of the data by using the first p out of m PCs (where p, m are integers and m is the total number of features in the data and, $p < m$) without any significant loss in the information. Since it is an unsupervised technique, PCA does not need (nor can benefit from) any prior information from the class labels. Hence, it is hypothesized that a supervised feature reduction method may lead to improved results over those obtained by Peiris *et al*. (2011). In mathematical terms, PCA attempts to find a linear mapping K that maximizes $K^T \text{cov}(\mathbf{X})K$ where $\text{cov}(\mathbf{X})$ is the sample covariance matrix of the data $\mathbf{X}$. The linear mapping in the data can be formed by the PCs (eigenvectors). The size of the covariance matrix is usually proportional to the dimensionality of the datasets. As a result, the computation of the eigenvectors might be complicated and computationally expensive for very high dimensional datasets.

Figure 6-5- and 6-6- depict the 3D-cluster pattern of the meta-features generated by PCA for subjects 804 and 820. In both figures, spectral features containing events are identified with red circles (microsleep states) and the features without an event are identified with blue circles (alert state). Figure 6-5 not only depicts how the meta-features are interpreted from a processed dataset using PCA, but also depicts the innate intricacies of the current microsleep classification problem on the whole. As evident from the plot, the meta-features corresponding to both the classes (microsleep and non-microsleep) are clustered on top of each other making the classification problem extremely complicated for any classifier to identify between the labels or even create a linear or a non-linear separation between them.

As stated above, PCA is the best method in the mean-square error sense as it provides a linear dimension reduction solution and is based on the covariance matrix of the variables. In essence, PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (PCs) of the original variables with the largest variance in an unsupervised manner. PCA was used as a baseline for comparison to other feature reduction methods due it is use in the previous lapse detection studies (Peiris, *et al.*, 2011; LaRocco, 2015).

To summarize, PCA is an extremely useful tool in the approximation of datasets with inherent low dimensionality, but when it comes to higher-dimensional datasets, non-linear feature reduction techniques may generally yield better features. It is hypothesized that PCA is not as useful in the microsleep detection problem because:

1. The covariance matrix is difficult to be evaluated in an accurate manner.

2. Even the simplest invariance could not be captured by PCA unless the training data explicitly provides this information.

3. Directions in the data maximizing the variance do not always maximize information.

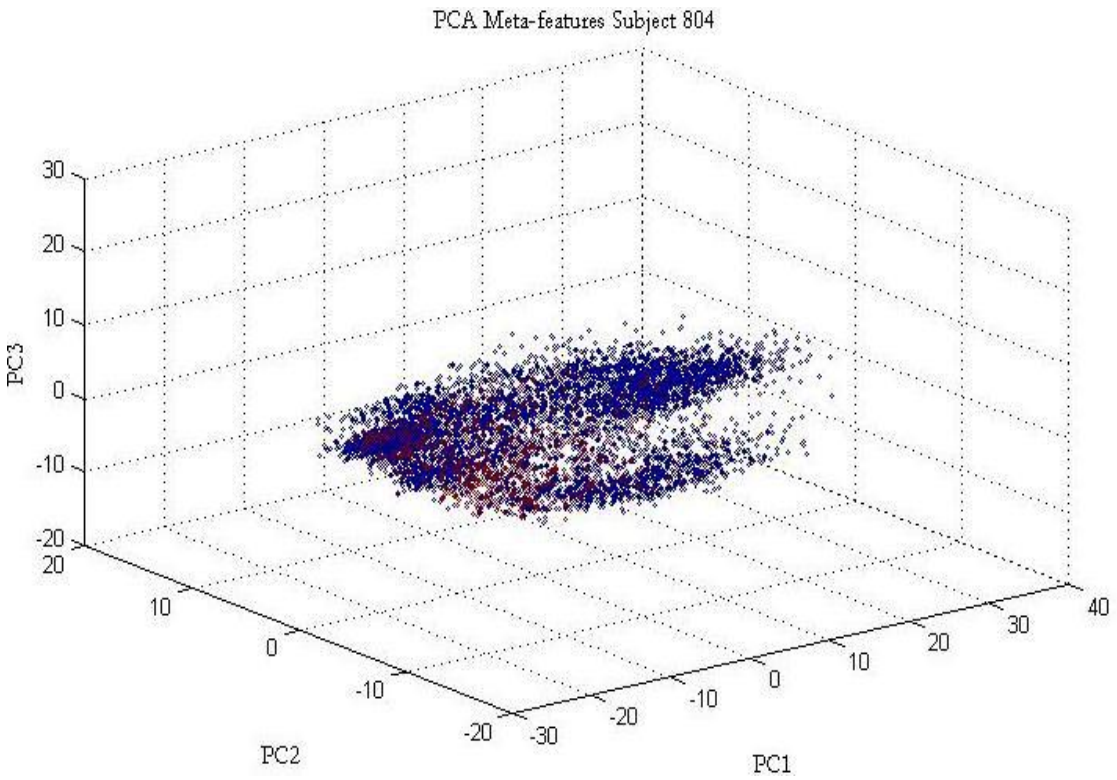


Figure 6-5 Visualizing the class distribution of the PCA meta-features on subject 804. The three axes represents the top 3 principal components PC1, PC2, and PC3 after PCA feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

To overcome the challenges discussed above, or maximally separate desired state/ class from noise, a new method, average distance between events and non-events (ADEN) was proposed by LaRocco *et al.* (2014).

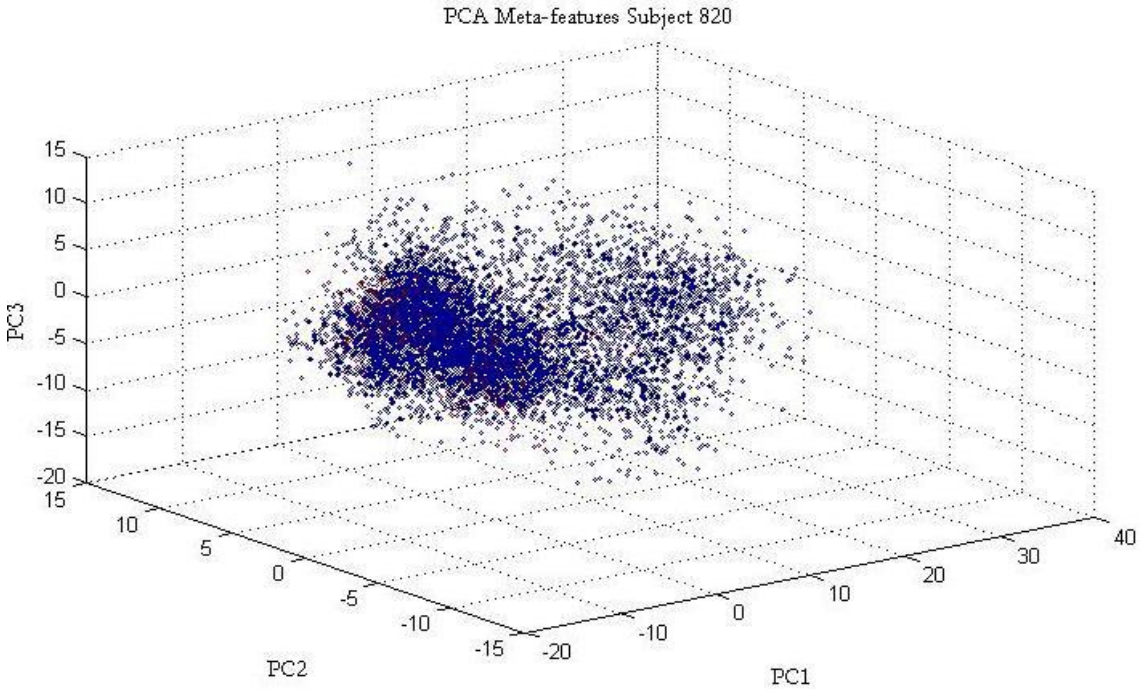


Figure 6-6 Visualizing the class distribution of the PCA meta-features on subject 820. The three axes represents the top 3 principal components PC1, PC2 and PC3 after PCA feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.2.2 Average distance between events and non-events

ADEN is a supervised feature reduction method in which all the features in each class are averaged together within each observation and the resultant vectors are then subtracted from each other after the largest distances are found. The features corresponding to the largest average distances are retained, and the rest discarded. Interestingly, despite being superior to PCA, ADEN, didn't outperform PCA on the microsleep detection problem (LaRocco, 2015).

Although from a theoretical standpoint of ADEN, the largest distances of the features should provide a better separation between both microsleep and alert states. However through visual analysis depicted in the Figure 6-7, it is evident that the algorithm could not clearly separate between the two classes. Ideally, one should expect all the microsleep events to be closer to the origin and the alert states to be further away from the origin using ADEN.

In mathematical terms, for every *U* features to be retained, the training data **X** consists of *f* features and **M** observations. Then, features corresponding to both the classes are separated into $\mathbf{x}_e$ (event) and $\mathbf{x}_n$ (non-event). Following this, data from each of the classes was averaged to form a mean feature vector, $\bar{\mathbf{x}}_e$ and $\bar{\mathbf{x}}_n$. Then the difference forms a single vector $\Delta\bar{\mathbf{x}}_f$, where,

$$\Delta\bar{\mathbf{x}}_f = abs\ (\bar{\mathbf{x}}_{e,f} - \bar{\mathbf{x}}_{n,f}) \tag{6.4}$$

"The difference between classes in ADEN is normalized by dividing the vector by Cohen's d, such that within-group variances in the training data can be accounted for. Training data **X** is reduced to a matrix of **U** features and **M** observations, with all remaining features based on the indices *f* of the **U** terms in $\Delta\bar{\mathbf{x}}_f$. The testing data would likewise be reduced to features, selected from the *f* indices corresponding to features in the training data" (LaRocco, *et al.*, 2014).

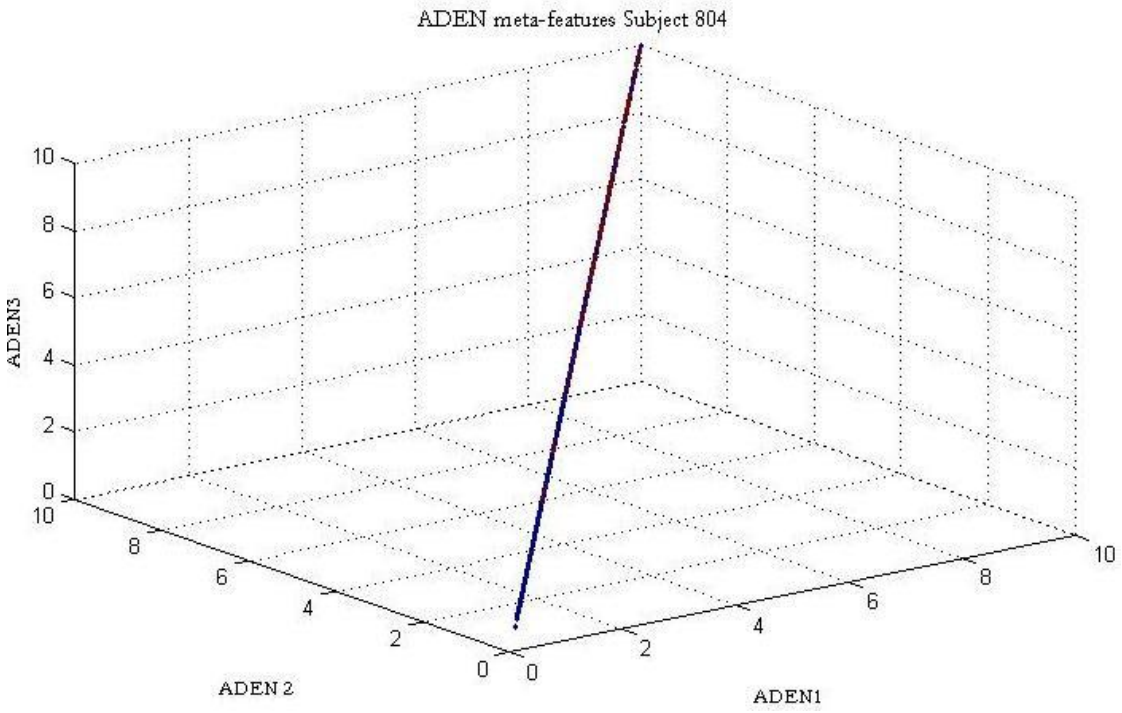


Figure 6-7 Visualizing the class distribution of the ADEN meta-features on subject 820. The three axes represents the top 3 (ADEN1, ADEN2 and ADEN3) of the 10 ADEN features. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

LaRocco, (2015) evaluated 2 other variants of ADEN, namely, normalized z-score transform ADEN (ZADEN) and Genetic algorithm-based ADEN (GADEN). Both these methods were applied to microsleep detection on Study A. However, it was reported that both these methods failed to outperform PCA on Study A dataset with a mean performance in terms of φ = 0.21 and φ = 0.32, respectively. However, PCA achieved the current benchmark performance on these datasets φ = 0.39 (Peiris *et al.*, 2011).

6.2.3 Kernel PCA

Kernel PCA (KPCA) is a reformulation of traditional linear PCA in a high-dimensional space that is constructed using a kernel function (Scholkopf *et al.*, 1998). KPCA achieves non-linear feature reduction through the use of Kernel functions. KPCA with a linear kernel is the same as traditional PCA. Since KPCA is a kernel-based feature reduction method, the mapping function from the KPCA is highly reliant on the type of the kernel function being used. There are multiple variants for a nonlinear kernel, such as Guassian and polynomial kernels (Shawe-Taylor and Christianini, 2004).

Constructing the Kernel matrix

The kernel matrix **K** is computed for the observations x_i and the entries are defined as

$$k_{ij} = \mathbf{K}(x_i, x_j), \tag{6.5}$$

where *k* represents the kernel function and *k* can be any function that generates a positive-semi-definite kernel **K** (Van der Maaten, 2009). Consequently, **K** is centered with the following modification to the entries

$$k_{ij} = k_{ij} - \frac{1}{n}\sum k_{il}\frac{1}{n}\sum k_{jl} + \frac{1}{n^2}\sum_{lm} k_{lm}. \tag{6.6}$$

In KPCA, the centering operation corresponds to subtracting the mean of the features in traditional PCA. The kernel centering operation determines that the features in higher-dimensional space are defined by the kernel function and contain a zero-mean. Finally, the principal eigenvectors v_i of the centred kernel matrix are computed by

$$\alpha_i = \frac{1}{\sqrt{\lambda_i}} v_i, \tag{6.7}$$

where α_i represents all the eigenvectors of the covariance matrix in the higher dimensional space and v_i represents the scaled versions of the eigenvectors of the kernel matrix. From this relationship it is evident that the eigenvectors of the covariance matrix can be scaled versions of the eigenvectors of the kernel matrix.

For obtaining the low-dimensional data representation **Y**, using KPCA, the data is typically projected onto the eigenvectors of the covariance matrix α_i. The resulting **Y** is denoted as

$$\Upsilon = \sum_j \alpha_1 K(x_j, x),\ \sum_j \alpha_2 K(x_j, x), \ldots\ldots,\ \sum_j \alpha_d K(x_j, x). \quad (6.8)$$

Many studies in machine learning have successfully demonstrated the application of KPCA techniques in speech processing and image processing (Lima *et al.,* 2004; Kim *et al.,* 2002). Subsequently, the KPCA approach was studied for the application on the microsleep detection problem, not only to compare its validity against the traditional PCA schemes, but also as it was thought that feature mapping in the microsleep detection problem is nonlinear in nature.

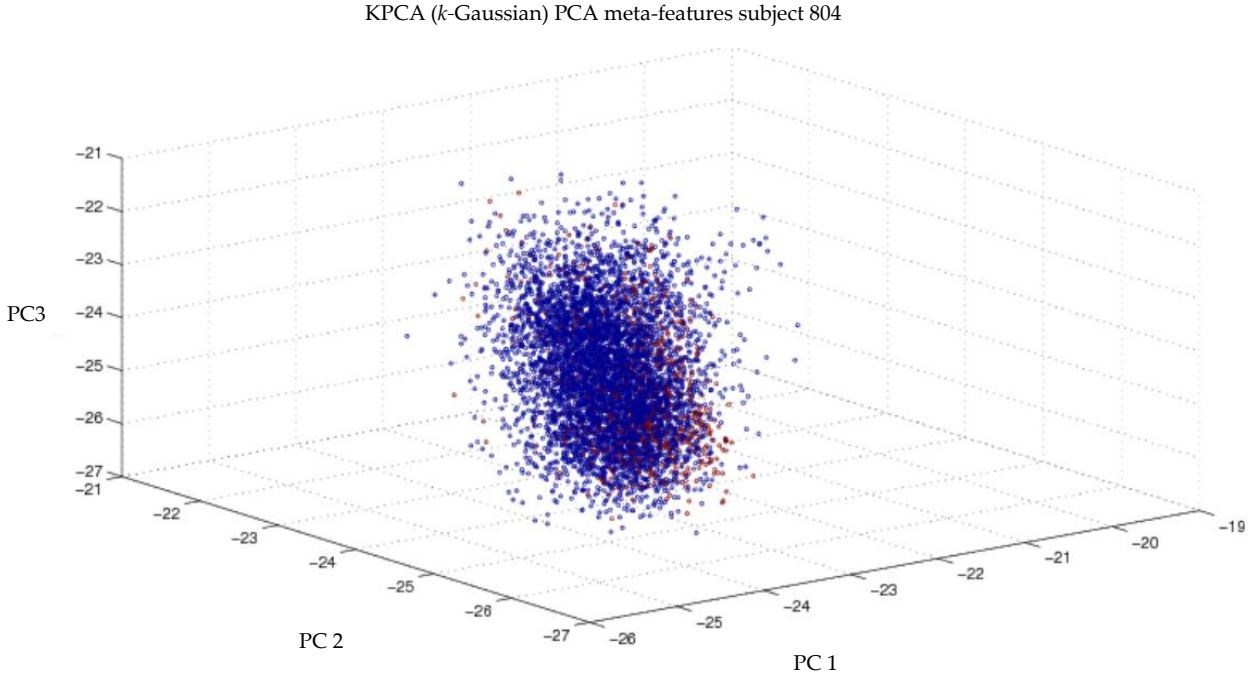


Figure 6-8 Visualizing the class distribution of the kernel PCA (K = Gaussian) meta-features on subject 804. The three axes represents the top 3 PCs (PC1, PC2 and PC3) of the 50 KPCA meta-features. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

The 3D scatter plot in the Figure 6-8 demonstrates the application of the kernel PCA technique using the Guassian kernel on the subject 804 in Study A. The three most significant meta-features (PCs) are plotted against the *x, y* and *z* axes. The red circles correspond to the microsleep states and the blue circles correspond to the alert state. Although, KPCA is considered to be a better feature reduction to traditional PCA (due to its kernel functions), the comparison of the scatter plots in Figure 6-5 and Figure 6-6, does not reveal any significant insights or improvement into the ideal separation between the classes (microsleep and alert). It is evident from both scatter plots that any classifier technique (be it linear or non-linear) could struggle in order to create an ideal separation between both the classes. That said, however, the scatter pattern in KPCA is a little less disarranged compared to Figure 6-6; the scatter pattern in traditional PCA. Partly due to this reason, and its reasonable

mean trustworthiness (explained earlier in this section) score T = 0.40, KPCA was evaluated as an alternative to the PCA based approach for the microsleep detection.

6.2.4 Probabilistic PCA

Probabilistic PCA (PPCA) is used to obtain a probabilistic formulation of PCA from a Gaussian latent variable model, which is closely related to statistical factor analysis; this allows a likelihood measure which in turn enables comparison with other probabilistic techniques, while facilitating statistical testing. PPCA has long been used as a method to estimate the principal axes when any data vector has one or more missing values (Roweis, 1997). PPCA has been applied in many signal processing applications and, interestingly enough, is not based on a probability model but is deterimed through maximum-likelihood estimation of the parameters in a latent model closely related to factor analysis (Tipping and Bishop, 1999).

PPCA is based on an isotropic error model (Ilin and Raiko, 2010). It seeks to relate a p-dimensional observation vector $\mathbf{y}$ to a corresponding k-dimensional vector of latent (or unobserved) variable $\mathbf{x}$, which is normal with zero-mean and covariance I(k). The relationship is established as

$$\mathbf{y}^T = \mathbf{W} \times \mathbf{x}^T + \mu + \varepsilon, \tag{6.9}$$

where $\mathbf{y}$ represents the row vector of observed variable, $\mathbf{x}$ represents the row vector of latent variables, and ε is defined as an isotropic error term. ε is Gaussian with zero-mean and covariance of $v \times I(k)$, where v is the residual variance. In the case of PPCA, the value of k needs to be smaller for the rank of the residual variance to be greater than zero. Traditional PCA, where the residual variance is zero, is the limiting case of PPCA. $\mathbf{W}$ is a matrix of observations relating to the latent and observed values.

Here, the values of $\mathbf{y}$ are conditionally independent, given the values of $\mathbf{x}$. Therefore, it is possible that the values of $\mathbf{x}$ can explain the correlations between the values of $\mathbf{y}$ and the error explains the variability unique to a particular $\mathbf{y}_j$.

In PPCA, it is assumed that the observable values are missing at random intervals throughout the data set, which means that whether a data value is missing or not, the algorithm does not depend on the latent variable $\mathbf{x}$ given the $\mathbf{y}$ values. Therefore, based on this, the observable variable y can be expressed as;

$$\mathbf{y} \approx N(\mu, \mathbf{W} \times \mathbf{W}^T + v \times I(k)). \tag{6.10}$$

Tipping and Bishop (1995), proposed that there is no closed-form solution for both W and v, and, therefore, their estimates are determined by iterative maximization of loglikelihood using an expectation minimization algorithm. Hence, for this reason, PPCA has been proposed as a powerful algorithm in several image processing, time-series prediction and pattern recognition tasks.

Figure 6-9 depicts the application of the PPCA algorithm to the subject 804 of Study A dataset. It is evident from the comparison of all the PCA-based methods described above, PPCA is a superior algorithm from visualization and comparison of all the scatter plots. Therefore, it can be hypothesized that based on the visualizations and the algorithmic strength of the PPCA, that it could outperform a microsleep detection system, which uses a traditional PCA algorithm. PPCA reported the best overall trustworthiness scores (mean from all the 8 subjects) with the highest mean T = 0.54. Additionally, there is strong evidence suggesting that PPCA-based approaches have worked well on time-series and pattern recognition problems (Tipping and Bishop (1995). For these reasons PPCA was considered as a superior non-linear technique and was evaluated as an alternative scheme for the traditional PCA-based approach for microsleep detection on the Study A dataset.

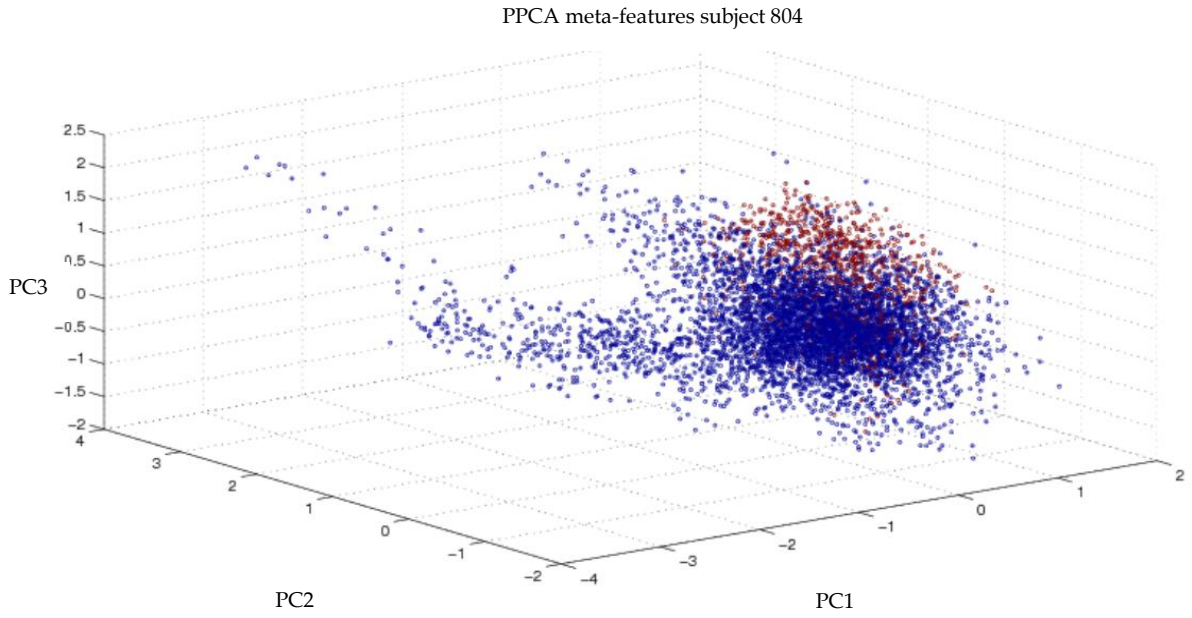


Figure 6-9 Visualizing the class distribution of the probabilistic PCA meta-features on subject 804. The three axes represents the top 3 PCs (PC1, PC2 and PC3) of the 50 PPCA meta-features. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.2.5 Classical multi-dimensional scaling

Multidimensional scaling (MDS) was proposed as a classical approach to the problem of finding underlying attributes or dimensions, via a visual representation of the pattern of proximities (i.e., similarities or distances) among a set of objects (Wickelmaier, 2003). The

MDS algorithm represents a collection of nonlinear techniques which can map the high-dimensional data into a much lower-dimensional representation whilst retaining the pairwise distances between the observable datapoints as much as possible (Van der Maaten, 2007). MDS is often considered as an extension of the Sammon mapping problem (Balasubramaniam and Schwartz, 2002).

MDS uses a stress function φ to express the quality of mapping, which is a measure of the error between the pairwise distances in both low-and high-dimensional representations (**Y** and **X**, respectively) of the observable data. The stress function is defined as

$$\varphi(\mathbf{Y}) = \mathbf{X}(kx_i - x_jk - ky_i - y_jk)^2, \tag{6.11}$$

where $kx_i - x_jk$ represents the Euclidean distance between the high-dimensional observations x_i and x_j and $ky_i - y_jk$ represents the Euclidean distance between the low-dimensional observations y_i and y_j, and k represents the nearest neighbours.

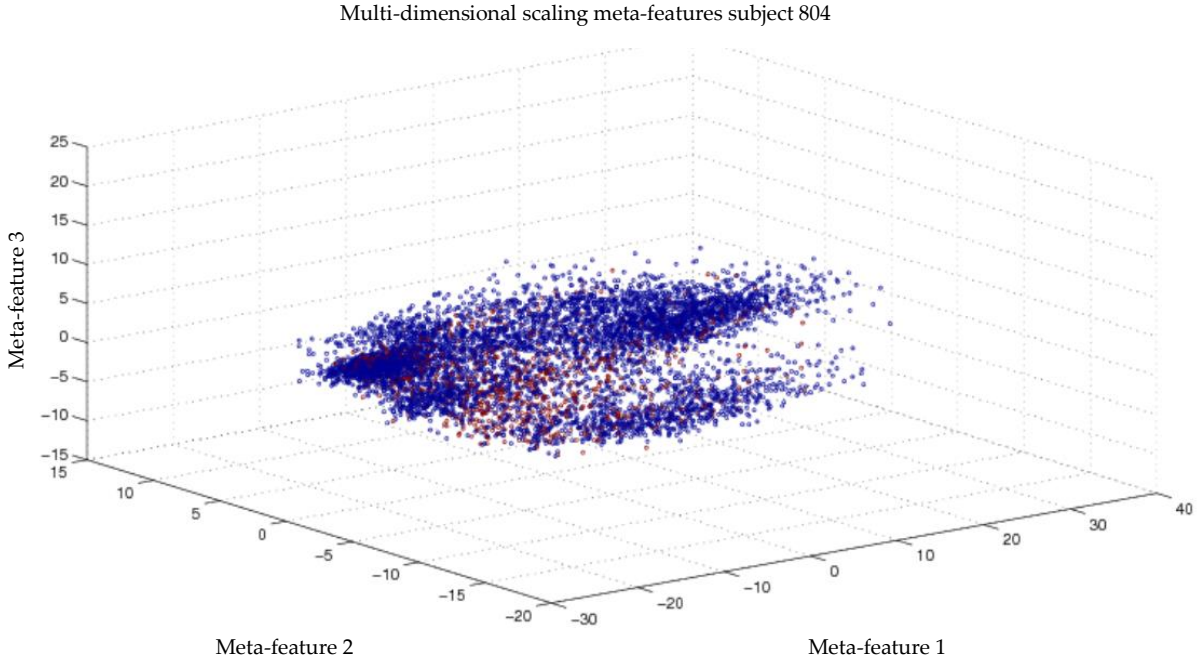


Figure 6-10 Visualizing the class distribution of the MDS meta-features on Subject 804. The three axes represents the top 3 meta-features of the 50 MDS features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

MDS is a widely accepted technique for data visualization of EEG, fMRI and other biomedical and biomolecular based analyses. (Tagaris *et al.*, 1998; Venkatarajan and Braun, 2004). Figure 6-10 depicts the 3D scatter plot of the MDS approach on the microsleep detection problem for subject 804 of Study A. Based on the visualization of the MDS algorithm, it is inconclusive to determine whether or not this approach could offer a better

seperation between the classes. Although, MDS works by preserving distances between points in the data, the individual clusters of each class in the above figure appear to be close to one another.

A significant disadvantage of this approach is that it is based on Euclidean distances, and does not take into account the distribution of the neighbouring observations. Therefore, to address this problem, another non-linear technique called Isomap was proposed (Balasubramaniam and Schwartz, 2002. Due to its poor trustworthiness scores and the inconclusive scatter-plot analysis, MDS algorithm was not investigated any further for microsleep detection on the Study A dataset. The mean Trustworthiness score using the classical MDS was reported at 0.15.

6.2.6 Isometric mapping

Isometric mapping (Isomap) is one of the earliest approaches for manifold learning and is widely accepted as an extension for MDS or KPCA methods (Balasubramaniam and Schwartz, 2002). Isomap seeks a low-dimensional embedding which maintains geodesic distances[10] between all the points. Isomap uses the same principles as the MDS algorithm. The intuition behind the Isomap is that it:

1) Obtain a matrix of proximities (distances between points in a dataset).

2) The distance matrix is a matrix of inner products.

3) An Eigen decomposition of the distance matrix gives a lower dimensional embedding.

A significant difference between Isomap and the MDS algorithm is the way in which the distance matrix is constructed.

In Isomap, the geodesic distances between the observations x_i are computed by constructing a neighbourhood graph G, in which every datapoint x_j is connected with its k nearest neighbours x_{ij} in the dataset **X**.

The shortest path between two points in the graph forms a good estimate (over estimate of the geodesic distance between these two points, and can easily be computed using Dijkstra's shortest-path algorithm (Djikstra, 1959)). In the Isomap algorithm, distances between points

[10] Geodesic or curvilinear distance is the distance between two points measured over the manifold. Manifold learning is an approach to non-linear feature reduction. Manifold learning algorithms are based on the idea that the dimensionality of many data sets is only artificially high.

are considered as the weight of the shortest path in a point-graph. The pairwise geodesic distance matrix is formed from the geodesic distances between all observations in **X**.

Finally, the low-dimensional representations y_i of the observations x_i in the low-dimensional space **Y** are computed by applying the MDS framework to obtain resulting distance matrix.

The Isomap algorithm has been applied on many datasets, in computer vison, biomedical engineering, and image processing with reasonable results (Ng *et al.*, 2001, Raymer *et al.*, 2000; Lim *et al., 2003*). Figure 6-11 depicts the isomap application of the microsleep detection problem on subject 804 from Study A. From the 3D cluster plot, it can be seen that the performance of the Isomap algorithm on subject 804 data is very poor. A possible reason behind this may be that the isomap algorithm might have failed to attain a topological stability due to being riddled with several erroneous connections while computing the neighbourhood graph, *G*.

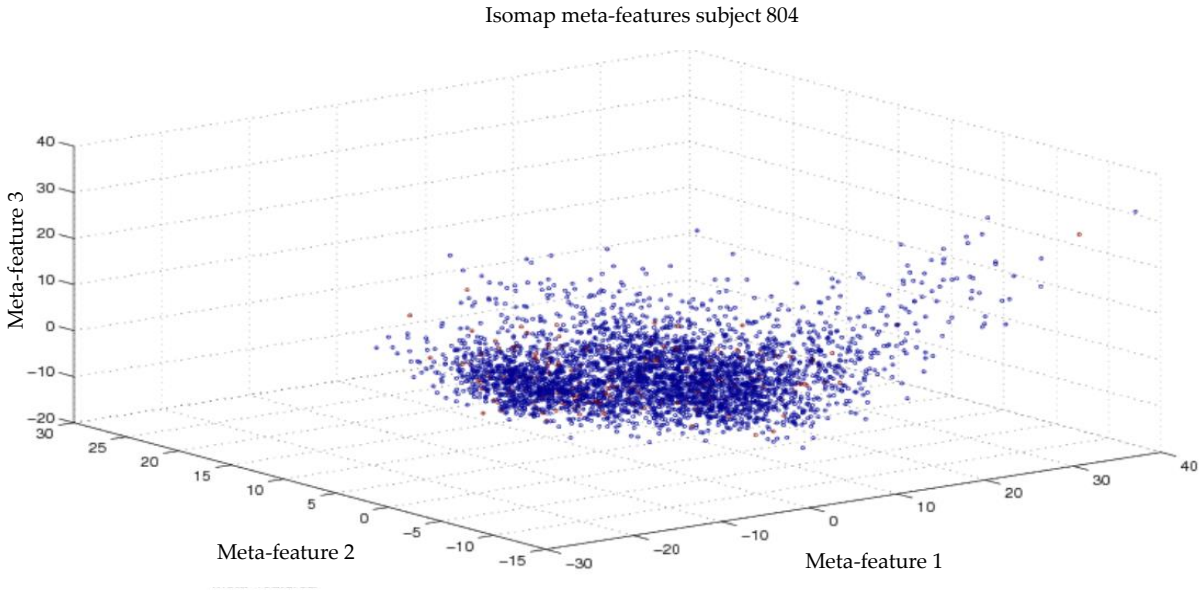


Figure 6-11 Visualizing the class distribution of the Isomap meta-features on Subject 804. The three axes represents the top 3 meta-features of the 50 Isomap features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

The Isomap algorithm was another technique which was discarded from further investigation for the microsleep detection on Study A dataset, due to its poor showing on the 3D scatter analysis and the trustworthiness scores (T = 0.10). Isomap reported the lowest trustworthiness scores of all the algorithms investigated in this Chapter.

6.2.7 Nearest neighbour estimation

The nearest neighbour estimation (NNE) algorithm is based on the estimation of the number of neighbouring observations that are covered by a hypersphere with radius *r*. The NNE algorithm does not explicitly count the number of observations inside a hypersphere with radius *r*, but computes the minimum radius *r* of the hypersphere necessary to cover *k* nearest neighbours (Van der Maarten, 2009). The nearest neighbour estimator computes

$$C(k) = \frac{1}{n} \sum T_k(x_i), \quad \text{where } c = \begin{cases} 1 & if \left\| x_i - x_j \right\| < r \\ 0 & if \left\| x_i - x_j \right\| > r \end{cases}, \tag{6.12}$$

where $T_k(x_i)$ signifies the radius of the smallest hypersphere with centre (x_i) which coves *k* neighbouring observations. The dimensionality of a dataset $\hat{d}$ is estimated by

$$\hat{d} = \frac{\log(C(k_2 - C(k_1))}{\log(k_2 - k_1)}. \tag{6.13}$$

The NNE algorithm has previously been used in applications with image processing and biomedical datasets (Kegl, 2002; Kim *et al.*, 2002).

Figure 6-12 depicts the NNE application to the subject 804 of Study A dataset for microsleep detection. From the scatter plot above, it is evident that the NNE algorithm was able to reasonably distinguish and separate both the classes, although a few errors were noted. Comparing the NNE scatter analysis in Figure 6-12 with the PCA in Figure 6-6, it can be hypothesized that NNE could serve as a strong contender for PCA as another linear technique. Albeit being simple in terms of its implementation, the visual analysis and the trustworthiness scores of the NNE algorithm showed that it is able to better separate the classes than several of the other computationally-intensive algorithms, such as isomaps. Hence, for this reason, NNE was investigated further in microsleep detection system. The trustworthiness values for the NNE-based algorithm on the Study A dataset was found at T= 0.40.

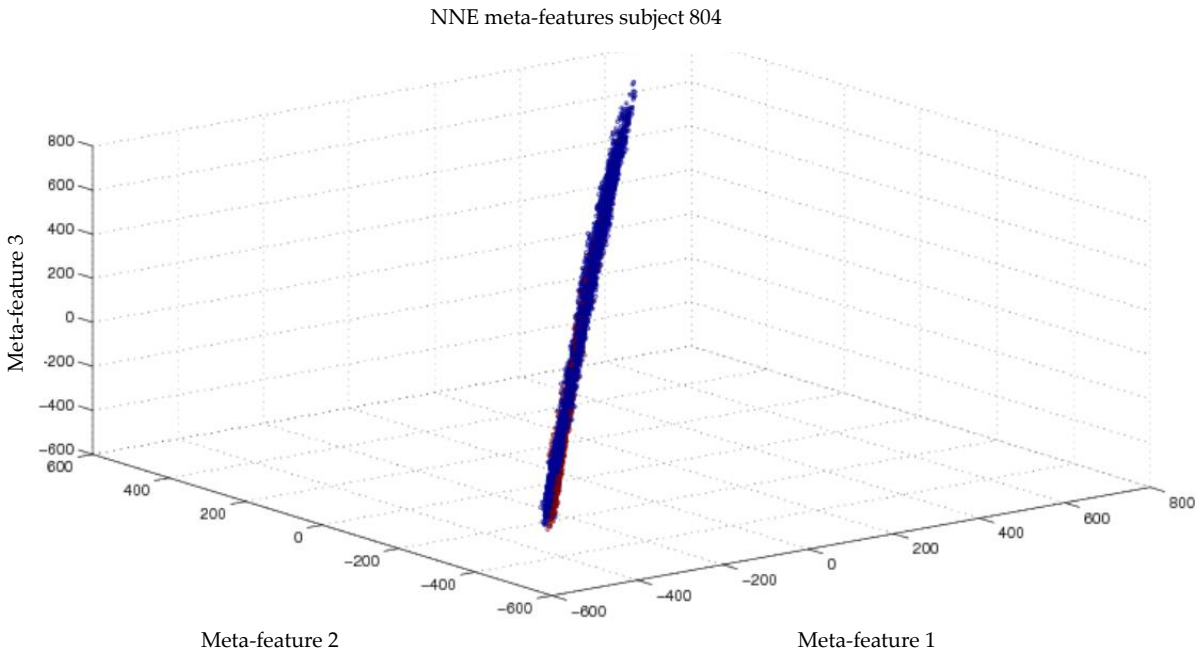


Figure 6-12 Visualizing the class distribution of the nearest neighbourhood estimator meta-features on subject 804. The three axes represents the top 3 meta-features of the 50 NNE features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.2.8 Stochastic neighbourhood embedding

Stochastic neighbourhood embedding is a probabilistic approach in which features described by high-dimensional vectors or by pairwise dissimilarities are placed in a low-dimensional space with their neighbour identities preserved. In essence, SNE is only slightly different to MDS in terms of the distance measure that it used a minimalist cost function. In SNE, a Gaussian is centred on each object in the high-dimensional space and a few given dissimilarities under this Gaussian are used to define a probability distribution over all the potential neighbours of the features (Hinton and Roweis, 2002).

In SNE, the matrix *P* denotes the distribution of all the individual probabilities p_{ij} for all the observations x_i and x_j generated by the same Gaussian. SNE models the similarity of datapoint x_i to datapoint x_j as the conditional probability p_{ij}, that x_i would pick x_j as its neighbour if neighbours were picked in proportion to their probability density under a Gaussian centered at x_i.

The probabilities p_{ij} are calculated using the Gaussian kernel function

$$w_{ij} = e^{\frac{\| x_i - x_j \|}{2\sigma^2}}, \tag{6.14}$$

where w_{ij} represents the individual weights of the observations and σ represents the variance of the Gaussian. In the next phase of the algorithm, the coordinates of the low-dimensional representations y_i are set to random values, close to zero.

Following this, the probabilities for the low-dimensional counterparts y_i and y_j of the high-dimensional datapoints x_i and, x_j a similar conditional probability q_{ij} can be generated by the same Gaussian are computed and stored in the matrix $\mathbf{Q}$ which also uses the same Gaussian kernel.

The SNE algorithm aims to minimize the difference between the probability distributions **P** and **Q**. Hinton and Roweis, (2002) state that in a perfect low-dimensional representation of the data, the matrices **P** and **Q** are identical. In SNE, a natural cost function is the sum of the Kullback-Leibler divergences[11], and is given by.

$$\varphi(Y) = \sum_{ij} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{6.15}$$

SNE aims to minimize the sum of this Kullback-Leibler divergences by applying methods such as gradient descent technique (Hinton and Roweis, 2002). Figure 6-13 represents the application of SNE to the Subject 804 from the Study A dataset. The 3D scatter plot analysis depicts that, due to the probabilities, distributions of the SNE majority of the class distributions are spread around in the corner of the image. Despite seeming better in comparison to some of the other approaches evaluated, the SNE scatter plot reveals inconclusive evidence as to the validity of this approach.

Despite these initial setbacks, SNE was another non-linear feature reduction technique evaluated in microsleep detection due to its reasonably high trustworthiness (T = 0.38).

[11] Kullback-Leibler divergences are a natural distance measure to measure the difference between two probability distributions (Kullback & Leibler, 1951).

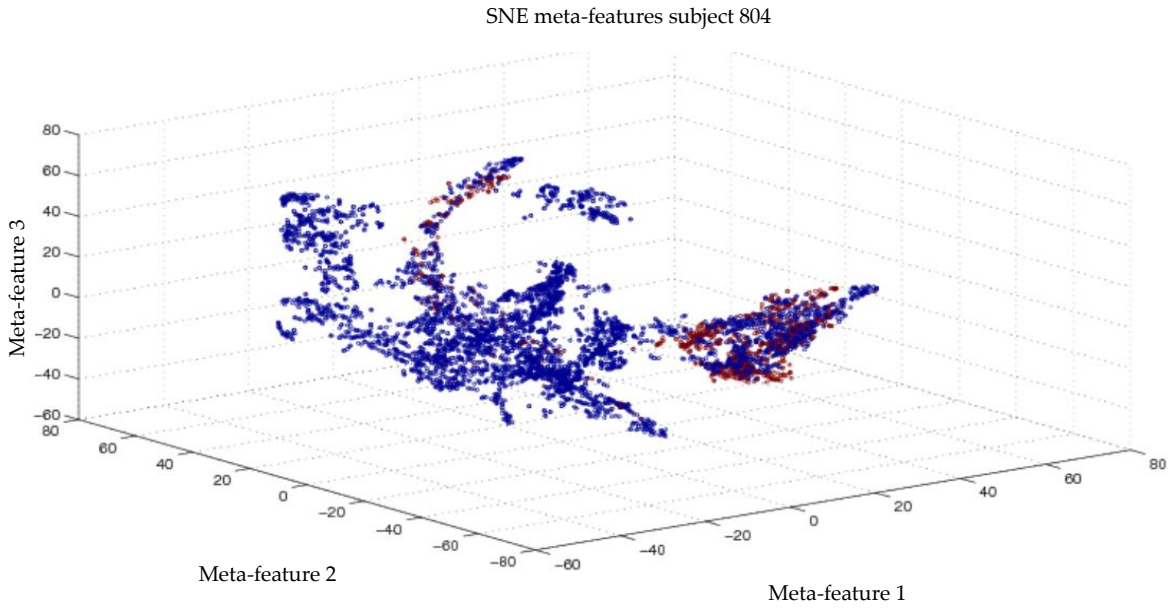


Figure 6-13 Visualizing the class distribution of the SNE meta-features on subject 804. The three axes represents the top 3 SNE meta-features of the 10 SNE features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.2.9 Autoencoder

Autoencoders are multi-layered and are a type of feed-forward neural networks possessing an odd number of hidden layers (DeMers and Cortell, 1993; Hinton and Salakhutdinov, 2006). The way an autoencoder works is by an unsupervised learning rule that applies backpropagation, setting the target values to be equal to the inputs $y_i = x_i$. Hence, autoencoder networks are trained to minimize the mean squared error (MSE) between the input and the output of the network.

When a set of data is passed through an autoencoder, the network initially compresses (encodes) input vector to *fit* in a smaller representation, and then tries to reconstruct (decode) the information back (Figure 6-14) . The premise of the training algorithm is to find the most efficient encoding representation for an input sequence. Autoencoders also try to convert a vector of *n*-dimensional space to an *m*-dimension, whilst retaining all the necessary information and, at the same time removing noise.

In order to train an autoencoder to model a nonlinear mapping between the high-dimensional and low-dimensional data representation, typically sigmoid activation functions are used. Autoencoders with linear activation functions resemble PCA (Kung *et al.*, 1994). The schematic structure of an autoencoder is shown in Figure 6-14.

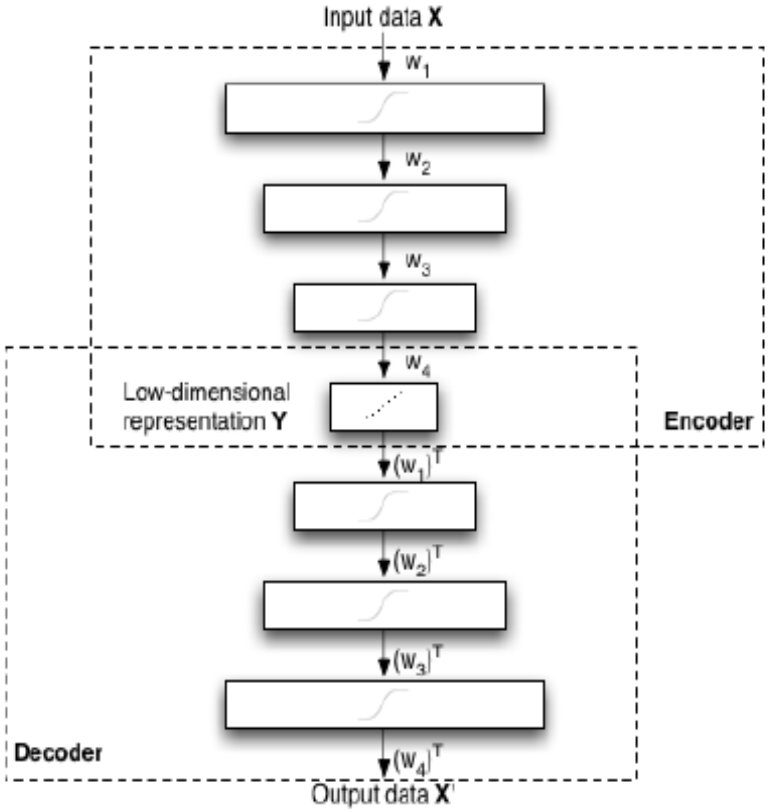


Figure 6-14 Structure of an autoencoder scheme (from Van der Maaten, 2009).

An autoencoder neural network was trained to reduce the EEG-features from all the subjects within the Study A. Figure 6-15 depicts the 3D cluster plot from the application of the autoencoder schemes to the Subject 804. It is evident from the class distribution of the cluster pattern that the autoencoder based-neural network could not perform well, which could have been the case of a possible over-fitting within the data. In addition, based on the trustworthiness scores (T = 0.12), this method was excluded from further investigation.

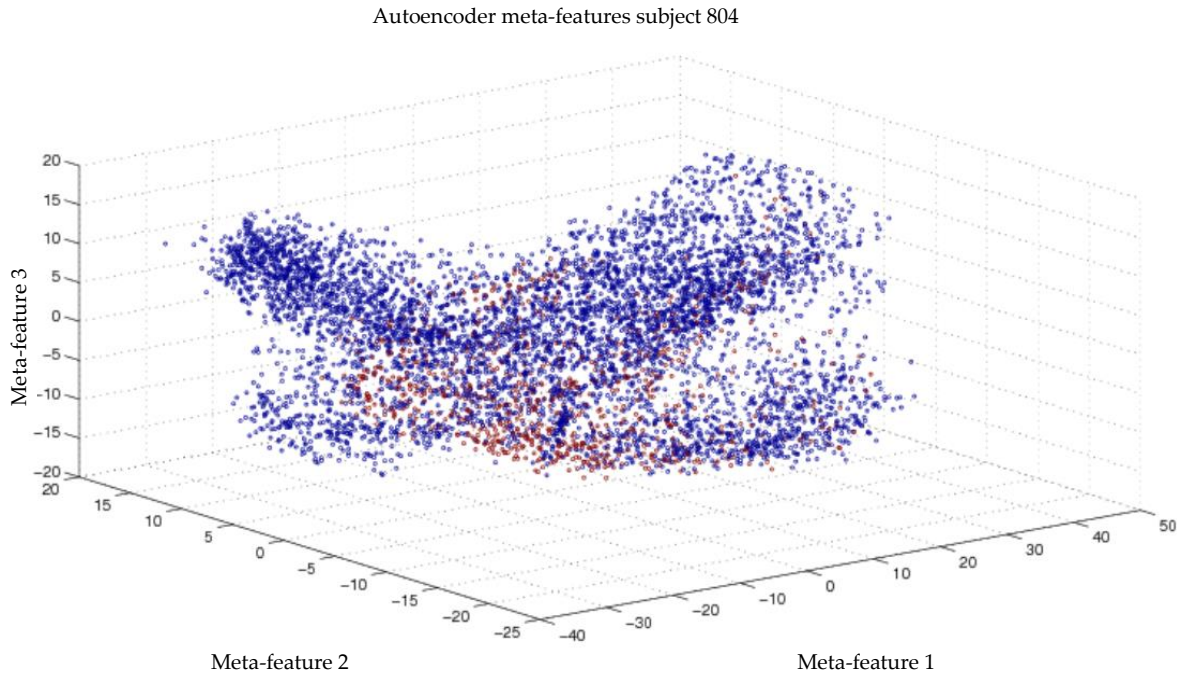


Figure 6-15 Visualizing the class distribution of the autoencoder meta-features on Subject 804. The three axes represents the top 3 meta-features of the 50 autoencoder features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.2.10 Stochastic proximity embedding

Stochastic proximity embedding (SPE) is considered to be a possible extension to the multi-dimensional scaling scheme. SPE runs an iterative algorithm to minimize the raw stress function ϕ of MDS.

$$\varphi(Y) = \sum_{ij}(d_{ij} - r_{ij})^2, \tag{6.16}$$

where r_{ij} is the proximity between the high-dimensional data points x_i and x_j, and d_{ij} is the Euclidean distance between their lower-dimensional counterparts y_i and y_j in the current approximation of the embedded space (Van der Maate, 2009).

SPE algorithm updates the current estimate of the low-dimensional data representation. SPE also has a behaviour comparable to that of the isomap in terms that, SPE can readily be applied in order to retain only distances in a neighbourhood graph *G* defined on the data by setting d_{ij} and r_{ij} to 0 if $(i, j) \notin G$.

The updating in the SPE algorithm is performed using the following update rules.

$$y_i = y_i + \lambda \frac{r_{ij} - d_{ij}}{2d_{ij} + \varepsilon}(y_i - y_j), \tag{6.17}$$

$$y_j = y_j + \lambda \frac{r_{ij} - d_{ij}}{2d_{ij} + \varepsilon}(y_j - y_i), \tag{6.18}$$

where λ is defined as a learning parameter. λ decreases with the number of iterations, and ε is a regularization parameter.

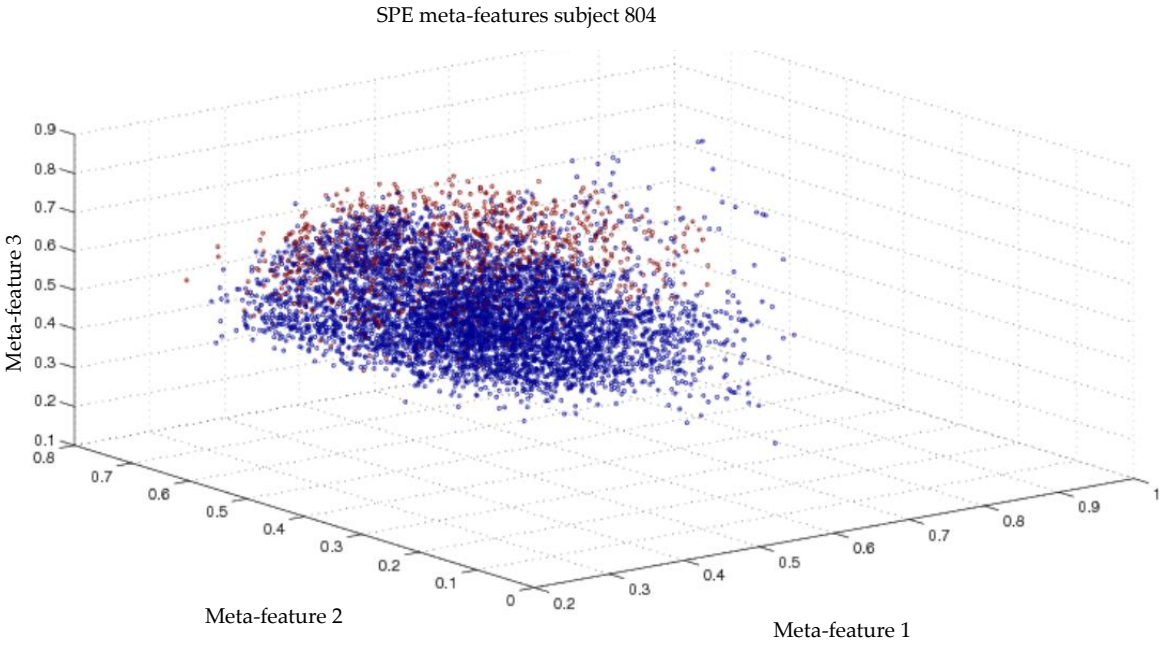


Figure 6-16 Visualizing the class distribution of the stochastic proximity embedding meta-features on Subject 804. The three axes represents the top 3 meta-features of the 50 SPR features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

Figure 6-16 provides an illustration of the 3D scatter plot from the application of the SPE to the subject 804 of Study A. The visualization results in addition the modest mean trustworthiness scores (T = 0.38) makes this approach eligible for further investigation. SPE is also computationally inexpensive and could accommodate a large number of iterations for updating its embedded coordinates.

6.2.11 Laplacian Eigenmaps

Lastly, the Laplacian Eigenmaps (LE) algorithm bears a resemblance to the Isomap in that it constructs a graph representation of all observations. The LE algorithm is based on the pairwise distance between the neighbours. LE computes a low-dimensional representation of

the data in which the distances between a datapoint and its *k* nearest neighbours are minimized (Van der Maaten, 2009).

The LE algorithm begins by constructing a neighbourhood graph *G* in which every observation x_i is connected to its *k* nearest neighbours. For all the points x_i and x_j in the graph *G* that are connected by an edge, the weight of the edge is computed using a Gaussian kernel function leading to a sparse adjacency matrix **W** (Belkin and Niyogi, 2001).

The cost function in the low-dimensional representations is minimized and is represented as

$$\varphi(Y) = \sum_{ij} (y_i - y_j)^2 w_{ij}. \tag{6.19}$$

A degree matrix **M** and a graph Laplacian *L* are computed for a graph *G* which allows for formulating the minimization problem as the Eigen problem (Anderson and Morley, 1985). The degree matrix **M** of **W** is a diagonal matrix, whose entries are the row sums of **W** ($m_{ii} = \sum_j w_{ij}$). Finally, the graph Laplacian *L* is computed by *L* = **M** – **W**. It can be shown that the following holds

$$\varphi(\mathbf{Y}) = \sum_{ij} (y_i - y_j)^2 w_{ij} = 2\mathbf{Y}^T L Y. \tag{6.20}$$

Therefore, minimizing $\varphi(Y)$ is proportional to minimizing $\mathbf{Y}^T L Y$. The low dimensional data representation **Y** can thus be found by solving the generalized eigenvector problem.

$$Lv = \lambda \mathbf{M} v. \tag{6.21}$$

The LE algorithms have been implemented in image processing and machine learning fields (Ng *et al.*, 2001; Shi and Malik, 2000). Figure 6-17 represents the 3D cluster plot of the LE algorithm on Subject 804 from Study A. The visualization of the scatter plot depicts results which are inconclusive as the class separation is unevenly spread in the plane. Additionally, the low mean trustworthiness scores (T = 0.15) resulted in not pursuing this method in the final microsleep detection.

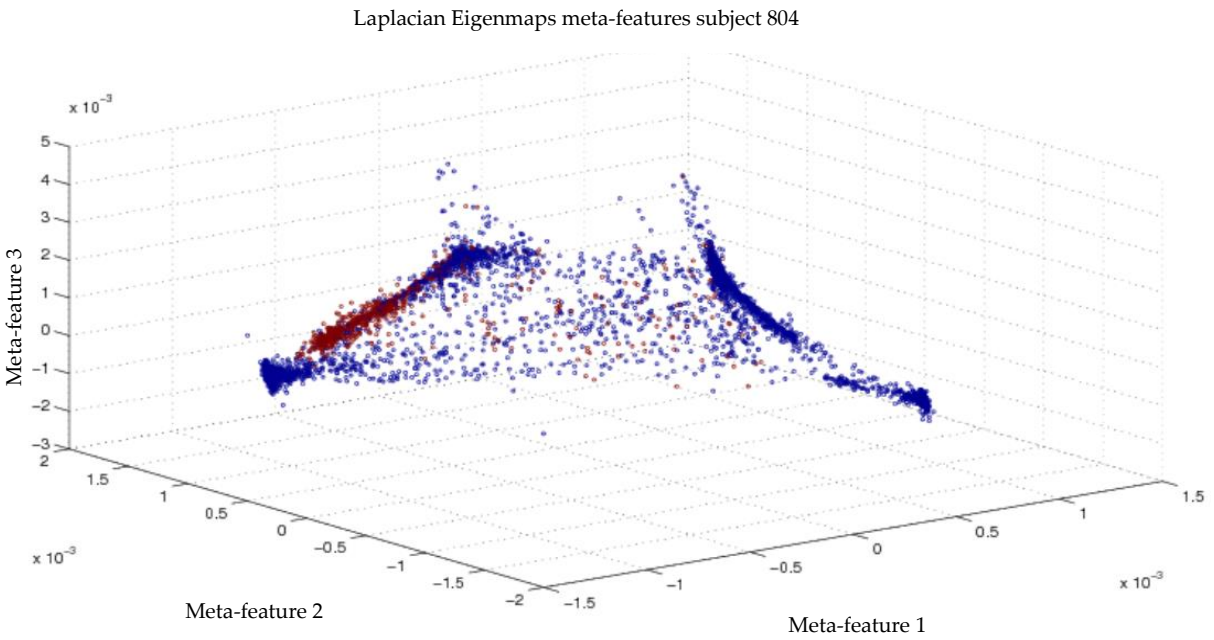


Figure 6-17 Visualizing the class distribution of the Laplacian Eigenmap meta-features on Subject 804. The three axes represents the top 3 meta-features of the 60 SPR features after the feature reduction. Blue circles represent the class 'alert state' and the red circles represent the 'microsleep state'.

6.3 Results and discussion

Section 6.2 presented a qualitative comparative study of techniques for linear and non-linear feature reduction methods analysed in this study. The performance of each feature reduction algorithm was scored by its trustworthiness metrics.

The trustworthiness (T) scores for the different feature reduction schemes evaluated on the Study A dataset (Subjects 804 to 820) are depicted in Table 6-3. Based on the scores from the trustworthiness matrix for all the subjects and the visual analysis of the 3D scatter plots, a decision was taken whether or not to implement a particular feature reduction method. Five feature selection modules were chosen for evaluation on the microsleep detection problem and another 5 were discarded.

All the PCA-based methods had superior T scores with the Probabilistic PCA being the highest and the simple PCA next. The other methods that reported reasonable T values were the NNE, SNE and SPE. Overall, 50% of the feature reduction methods were investigated further and the remainder discarded. Isomap and autoencoders, despite being neighbourhood-based approaches, did not provide any satisfactory results, be it in terms of the T values (analytically) or 3D scatter patterns (visual isnpection).

Table 6-3 Trustworthiness scores of the feature reduction modules evaluated.

	PCA	KPCA	PPCA	MDS	Isomap	NNE	SNE	Autoencoder	SPE	LE
Subject										
804	0.74	0.61	0.89	0.29	0.13	0.63	0.52	0.1	0.56	0.29
809	0.62	0.57	0.66	0.17	0.07	0.44	0.46	0.19	0.40	0.17
810	0.08	0.03	0.11	0.01	0	0.02	0.03	0.0	0.04	0.01
811	0.67	0.49	0.74	0.11	0.21	0.54	0.41	0.11	0.51	0.11
814	0.32	0.23	0.33	0.09	0.01	0.30	0.26	0.09	0.21	0.09
817	0.51	0.44	0.50	0.14	0.1	0.47	0.34	0.11	0.33	0.14
819	0.33	0.27	0.37	0.17	0.04	0.27	0.11	0.17	0.35	0.17
820	0.72	0.60	0.74	0.22	0.22	0.55	0.62	0.21	0.68	0.22
Mean	0.49	0.40	0.54	0.15	0.10	0.40	0.34	0.122	0.38	0.15
Evaluated for final microsleep detection on Study A (Y or N)	Y	Y	Y	N	N	Y	Y	N	Y	N

From Table 6-3 it is also evident that most of the nonlinear techniques did not outperform PCA on the Study A dataset, despite their ability to learn the structure of complex nonlinear manifolds. Interestingly however, both the linear feature reduction methods, PCA and NNE, had substantial T values (second best and fourth best) and have outperformed several of the non-linear techniques on what is thought to have been a non-linear dataset.

An interesting trend was observed in the results from the trustworthiness scores across all the algorithms. Certain subjects had higher T values, while the others had consistently lower performances, irrespective of the algorithm implemented. For example, Subject 810 and 819 had the lowest scores reported. A possible reason for the consistently lower performances on these subjects could be attributed to the presence of a potentially lower number of training labels (i.e., lapses) within the data from that subject.

This, notwithstanding, the main challenge moving forward is the performance of these feature reduction problems on the actual microsleep detection problem which are highlighted in Chapter 7.

Sensitivity analysis

Various parameters – number of nearest neighbours, variance, number of iterations and perplexity – have been evaluated. These sensitivity parameters are said to affect the meta-features generated from the feature reduction methods, and also the results obtained from these methods (van der Maarten, 2009). The value of the variance from 1 to 10 (1, 2, 3, 5 and 10) were evaluated for the KPCA method. PPCA was tested on a number of iterations from 10 to 100 (10, 50, 100, 200 and 300). The number of neighbours from 1 to 20 (1, 5, 10, 12, 15

and 20) were implemented for Isomap and NNE. Similarly, Laplacian Eigenmaps were evaluated on various values of variance and number of neighbours. Finally, SNE was evaluated on perplexity, (a smooth measure of the effective number of neighbours). The performance of SNE is fairly robust to changes in the perplexity, with typical values between 5 and 50). The trustworthiness scores of the feature reduction algorithms evaluated using the above sensitivity parameters are provided in Appendix B. Generally, it was found that these sensitivity parameters had very minimal effect on the performance of the meta-features, which was also evident from the minimal variance in their trustworthiness scores (around 2-3%). A possible reason for the minimal impact of the sensitivity analysis can be attributed to the complex class distributions inherent in the Study A dataset (which was also evident from the overlap between the classes).

6.4 Summary

This chapter provided a review of the feature reduction algorithms used in this study, in addition to highlighting the importance of data visualization techniques in machine learning. Finally, the majority of this chapter was aimed at providing a comparative study of techniques for feature reduction used in microsleep detection. From the results obtained, it was concluded that nonlinear techniques for dimensionality reduction are, despite their large variance, not yet capable of outperforming traditional PCA. In the future, research might focus on the development of new nonlinear techniques for dimensionality reduction that (i) do not suffer from trivial optimal solutions, and (ii) do not rely on neighbourhood graphs to model the (local) structure of the data manifold.

CHAPTER 7

Reservoir computing for microsleep detection

7.1 Introduction

As discussed in Chapter 1, the overall goal of this research was to detect the occurrence of unintentional microsleep events in occupations which require sustained attention. Therefore, a device capable of monitoring an individual in real-time and providing an alarm at the start of an impending microsleep episode or, preferably, an advance warning of an impending lapse, would be invaluable. The work described in this chapter utilizes EEG data collected from the continuous tracking task study (Study A) described earlier in Chapter 3. Features extracted from the EEG power spectrum, such as, power in the traditional bands (delta, theta, alpha, beta, etc.) and ratios between band powers, were evaluated for their ability to detect the underlying microsleeps events.

A crucial step in the development of a microsleep detection system discussed in this thesis is the development of a classifier subsystem using an advanced machine learning algorithm, such as, the reservoir computing. This classifier subsystem should be capable of detecting microsleeps with high temporal resolution in new subjects, using only data from their EEG feature matrices. Therefore, the major focus of this chapter is on evaluating the RC-based algorithms for microsleep detection from the cues identified from EEG.

In addition to RC methods, such as the ESNs and LSMs, several other machine learning techniques, which were considered to have the potential to contribute towards reliable EEG-based microsleep detection, are also evaluated. Other machine learning techniques implemented were: support vector machines (linear kernel (SVML) and polynomial kernel (SVMP), k-nearest neighbours classifier (kNN), and spiking neural networks. Details of these

machine learning algorithms and their microsleep detection performances are presented in this chapter. Results on the relative contribution of EEG features and derivations, to achieve the best overall lapse detection model are discussed. Finally, the variation in performance with the number of features used to form the detector model is also presented.

7.2 Forming the microsleep detection model for Study A

As discussed in Chapter 3, Study A data contained a sub-category of lapses in which clear behavioural signs of sleep (such as eye-lid closure and head-nodding), and tracking flat spots, overlapped. These events were termed *microsleeps* to emphasize their arousal-related nature, to distinguish them from EEG-defined microsleeps (Ogilvie, 2001), to highlight that they were identified using two independent and conservative measures and, lastly, to distinguish them from less definite lapses, i.e., albeit mostly were also microsleeps.

Lapses, as defined by the presence of either a video microsleep and/or a flat spot, were selected as the events to be detected by an EEG-feature-based microsleep detector. Detector performances were also assessed for systems trained to detect all gold standard indicators including flat spots, video microsleeps, and definite microsleeps. However, results showed that the system based on detecting lapses performed best and, hence, only the results based on detecting lapses are reported here. Additionally, previous studies on lapse detection (Davidson *et al.*, 2007 and Peiris *et al.*, 2011) confirmed this observation since the same phenomenon was observed in both studies. The resulting output of the microsleep detection model was generated at a frequency of 1 Hz – i.e., a temporal resolution of 1.0 s.

The performance of the microsleep detector was measured both in terms of its ability to detect the *lapse-of-responsiveness states* and *microsleep states* (in 1-s epochs). In this chapter, performance analysis of the feature reduction methods introduced in Chapter 6 is also presented. Additionally, variation in performance, considering the number of meta-features used to form the best overall microsleep detection model, is also discussed. A comprehensive comparative analysis of the resultant detector performances from the combination of all the feature reduction methods (PCA, PPCA, KPCA, NNE, SNE and SPE) to the classifier modules (cascaded-leaky ESN, LSM, LDA, SVMP, SNN and kNN) are presented.

7.3 Performance metrics

The same methods used by Peiris *et al.* (2011) were reproduced and the subsequent results obtained were considered as the benchmark for performance. Cross validation of the datasets was performed on all classifiers in addition to the leave-one-out approach, where

one subject out of the N=8 is set aside for testing, while the others are used for training the classifier. This procedure was repeated until each of the 8 subjects in the (N = 8) dataset had been utilized as a test subject. The results from each of the datasets were then averaged together to determine the final performance of the lapse detector.

The output of the classifier models was a binary matrix indicating whether each of the given states was a 'lapse' state or an 'alert' state. Next, the classifier output was checked to see if it equalled 1 during any portion of the gold standard 'event'. If this condition was met, the entire portion of the event signal corresponding to the gold standard event was marked as a true positive (TP) event. Similarly, if the detector output was -1 during the entirety of the gold standard lapse, the corresponding region of the event signal was marked as a false negative (FN) event (Parker, 2011).

The region of the event signal corresponding to the detector lapse event was checked. True positives provided correct identifications of events and true negatives provided correct identifications of non-events. False positives provide incorrect labelling of a non-event as an event and false negatives, the incorrect assessment of an event as a non-event. The number of TPs, FPs, TNs, and FNs in the event signal were counted and the following performance parameters calculated:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Sensitivity} = TP / (TP + FN)$$

$$\text{Specificity} = TN / (TN + FP)$$

$$\text{Selectivity} = \text{Positive predictive value} = TP / (TP + FP)$$

$$\text{Negative predictive value} = TN / (TN + FN)$$

$$\text{Phi} = (TP \times TN) - (FP \times FN) / \sqrt{(TP + TN)(TN + FN)(TP + FN)(FP + TN)}$$

Sensitivity is also referred to as the true positive rate or 'hit' rate, selectivity is referred to as the positive predictive value or precision, and accuracy is defined as the total rate of correct identifications in the entire dataset. All the parameter values are expressed as percentages and are recorded for each configuration of the modules.

Mean phi correlation was considered as the primary performance metric because of its largely independent nature of the group distributions in a class. In addition to the phi correlation, two other performance measures (which are independent of operating point) were also calculated: (1) area under the receiver-operator characteristic curve (AUC-ROC) (Fawcett, 2005) and (2) area under the precision-recall curve (AUC-PR) (Davis and Goadrich,

2006). The final overall performance of the detector model is assessed using these three performance measures.

7.4 Combining multiple classification models to form an overall detection model

Several studies have improved classification accuracy by combining the output of several individual classifier models to increase predictive performance over that from a single model (Witten and Frank, 2000; Wolpert, 2002). Thus, instead of using the output of a single classification model, several classifier outputs are combined to arrive at a consensus. Bagging, boosting, and stacking (stacked generalization) are three extensively used methods in machine learning literature used to combine the output of multiple models.

In this study, the output of multiple individual microsleep detectors were combined using stacked generalization. Stacking overcomes a substantial problem with the voting procedures in bagging and boosting, which do not clarify a specific base model, i.e., the output of individual microsleep detectors to trust (Peiris *et al.,* 2011). Stacking maximizes ensemble learning by using a meta-learner process. The following discussion outlines how a successful stacking methodology has been extended in this current study.

Stacked generalization

The stacking framework depicted in Figure 7-1 consists of level-0 and level-1 generalizers. The level-0 models are formed by base classifiers which are trained using the input data and the target output. The level-0 outputs are then presented as an input to the level-1 generalizer (meta-learner) which is also trainable.

For the classification phase of the stacking system, new cases are generated for the level-0 models, each producing a classification value at their output. Then, the resulting base model predictions are passed to the level-1 model and combined linearly. The linear combination scales the output of each model according to its weight, adds the new scaled model outputs, and applies a threshold to the added model output to obtain an overall prediction.

In this stacking framework, some of the test data are held back and used to train the level-1 model, while the level-0 models are trained on the rest of the data. Only after all of the level-0 models are trained, the data that was held out is classified using the level-0 models, which then form the training data for the subsequent level-1 model. As the held-out data is not used to train the level-0 models, their predictions are unbiased and, therefore, the level-1 training data accurately reflects the true performance of the level-0 models.

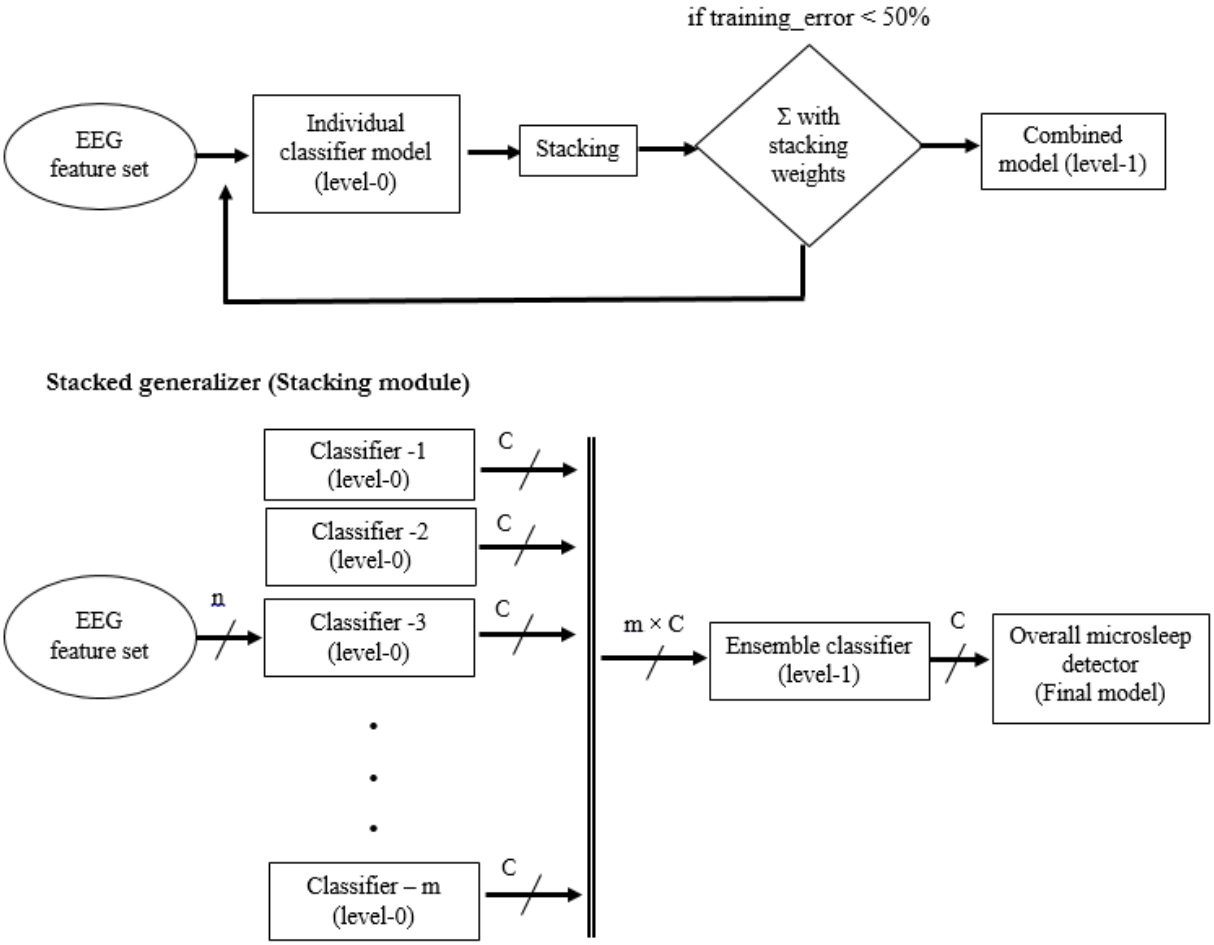


Figure 7-1 Diagram showing internal structure of the meta-learner used in the stacking framework. The outputs of the classifiers are denoted by C and *m* represents the number of level-0 models passed to the meta-learner before combining them to form an overall prediction.

However, this approach has a disadvantage as it deprives the level-1 models of some of the training data and, hence, this problem has to overcome by applying 8-fold cross-validation which ensures that all of the training data are used to train the level-1 model. Therefore, each instance of the training data is used in one test-fold of the cross-validation and predictions from the models built from the corresponding training fold are used to build the level-1 training set, thus generating a level-1 training set for each level-0 training set. This algorithm is similar in terms of its implementation to the stacking ensemble used in previous lapse detection studies (Peiris *et al.*, 2011; LaRocco, 2015)

The following steps were followed to create an overall microsleep meta-learner detection model from the stacking framework:

1. Reserve one of the 8 subjects for testing, and use the rest of the subjects for training.
2. Reserve one of the training subjects as pseudo-testing subject.

3. Train a single classifier model for each training subject.
4. Evaluate all classifiers on pseudo-testing subject.
5. Based on outputs from each classifier, use linear regression to weight each individual classifier.
6. Repeat steps (2) to (5) until all subjects are used as the pseudo-testing subject at least once.
7. Evaluate each configuration on the testing subject.
8. Evaluate the performance metrics (accuracy, sensitivity, specificity, selectivity and phi correlation).
9. Change the subject used for testing.
10. Exclude any data with training error greater than 50%.
11. Repeat steps (1) to (9) until each subject has been used at least once as the testing subject.
12. Average all performance metrics.

The stacking framework presented in Figure 7-1 was applied to the level-0 models of all the classifiers (LDA, SVM, cascaded-leaky-integrator ESN and LSMs) evaluated in this study, excluding the SNN and kNN classifiers (since these classifiers were not the focus of this study). Finally, the results obtained from the level-1 models of the stacking framework were also compared to the level-0 single classifier modules and are presented in the following sections.

7.5 Classifier models for microsleep detection

In this section, several classifier models including, LDA, SVM polynomial kernel (SVMP), k-nearest neighbour (kNN) classifier and SNNs were evaluated for microsleep detection on Study A dataset. In section 7.6, their performance is directly compared to the performance of the RC-based classifiers (ESNs and LSMs).

7.5.1 Linear discriminant analysis – detector performance

As discussed in §4.4, LDA is a simple statistical method for finding a linear combination of variables which best separates two or more classes (Ripley, 1996). Depending on which side of a boundary a new observation is assigned to, it is assigned a different group label. In itself,

LDA is not a classification algorithm, although it makes use of class labels. However, the LDA result is mostly used as part of a linear classifier.

LDA was used previously in microsleep detection (Peiris *et al.*, 2011, LaRocco, 2015) and provided the current benchmark for microsleep detection on Study A dataset. However, as was discussed in Chapter 6, several methods from the prior research, including LDA were re-evaluated to address a few inconsistencies and uncertainties raised during the pre-processing of the Study A dataset. Consequently, the results obtained from the re-implementation of the LDA-based classifiers were compared against the previous benchmarks established in Peiris *et al. (*2011).

Table 7-1 provides a summary of system performance for an LDA-based microsleep detector using PCA for feature reduction. The detector running a level-1 stacking framework provided the largest detector performance, as shown by the mean φ correlation at 0.41. These results are in agreement with the previous findings of Peiris *et al.* (2011) at φ = 0.39, and as stated §3.6.6, provided a benchmark for comparison of other classifier models used in this study. Approximately 280 PCs were used to achieve the optimal performance using LDA and stacking.

Table 7-1: Mean detector performances (sensitivity, specificity, selectivity, accuracy and φ) on the pruned EEG Study A dataset for LDA-based systems trained to detect microsleeps using 280 PCs.

Linear discriminants analysis Detector performance		
	Pruned EEG Study A	
Detector features	**Single classifier** (level-0 models)	**Stacked LDA modules** (level-1 models)
Sensitivity	0.64	0.73
Specificity	0.60	0.60
Selectivity	0.29	0.36
Accuracy	0.54	0.61
Mean Phi	0.31	0.41

The performance of a single detection model (level-0 model) created by grouping data of seven subjects together and validated using the remaining subject's data was investigated to compare its performance with the stacked approach. The highest performance in terms of φ was achieved by a single LDA classifier was reported at 0.31.

The effect of data pruning on LDA classifiers was also evaluated on Study A data. Pruning had a considerable impact on the LDA classifiers and, therefore, resulted in better detector performances. LDA on unpruned datasets, in comparison to the pruned dataset, suffered a loss in performance due to being plagued with noisy epochs. The highest mean phi reported

on the stacked LDA models for the unpruned Study A EEG was φ = 0.35, using 260 PCs. Table 7-2 provides an outline of the performance when various numbers of PCs are used to attain the highest mean phi on a LDA-based lapse detector.

Table 7-2: Mean detector performances (φ) on the variation of the number of PCs for the pruned and unpruned EEG Study A dataset.

Linear discriminants analysis detector performance		
No. of PCs	**Unpruned datasets** (level-1 models)	**Pruned datasets** (level-1 models)
150	0.33	0.39
200	0.34	0.39
260	0.35	0.40
280	0.34	0.41

7.5.2 k-Nearest Neighbour classifier – detector performance

As discussed in §4.5, the kNN algorithm implements learning based on the k-nearest neighbours of each query point. Since the neigbourhood approximation methods in Chapter 6 resulted in reasonable trustworthiness scores (Table 6-3), the kNN technique was evaluated to confirm if the Euclidean-distance-based classification could improve performance on microsleep detection. However, the best detector performance from the kNN algorithm was φ = 0.23, which was the lowest of all the algorithms investigated on Study A dataset. As a result, kNN detectors were not investigated in great depth.

Table 7-3: Mean detector performances (φ, AUC-ROC) on the pruned EEG Study A dataset for kNN classifier-based systems trained to detect microsleeps using PCA and PPCA feature sets.

k-Nearest Neighbour detector performance		
Pruned EEG datasets – Study A		
	Mean phi Mean±SE	**Mean AUC-ROC Mean±SE**
Principal components analysis (PCA)	0.23±0.05	0.64±0.04
Probabilistic PCA	0.20±0.06	0..60±0.04

Table 7-3 shows the detector performances of PCA and PPCA meta-features on the kNN classifier. Since both feature reduction algorithms resulted in poor detection rates, the rest of the feature reduction schemes were not implemented. Similarly, since the detection rates on the pruned datasets were lower, analysis on the unpruned datasets was not evaluated.

7.5.3 Spiking neural network – detector performance

An open source spiking neural network simulator, NEST[12], for Python was used to develop an SNN-based microsleep detector. Parameters for dynamics, size, and structure of neural systems were evaluated from the simulator before these were applied to the Study A dataset. The neuron model based on the Hodgkin-Huxley models (Hodgkin and Huxley, 1952) was applied to the SNNs. Synapse models using the short-term plasticity (Tsodyks & Markram, 1998) were evaluated on the Study A dataset. The current SNN-based lapse detector was constructed with 180 individual spiking neurons.

The application of SNN models to the pruned EEG dataset of Study A found peak detector performances with PCA and PPCA meta-features. In Table 7-4, detector performances are shown for all the feature reduction schemes used to train the SNN classifier. Despite its clear conceptual advantage over linear methods like LDA, the SNN failed to live up to its promise, as the highest φ achieved was 0.36. As the mean detection rates of the SNN models on pruned Study A dataset were below the previous benchmarks established using LDA (φ = 0.39), the performance of the unpruned datasets using SNN models was not evaluated.

Table 7-4: Mean detector performances (φ, AUC-ROC) on the pruned EEG Study A dataset for an SNN-based classifier systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Spiking neural network detector performance	
Pruned EEG datasets – Study A	
	Mean phi Mean±SE
Principal components analysis (PCA)	0.36±0.07
Probabilistic PCA	0.35±0.06
Kernel PCA	0.33±0.05
Nearest neighbour estimation (NNE)	0.28±0.06
Stochastic neighbourhood embedding (SNE)	0.34±0.05
Stochastic proximity embedding	0.33±0.06

7.5.4 Support vector machine

As described in Chapter 4, Support Vector Machines (SVMs) are a method of supervised learning based on the projection of a hyperplane (also called the decision border) into high dimensionality space (Ruping, 2001). The hyperplane divides the feature space into two

[12] http://www.nest-simulator.org/

classes and the classification result depends on which side of the hyperplane the class is located in. The primary goal of SVM classifier is to find such a hyperplane which divides the two classes, and, in addition, for which the distances between the hyperplane and the closest examples are maximized. Support vector machines can be transformed into a non-linear technique when the training algorithm associated with a non-linear kernel function is used.

A kernel function is a function that defines a new feature vector for a given sample by calculating the similarity between the sample *x* and another sample *y*. For any given data points $(\mathbf{x}_i, \mathbf{x}_j)$ the training algorithm is dependent upon the *kernel matrix K* and is represented by the equation.

$$K_{ij} = \mathbf{x}^T \mathbf{x}_j. \tag{7.1}$$

In the case of a non-linear mapping, the feature vectors $\mathbf{x}_i$ are replaced by some augmented feature vectors $\phi(\mathbf{x}_i)$, where, ϕ is a non-linear mapping function. Given this, the kernel function in the Equation 7-1 becomes

$$K(\mathbf{x}, \mathbf{z}) = \varphi(\mathbf{x}^{\mathrm{T}})\varphi(\mathbf{z}). \tag{7.2}$$

Polynomial kernel

Based on the above formulations, polynomial kernels usually take the form:

$$K(\mathbf{x}, \mathbf{z}) = (1 + \mathbf{x}^{\mathrm{T}}\mathbf{z})^d, \tag{7.3}$$

where, *d* is the parameter which determines the nature of the kernel. The case of $d = 1$ gives a linear kernel and the case of $d = 2$ gives a polynomial (quadratic) kernel.

In addition to the kernel functions, the current SVMP implementation also uses a soft-margin formulation which allows a few example classes to be misclassified in order for the SVM algorithm to increase its margins. This soft-margin formulation introduces a parameter *C*, which is used to control the classification rate. The parameter *C* takes the form of an optimization tool which can be used to control over-fitting within the data.

The library for support vector machines (LIBSVM) toolbox for C++, written by Chang and Lin (2011), was used to construct the current SVMP model suitable for the classification of microsleeps from the EEG on Study A.

SVMP – detector performance

Performance on both the pruned and unpruned Study A EEG dataset were evaluated for the SVM Polynomial-based microsleep detector. Detector performances are shown for all the feature reduction schemes which were used to train the SVMP classifier to obtain an overall prediction. Table 7-5 shows, the overall system performance of a SVMP-based lapse detector, in terms of its mean φ and AUC-ROC, on both the pruned and unpruned Study A EEG dataset. The highest performance for SVMP detectors was recorded at φ = 0.44 and AUC-ROC = 0.88 in both the PCA and PPCA feature reduction schemes on the pruned data. For the unpruned datasets a drop in performance was seen on both the PCA and PPCA models at φ = 0.40 and AUC-ROC = 0.86. All of the PCA-based feature reduction schemes outperformed the neighbourhood approximation schemes. The highest performance for a neighbourhood-based method was observed by the SNE technique, with φ = 0.35 on pruned Study A EEG dataset.

The SVMP classifier parameters *C* (soft margin constant which controls the cost of misclassification on the training data) and *d* (degree of polynomial kernel) were optimized using a grid search method (Hsu *et al.,* 2003), which tests the performance of all possible combinations from a given set of both parameters.

Table 7-5: Mean detector performances (φ, AUC-ROC and AC-PR) on the pruned and unpruned EEG Study A dataset for SVMP-based systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Support vector machine (polynomial) detector performance				
	Pruned EEG Study A		Unpruned EEG Study A	
Detector features	**Phi correlation Mean±SE**	**AUC-ROC Mean±SE**	**Phi correlation Mean±SE**	**AUC-ROC Mean±SE**
Principal components analysis (PCA)	0.44±0.05	0.88±0.04	0.40±0.05	0.84±0.04
Probabilistic PCA	0.44±0.06	0.88±0.03	0.40±0.06	0.85±0.04
Kernel PCA	0.38±0.05	0.84±0.03	0.33±0.05	0.77±0.03
Nearest neighbour estimation (NNE)	0.26±0.06	0.65±0.03	0.23±0.06	0.61±0.03
Stochastic neighbourhood embedding (SNE)	0.35±0.05	0.82±0.03	0.34±0.05	0.81±0.03
Stochastic proximity embedding (SPE)	0.34±0.06	0.82±0.04	0.34±0.06	0.81±0.04

7.6 Reservoir computing methods for microsleep detection

As discussed in Chapter 4, both the ESN and LSM classifiers were evaluated for microsleep detection. The ESN modules included analysis on the standard sigmoidal neuron-based ESNs and cascaded-leaky-integrator ESNs. Results of these classifiers are presented in this section.

7.6.1 Sigmoidal ESN – detector performance

Tables 7-6 and 7-7 provide a summary of system performance for microsleep detection based on a sigmoidal neuron-based ESN developed using 1000 neurons on the pruned and unpruned Study A EEG dataset, respectively.

Table 7-6: Mean detector performances (φ, AUC-ROC and AC-PR) on the pruned EEG study A dataset for sigmoidal ESN-based systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Detector features	Sigmoidal ESN detector performance		
	Pruned EEG study A dataset		
	Phi correlation (φ) Mean±SE (min, max)	AUC-ROC Mean±SE	AUC-PR Mean±SE
Principal components analysis (PCA)	0.31±0.01 (0.09,0.45)	0.75±0.02	0.38±0.06
Probabilistic PCA	0.30±0.01 (0.09,0.45)	0.75±0.03	0.39±0.07
Kernel PCA	0.26±0.02 (0.05,0.39)	0.70±0.03	0.31±0.07
Nearest neighbour estimation (NNE)	0.25±0.01 (0.06,0.40)	0.66±0.02	0.21±0.08
Stochastic neighbourhood embedding (SNE)	0.28±0.03(0.12, 0.42)	0.80±0.03	0.37±0.08
Stochastic proximity embedding (SPE)	0.27±0.02 (0.10,0.42)	0.78±0.04	0.35±0.05

Table 7-7: Mean detector performances (φ, AUC-ROC and AC-PR) on the unpruned EEG study A dataset for sigmoidal ESN-based systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Detector features	Sigmoidal ESN detector performance		
	Unpruned EEG study A dataset		
	Phi correlation (φ) Mean±SE (min, max)	AUC-ROC Mean±SE	AUC-PR Mean±SE
Principal components analysis (PCA)	0.30±0.07 (0.09,0.47)	0.74±0.03	0.38±0.09
Probabilistic PCA	0.30±0.08 (0.17,0.69)	0.75±0.04	0.39±0.09
Kernel PCA	0.25±0.06 (0.12,0.47)	0.72±0.03	0.32±0.09
Nearest neighbour estimation (NNE)	0.23±0.07 (0.06,0.40)	0.64±0.03	0.21±0.10
Stochastic neighbourhood embedding (SNE)	0.28±0.05(0.12, 0.52)	0.80±0.05	0.38±0.07
Stochastic proximity embedding (SPE)	0.25±0.06 (0.10,0.54)	0.77±0.03	0.35±0.06

Detector performances are shown for all feature reduction schemes used to train the sigmoidal ESN classifier to obtain an overall prediction. Detectors based on PCA, and probabilistic PCA (PPCA) provided the best generalization performances with φ~0.31 and φ~0.30, respectively on the pruned EEG datasets of the Study A.

7.6.2 Cascaded-leaky-integrator-ESN – detector performance

As discussed in Chapter 5, several global control parameters were required to be optimized to attain higher detection rates. Table 7-8 depicts the global control parameters used to obtain optimal microsleep detection performance.

Table 7-8 Optimal cascaded-leaky-integrator ESN global control parameters for microsleep detection.

Parameter	Value	Parameter	Value
Spectral radius	0.88	Leakage rate	0.05
Input scaling	0.2	No. of neurons	5-10
Bias Scaling	0.5		

Tables 7-9 and 7-10 provide a summary of system performance for a microsleep detector based on a cascaded-leaky-integrator ESN on the pruned and unpruned Study A EEG datasets, respectively. Detector performances are shown for all the feature reduction schemes used to train the cascaded-leaky-integrator ESN classifier to obtain an overall prediction. Detectors based on PCA, and probabilistic PCA (PPCA) provided the best generalization performances with φ~0.51 and φ~0.44, respectively on the pruned EEG datasets of the Study A.

The high performances observed in terms of φ were also confirmed by observing that PCA and PPCA based feature sets having the largest AUC-ROC and AUC-PR values as shown in Tables 7-9 and 7-10. As previously mentioned, AUC-ROC and AUC-PR, being threshold-independent measures, highlight these methods as providing the best performance of the 6 feature sets evaluated on cascaded-leaky-integrator ESN detectors. The highest AUC-ROC and AUC-PR values were seen on the PPCA-based meta-features at 0.91 and 0.47, followed by the PCA meta-features at 0.88 and 0.44, respectively.

Concerning the unpruned EEG datasets, the best detection performances in terms of φ, AUC-ROC, and AUC-PR were observed on the PPCA meta-features and these were reported to be 0.44, 0.88, and 0.45, respectively. PCA meta-features reported the next best values at 0.43, 0.86, and 0.41, respectively. The performance of a PCA-based sigmoidal-ESN model developed using 1000 neurons was also compared to the performance of the cascaded-leaky-integrator-ESN approach configured using only 8 neurons. The results from this comparison

are presented in Table 7-12. On the pruned datasets, sigmoidal ESNs achieved highest performance at φ = 0.31 compared to φ = 0.51 with the cascaded-leaky-integrator ESNs.

Table 7-9: Mean detector performances (φ, AUC-ROC and AC-PR) on the pruned EEG Study A dataset for cascaded-leaky-integrator ESN-based systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Detector features	Cascaded-leaky-integrator ESN detector performance		
	Pruned EEG study A dataset		
	Phi correlation (φ) Mean±SE (min, max)	AUC-ROC Mean±SE	AUC-PR Mean±SE
Principal components analysis (PCA)	0.51±0.07 (0.15,0.65)	0.88±0.03	0.44±0.09
Probabilistic PCA	0.47±0.08 (0.17,0.69)	0.91±0.04	0.47±0.09
Kernel PCA	0.36±0.06 (0.12,0.47)	0.82±0.03	0.39±0.09
Nearest neighbour estimation (NNE)	0.29±0.07 (0.06,0.40)	0.69±0.03	0.23±0.10
Stochastic neighbourhood embedding (SNE)	0.38±0.05(0.12, 0.52)	0.85±0.05	0.41±0.07
Stochastic proximity embedding (SPE)	0.37±0.06 (0.10,0.54)	0.81±0.03	0.40±0.06

Table 7-10: Mean detector performances (φ, AUC-ROC and AC-PR) on the unpruned EEG Study A dataset for cascaded-leaky-integrator ESN-based systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Detector features	Cascaded-leaky-integrator ESN detector performance		
	Unpruned EEG study A dataset		
	Phi correlation (φ) Mean±SE (min, max)	AUC-ROC Mean±SE	AUC-PR Mean±SE
Principal components analysis (PCA)	0.43±0.09 (0.13,0.62)	0.86±0.03	0.41±0.09
Probabilistic PCA	0.44±0.06 (0.14,0.62)	0.88±0.04	0.45±0.09
Kernel PCA	0.32±0.07 (0.10,0.42)	0.81±0.03	0.39±0.09
Nearest neighbour estimation (NNE)	0.27±0.08 (0.06,0.38)	0.69±0.03	0.20±0.10
Stochastic neighbourhood embedding (SNE)	0.38±0.05 (0.12, 0.52)	0.85±0.05	0.41±0.07
Stochastic proximity embedding (SPE)	0.37±0.04 (0.10,0.54)	0.81±0.03	0.40±0.06

Table 7-11 illustrates the effect that global control parameters have on a cascaded-leaky-integrator ESN's performance. Larger spectral radius values generally gave a higher detection performance. The optimal performance in terms of φ~0.51 was attained with a spectral radius of 0.88. This, in accordance with the discussion in §6.3, suggests that the current microsleep detection problem is indeed non-linear in nature. Lower leakage rates ~0.05 were found to have the best detection performances with φ values at 0.51.

Table 7-11: Mean detector performances (φ) on variation of the global control parameters (spectral radius, leakage rate and number of neurons) of the cascaded-leaky-integrator ESN classifier on the pruned Study A dataset using 50-60 PCs from the PCA.

Performance of global control parameters on cascaded-leaky-integrator ESN detectors (φ)					
Spectral radius	**Mean Phi**	**Leakage rate**	**Mean Phi**	**No. of neurons**	**Mean Phi**
0.60	0.44	0.03	0.47	3	0.38
0.70	0.47	0.05	0.51	4	0.45
0.80	0.49	0.10	0.49	5	0.48
0.88	0.51	0.20	0.47	8	0.51
0.90	0.50	0.30	0.44	10	0.49
1.0	0.45	0.40	0.42	12	0.48

Table 7-12: Mean detector performances (sensitivity, selectivity, accuracy and φ) on the pruned EEG Study A dataset for sigmoidal ESN and cascaded-leaky-integrator-ESN-based systems trained to detect lapses using 50-60 PCs from the PCA.

Sigmoidal ESN and cascaded-leaky-integrator-ESN detector performance		
	Pruned EEG study A	
Detector features	**Sigmoidal ESN** (level-1 models)	**Cascaded-leaky ESN** (level-1 models)
Sensitivity	0.62	0.86
Selectivity	0.28	0.34
Accuracy	0.41	0.71
Mean Phi	0.31	0.51

Closer analysis showed a few monotonic falloffs in the performance (φ) either side of $p = 8$ (p is the number of neurons). Consequently, optimal performance was observed when the number of neurons was set to 8. All of the findings in Table 7-12 were a result of rigorous fine-tuning of the cascaded-leaky-integrator network heuristics based on the PCA feature sets of the pruned Study A dataset.

Figure 7-2 depicts the fluctuation in performance given the number of PCs in terms of mean phi correlation, AUC-ROC, and AUC-PR for a cascaded-leaky-integrator ESN-based microsleep detector. The results show that detector performance for PCA feature sets reached a saturation point after approximately 50 PCs and that adding additional PCs to the model caused over-fitting and reduced the overall performance.

The results shown in Figure 7-2, however, are not in complete agreement with other feature reduction models, as some of the other models needed more than 50 PCs to reach a saturation point. Therefore, it was decided not to include all of the PCs in the subsequent

analysis for the construction of the microsleep detector models as this avoids the added complication of determining the number of PCs to be used for model formation.

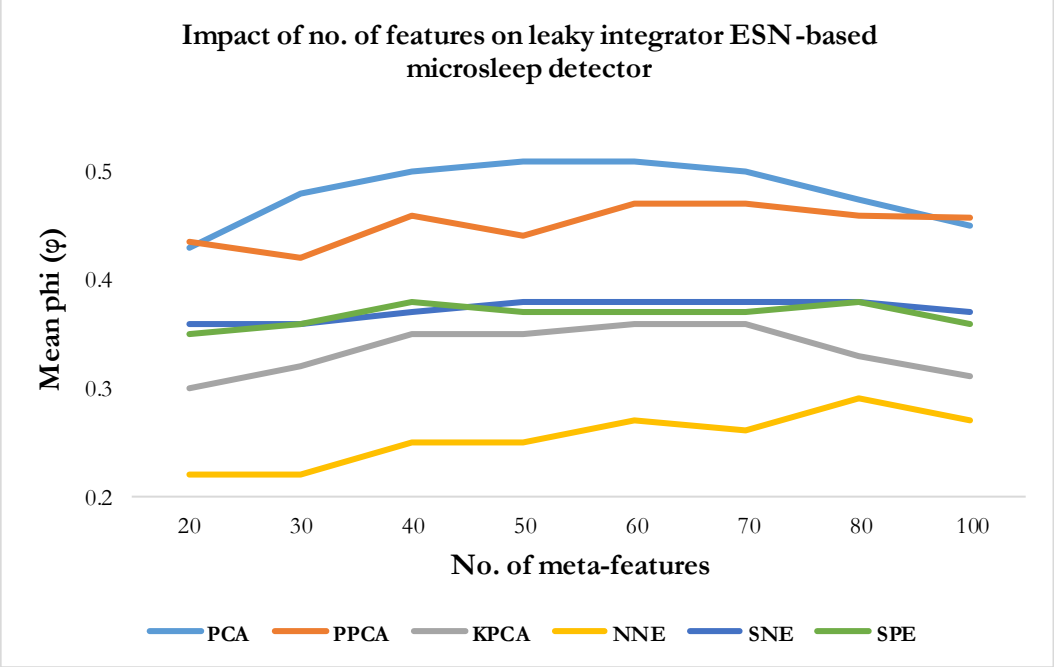


Figure 7-2 Performance analysis on variation of number of meta-features in relation to lapse detection performance (measured in terms of mean phi). Feature reduction modules evaluated were, PCA, PPCA, KPCA, NNE, SNE, and SPE.

7.6.3 Liquid state machines – detector performance

Lapse detection performances on the pruned and unpruned Study A EEG dataset were evaluated for the LSM microsleep detector. Detector performances are shown for all of the feature reduction schemes used to train the LSM classifier. Tables 7-13 provides an overall summary of system performance for a microsleep detector based on LSM on both the pruned and unpruned Study A EEG dataset. On the pruned EEG datasets, detectors trained using the probabilistic PCA meta-features reported the highest detection scores in terms of both φ and AUC-ROC at 0.42 and 0.84, and this performance was only slightly better than the PCA models at 0.40 and 0.82, respectively.

Furthermore, concerning the unpruned EEG datasets, the highest detector scores were also achieved using PPCA meta-features, at φ = 0.41, AUC-ROC = 0.84 and φ = 0.38, AUC-ROC = 0.82, respectively, followed by the PCA meta-features. Similar to the cascaded-leaky-integrator ESN approach, several internal parameters (control parameters) needed to be optimized for the LSM classifiers. These parameters included, number of neurons in the reservoir, leaking rate, and action potential. Most of the values for these control parameters

of the LSM classifier were very similar to the ESNs, apart from the number of neurons which were set in the range of 100 – 120.

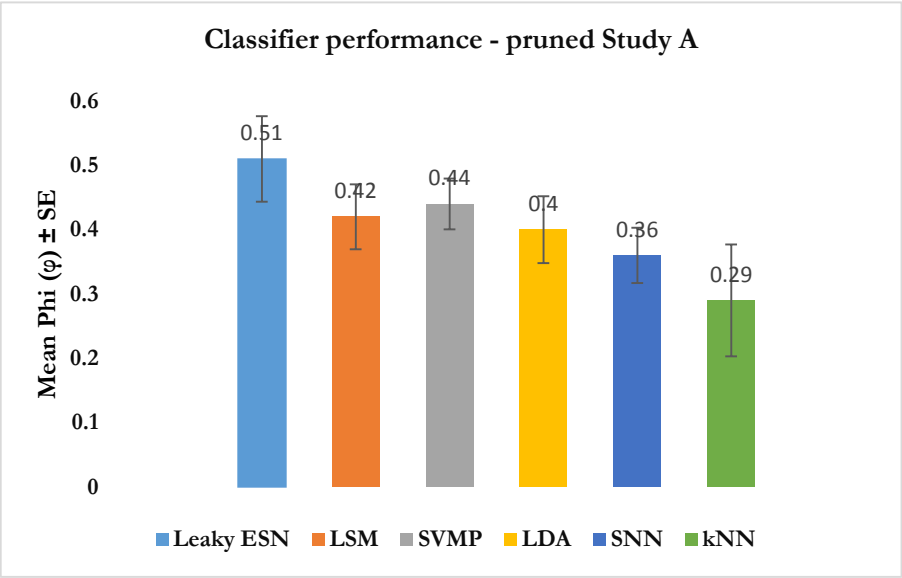
Table 7-13: Mean detector performances (φ and AUC-ROC) on the pruned and unpruned EEG Study A dataset for LSM-based systems trained to detect microsleeps using PCA, PPCA, KPCA, NNE, SNE, and SPE feature sets.

Liquid state machine detector performance				
	Pruned EEG Study A		Unpruned EEG Study A	
Detector features	Phi correlation Mean±SE	AUC-ROC Mean±SE	Phi correlation Mean±SE	AUC-ROC Mean±SE
Principal components analysis (PCA)	0.40±0.05	0.83±0.03	0.38±0.05	0.82±0.04
Probabilistic PCA	0.42±0.06	0.84±0.04	0.41±0.06	0.84±0.04
Kernel PCA	0.33±0.05	0.79±0.03	0.30±0.05	0.77±0.03
Nearest neighbour estimation (NNE)	0.29±0.06	0.69±0.03	0.29±0.06	0.69±0.03
Stochastic neighbourhood embedding (SNE)	0.38±0.05	0.82±0.03	0.37±0.05	0.81±0.03
Stochastic proximity embedding (SPE)	0.35±0.06	0.81±0.04	0.35±0.06	0.81±0.04

7.7 Overall lapse detection model – performance

Figure 7-3 provides an overall summary of system performances of the six classifier schemes evaluated for the lapse detection on Study A, on both pruned and unpruned datasets. One aim of this study was to show the effect that *data pruning* has on classification accuracy. Consequently, a comparative analysis for the performance is provided between the pruned and unpruned datasets for the six classifier modules. Only the best detection performances in terms of φ are depicted for all the classifiers evaluated on the pruned EEG Study A dataset. Additionally, these results also only depict the performance from the feature reduction schemes which contributed to the maximum φ values.

Analysis of these results indicate that the detection performance of the cascaded-leaky-integrator ESN classifiers was substantially higher than any other classifier schemes evaluated for microsleep detection (φ = 0.51). Furthermore, the cascaded-leaky-integrator ESNs also had the highest φ on the unpruned datasets (φ = 0.44). The second best classifier to have the high φ values on both the pruned and unpruned datasets was SVMP (φ = 0.44 and φ = 0.40, on the pruned and unpruned datasets, respectively). This effect can also be seen in the AUC-ROC and AUC-PR analysis. Classifier modules including, cascaded-leaky-ESN, LSM, SVMP, LDA, SNN, and kNN are depicted on the X-axis and their corresponding maximum φ values are depicted on the Y-axis.



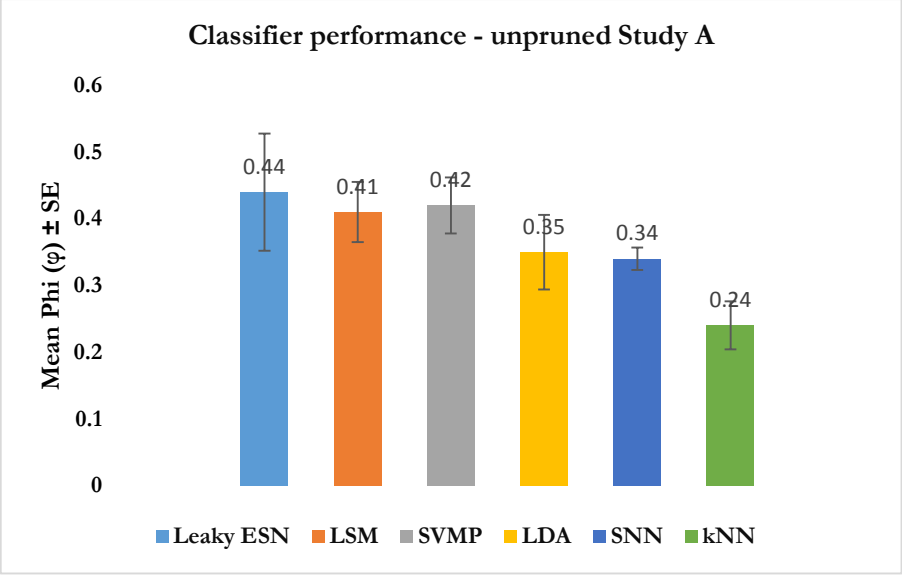


Figure 7-3 Comparison of detector performances on pruned and unpruned datasets across all classifiers and for maximal feature reduction in each case. Standard error is depicted by the vertical bars on top of the classifier results for both graphs.

7.7.1 Within-subject performance trends

The results presented in §7-5 and §7-6, were averaged from a range of individual performance values for each of the 8 subjects. From the initial observations, it was found that certain subjects had consistently low detection values across multiple system configurations (six classifiers and six feature reduction schemes). Even across the Study A feature sets, certain subjects consistently scored higher or lower than the average phi value. Understanding the reasons for successful individual subject classification was thought to provide insights into improving the overall detection performance. Consequently, 8-fold

cross-validation was used to validate the classifier models on individual subjects. In this type of cross-validation scheme, the data set was divided into 8 subsets and the holdout method was repeated 8 times. Each time, one of the 8 subsets was used as the test set and the rest of the subsets were put aside to form a training set. Finally, the average error across all 8 trials was computed.

Figure 7-4 provides a subject-wise summary of system performance of stacked-LDA classifier using PCA for the lapse detection on Study A pruned EEG dataset. As discussed earlier, one of the aims of this study was to show the effect of within-subject classifier variances on the overall lapse detector. The correlation values for each of the 8 subjects are depicted in terms of phi. The highest phi values across all of the subjects on the pruned Study A dataset were seen in Subject 804 ($\varphi = 0.78$), while the lowest phi values were seen on Subject 810 ($\varphi = 0.09$). Overall, there were 5 subjects (804, 809, 814, 817, and 820) who consistently scored higher phi values than the mean phi of the overall lapse detector. The detection performance on the rest of the 3 subjects (810, 811, and, 819) was found to sub-optimal across all the classifier modules evaluated.

The ROC analysis presented few interesting results as several of the classifier modules with low phi values have reported reasonably higher AUC-ROC values. These trends can be attributed to several other factors such as the number of true positive and true negative labels computed on each of these subjects. For instance, although demonstrating the lowest phi values across all the subjects, Subject 810 reported higher true positive rates (TPR = 0.85) than subjects such as 817, 819 and 820 (TPR = 0.30, 0.47 and, 0.71, respectively). Overall, the highest ROC values were reported in Subject 814 (AUC-ROC = 0.92) and the lowest in Subject 819 (AUC-ROC = 0.63).

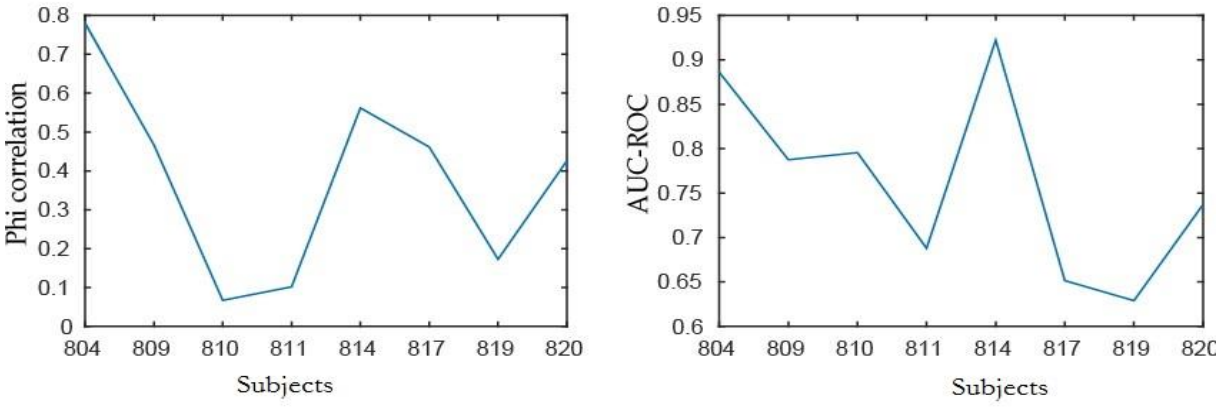


Figure 7-4 Comparison of subject-wise mean detector performances on pruned and unpruned datasets for all of the classifiers implemented. On the left hand side are the microsleep detector performances on the pruned datasets using phi correlation as a metric and on the right hand are the detection performances for the pruned datasets using AUC-ROC as a metric. .

Table 7-14: Subject-wise detector performances (φ) on the pruned Study A dataset for stacked-LDA, SVMP and cascaded-leaky-integrator ESN based systems trained to detect lapses using PCA meta-features.

	Subject-wise detector performances		
	Pruned EEG study A dataset		
Subject	LDA Phi (φ)	SVMP Phi (φ)	Cascaded-leaky ESN Phi (φ)
804	0.78	0.81	0.84
809	0.48	0.51	0.50
810	0.07	0.07	0.09
811	0.11	0.15	0.13
814	0.52	0.62	0.67
817	0.48	0.55	0.62
819	0.19	0.24	0.21
820	0.41	0.49	0.54

Table 7-14 presents the subject-wise performance analysis of three different types of classifier systems trained to detect lapses using PCA meta-features (PCs). Results suggest that the Subject 804 was the overall best performing subject across all the classifier systems with a highest φ = 0.84, gained using PCA-based cascaded-leaky-integrator ESNs. Subject 810 had consistently lower phi values amongst all the other subjects evaluated, across all the classifier modules. The lowest performing subject overall was identified to be Subject 810, with φ = 0.09 using the stacked LDA approach.

The reasons for the consistently lower performances on certain subjects such as Subject 810 is attributed to the presence of a potentially lower number of training labels (i.e., lapses) within the data from that subject. For example, as depicted in Table 6-2, the total number of lapses and definite microsleeps identified in Subject 810 were less than 0.001% and 0.008% (8 and 64 epochs, respectively) in the total 7198 epochs. These lower numbers of training labels may not be sufficient for the classifier models to provide a reliable detection of lapses. As a result, when these subjects are used to train the classifier models in the LOOCV analysis, the overall performance of the classifier systems may be affected.

7.7.2 Lapses versus video microsleeps, definite microsleeps, and flatspots

In addition to lapses-of-responsiveness states, other gold-standard metrics such as definite microsleeps, flatspots, and video microsleeps were also evaluated on the Study A dataset. Table 7-15 reports the performance analysis of the stacked LDA, SVMP, and cascaded-leaky-integrator ESN classifiers trained using PCA meta-features on both the pruned and unpruned Study A EEG dataset. On the Study A EEG dataset, the best detection

performances with respect to the gold-standard labels evaluated were, in decreasing order: (*i*) lapses, (*ii*) video microsleeps, (*iii*) definite microsleeps and, (*iv*) flatspots.

Table 7-15: Subject-wise detector performances (φ) on the pruned Study A dataset for stacked-LDA, SVMP and cascaded-leaky-integrator ESN-based systems trained to detect lapses using PCA or PCA meta-features.

Detector performances on different gold standards				
	Pruned EEG study A dataset			
Classifier	**Flatspots** Phi (Mean±SE)	**Video microsleeps** Phi (Mean±SE)	**Definite microsleeps** Phi (Mean±SE)	**Lapses** Phi (Mean±SE)
Stacked LDA	0.31±0.06	0.40±0.03	0.34±0.04	0.41±0.06
SVMP	0.33±0.04	0.42±0.05	0.38±0.03	0.44±0.06
Cascaded-leaky ESN	0.33±0.08	0.49±0.06	0.43±0.06	0.51±0.07
Unpruned EEG study A dataset				
Stacked LDA	0.28 ±0.05	0.36±0.04	0.30±0.06	0.35±0.05
SVMP	0.31±0.04	0.40±0.02	0.34±0.04	0.40±0.03
Cascaded-leaky ESN	0.32±0.07	0.43±0.05	0.38±0.05	0.43±0.06

The lapse-in-responsiveness states had the best detection rates across all the classifier modules and gold standards evaluated on both the pruned and unpruned Study A EEG dataset. The only exception to this case, was the stacked-LDA classifier which indicated a marginally higher mean phi value on the video-based microsleeps $\varphi = 0.36\pm0.06$ compared to $\varphi = 0.35\pm0.05$ on lapses. Video microsleeps had the second best mean detection rates across all the classifiers on both the pruned and unpruned datasets. All of the classifier modules performed least on flatspots for both datasets (pruned and unpruned).

In addition, cascaded-leaky-integrator ESN modules reported the highest mean phi values across all of the gold-standards evaluated on both the pruned and unpruned Study A EEG dataset. On the pruned EEG datasets, the highest and lowest mean performances on the definite microsleeps were identified as $\varphi = 0.43\pm0.06$ and $\varphi = 0.34\pm0.04$ for cascaded-leaky integrator ESNs and stacked LDA classifiers, respectively. These values can be compared to the highest and lowest mean $\varphi = 0.51\pm0.07$ and $\varphi = 0.41\pm0.06$ for lapses.

As expected, a drop in performance was visible across all of the gold-standards for the unpruned datasets. In addition, trends similar to the pruned datasets were observed on all the gold-standards. On the unpruned datasets, cascaded-leaky integrator ESNs and stacked-LDA classifiers reported the highest and lowest mean performances on definite microsleeps at $\varphi = 0.38\pm0.05$ and $\varphi = 0.30\pm0.04$, respectively, compared to $\varphi = 0.38\pm0.05$ and $\varphi = 0.30\pm0.04$ for lapses.

These results suggest that the labels containing video microsleeps can provide more reliable and accurate detection rates across all the classifier schemes. In addition, flatspots alone are simply not enough to train the classifier modules capable of detecting the lapses. Definite microsleeps, despite containing attributes of both the tracking (flatspots) and video microsleeps, did not perform well, compared to lapses (tracking or video microsleeps).

A possible reason may be that the rate of video microsleeps was considerably higher than that for definite microsleeps (1.97 times), which appears to be primarily because microsleeps cause events other than flat spots in tracking, such as the response cursor drifting away from the target. Lapses on the contrary, include either features from tracking or video microsleeps and contain labels which give a true determination of location of the microsleeps, leading to accurate lapse detection models.

In addition, another reason why lapse detection was more accurate compared to other gold-standards may be because: (1) most of the lapses were in fact microsleeps and (2) therefore, there were substantial number of missed microsleeps in the 'definite microsleeps' which were labelled alert/ responsive which confused and subsequently degraded the performance of the trained classifier.

7.8 Overall trends

7.8.1 Single classifiers vs stacking

The advantage of creating multiple classifier models and the resulting increase in the classifier performances was revealed in §7.6 and §7.7 All of the classifiers evaluated (LDA, cascaded-leaky-integrator ESN, LSM, SVMP), apart from SNNs and kNNs, used stacking framework and showed a substantial increase in phi values compared to their base (level-0) models. Stacked LDA modules reported the second greatest increase in detector performance after cascaded-leaky ESNs, with mean phi correlation of 0.41 compared to mean phi of 0.31 on the single classifier modules.

Using eight level-0 models and a meta-learner, the cascaded-leaky-integrator ESN model resulted in a mean phi correlation of 0.51 for the PCA detector, whereas lumping all the features from seven subjects to create a single model and validating the lumped model on the eighth subject resulted in a mean phi of 0.38. A reason for the performance of the cascaded-leaky-integrator ESNs to be higher than any other model may be because the leaky ESN approach was in itself built on a multiple classifier scheme (similar to mini-bagging) depicted in Figure 5-3. The individual classifier output from the cascaded-leaky-integrator structures (Figure 5-3) was later applied to the stacking framework (Figure 7-1) making the

overall classifier model robust. Therefore, the predictive outputs of the cascaded-leaky ESN models are much stronger due to rigorous learning that occurs due these meta-learner models.

Generally, lumping of data prior to model formation resulted in the model being biased towards certain subjects, particularly the ones with the most lapses, resulting in a loss of generalization ability. Therefore, to prevent such bias, a stacking approach was applied. The stacking approach resulted in better performance as the level-0 models were adjusted by the meta-learner according to how well they generalized over the training set. It was also expected that the mean detector performance would increase proportionally with the size of the training set.

7.8.2 Feature reduction

An analysis of the number of features required to form an optimal lapse detector revealed that non-linear classifiers, such as the cascaded-leaky-integrator ESN, SVMP and, LSM require only 50-60 meta-features to attain maximal detection rates. Whereas, the linear approaches, like the stacked LDA, require substantially higher numbers of meta-features, around 260-280, to attain maximal detection rates.

Investigation into the meta-features that contributed most to the lapse detector (which showed the highest level of performance out of all the detectors) revealed that the PCA meta-features contributed the most to the overall detection model. This result fits with previous trustworthiness scores (Table 6-3), and has demonstrated a correlation between the trustworthiness scores and the meta-feature performance. It was surprising to find that that PPCA approach, shown to have one of the highest trustworthiness scores, could not perform better than PCA on lapse detection.

On the whole, the neighbourhood approximation methods (feature reduction modules including NNE, SNE and SPE) were found to generate poor models for microsleep detection with the lowest performance on Study A. This happened despite good trustworthiness scores.

A possible explanation for the low performance of these models could be associated with the neighbourhood methods, which were unable to separate between the classes (microsleeps and alert state) from the Euclidean distance approximations.

The clustering pattern, as evident from the Figure 6-3, and the subsequent feature reduction visualizations in Chapter 6, where the data belonging to both the classes (microsleep and alert state) was clustered together, showed the true complexity of these EEG datasets. As is

evident in the Figures 6-5 – 6-17, features from both the classes were projected extremely close to one another making a linear or non-linear separation between them extremely difficult.

7.8.3 Pruned vs unpruned datasets

As mentioned earlier in Chapter 6, following the noise filtering and ICA-based eye blink removal, EEG epochs with z-scores >30 were excluded as artefacts, and were not used for training and testing the lapse detection models. This process was referred to as data pruning or data pruning. Over 8.5% of the epochs were excluded on an average per subject. In comparison, performances on the unpruned technique, were on complete datasets without any epochs removed.

It was established from the results presented in §7.5 and §7.6, data pruning had a substantial impact on the detection performance of all the classifier modules. Additionally, the unpruned datasets were found to have biased the detector towards a lower level of performance. Consequently, the elimination of noisy epochs caused the system to display a considerably higher level of performance

7.9 Discussion

Chapter 5 described how the classifier models involving reservoir computing techniques were used to create detection models capable of detecting both lapses and microsleeps from the EEG. Firstly, an overview of the ESN classifiers using both sigmoidal activation and cascaded-leaky-integrator neurons was provided. This was followed by defining the *global control parameters* used to control the internal state dynamics of the system and accurately model the target output. Next, the internal structure of the cascaded-leaky-integrator-ESN classifier used for the microsleep detection was explained in significant detail. Then, the algorithm behind the implementation of the LSM classifiers was explained. Following this, implementations behind other classifier models such as, LDA, SVMP, SNN, and kNN which were used for microsleep detection was presented.

This chapter presented the concept behind an overall microsleep model creation and ensemble learning. In addition, the performance metrics used to evaluate the classifier systems were defined. Finally, results for the various microsleep detector models created using the meta-features from the feature reduction algorithms discussed in Chapter 6 (Table 6-3) were presented and the performance of the different models compared.

7.9.1 Study A performance – reservoir computing

Results were presented for microsleep detectors which were trained using the meta-features from both linear techniques, non-linear techniques, and combinations thereof for the pruned and unpruned Study A EEG dataset. An analysis of the results showed that the best detector performance (in terms of the highest mean φ) was achieved using the detector model created using cascaded-leaky-integrator ESN models with 50 PCs from the PCA for pruned data. This model showed a mean correlation of $\varphi = 0.51 \pm 0.07$ (AUC-ROC = 0.91 ± 0.04; AUC-PR = 0.47 ± 0.09). This result was the best result achieved in this project. The nonparametric Wilcoxon signed-rank test also confirmed that microsleep detection by the cascaded-leaky-integrator ESN was statistically superior to that of the previous LDA-based approach (Peiris, *et al.*, 2011) on the pruned Study A dataset (Phi = 0.51 vs. 0.39, z = 2.54, p = 0.012 (2-tail)).

Cascaded-leaky-integrator ESN classification models created using the meta-features from the neighbourhood approximation methods (NNE, SPE, SNE) had lower mean performances than the PCA-based meta-features ($\varphi = 0.51 \pm 0.05$ for PCA and $\varphi = 0.47 \pm 0.08$ for PPCA). The performance of the detectors showed a marginal decrease when more meta-features (PCs) were added to the classifier (see Figure 7-2) suggesting a possible over-fitting within the model. For the majority of the feature reduction algorithms the optimal number of meta-features required to attain a max φ were found in the range of 50–60.

The effect of *global control parameters* was also presented in Table 7-11. The optimal parameters used to attain the best microsleep detection rate (mean $\varphi \approx 0.51$) were a spectral radius of 0.88, a total number of neurons maintained at 8 and a leaking rate of 0.05 (see Table 7-11). In terms of the cascaded-leaky-integrator ESN classifiers, probabilistic PCA was found to be the only feature reduction algorithm with mean φ values marginally close to the PCA-based detectors. The lowest performance using the cascaded-leaky-integrator ESNs was seen on the unpruned datasets using the NNE-based meta-features with $\varphi = 0.27 \pm 0.07$, followed by kernel PCA, SPE and SNE at $\varphi = 0.32 \pm 0.07$, $\varphi = 0.38 \pm 0.05$, $\varphi = 0.37 \pm 0.04$, respectively. Nevertheless, the results in Tables 7-9 to 7-10 demonstrate the dominance of cascaded-leaky-integrator ESN classifiers, where simple, 8-leaky neuron ESN structures were uniquely configured to outperform their counterparts that used 1000 sigmoidal neurons.

The LSM detectors showed the third highest performance overall ($\varphi = 0.42 \pm 0.06$; AUC-ROC = 0.84 ± 0.04; AUC-PR = 0.41 ± 0.09). Interestingly, however, LSMs were the only classifier modules evaluated to have a better performance with PPCA meta-features, followed by the PCA-based LSM detectors, followed by the LSM detector based on SNE.

Table 7-16 provides the detector performances with respect to the sensitivity and specificity values of the microsleep detectors on the pruned Study A dataset. Results from the Table 7-

16 indicate that the classifiers such as cascaded-leaky-integrator ESN-based and SVMP were moderately sensitive and highly specific compared to LDA, which demonstrated a higher sensitivity and lower specificity. The LSM models reported the lowest specificity and highest sensitivity compared to the other classifier modules.

The results from the Table 7-16 imply that even the consistently performing classifiers, such as cascaded-leaky-integrator ESNs, and SVMP have a greater susceptibility to missing microsleep events, whereas the classifiers such as LDA and LSM are more vulnerable to false positives. The consistently performing cascaded-leaky-integrator ESNs also exhibited relatively poor precision (PR = 0.47; Figure 7-5), particularly for those subjects who lapsed only a few times during their two hours. Low precision would often be tolerable in a microsleep detection system, as false alarms have low cost and are preferable to missed lapses.

Table 7-16: Mean detector performances (sensitivity and specificity) on the pruned EEG Study A dataset for cascaded-leaky-integrator ESN-based and LSM-based systems.

Detector performances				
	Pruned EEG Study A			
Detector features	**Cascaded-leaky-integrator ESN**	**LSM**	**SVMP**	**LDA**
Sensitivity	0.63	0.75	0.68	0.73
Specificity	0.87	0.56	0.89	0.60

On the unpruned datasets, the best detector performance was also achieved by the cascaded-leaky-integrator ESN modules. However, this was achieved using the PPCA meta-features instead. The PPCA method also had the highest performance achieved on the unpruned datasets in comparison to any other ML method evaluated for microsleep detection ($\varphi = 0.44 \pm 0.06$; AUC-ROC = 0.88 ± 0.04; AUC-PR = 0.45 ± 0.09). Nevertheless, the PCA-based cascaded-leaky ESN detector achieved φ values marginally closer to the PPCA method ($\varphi = 0.43 \pm 0.06$) and was the second best method overall on the unpruned datasets. Figure 7-5 presents the results from the AUC-ROC and AUC-PR analysis of PPCA-based cascaded-leaky-integrator ESN on pruned Study A data.

Another interesting observation was the drop in the performance from pruned data ($\varphi = 0.51$) by the PCA-based cascaded-leaky-integrator ESNs to the unpruned data ($\varphi = 0.44$). This represented the largest drop in performance compared with any other method. This, in turn, suggests that the cascaded-leaky ESN-based classifiers, albeit being effective in terms of modelling the changes in low frequency signals, are extremely prone to over-fitting in the presence of noisy labels/epochs. This can also be verified by the drop in the values of AUC-ROC and AUC-PR for these methods (see Table 7-10).

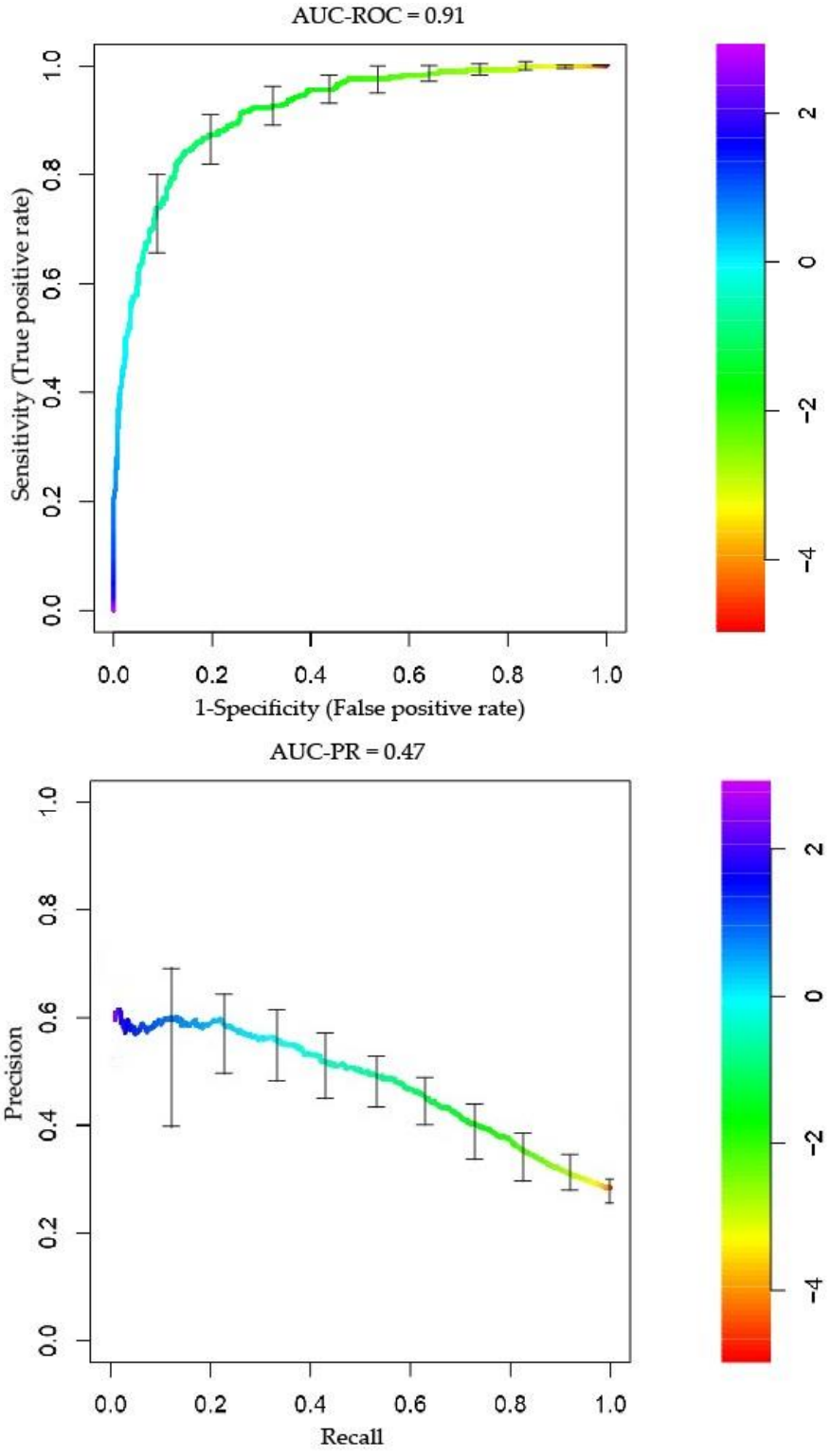


Figure 7-5 Graph on the top illustrates the ROC analysis in terms of false positive rate and sensitivity (plotted on the X and Y-axes respectively), and the graph on the bottom indicates the AUC-PR analysis precision and recall (plotted on the X and Y-axes respectively), for the PPCA-based cascaded-leaky-integrator ESNs. The corresponding AUC values are measured at 0.91. The vertical bars indicate standard error on both graphs. The colour scale on the right hand side of both the graph gives an indication, as to which classification threshold results in a certain point on the curve, i.e., a certain pair of sensitivity and false positive rate or a certain pair of precision or recall.

Of course, the phenomenon of over-fitting is not necessarily solely associated with RC. It is thought to be because the LSM models have shown the least drop in performance on the unpruned datasets, despite being filled with noisy EEG epochs and artefacts. The unpruned datasets only contributed to a marginal drop in φ of 0.42 to 0.41. PCA was a close second on this dataset with $\varphi = 0.38$.

In essence, the RC method using cascaded-leaky-integrator ESN has proven to be the best classification technique implemented on microsleep detection on the Study A dataset. In addition, the cascaded-leaky-integrator ESN method developed as part of this research has shown to produce the best results by creating a new benchmark for performance on Study A EEG data: $\varphi \approx 0.51 \pm 0.07$, AUC-ROC = 0.91 ± 0.04, AUC-PR = 0.47 ± 0.09, on the pruned datasets and $\varphi = 0.44 \pm 0.06$, AUC-ROC = 0.88 ± 0.04, AUC-PR = 0.45 ± 0.09 on the unpruned datasets.

7.9.2 Study A performance – linear discriminant analysis

Analysis on the stacked LDA modules was undertaken to reproduce the previous benchmark, resulting in phi value ($\varphi = 0.41$) marginally higher than those previously reported in Peiris *et al.* (2011). This was the result of rigorous testing of the classifier methods. The only substantial difference found in terms of the LDA implementation was the number of PCs used for the analysis by Peiris *et al.* (2011). They reported that 50 PCs were used to train the stacked LDA classifiers. However, the current implementation included an exhaustive search to find the optimal features used for the LDA based microsleep detection. This analysis reported that as the number of PCs increased (280-290), the performance of the stacked LDA classifier increased resulting in a $\varphi = 0.41$.

Comparison between the single classifier and level-1 meta-learners of the LDA demonstrated the dominance of ensemble learning techniques. Another interesting observation from the LDA analysis was the drop in performance on the unpruned EEG data. The φ values dropped substantially on the unpruned datasets ($\varphi = 0.41$ to 0.35).

7.9.3 Study A performance – support vector machine

The performances of the SVMP-based classifiers were found to be second best in terms of overall microsleep classification after the cascaded-leaky-integrator ESNs, on both the pruned and unpruned datasets with mean of $\varphi = 0.44 \pm 0.06$ and $\varphi = 0.40 \pm 0.05$, respectively This detection rate of the SVMP was attained using both the PCA and PPCA meta-features. SVMP was the only classifier technique which showed a similar lapse detection performance for both the PCA and PPCA feature reduction techniques.

In addition, SVMP was the only technique to show similar performance on both the pruned and unpruned EEG data, irrespective of whether PCA or PPCA-based meta-features were used. Similar to the cascaded-leaky-integrator ESNs, the lowest performance using the SVMP-based methods was seen on the unpruned datasets using the NNE-based meta-features with $\varphi = 0.25 \pm 0.07$, followed by kernel PCA, SPE, and SNE at $\varphi = 0.35 \pm 0.07$, $\varphi = 0.38 \pm 0.05$, $\varphi = 0.33 \pm 0.04$, respectively.

Overall, the SVMP implementation was ranked second after the cascaded-leaky-integrator ESN-based classifiers for microsleep detection on Study A. The LSM approach follows the SVMP at #3, in performance ranking for Study A microsleep detection, followed by the stacked LDA classifiers at #4.

7.9.4 Study A performance – spiking neural networks

Ranked #5 classifier in terms of lapse detection, SNNs were also found to have the PCA-method outperforming the rest of the feature reduction schemes. The highest and lowest φ for SNNs were found to be $\varphi = 0.36$ (PCA) and $\varphi = 0.28$ (NNE) on the pruned study-A EEG dataset. Interestingly, however, despite the low detection scores, SNNs modules had the lowest number of false positives (high specificity and selectivity but low sensitivity) and with an AUC-ROC = 0.82, compared to any other single classifier (level-0 model) techniques implemented on Study A.

7.9.5 Study A performance – kNN

The kNN algorithm was the lowest performing algorithm reported to date on the microsleep detection with the performance scores less than half of many of the other detectors. The highest and lowest φ reported from kNN were 0.23 and 0.20 for PCA and PPCA algorithms, respectively.

7.10 Summary

This chapter presented a review of lapse detection systems implemented, and comparisons between various types of classifiers and feature reduction models. In the course of this research, six classifier techniques and six feature reduction techniques (~27 combinations of which) were developed for lapse detection and evaluated on both pruned and unpruned EEG from Study A. Several of these techniques outperformed the previous benchmarks for lapse detection (see §3.6.6).

As discussed in §3.2.5, Davison *et al.* (2015) used normalized EEG log-power spectrum inputs to train a long short-term memory (LSTM) recurrent neural network (RNN) based lapse detector which demonstrated an overall performance level of AUC-ROC = 0.84 ± 0.02, AUC-PR = 0.41 ± 0.08, and mean phi = 0.38 ± 0.05, on the unpruned EEG datasets. On the same datasets, the current methods using PPCA-based cascaded-leaky-integrator ESNs have set new benchmarks, demonstrating an overall performance level of AUC-ROC = 0.88 ± 0.04, AUC-PR = 0.45 ± 0.09; and mean phi = 0.44 ± 0.06.

Classifier methods using LSMs and SVMP were also found to outperform the previous LSTM benchmarks with an overall performance level of AUC-ROC=0.84 ± 0.04, AUC-PR = 0.42 ± 0.08, and mean phi = 0.41 ± 0.06, and AUC-ROC = 0.85 ± 0.03, AUC-PR = 0.41 ± 0.08, mean phi = 0.40 ± 0.05, respectively. Figure 7-6 depicts a comparison between the previous and current methods implemented for microsleep detection on Study A dataset. Figure 7-6 also depicts the overview of the lapse detection techniques evaluated on the unpruned datasets.

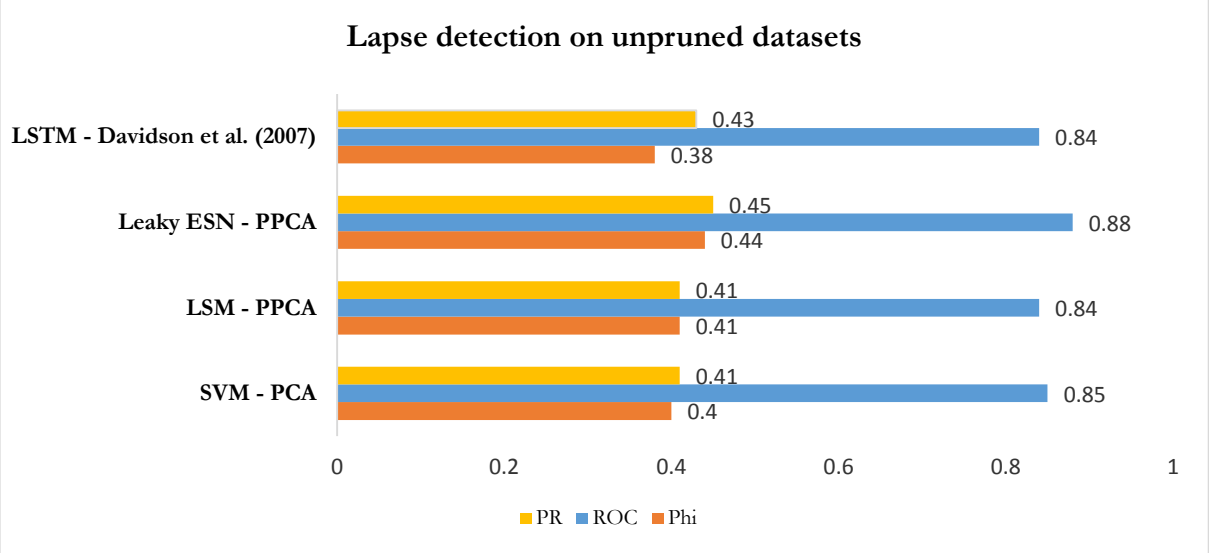


Figure 7-6 Lapse detection on Study A unpruned dataset. Three independent parameters (phi correlation, receiver operator characteristics and precision recall) are depicted as percentages (scale 1: 100).

On the pruned Study A EEG dataset, the detection performances were higher than for unpruned data. Cascaded-leaky-integrator ESNs present a new benchmark performance with an overall performance level of AUC-ROC = 0.84 ± 0.04; AUC-PR = 0.42 ± 0.08, phi = 0.44 ± 0.08.

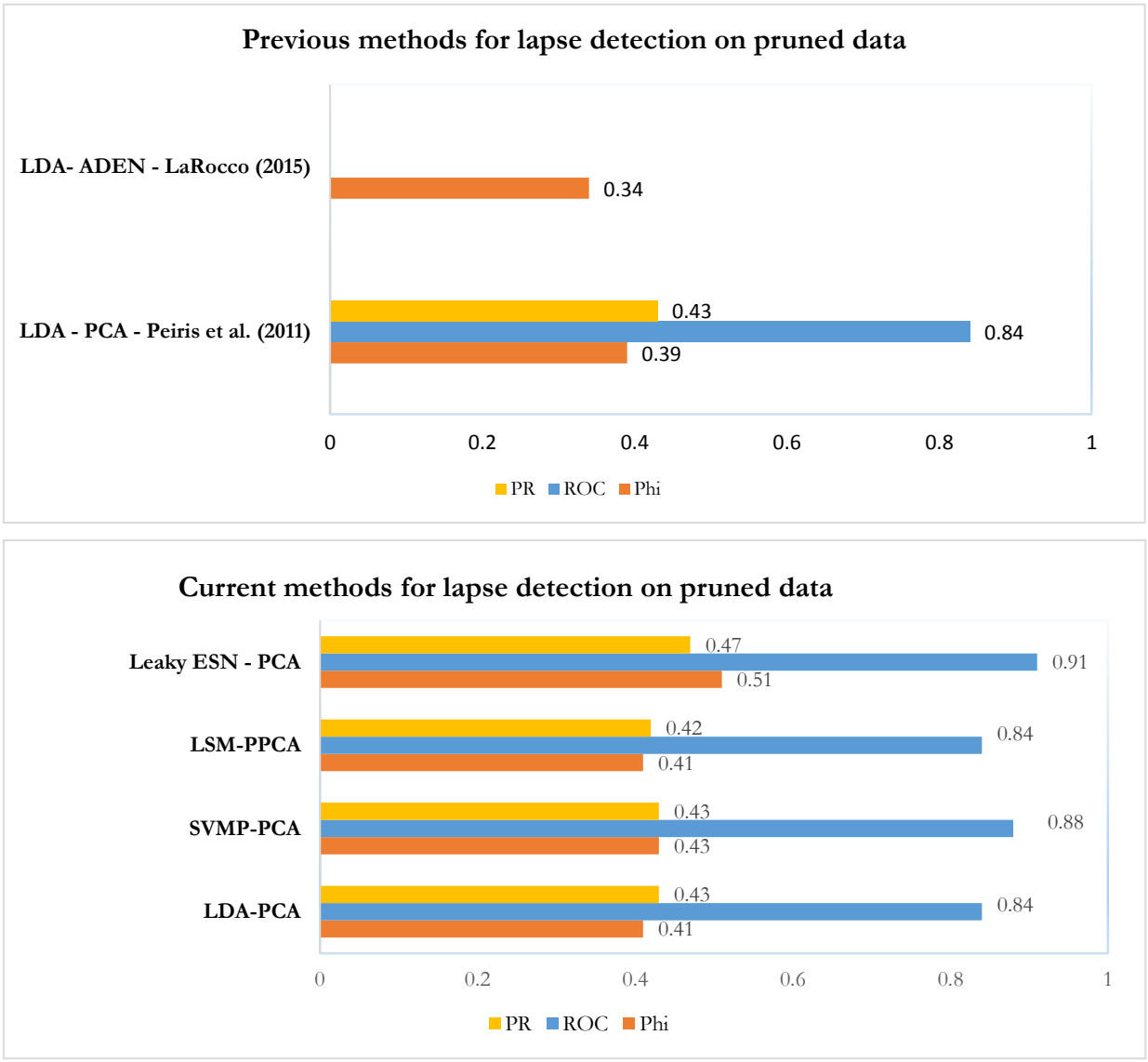


Figure 7-7 Comparison between previous (top) and current (bottom) for the lapse detection on Study A pruned dataset. Three independent parameters (phi correlation, receiver operator characteristics and precision recall) are depicted as percentages (scale 1: 100).

Figure 7-7 depicts a comparison between the previous and current lapse detection methods implemented on pruned Study A EEG dataset. A possible reason for overall detector performances being relatively low may be due to the majority of lapses being only of few seconds in duration. These short lapses may have been of insufficient duration to cause substantial changes in the EEG which could be extracted via various signal processing methods (Peiris *et al.*, 2011). In addition, it was observed that the overall performances of the lapse detection were expected to increase if the subjects with lower detection rates (Subject 810, 817, and 819) were completely excluded from the analysis. Furthermore, since the behavioural gold standard was human-rated, errors/ noise in the gold standard rating process is more than likely to have reduced the classifier performances.

Due to the type of tracking task used, it was not possible to use flatspots alone to determine a subject's performance. However, the combination of independently rated facial video to the flatspots could achieve a better performance. Overall, the levels of performance and reliability demonstrated by the lapse detection models are still deemed insufficient to be implemented in the real world for real-time detection. However, they are encouraging since the current methods provided detecting rates on a second-scale resolution.

CHAPTER 8

Conclusions and future research

8.1 Key findings and discussion

The main motivation behind this research was EEG-based characterization and early detection of lapses of responsiveness states and microsleeps during a sustained attention task. Additionally, this research focused on contributing towards an increased understanding of the characteristics of lapses in responsiveness in healthy non-sleep-deprived subjects. An overall goal was to develop a real-time lapse detection, and preferably, a predictive system capable of continuously monitoring an individual and providing warnings of impending lapses. Thus, allowing for a precautionary measure to maintain safety. Most of this research focused towards exploring the abilities of RC-based techniques toward achieving this goal.

To validate the novel cascaded-leaky integrator ESN approach designed for microsleep detection, and to help determine global parameters that were to be used on the cascaded-leaky ESN approach, a preliminary study was also conducted which employed the gold standard data from 9 subjects, collected in a BCI-controlled setting. The BCI data collected by Brunner *et al.* 2008 used a cue-based experimental paradigm comprising four different motor imagery task, (Brunner *et al.*, 2008). Several feature reduction and classifier schemes were applied to this study. PCA was evaluated for feature reduction. Classifier schemes were LDA, SVMP, ESN, where the latter used both sigmoidal neurons and cascaded-leaky-integrator neurons.

In this preliminary study, performance of classifier systems were evaluated in terms of the mean kappa value using 10 × 10-fold cross-validations (BCI competition requirements). The

results indicated that SVMP classifiers yielded the best averaged mean kappa (κ = 0.52), followed by the cascaded-leaky ESN based approach with a mean kappa κ = 0.49; the LDA-based approach resulted in a mean kappa of 0.34. The results from this study were then compared to previous implementations by other researchers on the same BCI competition datasets (Ang *et al.,* 2012; Pszachewicz, 2013; Song *et al.,* 2012; Kam *et al.,* 2013 and Alonso *et al.,* 2013)). Both SVMP and cascaded-leaky-ESN methods outperformed several other implementations and were moderately close to the BCI competition winning algorithm using CSP, with a mean kappa of 0.57 (Alonso *et al.,* 2013).

However, the main purpose of the preliminary study was to determine how the signal power would affect the performance of a classifier on which the various detection systems/configurations of a gold dataset for which events were precisely known.

The primary study (Study A) comprised more comprehensive physiological and performance data, collected from 15 normal non-sleep-deprived subjects performing a continuous tracking task during normal work hours (Peiris *et al.,* 2011). Fourteen of these 15 subjects reported one or more microsleeps, with an overall rate of 39.3 ± 12.9 microsleeps per hour (mean ± SE) and microsleep duration of 3.4 ± 0.5 s (Peiris *et al.,* 2006b). The previous best implementations used on data from this study reported mean microsleep detection performances of $\varphi = 0.38 \pm 0.05$ for the unpruned datasets and $\varphi = 0.39 \pm 0.06$ on pruned datasets, using LSTM and LDA-based approaches, respectively (Davidson *et al.,* 2007; Peiris *et al.,* 2011). Consequently, the primary goal of this research was to improve these detection rates. Therefore, various combinations of feature reduction and classification schemes were designed to detect these impending lapses of responsiveness states and microsleeps from the underlying EEG, with second scale resolution.

The best lapse detector performance on the pruned Study A dataset was achieved using a detector based on a novel ESN with cascaded-leaky-integrator architecture, using 50-60 PCs from PCA feature reduction (mean $\varphi = 0.51 \pm 0.07$). In comparison, the best LSM-based detector had a performance level of $\varphi = 0.42 \pm 0.06$. The mean performance of lapse detectors based on stacked LDA framework ($\varphi = 0.41 \pm 0.06$) was marginally higher than the previous findings from Peris *et al.* (2011) due to rigorous testing of the system with an increased number of features. The classifier models based on SNNs and kNNs gave the lowest overall performances. The PCA-based feature reduction modules showed the highest overall performances out of all the 6 feature-reduction schemes evaluated. This high performance of PCA-modules was constant across all classifier schemes. PPCA-based methods followed the PCA schemes in terms of the best lapse detection performances. The results from the feature reduction analysis suggests a possible over-fitting in the classifiers for the non-linear feature reduction-based algorithms.

As with pruned Study A EEG dataset, ESNs with cascaded-leaky-integrator neuron models outperformed all the other classifier schemes on the unpruned EEG datasets. That said, in comparison to pruned data, a substantial drop in lapse detection performance was noted for many classifiers. Unlike the pruned datasets, the highest lapse detection performance was seen in cascaded-leaky-integrator ESNs using PPCA meta-features ($\varphi = 0.44 \pm 0.09$). LSM-based detectors using the same PPCA meta-features, followed the cascaded-leaky ESNs with a $\varphi = 0.41 \pm 0.06$, followed by PPCA-SVMP classifiers with a $\varphi = 0.40 \pm 0.06$.

Analysis also showed that creating multiple lapse detection models (ensemble learning) and combining them to form an overall detector usually resulted in a substantial improvement in performance over creating a single model by concatenating data of all the subjects. It was also established that the lapse detection schemes reported the highest performances compared to the other gold standards of flatspots, video microsleeps, and definite microsleeps.

The work performed in this research resulted in higher detection rates, compared to all previous methods. Notwithstanding, the state detection performances were still deemed moderate at best, and there was a large degree of intra- and inter-subject variability.

8.2 Review of hypotheses

Hypothesis 1: Non-linear classification techniques will be able to detect microsleeps with greater accuracy than a linear classifier, such as linear discriminant analysis.

No evidence was found to support this hypothesis. Only, three of the five non-linear classifier models (cascaded-leaky-integrator-ESNs, LSMs and, SVMP), demonstrated higher mean performances compared to the previous LDA-based linear model on the pruned datasets. An LDA-based microsleep detector was able to demonstrate its dominance over other non-linear techniques such as, SNN and kNN.

Hypothesis 2:

A reservoir-computing paradigm, referred to as a cascaded-leaky-integrator ESN, will be able to detect microsleeps with greater accuracy than other linear and non-linear classifiers, such as linear discriminant analysis and polynomial-based support vector machines.

Strong evidence was found to support this hypothesis. By analyzing the spatio-temporal characteristics of the EEG data, both the ESN-based cascaded-leaky-integrator-and-fire neuron model and LSM-based spiky neural network models have shown to aid in the

enhanced detection of lapses and microsleeps by outperforming the previous methods, on both the pruned and unpruned EEG. The current cascaded-leaky ESN-based architecture using only 5-8 neurons, provided a new benchmark performance for the lapse and microsleep detection on Study A data. However, the lower performances in the SNN models could be attributed to the fact that these models were trained on single classifier schemes and not on meta-learner schemes.

8.3 Critique

The value of this research has been well demonstrated in the quantification of characteristics of lapses and EEG-based detection of lapses, with high temporal resolution (Davidson *et al.*, 2007; Peiris *et al.*, 2011; LaRocco 2015). To arrive at more definitive conclusions, EEG data from a larger number and wider range of subjects with a larger number of lapses/microsleeps may help the training and validation of detection models. This may have enabled superior detection performances using the RC-based models.

The work documented in §7.6.2 and §7.6.3 is the first to provide detailed information on the application of a cascaded-leaky-integrator neuron-based ESN and LSM models for EEG-based application, characterization, and detection of lapse-of-responsiveness states and microsleeps during an extended continuous tracking task in non-sleep-deprived subjects. Although modest, the cascaded-leaky-integrator ESN-based microsleep detector (described in Chapter §7.6.2) has shown the best performance achieved to date on microsleep detection, including over more complex systems despite its relatively simple structure, and dynamic reservoir of only 5-8 neurons.

8.4 Future work

Despite the findings presented here, considerable future work in the field of EEG-based lapse and microsleep detection remains. Some of the key areas identified for further investigation in this area of research are presented in this section.

8.4.1 Signal processing techniques

Several concepts within the signal processing domain were identified for application towards microsleep detection from Study A dataset. Some of these techniques are presented below.

8.4.1.1 Higher order spectra

Additional signal processing enhancements could be made to the current lapse detection system using higher-order spectra, which might make a substantial contribution towards improved lapse detection system. Higher-order spectra have been shown to be of value in other EEG-based classification models, such as in the detection of epileptic activity in the EEG (Chua *et al.,* 2007).

8.4.1.2 Other electrophysiological measures

It was decided to exclude EOG signals from model forming due to time restrictions on the project and therefore only EEG analysis was carried out. Therefore, microsleep detection models based on eye movements (EOG) alone could be implemented and their performance compared to the EEG-based detectors constructed during this project. It is possible that a combination of EEG and EOG feature-based detector could perform better than a detector based on either of the features alone. Notwithstanding, vertical EOG has already been presented to the classifiers by way of frontal channels of EEG, especially in the unpruned EEG.

8.4.1.3 Higher frequency activity in EEG

The limitations of the current feature sets used included a limited number of subjects, EEG recording sessions, and variable quality of EEG features for each individual. Despite this, a more thorough analysis of the higher frequency activity might provide new insights into the underlying brain-states. Therefore, the possibility of microsleep-related activity in the gamma and beta bands could warrant further investigation. Additionally, utilizing spatio-temporal features for classification could also potentially provide more insight concerning brain-states.

8.4.1.4 Inclusion of spatial information

In order to develop state-of-the-art features for microsleep detection, it must be possible to obtain feedback on the quality of the feature sets. The existing feature set uses power spectral features (PSD) estimated over a 2-s sliding window with 50% overlap. The length of the sliding window limits the temporal resolution of the feature set, which might be causing information relevant to microsleep detection to be discarded. Therefore, alternative methods for extracting features with similar qualities to PSDs, but with better temporal resolution, using stationary subspace analysis (SSA), and where a multivariate time series is decomposed into stationary and non-stationary components, could be explored (Müller *et al.*, 2011).

8.4.2 Study C: EEG + fMRI

The use of functional-MRI (fMRI) during such a study as the CTT should also be given due consideration as this should provide insight on brain mechanisms underlying lapses. Such a project has already been carried out by a group of members of the Lapse Research Programme (Poudel *et al.*, 2006). This dataset was labelled as Study C.

This study recruited 20 participants to complete a 50-min two-dimensional tracking task in a 3-Tesla MRI scanner (Poudel *et al.*, 2008). Data was collected from several physiological measures including functional MRI (fMRI), eye video, vertical EOG, and EEG from 64 channels, whilst the participants were responding to the tracking target. To confirm whether the participants were normally rested, sleep diaries and wrist actigraphs were recorded each for six days preceding the MRI session. The gold standard was defined by manually rating the tracking response and eye video, with microsleeps defined as poor tracking with full or partial eye closure lasting less than 15 s.

Innes *et al*. (2010) reported that 16 of the 20 subjects in this study had microsleeps during the session, with 14 having more than 43 h^{-1}. Due to the nature of this experiment, the overall rate of microsleeps was found to be high but variable at 79.1 ± 66.2 h^{-1} (range 0–225 h^{-1}) with an average duration of 3.3 ± 1.6 s. Innes *et al.* (2010) also suggested a moderate correlation between the number and duration of microsleeps and self-rated sleep propensity. Interestingly, they also suggested that there was no correlation between the number of microsleeps and circadian type, or the quality, duration, or efficiency of sleep.

From the fMRI data, Poudel *et al.* (2009, 2014) identified decreased blood-oxygen-level dependent (BOLD) activity in the thalamus (related to arousal) and the posterior cingulate gyrus and medial frontal cortex (part of the default-mode network which deactivates during the transition to sleep). An unexpected increase in activity in sensory-motor areas, possibly related to some form of compensation mechanism was also found. They also discovered that the posterior theta-band activity in the EEG was correlated with the occurrence of behavioural microsleeps (Poudel *et al.*, 2010). Another distinction with the Study C dataset is that they contain data from 64 EEG channels and a larger group of subjects than in Study A. Therefore, analysis on Study C could provide a better understanding on characterization and detection of lapses.

8.4.3 Real-time implementation

Since an end goal of this research is the development of a real-time microsleep detection system for implementation in various high-performance occupations, such a system should be made as compact as possible for a user. Therefore, real-time hardware implementations of

the current algorithms on a GPGPU (Keith and Weddell, 2013), or FPGA (Muthuramalingam *et al.*, 2008) may serve a great purpose in this regard. Additionally, Schrauwen *et al.* (2008) reported a successful implementation of a LSM on FPGAs.

Another member of NeuroTech's Lapse Research Programme, Knopp (2015), focused on developing a head-mounted wearable EEG device for lapse detection. This device was fitted with the ability to capture data from multiple physiological signals of the wearer. In addition, this device also provided an extensible software framework for implementing signal processing algorithms. One of the most important extensions to this device would be to implement the current feature extractors, feature reduction, and classification schemes, including cascaded-leaky ESNs by employing a GPGPU to continuously detect physiological changes in the brain underlying microsleeps, and to trigger an alert to rouse an individual.

8.4.4 Prediction of microsleep events in real time

Davidson *et al.* (2007) and Poudel *et al.* (2010) have shown that changes in the theta and alpha bands occur up to 4 s before microsleeps. As a result, these changes in spectral features could allow for EEG-based prediction of microsleeps. Therefore, future work in this area could entail the development of a state-of-the-art microsleep prediction system using the current cascaded-leaky ESN configurations to find the *hidden* changes in the EEG as the brain state changes from alert to non-responsive. This microsleep prediction system should be capable of continuously monitoring an individual's state of responsiveness and predict microsleeps in real-time, alerting a user in advance, potentially avoiding a catastrophe.

8.4.5 Cross validation

One of the main inadequacies identified in this research was its reliance on the leave-one-out cross-validation methods for validating the models. As the current microsleep detection problem is one with highly imbalanced class distributions, the leave-one-out approach may have actively contributed towards a bias within the correlation between predicted and actual values of the output. Therefore, implications of other forms of cross-validation techniques, such as balanced K-fold (in which the training/testing is performed after breaking the data into K groups) or bootstrapping (Hastie and Tibshirani, 2009) could also be evaluated.

8.4.6 Combination systems/ approaches

Finally, an overall automated detection model could be designed by combining the best results from all the 6 feature reduction methods and 6 classification models evaluated during this project. Such a system could not only enhance the performance of the overall lapse

detection, but could also pave the way for a new technique into a personalized microsleep detection system (LaRocco, 2015).

Future work still remains a challenge in the field of microsleep detection, but the inclusion of novel techniques such as RC-based cascaded-leaky-integrator ESNs could lead in a promising direction.

References

Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H. T. Learning from Data: A Short Course. AMLBook.com, 2012.

Akerstedt, T., Kecklund, G., Knutss, A. Manifest sleepiness & the spectral content of the EEG during shift work, Sleep, 1995, 14: 221-225.

Åkerstedt, T. & Gillberg, M. Subjective & objective sleepiness in the active individual. International Journal of Neuroscience, 1990, 52: 29-37.

Alba, N. A., Sclabassi, R. J., Mingui, S., & Cui, X. T. Novel Hydrogel-Based Preparation-Free EEG Electrode. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2010, 18(4), 415-423

Alexandr, A, & Indyk, P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. 47th Annual IEEE Symposium Foundations of Computer Science, 2006; 459-468.

Álvarez-Estévez, D., Fernández-Pastoriza, J. and Moret-Bonillo, V. A continuous evaluation of the awake sleep state using fuzzy reasoning', In Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 5539–5542.

Amara G. An Introduction to Wavelets. IEEE Computational Science and Engineering, Summer 1995, vol. 2, num. 2.

Agustina, G. C., Lorena, O., and Eric, L. Automatic detection of drowsiness in EEG records based on multimodal analysis. Medical Engineering and Physics, 2014, 2(36): 244-249.

Aldrich, M. S. Automobile accidents in patients with sleep disorders. Sleep, 1989, 12(6):487–494.

Alexandr, A. & Indyk, P. Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. 47th Annual IEEE Symposium, Foundations of Computer Science, 2006, 117-122.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., & Zhang, H. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Frontiers in Neuroscience, vol. 6, p. 39, 2012.

Arnedt, J. T., Wilde, G. J., Munt, P. W. & MacLean, A. W. How do prolonged wakefulness & alcohol compare in the decrements they produce on a simulated driving task? Accident Analysis & Prevention, 2001, 33: 337-344.

Arruda, J. E., Walker, K. A., Weiler, M. D. and Valentino, D. A. Validation of a right hemisphere vigilance system as measured by principal component and factor analyzed quantitative electroencephalogram. International Journal of Psychophysiology, 1999, 32: 119-128.

Astolfi, L., Cincotti, F., Mattia, D., Babiloni, F., Marciani, M. G.,Fallani, F. Mattiocco, M. Miwakeichi, F., Yamaguchi. Y., Martinez, S., Salinari, P. Tocci, , 2006 Removal of ocular artifacts for high resolution EEG studies: a simulation study, Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 976979 , 1-42440-033-3 York City, USA, Aug 30-Sept 3, 2006.

Baudry, M. Synaptic plasticity and learning and memory: 15 years of progress. Neurobiology of Learning and Memory, 1998, 70: 113–118.

Bawa, M., Condie, T., Ganesan, P., LSH Forest: Self-Tuning Indexes for Similarity Search. Proceedings of the 14th international conference on World Wide Web Pages 651-660, 2005.

Belkin, M., & Niyogi, P. Laplacian Eigenmaps and spectral techniques for embedding and clustering. In Advances in Neural Information Processing Systems, volume 14, pages 585–591, Cambridge, MA, USA, 2002. The MIT Press.

Berka, C., Levendowski, D., Cvetinovic, M. Real-time analysis of EEG indices of alertness, cognition, & memory with a wireless EEG headset. International Journal of Human Computer Interaction, 2004; 17:151– 70.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R. & Lopez, M. E. Real-time system for monitoring driver vigilance. IEEE Transactions on Intelligent Transportation Systems, 2006, 7: 63-77

Beierholm, U., Nielsen, C. D., Ryge, J., Alstrom, P. and Kiehn, O. Characterization of reliability of spike timing in spinal interneurons during oscillating inputs. Journal of Neurophysiology, 2001, 86: 1858–1868.

Binnie, C. D & Osselton, J. W. Origins & technique, Clinical Neurophysiology: EMG, nerve conduction & evoked potentials / EEG technology. Butterworth-Heinemann Press Ltd, Oxford, 1995.

Bittner, R., Smrcka, P., Vysoký, P., Haná, K., Pousek, L. & Scheib, P. Detecting of fatigue states of a car driver, Springer, Berlin/Heidelberg, 2000.

Bishop, C. M. Neural Networks for Pattern Recognition, Oxford University Press, 1995.

Bjerner, B. Alpha depression and lowered pulse rate during delayed actions in a serial reaction test. Acta Physiologica Scandinavica, 1949, 19:1–93.

Boyle, L. N., Tippin, J., Paul, A., & Rizzo, M., Driver performance in the moments surrounding a microsleep. Transportation Research Part F: Traffic Psychology Behaviour, 2008, 11: 126–36.

Breiman, L. Bagging predictors. Machine Learning 1996, 24:2:123-140.

Bremer, G., Smith, J., Karacan, I. Automatic detection of the K-Complex in sleep electroencephalograms. IEEE Transactions in Biomedical Engineering, 1970, 17: 314–323.

Bronzino, J. D., Berbari, E. J., Johnson, P. L. & Smith, W. M. Bioelectronics & instruments. In: R. C. Dorf (Ed) The electrical engineering h&book. CRC Press, 2000.

Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E., & McCarley, R. W. Control of sleep and wakefulness. Physiological Reviews, 2012, 92(3), 1087–1187.

Brownlee, J. Clever Algorithms. Nature-Inspired Programming Recipes. LuLu, 2011 p. 436,

Buckley, R. J. Sustained Attention Lapses & Behavioural Microsleeps During Tracking, Psychomotor Vigilance, & Dual Tasks. Masters Thesis at the University of Canterbury, 2013.

Burgsteiner, H. On learning with recurrent spiking neural networks and their applications to robot control with real-world devices, (2005), Graz University of Technology

Burgsteiner, H. Mark, K. Leopold, Alexander, Steinbauer, G Movement Prediction from Real-World Images Using a Liquid State Machine. Industrial engineering and other applications of applied intelligent systems, 2005: 121-130.

Buteneers, P., Verstraeten, D., Nieuwenhuyse, B. V,, Stroob&t, D., Raedt, R., Vonck, K., Boon, P. & Schrauwen B. Real-time detection of epileptic seizures in animal models using reservoir computing. Epilepsy Research, 2012, 103(2-3):124-34.

Buysse D. J., Reynolds, C. F., Kupfer, D. J., Thorpy, M. J., Bixler, E. O. & Manfredi. R. Clinical diagnoses in 216 insomnia patients using the international classification of sleep disorders (ICSD), DSM-IV and ICD-10 categories: a report form the APA/NIMH DSM-IV trial. Sleep. 1994;17:630–637.

Byrne, P. Battle against microsleep, 2002, Online. Available: http://www.eforecourt.com/body.htm/pid/452/catagory/general

Carlson, N. R. Physiology of behaviour, Pearson Education Inc., Boston, 2004.

Carskadon, M. A. & Rechtschaffen, A. Monitoring & staging human sleep: methodology. In: M. Kryger, T. Roth & W. C. Dement (Eds) Principles & practice of sleep medicine. W. B. Saunders Company, Philadelphia, 2000: pp. 15-25.

Caton, R. The electric currents of the brain. British Medical Journal, 1875, 278.

Chan, F. H., Qiu, W., & Poon, L. Evoke potential estimation using modified time-sequence adaptive filter, Medical & Biological Engineering & Computing, 36 4, 1998, 407414.

Chee, M. W. L., Tan, J. C., Zheng, H., Parimal, S., Weissman, D. H., Zagorodnov, V., & Dinges, D. F. Lapsing during sleep deprivation is associated with distributed changes in brain activation. Journal of Neuroscience, 2008, 28(21), 5519-5528.

Chee, M. W. L., & Choo, W. C. Functional imaging of working memory after 24 hr of total sleep deprivation. Journal of Neuroscience, 2004, 24: 4560-4567.

Chua, C. K., Chandran, V., Acharya, R., Lim C. M. Higher Order Spectral (HOS) analysis of epileptic EEG signals. Proceedings of International Conference of IEEE Engineering in Medicine & Biology Society, 2007, 6496-6499.

Cohen, J. (1960). A coefficient of agreement for nominal scale. Cohen J. Educat Psychol Measure; 20: 37-46.

Colten, H. R., Altevogt, B. M. Sleep Disorders & Sleep Deprivation. An Unmet Public Health Problem. Institute of Medicine (US) Committee on Sleep Medicine & Research National Academies Press (US); 2006.

Conradt, R., Brandenburg, U., Penzel, T., J Hasan, J., Värri, A., & Peter, J. Vigilance transitions in reaction time test: a method of describing the state of alertness more objectively. Clinical Neurophysiology, 1999, 110(9), 1499-1509.

Coren, S. Sleep Thieves. New York, free press, 1996.

Coufal, D. Redundant rules detection in EEG fuzzy classifier. Intelligent Engineering Systems, 2009, 177-181.

Cristianini, N. and Shawe-Taylor, J. An introduction to support vector machines and other kernel-based learning methods.Cambridge university press, 2000.

Davidson, P. R., Jones, R. D. and Peiris, M. T. R. EEG-based Lapse Detection with High Temporal Resolution. IEEE Transactions in Biomedical Engineering, 2007, 54: 832-839.

Davies, D. R. and Parasuraman, R. The psychology of vigilance, Academic Press, London, 1982.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N. and Hobart, G. Human brain potentials during the onset of sleep. Journal of Neurophysiology, 1938, 1: 24-38.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N. and Hobart, G. Changes in human brain potentials during the onset of sleep. Science, 1937, 86: 448-450.

Davis, J. & Goadrich, M. The relationship between Precision-recall and ROC curves, Proceedings of the 23rd international conference on machine learning, p.233-240, 2006.

De Gennaro, L., Ferrara, M., Curcio, G. & Cristiani, R. Antero-posterior EEG changes during the wakefulness–sleep transition. Clinical Neurophysiology, 2001, 112: 1901-1911.

De Gennaro, L., Ferrara, M., Ferlazzo, F. & Bertini, M. Slow eye movements & EEG power spectra during wake-sleep transition. Clinical Neurophysiology, 2000, 111: 2107-2115.

DeMers, D., and Cottrell, G. Non-linear dimensionality reduction. In Advances in Neural Information Processing Systems, volume 5, pages 580–587, San Mateo, CA, USA, 1993.

Domingos, P. A few useful things to know about machine learning. Communications. ACM 55, 78-87, 2012.

Doran, S. M., Van Dongen, H. P. & Dinges, D. F. Sustained attention performance during sleep deprivation: evidence of state instability. Archives Italiennes de Biologie, 2001, 139: 253-267.

Dorrian, J., Rogers, N. L., & Dinges, D. F. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In C. Khushida (Ed.), Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects (pp. 39-70). New York: Marcel Dekker Inc. (2005).

Dickinson, D. L. and Drummond, S. P. A. The effects of total sleep deprivation on Bayesian updating, Judgement and Decision Making, 2008, 3(2):181-190.

Diekelmann, S. and Born, J. The memory function of sleep. Nature Reviews Neuroscience, 2010, 11: 114–126.

Dinges, D. F. & Kribbs, N. B. Performing while sleepy: Effects of experimentally-induced sleepiness. In: T. H. Monk (Ed) Sleep, sleepiness & performance. John Wiley & Sons, Chichester, 1991.

Dinges, D. F., Mallis, M. M., Maislin, G. & Powell, J. W. Evaluation of techniques for ocular measurement as an index of fatigue & as the basis for alertness management. Final Report DOT HS 808 762. National Highway Traffic Safety Administration, Washington, DC, 1998.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C. & Pack, A. I. Cumulative sleepiness, mood disturbance, & psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep, 1997, 20: 267-277.

Dinges, D. F. & Powell, J. W. Microcomputer analysis of performance on a portable, simple visual RT task during sustained operation. Behavior Research Methods, Instruments, & Computers, 1985, 17: 652-655.

Dijkstra, E. W. A note on two problems in connexion with graphs. Numerische Mathematik, 1:269–271, 1959.

Dorrian, J., Rogers, N. L. and Dinges, D. F. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: C. A. Kushida (Ed) Sleep deprivation - clinical issues, pharmacology, and sleep loss effects. Marcel/Dekker, New York, 2005.

Drummond, S. P. A., Brown, G. G., Stricker, J. L., Buxton, R. B., Wong, E. C. & Gillin, J. C. Sleep deprivation-induced reduction in cortical functional response to serial subtraction. NeuroReport, 1999, 10: 3745–3748.

Duffy, F. H., Iyer, V. G. & Surwillo, W. W. Clinical electroencephalography & topographic brain mapping, Springer-Verlag, 1989.

Du Bois-Reymond E. H. Studies on Animal Electricity. Vol.1 (In German). G Reimer, Berlin, DE, 1848.

Dy, J. G. and Brodley, C. E. Feature selection for unsupervised learning. Journal of Machine Learning Research, 2004 5, 845-889.

Erwin, C., Volow, M., & Gray, B. Psychophysiologic Indices of Drowsiness," SAE Technical Paper, 1973, 10.4271/730122.

Fairclough, S. H. and Graham, R. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. Human Factors, 1999, 41: 118-128.

Fatemian, S. Z. and Hatzinakos, D. "A new ECG feature extractor for biometric recognition", 16th International Conference on Digital Signal Processing, pp. 1-6, 2009.

Fawcett, T. An introduction to ROC analysis. Pattern Recognition Letters, 27(8):861–874, 2006.

Federal Aviation Administration (FAA). OpSpec A332 Ultra Long Range (ULR) Flag Operations in Excess of 16 Hours Block-to-Block Time. FAA, Washington, DC, 2007.

Fell, J., Roschke, J., Mann, K. & Schaffner, C. Discrimination of sleep stages: a comparison between spectral & nonlinear EEG measures. Electroencephalography & Clinical Neurophysiology, 1996, 98: 401-410.

Flores, M., Armingol, J., and de la Escalera, A. Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions. Journal of Advanced Signal Processing, 2010:438205

Foucher, J.R., Otzenberger, H., Gounot, D. Where arousal meets attention: a simultaneous fMRI and EEG recording study. NeuroImage, 2004, 22: 688-697.

Freund, D. M., Knipling, R. R., L&sburg, A. C., Simmons, R. R. & Thomas, G. R. A holistic approach to operator alertness research. 950636. U.S. Department of Transportation, Washington, D. C., 1995.

Friedl, K. E., Mallis, M. M., Ahlers, S. T., Popkin, S. M., & Larkin, W. Research requirements for operational decision-making using models of fatigue and performance. Aviation Space & Environmental Medicine, 2014, 75; A192–9.

Gerstner, W. and Kistler, W. Spiking Neuron Models. Single Neurons, Populations, Plasticity. Cambridge University Press, Cambridge, MA, 2002.

Ghongade, R. and Ghatol, A. Performance analysis of feature extraction schemes for artificial neural network based ECG classification. International Conference on Computational Intelligence and Multimedia Applications, 2007.

Ghosh-Dastidar, S. & Adeli, H. Spiking Neural Networks. International Journal of Neural Systems, 2009, 19: 295.

Golz, M., & Sommer, D. Monitoring of drowsiness & microsleep. Proceedings of International Conference of IEEE Engineering in Medicine & Biology Society, 2010 32, 1787-1791.

Golz, M., Sommer, D., Chen, M., Trutschel, U., & M&ic, D. Feature fusion for the detection of microsleep events. Journal of VLSI Signal Processing, 2007, 49(2), 329-342.

Golz, M., Sommer, D., & M&ic, D. Microsleep detection in electrophysiological signals. Proceedings of International Workshop on Biosignal Processing & Classification, 2005, 1: 102-109.

Golz, M., Sommer, D., Seyfarth, A., Trutschel, U., & Moore-Ede, M. Application of vector-based neural networks for the recognition of beginning microsleep episodes with an eyetracking system. Proceedings of the Computational Intelligence: Methods & Applications, 2001, 2, 1-5

Goovaerts, G., Denissen, A., Milosevic, M., Van Boxtel, G., and Van Huffel, G. Advanced EEG processing for the detection of drowsiness in drivers. Proceedings of International Conference on Biosignals. 2014, pp. 205-212.

Gerstner, W. and Kistler, W. M. Spiking Neuron Models. Single Neurons, Populations, Plasticity. Cambridge University Press, 2002.

Griffiths, R. R., Evans, S. M., Heishman, S. J., Preston, K. L., Sannerud, C. A., Wolf, B., Woodson, P. P. Low-dose caffeine physical dependence in humans. Journal of Pharmacology & Experimental Therapeutics, 1990, 255:1123–1132.

Hamilton, P., Wilkinson, R.T., Edwards, R.S. A study of four days partial sleep deprivation. in: W.P. Colquhoun (Ed.) Aspects of Human Efficiency – Diurnal Rhythm & Loss of Sleep. The English Universities Press Ltd, ; 1972: 101–114.

Haraldsson, P. O., Carenfelt, C., Diderichsen, F., Nygren, A. Clinical symptoms of sleep apnoea syndrome & automobile accidents. Otolaryngology, 52, 1990, pp. 57–62

Harrison, Y. and Horne, J. A. Occurrence of 'microsleeps' during daytime sleep onset in normal subjects. Electroencephalography and Clinical Neurophysiology, 1996, 98: 411-416.

Hastie, T., Tibshirani, T. and Friedman, J. The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition, 2009: 9780387848587.

Hilbe, J. M. Logistic Regression Models. Chapman & Hall/CRC Pres, 2009.

Hill, T. & Lewicki, P. Statistics: methods and applications: a comprehensive reference for science, industry, and data mining. StatSoft, 2006.

Hinton, G. E., and Salakhutdinov, R., Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786):504-507.

Hinton G.E. and Roweis.S.T. Stochastic Neighbor Embedding. In Advances in Neural Information Processing Systems, volume 15, pages 833–840, Cambridge, MA, USA, 2002. The MIT Press.

Hodgkin, A. L. & Huxley, A. F. Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. Journal of Physiology,116(4):449–472, 1956.

Hopfield, J. J. Hopfield network. Scholarpedia, 2(5):1977,

Hori, T., Hayashi, M. and Morikawa, T. Topographic EEG changes and the hypnagogic experience. In: R. D. Ogilvie (Ed) Sleep onset: Normal and abnormal processes. American Psychological Association, Washington, 1994.

Horn, D. & Opher, I. Collective Excitation Phenomena and Their Applications in Maass, W. & Bishop, C. M. (eds.) Pulsed Neural Networks, MIT-press, 1999.

Horne, J. A. & Reyner, L. A. Driver sleepiness. Journal of Sleep Research, 1995, 4: 23-29.

Horne, J. Why we sleep, Oxford University Press, New York, 1988.

Huang, R., Tsai, L. and Kuo, C. J. Selection of valid and reliable EEG features for predicting auditory and visual alertness levels. Proceedings of the National Sciences Council, 2001, 25: 17-25.

Huang, R.S., Jung, T.P., Delorme, A. & Makeig, S. Tonic and phasic electroencephalographic dynamics during continuous compensatory tracking. Neuro Image, 2008, 39(4), 1896-1909.

Itoi, A., Cilveti, R., Voth, M. Can Drivers Avoid Falling Asleep at the Wheel? AAA Foundation for Traffic Safety, 1993.

Innes, C. R. H., Poudel, G. R., Signal, T. L., & Jones, R. D. Behavioural microsleeps in normally-rested people. Proceedings of International Conference of IEEE Engineering in Medicine & Biology Society, 32, 4448-4451, 2010.

Jaeger, H. Echo State Networks. Scholarpedia, 2007, 2(9):2330.

Jaeger, H. Reservoir Riddles: Suggestions for Echo State Network Research (Extended Abstract of Invited Talk) 2005, 1460-1462.

Jaeger, H., Haas, H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. Science, 2004, 78-80.

Jaeger, H. Adaptive nonlinear system identification with echo state networks. In Advances in Neural Information Processing Systems 15, S. Becker, S. Thrun, K. Obermayer (Eds), MIT Press, Cambridge, 2003, pp. 593-600.

Jaeger, H. Tutorial on training recurrent neural networks, covering BPTT, RTRL, EKF and the "echo state network" approach. GMD Report 159, German National Research Center for Information Technology, 2002b, p.48.

Jaeger, H. Short term memory in echo state networks. GMD Report 152, German National Research Center for Information Technology, 2001a, p.60.

Jaeger, H. The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, German National Research Center for Information Technology, 2001b, p. 43 pp.

Jones, R. D., Poudel, G. R., Innes, C. R. H., Davidson, P. R., Peiris, M. T. R., Malla, A., Signal, T. L., Carroll, G. J., Watts, R., & Bones, P. J. Lapses of responsiveness: Characteristics, detection, & underlying mechanisms. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 32, 1788-1791, 2010.

Jonmohamadi, Y., Poudel, G., Innes, C. , Jones, R. Voxel-ICA for reconstruction of source signal time-series & orientation in EEG & MEG. Australasian Physical & Engineering Sciences in Medicine, 2014, 37, 457-64.

Joshi, P., and Maass, W. Movement generation with circuits of spiking neurons. Neural Computation. 2005; 17(8):1715-38.

Ju, H., Xu, J., Chong, E., & VanDongen, A. Effects of synaptic connectivity on liquid state machine performance, Neural Networks, 2013, Volume 38: 39-51

Jung, T.P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V. & Sejnowski, T. J. Removing electroencephalographic artifacts by blind source separation. Psychophysiology, 2000a, 37: 163-178.

Jung, T., Makeig, S., Stensmo, M. & Sejnowski, T. Estimating alertness from the EEG power spectrum. IEEE Transactions on Biomedical Engineering, 1997, 44: 60-69.

Jung T, P., Krishnamurthy, A. Deriving Gestural Scores from Articulator-Movement Records using Weighted Temporal Decomposition. IEEE Transaction on Speech, Audio Processing, 1996. 4(1), 2-18.

Kahnemann, D. Attention and effort, Prentice Hall, Englewood Cliffs, New Jersey, 1973.

Kam, T. E., Suk, H. I., & Lee, S. W. Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification. Neurocomputing, vol. 108, pp. 58–68, 2013.

Kiymik, M., Akin, M., & Subasi, A. Automatic recognition of alertness level by using wavelet transform and artificial neural network. Journal of Neuroscience Methods, 2004, pp. 231-240.

Knipling, R. R. & Wang, J. S. Revised estimates of the US drowsy driver crash problem size based on general estimates system case reviews in 39th Annual Association for the Advancement of Automotive Medicine, Chicago, IL, 1995. 451-466.

Knipling, R. R., & Wierwille, W. W. (1994). Vehicle-based Drowsy Driver Detection: Current Status & Future Prospects. IVHS America fourth annual meeting, 1994, pp. 1-24.

Knopp, S. A multi-modal device for application in microsleep detection. Doctoral Thesis at the University of Canterbury, 2015.

Knopp, S. J., Bones, P. J., Weddell, S, J., Innes R. H., Jones, R, D. A miniature head-mounted camera for measuring eye closure. In proceedings of 27th International Conference on Image and Vision Computing, 2012, pp: 312-318.

Kim, H., Rosen, J. Epileptic Seizure Detection - An AR Model Based Algorithm for Implantable Device. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 2010, 32,5541-5545.

Kasabov, N., Feigin, V., Hou, Z. G., Chen, Y., Liang, L., Krishnamurthi, R. & Parmar, P. Evolving spiking neural networks for personalised modelling, classification and prediction of spatio-temporal patterns with a case study on stroke. Neurocomputing, 2014, 134, 269-279.

Kasabov, N. To spike or not to spike: A probabilistic spiking neural model. Neural Networks, 2010, Vol. 23, 16-19.

Kegl. B. Intrinsic dimension estimation based on packing numbers. In Advances in Neural Information Processing Systems, volume 15, pages 833–840, Cambridge, MA, USA, 2002. The MIT Press.

Kim, K .I. Jung, K. & H.J. Kim. Face recognition using kernel principal component analysis. IEEE Signal Processing Letters, 9(2):40–42, 2002

Kirsch, A. D. Report on the statistical methods employed by the U.S. Federal Aviation Administration in its cost/benefit analysis of the proposed flight crewmember duty period limitations, flight time limitations & rest requirements. Washington, DC. United States Federal Aviation Administration, 1996, 28081.

Krajewski, J., Batliner, A., & Wiel, R. Multiple classifier applied on predicting microsleep from speech. Proceedings of the International Conference on Pattern Recognition, 2008, 19: 1-4.

Kung, S. Y., Diamantaras, K. I., & Taur, J. S. Adaptive Principal component Extraction (APEX) and applications. IEEE Transactions on Signal Processing, 42(5):1202–1217, 1994

Lal, S. K. L. & Craig, A. A critical review of the psychophysiology of driver fatigue. Biological Psychology, 2001, 55: 173-194.

Lal, S. K. L. & Craig, A. Driver fatigue: Electroencephalography & psychological assessment. Psychophysiology, 2002, 39: 313-321.

LaRocco, J. Detection of microsleeps from the eeg via optimized classification techniques. Doctoral Thesis at the University of Canterbury, 2015.

LaRocco, J., Innes, C.R.H., Bones, P.J., Weddell, S., & Jones, R.D. Optimal EEG feature selection from average distance between events and non-events Proceedings of Annual International Conference on Engineering in Medicine and Biology Society, 2014, 36, 2641-2644

Laufs, H., Holt, J., Elfont, R., Krams, M., Paul, J., Krakow, K., Kleinschmidt, A. Where the BOLD signal goes when alpha EEG leaves. NeuroImage, 2006, 31: 1408-1418.

Le Van Quyen, M., Martinerie, J., Baulac, M. and Varela, F. Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings. Neuroreport. Jul 13;10(10):2149-55 (1999).

Le Van Quyen, M., Martinerie, J., Navarro, V., Boon. P., D'Havé, M., Adam, C., Renault, B., Varela, F. & Baulac, M. Anticipation of epileptic seizures from standard EEG recordings. Lancet. Jan 20;357(9251):183-8 (2001).

Lenne, M. G., Triggs, T. J. and Redman, J. R. Interactive effects of sleep deprivation, time of day, and driving experience on a driving task. Sleep, 1998, 21: 38-44.

Leong, W. Y., Mandic, D. P., Golz, M., & Sommer, D. Blind extraction of microsleep events. Proceedings of the International Conference on Digital Signal Processing, 2007, 15, 207-210.

Li, C., Zheng, V., and Tai, C., Detection of ECG characteristic points using wavelet transforms, IEEE Transactions on Biomedical Engineering,1995, 42, 21–28,

Lin, C.-T., Wu, R.-C., Liang, S.-F., Chao, W.-H., Chen, Y.-J. and Jung, T.-P. EEG-based drowsiness estimation for safety driving using independent component analysis. IEEE Transactions on Circuits and Systems, 2005, 52: 2726-2738.

Lin, C.-T., Ko, L.-W., Chung, I.-F., Huang, T.-Y., Chen, Y.-C., Jung, T.-P. and Liang, S.-F. Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks. IEEE Transactions on Circuits and Systems, 2006, 53: 2469-2476.

Lins, O.G., Picton, T.W., Berg, P., Scherg, M. Ocular artifacts in recording EEGs & event-related potentials: II. Source dipoles & source components. Brain Topography, 1993, 6 (1), 65–78.

Lorist, M. M., Snel, J., Kok, A. Influence of caffeine on information processing stagesin well-rested & fatigued subjects. Psychopharmacology, 1994, 113:411–421.

Lukoševičius, M., Jaeger, H. On self-organizing reservoirs and their hierarchies. Jacobs University technical report Nr. 25, 2010.

Lukoševičius, M. Jaeger, H. Reservoir Computing approaches to Recurrent Neural Network training. Computer Science Review, 2009, 3(3), 127-149.

Lukoševičius, M. Jaeger, H. Overview of Reservoir Recipes. A survey of new RNN training methods that follow the Reservoir paradigm. School of Engineering and Science, Jacobs University Bremen gGmbH Technical, 2007 Report No. 11.

Luna, T. Fatigue in context: USAF mishap experience. [Abstract]. Aviation Space & Environmental Medicine, 2003, 74:388.

Maass, W., Joshi, P., and Eduardo, D. Principles of real-timecomputing with feedback applied to cortical microcircuit models. In Advancesin Neural Information Processing Systems. MIT Press, Cambridge, 2006. 835-842. .

Maass, W., Legenstein, R. A., and Markram, H. Computational models for generic cortical microcircuits. In Computational Neuroscience: A Comprehensive Approach, pages 575{605. Chapman & Hall/CRC, (2004).

Maass, W., Natschager, T., Markram, H. Real time computing without stable states: a new dramework for neural comptation without stable states: a new framework for neural computation based on perturbations. Neural Computation, 2002a 14(11):2531-2560.

Maass, W., Legenstein, R. A., and Markram, H. A new approach towards vision suggested by biologically realistic neural micro-circuit models. In Proc. of the 2nd Workshop on Biologically Motivated Computer Vision, Lecture Notes in Computer Science, 2002b.

Maass, W. Networks of spiking neurons: The third generation of neural network models. Neural Networks, 1997, 10: 1659–1671.

Makeig, S., Jung, T. P., & Sejnowski, T. J. Awareness During Drowsiness: Dynamics & Electrophysiological Correlates. Canadian Journal of Experimental Psychology, 2000, 54: 266-273.

Makeig, S, Jung, T. P., & Sejnowski, T. J. Using feedforward neural network to monitor alertness from changes in EEG correlation & coherence. Advances in Neural Information Processing Systems, 1996, 8: 931-937.

Makeig, S & Jung, T. P. Changes in Alertness Are a Principal Component of Variance in the EEG Spectrum, NeuroReport, 1995, 7(1), 213-216.

Makeig, S. & Inlow, M. Lapses in alertness: coherence of fluctuations in performance & EEG spectrum. Electroencephalography & Clinical Neurophysiology, 1993, 86: 23-35.

Malla, A. M., Davidson, P. R., Bones, P. J., Green, R., & Jones, R. D. Automated video-based measurement of eye closure for detecting behavioral microsleep. Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society, 2010, 32, 6741-6744.

Mallis, M. M. (1999) Evaluation of Techniques for Drowsiness Detection: Experiment on Performance-Based Validation of Fatigue-tracking Technologies, Drexel University.

Mardi, Z., Ashtiani, S. N. M., and Mikaili, M. EEG-based Drowsiness Detection for Safe Driving Using Chaotic Features and Statistical Tests. Journal of Medical Signals and Sensors, 2014, 1(2), 130-137.

Martin, G., Sommer, D., Chen, Mo., Mandic, D., Trutschel, U. Feature Fusion for the Detection of Microsleep Events. Journal of VLSI Signal Processing, 2007, 49, 329–342.

Martin, G., Sommer, D., Holzbrecher,M., Schnupp,T. Detection and prediction of driver's microsleep events. In conference proceedings of 14th International Conference of Road Safety on Four Continents, 2008, 1-11.

Matousek, M. & Petersen, I. A method for assessing alertness fluctuations from EEG spectra. Electroencephalography & Clinical Neurophysiology, 1983, 55: 108-113.

May, J. F. and C. L. Baldwin .Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation Research Part F: Psychology and Behaviour, 2009, 12(3): 218-224.

McCartt, A. T., Rohrbaugh, J. W., Hammer, M. C. & Fuller, S. Z. Factors associated with falling asleep at the wheel among long-distance truck drivers. Accident Analysis & Prevention, 2000, 32: 493-504.

Mitler, M. M., Gujavarty, K. S. & Browman, C. P. Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. Electroencephalography & Clinical Neurophysiology, 1982, 53: 658-661.

Moorcroft, W. H. Functions of sleep and NREMS. In: W. H. Moorcroft (Ed.) Understanding Sleep and Dreaming. Springer, New York, NY, 2013: 231–259.

Muhammad, A., Badruddin, N., & Drieberg, M. A simulator based study to evaluate driver drowsiness using electroencephalogram. Proceedings of International Conference on Intelligent and Advanced Systems, 2014, 5: 1-5.

Murphy, C. E. and Trivedi, M. M. Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. IEEE Transactions on Intelligent Transport Systems; 2010, 11:300–311.

National Highway Traffic Safety Administration (NHTSA). Drowsy driving & automobile crashes. Washington, DC: United States Department of Transportation, DOT HS: 808-707, 1998.

National Highway Traffic Safety Administration (NHTSA). National Automotive Sampling System - Crashworthiness Data System 2012 Analytical User's Manual. Washington, DC: United States Department of Transportation, 2013, 811- 830

National Highway Traffic Safety Administration (NHTSA). Traffic Safety Facts: Drowsy Driving. Washington, DC: United States Department of Transportation, 2011, 811-449.

National Highway Traffic Safety Administration (NHTSA), National Motor Vehicle Crash Causation Survey: A Report to the United States Congress. Washington, DC: U.S. Department of Transportation, 2008, 811-059.

Nebes, R. D and Brady, C. B Focused and divided attention in Alzheimer's disease. Cortex, 1989, 25(2); 305-15.

Ng. A., Jordan, M. and Weiss, Y. On spectral clustering: Analysis and an algorithm. In Advances in Neural Information Processing Systems, volume 14, pages 849–856, Cambridge, MA, USA, 2001. The MIT Press.

Niedermeyer, E. The normal EEG of the waking adult. In: E. Niedermeyer & F. H. Lopes da Silva (Eds) Electroencephalography: Basic principles, clinical applications & related fields. Williams & Wilkins, Baltimore, 1999: pp. 149-173.

Ogilvie, R. D. The process of falling asleep. Sleep Medicine Reviews, 2001, 5: 247-270.

Ogilvie, R. D., McDonagh, D. M., Stone, S. N. & Wilkinson, R. T. Eye movements & the detection of sleep onset. Psychophysiology, 1988, 25: 81-91.

Ogilvie, R. D. & Wilkinson, R. T. The detection of sleep onset: Behavioural & physiological convergence. Psychophysiology, 1984, 21: 510-520

Oken BS, Zajdel D, Kishiyama S, Flegal K, Dehen C, Haas M, Kraemer DF, Lawrence J, Leyva J. Randomized, controlled, six-month trial of yoga in healthy seniors: effects on cognition and quality of life. Alternative Therapies In Health And Medicine. 2006, 12(1):40-7.

Pal, N., Chuang, C., Ko, L., Chao, C., Jung, T, Liang, S and Lin, C. EEG based subject and session independent drowsiness detection; an unsupervised approach. Journal on Advances in Signal Processing, Vol. 2008, P. 192.

Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I. and Bekiaris, E. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. Clinical Neurophysiology, 2007, 118: 1906-1922.

Parasuraman, R. & Davies, D. R. Varieties of attention, Academic Press, Orl&o, 1984.

Parasuraman, R., Warm, J. S., & See, J. E. Brain systems of vigilance. In The Attentive Brain, 1998, pp. 221–256.

Parker, C. An Analysis of Performance Measures for Binary Classifiers, Proc. 2011 IEEE International Conference on Data Mining, pp. 517-526

Peiris, M. T. R, Davidson, P. R., Bones, P. J., & Jones, R. D. Detection of lapses in responsiveness from the EEG. Journal of Neural Engineering, 2011, 8 (016003)(1), 1-15.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., & Bones, P. J. Event-based detection of lapses of responsiveness. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine & Biology Society, 2008, 30, 4960-4963.

Peiris, M. T. R. Lapses in Responsiveness: Characteristics & Detection from the EEG. Doctoral Thesis at the University of Canterbury, 2008.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., & Bones, P. J. Detecting behavioral microsleeps from EEG power spectra. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 2006a 28, 5723-5726

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., & Bones, P. J. Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep deprived subjects. Journal of Sleep Research, 2006b 15(3), 291-300.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Bones, P. J., & Myall, D. J. Fractal dimension of the EEG in detection of behavioural microsleeps. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 2005a, 27, 5742-5745.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., Signal, T. L., Parkin, P. J., van den Berg, M., & Bones, P. J. Identification of vigilance lapses using EEG/EOG by expert human raters. Proceedings of the Annual International Conference of IEEE Engineering in Medicine & Biology Society, 2005b, 27, 5735-5738.

Philibert, I. Sleep loss and performance in residents and nonphysicians: a meta-analytic examination, Sleep, 2005, 28: 1392–402.

Philip, P., Vervialle, F., Le Breton, P., Taillard, J. & Horne, J. A. Fatigue, alcohol, & serious road crashes in France: factorial study of national data. British Medical Journal, 2001, 322: 829-830.

Pomfrett, C. J. D., & Pearson, A. J. EEG monitoring using bispectral analysis. Proceedings of IEEE Colloquium on New Measurements & Techniques in Intensive Care, 1996, 179, 1-3.

Ponulak, F & Kasinski A. Introduction to spiking neural networks: Information processing, learning and applications. Acta Neurobiol Exp (Wars), 2011, 71(4):409-33.

Portas, C. M., Rees, G., Howseman, A. M., Josephs, O., Turner, R., & Frith, C. D. A specific role for the thalamus in mediating the interaction of attention and arousal in humans. Journal of Neuroscience, 1998, 18: 8979-89

Porte, H. S. Morphology of slow eye movement at sleep onset. Sleep, 2003, 26: A71

Posner, M. I. and Petersen, S. E. The attention system of the human brain. Annual Review of Neuroscience, 1990, 13:25–42.

Poudel, G. R., Innes, C. R. H., Bones, P. J., Watts, R., & Jones, R. D. Losing the struggle to stay awake: Divergent thalamic & cortical activity during microsleeps. Human Brain Mapping, 2014, 35, 257-269.

Poudel, G. R., Innes, C. R. H., Bones, P. J., & Jones, R. D. The relationship between behavioural microsleeps, visuomotor performance & EEG theta. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 2010, 32, 4452-4455.

Poudel, G. R.Functional magnetic resonance imaging of lapses of responsiveness during visuomotor tracking. Doctoral Thesis at the University of Otago, 2009.

Poudel, G. R., Jones, R. D., & Innes, C. R. H. A 2-D pursuit tracking task for behavioural detection of lapses. Australasian Physical & Engineering Sciences in Medicine, 2008, 31, 528-529.

Poudel, G. R., Jones, R. D. & Davidson, P. R. Proposed multimodal study of lapses in responsiveness with simultaneous haemodynamic, electrophysiological, & behavioural recording in The fMRI Experience, Melbourne, 2006.

Poudel, G. R., Innes, C. R. H., Bones, P. J., & Jones, R. D. The relationship between behavioural microsleeps, visuomotor performance & EEG theta. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 32, 4452-4455, 2010.

Ranzato, M., Taylor, P. E., House, J. M., Flagan, R. C., LeCun, Y., and Perona, P.: Automatic recognition of biological particles in microscopic images, Pattern Recognition Letters, 2007, 28, 31–39.

Raymer, M.L., Punch, W.F., Goodman, E.D., Kuhn, L.A. and Jain, A.K., Dimensionality reduction using genetic algorithms. IEEE Transactions on Evolutionary Computation, 4:164–171, 2000.

Rechtschaffen, A. and Kales, A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. University of California, Brain Information Service/Brain Research Institute, Los Angeles, 1968.

Repovš G: Dealing with Noise in EEG Recording and Data Analysis. Informatica Medica Slovenica 2010; 15(1).

Reyner, L. A. & Horne, J. A. Evaluation "in-car" countermeasures to sleepiness: cold air & radio. Sleep, 1998, 21: 46-50.

Ripley, B. D. Pattern recognition and neural networks, Cambridge University Press, Cambridge, 1996.

Roth, M., Shaw, J. an dGreen, J. The form, voltage distribution and physiological significance of the K-complex. EEG Clinical Neurophysiology, 1956, 8: 385–402.

Roweis, S. "EM Algorithms for PCA and SPCA." In Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems. Vol.10 (NIPS 1997), Cambridge, MA, USA: MIT Press, 1998, pp. 626–632.

Ruping, S. Incremental learning with support vector machines. Proceedings of the IEEE International Conference on Data Mining, 2001, 1, 641-642.

Russell, S. J. and Norvig, P. Artificial Intelligence: A Modern Approach (2 ed.). Pearson Education.

Sagberg, F. Road accidents caused by drivers falling asleep. Accident Analysis & Prevention, 1999, 31: 639-649.

Samuel, A. L. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 1959, 3(3):211-229. .

Santamaria, J. and Chiappa, K. H. The EEG of Drowsiness in Normal Adults. Journal of Clinical Neurophysiology, 1987, 4: 327-382.

Schacter, D.L. & Dodson, C.S. Misattribution, false recognition, and the sins of memory. Philosophical Transactions of the Royal Society (B), 2001, 356, 1385-1393.

Schrauwen, B., Defour, David, J., Verstraeten, D, Van Campenhout J. M.. The introduction of time-scales in reservoir computing, applied to isolated digits recognition. In proceedings of the 17th International Conference on Articial Neural Networks, 2007, vol. 4668 of LNCS, pp. 471-479.

Segalowitz, S., J., Velikonja, D., Baker, J. Attentional capacity in walking arousal. In R. D. Oglivie (Ed) Sleep onset: Normal and abnormal processes: American Psychological Association, Washington, 1994.

Sejnowski, T. J. and Destexhe, A. Why do we sleep? Brain Research, 2000, 886: 208-223.

Sharbrough, F. W., Electrical fields & recording techniques. In Daly & Pedley, 1990, 37.

Shawe-Taylor, J. and Cristianini, N. Kernel methods for pattern analysis. Cambridge university press, 2004.

Shen, J., Barbera, J. and Shapiro, C. M. Distinguishing sleepiness and fatigue: focus on definition and measurement. Sleep Medicine Reviews, 2006, 10: 63-76.

Shen, W., Sun, H., Cheng, E., Zhu, Q., Li, Q. Effective driver fatigue monitoring through pupil detection and yawing analysis in low light level environments. International Journal on Digital Technological Applications; 2012, 6: 372–383.

Siegel, J. M. Clues to the functions of mammalian sleep. Nature, 2005, 437, 1264-1271.

Spiegel, K., Leproult, R. and Van Cauter, E. Impact of sleep debt on metabolic and endocrine function. Lancet, 1999, 354: 1435– 1439.

Silber, M, H., Ancoli-Israel, S., & Bonnet, M. H. The visual scoring of sleep inadults. Journal of Clinical Sleep & Medicine, 2007, 3:121-131.

Smith, P., Shah, M., and Vitoria, L. N. Determining driver visual attention with one camera. IEEE Transactions on Intelligent Transport Systems; 2003, 4:205–218.

Sörnmo, L., & Laguna, P. Bioelectrical signal processing in cardiac and neurological applications (1st ed.), Elsevier, 2005, 0-12437-552-9.

Stampi, C. Sleep & circadian rhythms in space. Journal of Clinical Pharmacology, 1994, 34: 518–534.

Steriade, M., McCormick, D. A. and Sejnowski, T. J. Thalamocortical oscillations in the sleeping and aroused brain. Science, 1993, 262: 679-685.

Stern, J.A., Boyer, D., Schroeder, D. Blink rate: a possible measure of fatigue. Human Factors. 1994, 36(2):285-97.

Stern, R. M., W. J. Ray, et al. Psychophysiological Recording Oxford, Oxford University Press, 2001.

Stricker, J. L., Brown, G. G., Wetherell, L. A., & Drummond, S. P. A. The impact of sleep deprivation and task difficulty on networks of FMRI brain response. Journal of the International Neuropsychological Society, 2006: 12, 591-7.

Subasi, A., Alkan, A., Koklukaya, E., & Kiymik, M. K. Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing. Neural Networks, 2005, Vol.18, No.7, pp. 985-997.

Tefft, B. C., Williams, A. F., & Grabowski, J. G. Driver licensing & reasons for delaying licensure among young adults ages 18–20, United States, 2012. Injury Epidemiology, 2014, 1(1), 4–4.

Tipping, M. E., and C. M. Bishop. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society. Series B (Statistical Methodology), Vol. 61, No.3, 1999, pp. 611–622.

Tirunahari, V. L., Zaidi, S. A., Sharma, R., Skurnick, J., Ashtyani, H. Microsleep and sleepiness: a comparison of multiple sleep latency test and scoring of microsleep as a diagnostic test for excessive daytime sleepiness. Sleep Medicine, 2003, 4(1):63–67

Toppi, J., Astolfi, L., Poudel, G. R., Babiloni, F., Macchiusi, L., Mattia, D., & Jones, R.D. Time-varying functional connectivity for underst&ing the neural basis of behavioral microsleeps. Proceedings of Annual International Conference of IEEE Engineering in Medicine & Biology Society, 34, 6192 - 6195, 2012.

Torsvall, L. & Åkerstedt, T. Sleepiness on the job: continuously measured EEG changes in train drivers. Electroencephalography & Clinical Neurophysiology, 1987, 66: 502-511.

Torsvall, L. & Åkerstedt, T. Extreme sleepiness: quantification of EOG & spectral EEG parameters. International Journal of Neuroscience, 1988, 38: 435-441.

Townsend, R.E. and Johnson, L.C. Relation of frequency-analyzed EEG to monitoring behavior. Electroenceph. clin. Neurophysiology, 47:272-279, 1979.

Tsodyks, M., Pawelzik, K., & Markram, H. Neural Networks with Dynamic Synapses. Neural Computation. 10(4): 821-835, 1998.

Tsuchida, K., Nakatani, M., Hitachi, K., Uezumi, A., Sunda, Y., Ageta, H., Inokuchi, K. Activing signaling as an emerging target for therapeutic interventions. Cell Communication and Signaling 2009, 7:15.

Turing, A. M. Computing Machinery and Intelligence. Mind 49: 433-460, 1950.

Vallet, P.G., Charnay, Y., Bouras, C. Constantinidis, J. Distribution & colocalization of delta sleep inducing peptide (DSIP) with corticotropin-like intermediate lobe peptide (CLIP) in the human hypophysis. Neuroscience Letters, 1988, 90, pp. 78–82

Valley, V. & Broughton, R. The physiological (EEG) nature of drowsiness & its relation to performance deficits in narcoleptics. Electroencephalography & Clinical Neurophysiology, 1983, 55: 243-251.

Van der Maaten, L. J. P. Feature Extraction from Visual Data. PhD Thesis (cum laude), Tilburg University, The Netherlands, 2009.

Vanlaar, W., Simpson, H., Mayhew, D., Robertson, R. Fatigued and Drowsy Driving: A survey of attitudes, opinions and behaviors. Journal of Safety Research, 2008, 39: 303-309

Venna, J. Dimensionality reduction for visual exploration of similarity structures. Doctoral thesis, at Helsinki University of Technology, 2007.

Verstraeten, D., Dambre, J., Dutiot, X. & Schrauwen, B. Memory versus non-linearity in reservoirs. In Proceedings of the IEEE International Joint Conference on Neural Networks, 2010, pp. 1-8.

Verstraeten, D. Reservoir computing: computation with dynamical systems. PhD thesis, Doctoral University. Faculty of Engineering, Ghent, Belgium, 2009.

Verstraeten, D., Schrauwen, B., D'Haene, M., Stroobandt, D. An experimental unification of reservoir computing methods. Neural Networks, 2007, 20(3):391-403.

Verstraeten, D., Schrauwen, B. & Stroobandt, D. Reservoir-based techniques for speech recognition. In Proceedings of the IEEE International Joint Conference on Neural Networks, 2006, pp. 1050-1053.

Viola, P. and Jones, M.J. 'Robust real-time face detection', International Journal of Computer Vision, 2004, Vol. 57, pp. 137–154.

von Economo, C. Sleep as a problem of localization. Journal of Nervous and Mental Disease, 1930, 71: 249–259.

Vreeken, J. Spiking Neural Networks, an Introduction. Technical Report. Adaptive Intelligence Laboratory, Institute for Information and Computing Sciences, Utrecht University, 2003.

Vuckovic, A., Radivojevic, V., Chen, A. C. and Popovic, D. Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. Medical Engineering and Physics, 2002, 24: 349-360.

Walsh, J. K., Muehlbach, M. J. and Schweitzer, P. K. Simulated assembly line performance following ingestion of cetirizine or hydroxyzine. Annals of Allergy, 1992, 69: 195-200.

Wang, D., Miao, D., and Blohm, G. Multi-class motor imagery EEG decoding for brain-computer interfaces" Frontiers in Neuroscience, 2012, 6: l-13.

Warren, N. and Clark, B. Blocking in mental and motor tasks during a 65-hour vigil. Journal of Experimental Psychology, 1937, 21 (1):97.

Weinger, M. B., & Ancoli-Israel, S. Sleep deprivation and clinical performance. Journal of the American Medical Association, 2002, 287: 955–7.

Wierwille, W. W. & Ellsworth, L. A. Evaluation of driver drowsiness by trained raters. Accident Analysis & Prevention, 1994, 26: 571-581.

Wilkinson, R. T., & Houghton, D. Portable four-choice reaction time test with magnetic tape memory. Behavior Research Methods, Instruments, and Computers, 1975, 7: 441-446.

Wilkinson RT. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. Journal of Experimental Psychology, 1961, 62(3):263–271.

Williams, H. L., Lubin, A., & Goodnow, J. J. Impaired performacne with acute sleep loss. Psycholology Monographs: General & Applied. Nursing Research, 1959. 73, 1-25.

Wolpert, D. Stacked generalization. Neural Networks, 2002, 5: 241-259.

Woergoetter, F., and Porr B. Reinforcement learning. Scholarpedia, 2008, 3(3):1448.

Xue, T. Z., Nan, N. Z., Fan, M., Yong, J. H. Head Pose Estimation Using Isophote Features for Driver Assistance Systems. Proceedings of the IEEE Intelligent Vehicles Symposium; 2009, pp. 568–572

Zhang, Z and Parhi, K. K. Low-Complexity Seizure Prediction From iEEG/sEEG using Spectral Power and Ratios of Spectral Power, IEEE Transactions on Biomedical Circuits and Systems, 10(3), pp. 693-706, June 2016.

Zhang, Z and Parhi, K. K., Seizure Prediction using Polynomial SVM Classification, Proc. of 2015 IEEE Engineering in Medicine and Biology Society Conference (EMBC), pp. 5748-5751, Milan, Italy, August 2015, 2015(a).

Zhang, Z and Parhi, K. K., Seizure Detection using Regression Tree Based Feature Selection and Polynomial SVM Classification, Proc. of 2015 IEEE Engineering in Medicine and Biology Society Conference (EMBC), pp. 6578-6581, Milan, Italy, August 2015, 2015(b).

Zisapel, N. The Role of Melatonin in Sleep Regulation. Neuroendocrine Correlates of Sleep/Wakefulness, Chapter 10, Springer, 2005, 295-309.

Appendix A - Simulated EEG-events study

In this study, to approximate microsleep-type transient events, simulated artificial events were superimposed on subsets of EEG data from Study A. The main objectives behind this study were to:

- determine the sensitivity of various classifier configurations with the presence of Gaussian noise sources and,
- determine the detection performance of various detection systems/ configurations on a gold-standard dataset for which the events were precisely known.

An 'event' in this artificial data was a 15-Hz sinusoidal burst, lasting 2.0 s. EEG of 300 s (5 min) duration were taken from each of the 8 subjects. A total of 34 spectral features were extracted at 1.0-s intervals from each channel, resulting in 544 features. The sine wave was scaled relative to the EEG signal to generate data with amplitude signal-to-noise ratios of SNR = 16, 3, 1, 0.3 & 0.03, to allow exploration of effect of the reduction of SNR would have on classification performance.

To emulate a microsleep event, six of the 2-s segments had 15-Hz sinusoidal bursts, equating to 2% of the epochs being events and 98% non-events. This highly unbalanced dataset parallels that of the relatively rare, occurrence of microsleeps (Peiris *et al.*, 2011). Additionally, five balanced datasets, each with 150 events and 150 non-events, were created with SNRs identical to those of the unbalanced datasets. Examples of an artificial event (15 Hz sinusoidal burst) superimposed on the ambient EEG are depicted in the Figure A.1.

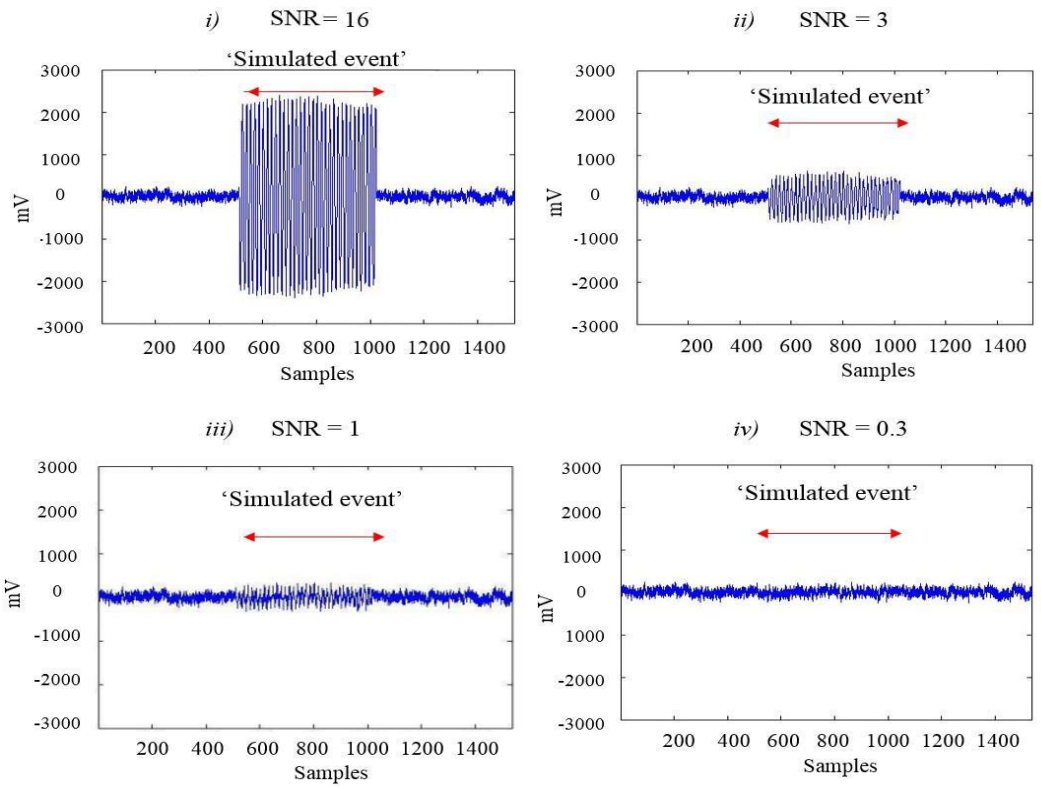


Figure A-1 An artificial event (15-Hz sinusoidal burst) superimposed on the ambient EEG. Sub-figure (*i*) depicts a combination of EEG and a simulated event (SNR=16). Sub-figure (*ii*) depicts a combination of EEG and a simulated event (SNR=3). Sub-figure (*iii*) depicts a combination of EEG and a simulated event (SNR = 1). Sub-figure (*iv*) depicts a combination of EEG and a simulated event (SNR=0.3).

A.1 Validation of the classifier systems on definitive gold standard

This study on the simulated EEG-events allowed for performance on known SNRs to be compared. Unlike simulated events, real microsleeps, often tend to have a variable and unknown SNR exacerbated due to class imbalances. Additionally, it is also important to differentiate EEG noise from 'noise' due to class imbalances. Given a perfect recording medium, both the class imbalances and degrees of white Gaussian noise will always exist. Therefore, a detection system will be challenged to produce accurate results on both fronts, however such challenges are from two completely different sources. A major difference between the artificial event datasets and the rated EEG feature sets was that the target events in the artificial sets were consistent in duration, total number, and amplitude for all subjects. This is unlike the rated EEG feature sets discussed in Chapter 6. The artificial event datasets offered a view of uniform and consistent features across the same dataset.

Several feature selection/reduction, and classifier algorithms discussed in chapter 4 were tested and validated on the simulated dataset. Feature reduction/selection modules included algorithms such as, PCA or ADEN and event-based feature selection (EFS). Both the LDA

and ESN modules (sigmoidal and cascaded-leaky neuron) were trained with features from PCA, ADEN, and EFS.

As stated in chapter 4, ESNs transform input to an excited state where they can be linearly mapped to outputs using a training set, and where state vectors represent intermediary functions that facilitate such mapping. In this simulation, results from the sigmoidal ESN approach provided insight that warranted further investigation and led to a more complex architecture that utilized cascaded-leaky integrator neurons. The cascaded-leaky neuron models are highly effective when modelling continuously slow transforming systems. The major advantage with cascaded-leaky ESNs is that even small networks (e.g., 5 cascaded-leaky integrator neurons) are capable of achieving superior performance. However, one of their drawbacks is the possible over-fitting within the network, especially if the global parameters of the network and the reservoir size are not chosen carefully. Therefore, several system parameters from simple Euler approximations were adjusted to 'tune' the network for optimum performance.

Moreover, as the weight matrices are randomly generated and fixed for the entire training cycle, this model exhibits small variations in terms of the results obtained. This is considered to be the typical ESN behaviour, because of the (pseudo) randomly-generated reservoir. Therefore, this warranted further development of the system, requiring interconnections of several individual networks to form a combined classifier (ensemble learning method). Class hypotheses from each of these classifiers were then combined and the mean of the individual votes calculated for each of the classifiers. Calculating the mean of the vote combination was used because it averages out vote fluctuations due to any single classifier's biases. Accordingly, 500 individual classifier models with 5 cascaded-leaky neuron ESNs each were trained on the simulated data to form a combined classifier model as depicted in Figure 5-3.

A.2 Results

Multiple tests were performed on all the classifier module and the same methods used for the performance analysis in prior research by Peiris *et al. (*2011) & Davidson *et al. (*2007) were reproduced. Classification performance was determined by leave-one-out cross-validation of the artificial datasets corresponding to the 8 subjects. Performance metrics used were mean accuracy, sensitivity, specificity, selectivity, and phi correlation. Phi correlation was used as the primary performance metric because of its reasonable independence from class distributions, in addition to being the best integrated measure of the other performance metrics (LaRocco, 2015). Phi values ranged from +1 to -1 with +1 being perfect positive correlation, the value -1 meaning a perfect negative correlation and the value 0 being no correlation.

LDA was used to form classification models as baselines for performance. Tests included cross-validation on both the single & multiple classifier modules (ensemble learning) on both the balanced and unbalanced datasets, a hypothesis was that ensembles would perform noticeably better than with single classifiers in almost all cases.

PCA was found to be the least performing feature reduction technique in terms of its phi correlation on both the balanced and unbalanced datasets. Whereas, ADEN, MAD and EFS performed comparably well on datasets with low signal-to-noise (SNR = 0.3). In addition, none of the methods tested were able to correctly classify the datasets with the lowest signal-to-noise (SNR = 0.03), balanced or unbalanced. Figure 6-8 depicts the performance metrics for all feature selection/reduction modules (PCA, ADEN with 10 features, ADEN with 1 feature, and EFS) on the SNR = 0.3 (left) and SNR = 0.03 unbalanced datasets (right).

While not presented in this appendix, ESNs with cascaded-leaky integrator modules performed the highest upon the balanced datasets. However, differences in performance were not as high as in the unbalanced datasets. This may be because while performing ADEN, averaging a larger number of non-events decreases background noise than a smaller subset like in the balanced data (LaRocco, 2015). Cascaded-leaky-integrator ESNs, when used with EFS method, were found to marginally dominate all other modules, in terms of performance, by achieving the highest phi correlation of 0.92 on the dataset with low signal-to-noise (SNR = 0.3) compared to any other classifier.

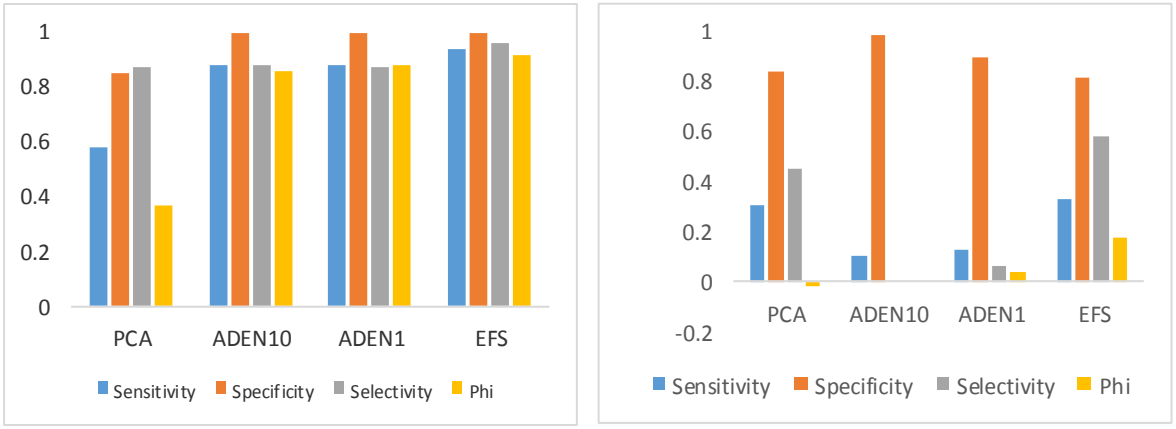


Figure A-2 Performance on feature selection modules with ESN with cascaded-leaky-integrator neurons for the SNR = 0.3, unbalanced 'hard dataset' (left) and SNR = 0.03, 'unbalanced very hard dataset' (right).

ADEN was another feature selection method which performed consistently with the exception of the lower signal-to-noise datasets (SNR = 0.03). Both the ESN modules, cascaded-leaky and sigmoidal connections performed better compared to LDA. More specifically, on the datasets with SNR of 0.3, LDA using PCA, had no appreciable performance in contrast to cascaded-leaky-integrator ESN using PCA for which phi was 0.37.

Individual performance of the cascaded-leaky-integrator ESN and LDA classifier modules on different SNRs is presented in the Figure A-3.

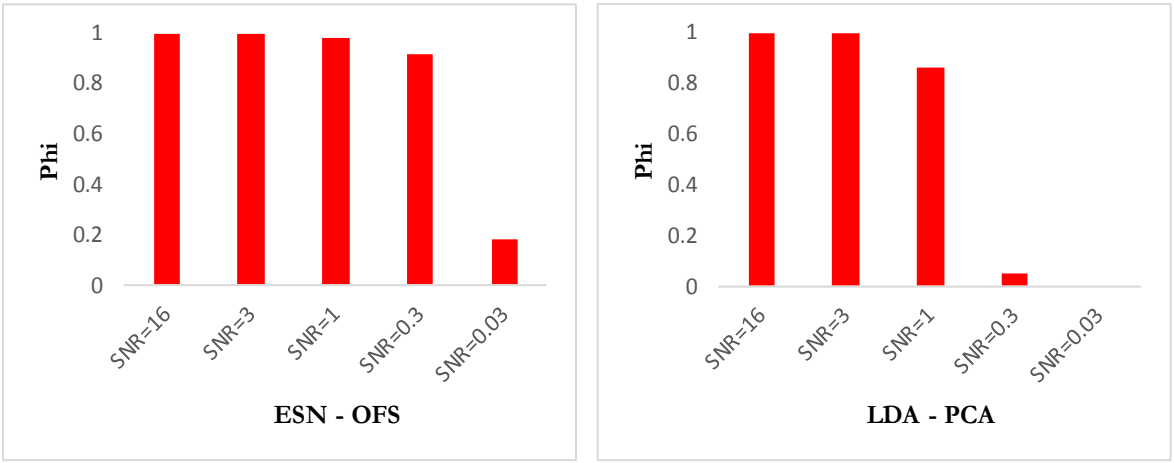


Figure A-3 Classification performance for cascaded-leaky ESN-OFS (left) and LDA-PCA (right) on the unbalanced data (SNR = 16.0 to 0.03).

As the simulated event decreased in amplitude, became harder for a classifier to discern from the background EEG. It appeared as if the amplitude of the simulated event has dropped off to an extent which was not distinguishable from the background EEG, and this phenomenon is evident in Figure A-4. The drop in performance occurred independently of whether the data was balanced or unbalanced. A possible reason for this sudden decrease in performance may be the EEG data used in this 'simulated' study has been contaminated with other noise sources such as the ocular and motion artefacts.

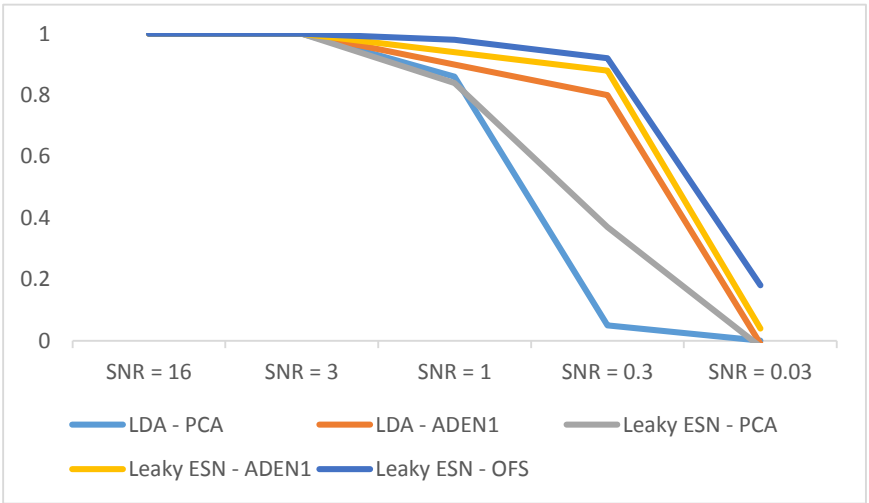


Figure A-4 Performance of various pre-processing and classifier methods across the different SNRs with respect to the background EEG.

A.3 Summary

As expected, by decreasing the amplitude of the simulated event, it became harder for the classifier to differentiate from the background EEG. For the SNR = 0.03 datasets, EFS with cascaded-leaky-integrator ESN was the only module which resulted in a marginal phi correlation value at 0.18. Additionally, ESN with the cascaded-leaky-integrator neurons method was found out to be a superior classifier, compared to LDA in terms of its classification performance. LDA on the (SNR = 0.3) dataset has shown very poor performance, in contrast to ESN with cascaded-leaky-integrator neuron on the same dataset. Results from all figures in the previous section indicate that the application of cascaded-leaky-integrator neuron scheme to the ESNs lead to a consistent and higher performance scores in this study. Therefore, from this experiment it is evident that the ESNs with cascaded-leaky integrator neuron based approaches may prove more suitable for the actual microsleep detection task.

Another assumption was that ensemble learning would perform noticeably better than single classifiers. However, neither the single classifier nor the ensembles could correctly identify the simulated events in the lower signal-to-noise SNR = 0.03 unbalanced datasets. While statistical significance tests were not performed, many of the results were only slightly better than single classifiers. On the artificial data, ADEN, with any single classifier or ensemble, yielded the highest consistent performance across all the SNRs. The major advantage with ADEN was its ability to select features based on *a priori* knowledge of class differences compared to that of PCA which selects meta-features that may be uncorrelated to the target event. Unlike an actual microsleep, the 15-Hz simulated event were found to be substantially more consistent than background EEG. Therefore, all system configurations investigated had to be re-evaluated before any could be eliminated.

Appendix B – Trustworthiness scores

Table B-1 Mean trustworthiness scores for KPCA modules with variances of 1, 2, 3, 5 and 10 on the subjects 804 - 810 of Study A.

Gamma (variance)	Trustworthiness
1	0.4010
2	0.4010
3	0.4010
5	0.4009
10	0.4009

Table B-2 Mean trustworthiness scores for PPCA modules with no. of iterations of 10, 50, 100, 200 and 300 on the subjects 804 - 810 of Study A.

No. of iterations	Trustworthiness
10	0.5412
50	0.5412
100	0.5412
200	0.5441
300	0.5412

Table B-3 Mean trustworthiness scores for Isomap modules with no. of neighbours of 1, 5, 10, 15 and 20 on the subjects 804 - 810 of Study A.

No. of neighbours	Trustworthiness
1	0.1037
5	0.1037
10	0.1038
15	0.1038
20	0.1038

Table B-4 Trustworthiness scores for Laplacian Eiganmaps modules with variances of 1, 2, 3, 5 and 10 and no. of neighbours of 1, 5, 10, 15 and 20, respectively on the subjects 804 - 810 of Study A.

No. of neighbours	Variance	Trustworthiness
1	1	0.1511
5	2	0.1511
10	3	0.1510
15	5	0.1510
20	10	0.1509

Table B-5 Trustworthiness scores for SNE modules with perplexity of 5, 10, 25, 40 and 50 on the subjects 804 - 810 of Study A.

Perplexity	Trustworthiness
5	0.3442
10	0.3442
25	0.3444
40	0.3443
50	0.3441