THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Design and optimisation of time-frequency analysis for multichannel neonatal EEG background features in term neonates with hypoxic ischaemic encephalopathy: characterisation, classification and neurodevelopmental outcome prediction

Md Abdul Awal

Bachelor of Science (Electronics and Communication Engineering)

Master of Science (Biomedical Engineering)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2017

Faculty of Medicine

Abstract

Neonatal Hypoxic Ischaemic Encephalopathy (HIE) is a major cause of morbidity and mortality in newborns throughout the world with an incidence of 1-8/1000 newborns in developed countries and up to 26/1000 newborns in developing countries. It causes one million neonatal deaths globally per year. Electroencephalography (EEG) is a useful method for assessment of brain activity in newborns with HIE. EEG performs well in the early diagnosis and classification of HIE severity and in predicting neurodevelopmental outcome. A normal EEG is highly predictive of a normal outcome and various abnormal EEG features are associated with neurological abnormalities or death.

A systematic review and meta-analysis of the literature was undertaken to determine which specific background features of the EEG best predict outcome (**Aim 1**). Automatic detection and classification would be useful for clinicians and reduce workload and subjectivity. Signal processing approaches particularly in the joint time-frequency domain and machine learning techniques have been used to characterise (**Aim 2**), and to detect and classify (**Aim 3**) these identified features. The final goal (**Aim 4**) was to identify the combination of EEG signal features that best predict neurodevelopmental outcome in a cohort of term neonates with HIE.

The first contribution of this project was to identify the specific EEG background patterns, in term neonates with HIE, that best predict neurodevelopmental outcome through a systemic review of the published literature. A meta-analysis was performed to establish the prognostic value of the identified EEG background patterns and the pooled sensitivity and specificity were calculated. A significant problem was identified: the use of different definitions of the different abnormal EEG features with respect to voltage level, phase or frequency. Agreement on definitions is necessary for the effective implementation and use of EEG in NICUs, I have called for the adoption of specific definitions.

The second contribution was to the optimal use of time-frequency distributions (TFD) for the characterisation of EEG signals and other non-stationary signals. Two optimisation methods, one a global method using a hybrid genetic algorithm and a second, a local optimisation method using a locally optimised spectrogram, have been proposed to optimise the TFD. These contributions present a data-adaptive kernel and analysis-window that reduces the presence of cross-terms and enhances concentration and resolution. The global optimisation method uses single global parameters whereas the local optimisation method divides the TFD into small grids and performs local optimisation. Both methods are applied automatically, and no user input is necessary. Using both simulated signals and real EEG, both methods improved characterisation. The optimised TFDs are not only suitable for

characterisation of non-stationary signals, but also have applications in signal detection and classification.

The third contribution was to the design and optimisation of classification of multi-channel EEG background patterns in term neonates with HIE. Two classification methods utilising a single feature subset and a class-specific feature subset have been designed and optimised to classify five EEG background patterns: burst, suppression, normal, seizure and artefact. These patterns are clinically relevant to treatment and prediction of neurodevelopmental outcomes. Various time domain, frequency domain and joint time-frequency domain features are first extracted and then concatenated to produce a long feature set. A hybrid feature selection (HFS) algorithm was proposed to simultaneously select the prominent feature subset as well as the support vector machine (SVM) tuning parameters. Three fusion techniques, channel fusion, feature fusion and decision fusion, have been used in the multi-channel classification. Finally, two outputs of a decision support system (DSS), *'EEG state'* and *'EEG quality'*, are presented as useful parameters to assist clinical staff in management of babies with HIE.

The fourth contribution was the conduct of a study to test the relationship between the identified EEG features in term neonates with HIE and neurodevelopmental outcome at the age of 2 years. A highly discriminative and non-redundant feature subset as well as SVM parameters have been selected from this high dimensional feature set using a HFS algorithm. This algorithm is applied to the statistically most unbiased Leave-One-Subject-Out (LOSO) cross-validation and an optimised model is created using the most consistent feature subset. This model predicts good/poor neurodevelopmental outcome with 83.77% accuracy when applied to a separate dataset and improved the prediction accuracy of other approaches/studies by 5–10%. A DSS has been built as a potential application of the model to visualise the *'probable long-term neurodevelopmental outcome'* in a continuous probabilistic fashion. These results provide strong support for a future objective decision support tool for the early prediction of neurodevelopmental outcome for babies with HIE.

In summary, this project makes significant contributions to optimise use of TFDs and SVM for EEG signal characterisation and the detection and classification of abnormal EEG patterns used in the prediction of neurodevelopmental outcome and suggests future directions for research and translation to clinical practice.

Declaration by author

This thesis *is composed of my original work, and contains* no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted *to qualify for the award of any* other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Journal papers:

[1] M. A. Awal, M. M. Lai, G. Azemi, B. Boashash, and P. B. Colditz, "EEG background features that predict outcome in term neonates with hypoxic ischaemic encephalopathy: A structured review," Clinical Neurophysiology, vol. 127, pp. 285-296, 2016.

[2] M. A. Awal and B. Boashash, "An automatic fast optimization of Quadratic Time-frequency Distribution using the hybrid genetic algorithm," Signal Processing, vol. 131, pp. 134-142, Feb, 2017.

[3] M. A. Awal, S. Ouelha, S. Dong, and B. Boashash, "A robust high-resolution time-frequency representation based on the local optimization of the short-time fractional Fourier transform," Digital Signal Processing, vol. 70, pp. 125-144, Aug, 2017.

Book Chapter:

[1] B. Boashash, M. A. Awal and S. Dong, “Gabor Spectrogram, WVD, and Energy Atoms,” in Time-Frequency Signal Analysis and Processing, ed. B. Boashash, Elsevier, 2nd Ed., 2015.

Conference papers:

[1] M. A. Awal, P. B. Colditz, B. Boashash, and G. Azemi, "Detection of neonatal EEG burst-suppression using a time-frequency approach," in, 2014 8th International Conference on Signal Processing and Communication Systems (ICSPCS), 2014, pp. 1-6.

[2] M. A. Awal, M. Khlif, S. Dong, G. Azemi, P. B. Colditz, and B. Boashash, "Detection of neonatal EEG burst-suppression using time-frequency matching pursuit," in Australian Biomedical Engineering Conference (ABEC 2015), 2015.

Publications included in this thesis

Journal papers:

[1] M. A. Awal, M. M. Lai, G. Azemi, B. Boashash, and P. B. Colditz, "EEG background features that predict outcome in term neonates with hypoxic ischaemic encephalopathy: A structured review," Clinical Neurophysiology, vol. 127, pp. 285-296, 2016.

- Incorporated in Chapter 2.

Contributor	Statement of contribution
M. A. Awal	Formed the concept (60%)

	Conducted the analysis (80%) Wrote the first draft (90%) Reviewed and edited the manuscript (60%)
M. M. Lai	Conducted the analysis (20%) Wrote the first draft (10%) Reviewed and edited the manuscript (10%)
G. Azemi	Reviewed and edited the manuscript (5%)
B. Boashash	Reviewed and edited the manuscript (5%)
P.B. Colditz	Formed the concept (40%) Critically reviewed and edited the manuscript (20%)

[2] M. A. Awal and B. Boashash, "An automatic fast optimization of Quadratic Time-frequency Distribution using the hybrid genetic algorithm," Signal Processing, vol. 131, pp. 134-142, Feb, 2017.

- Incorporated in Chapter 3.

Contributor	Statement of contribution
M. A. Awal	Formed the concept (90%) Conducted the analysis (100%) Wrote and edited the manuscript (80%)
B. Boashash	Formed the concept (10%) Critically reviewed and edited the manuscript (20%)

[3] M. A. Awal, S. Ouelha, S. Dong, and B. Boashash, "A robust high-resolution time-frequency representation based on the local optimization of the short-time fractional Fourier transform," Digital Signal Processing, vol. 70, pp. 125-144, Aug, 2017.

- Incorporated in Chapter 3.

Contributor	Statement of contribution
M. A. Awal	Formed the concept (80%) Conducted the analysis (80%) Wrote the first draft (100%) Reviewed and edited the manuscript (60%)
S. Ouelha	Conducted the analysis (20%) Reviewed and edited the manuscript (15%)
S. Dong	Reviewed and edited the manuscript (5%)
B. Boashash	Formed the concept (20%) Critically reviewed and edited the manuscript (20%)

Book Chapter:

[1] B. Boashash, M. A. Awal and S. Dong, “Gabor Spectrogram, WVD, and Energy Atoms,” in Time-Frequency Signal Analysis and Processing, ed. B. Boashash, Elsevier, 2nd Ed., 2015.

- Incorporated in Chapter 4.

Contributor	Statement of contribution
B. Boashash	Formed the concept (50%) Wrote the article (50%) Critically reviewed and edited the article (30%)
M. A. Awal	Formed the concept (50%) Collected neonatal EEG data (100%) Marked the data with neurologists (100%) Conducted the analysis (100%) Wrote the article (50%) Edited the article (50%)
S. Dong	Critically reviewed and edited the article (20%)

Conference papers:

[1] M. A. Awal, P. B. Colditz, B. Boashash, and G. Azemi, "Detection of neonatal EEG burst-suppression using a time-frequency approach," in, 2014 8th International Conference on Signal Processing and Communication Systems (ICSPCS), 2014, pp. 1-6.

- Incorporated in Chapter 3.

Contributor	Statement of contribution
M. A. Awal	Formed the concept (100%) Collected neonatal EEG data (100%) Marked the data with neurologists (100%) Conducted the analysis (100%) Wrote and edited the manuscript (70%)
P. B. Colditz	Reviewed and edited the manuscript (10%)
B. Boashash	Provided technical advice (50%) Reviewed and edited the manuscript (10%)
G. Azemi	Provided technical advice (50%) Reviewed and edited the manuscript (10%)

[2] M. A. Awal, M. Khlif, S. Dong, G. Azemi, P. B. Colditz, and B. Boashash, "Detection of neonatal EEG burst-suppression using time-frequency matching pursuit," in Australian Biomedical Engineering Conference (ABEC 2015), 2015.

- Incorporated in Chapter 4.

Contributor	Statement of contribution
M. A. Awal	Formed the concept (90%)

	Collected neonatal EEG data (100%) Marked the data with neurologists (100%) Conducted the analysis (70%) Wrote and edited the manuscript (60%)
M. S. Khlif	Formed the concept (10%) Conducted the analysis (30%) Reviewed and edited the manuscript (15%)
S. Dong	Reviewed and edited the manuscript (5%)
G. Azemi	Reviewed and edited the manuscript (5%)
P. B. Colditz	Reviewed and edited the manuscript (5%)
B. Boashash	Reviewed and edited the manuscript (10%)

Contributions by others to the thesis

The following researchers contributed to this thesis in terms of the concept and design of the studies; technical work; analysis and interpretation of the results; and drafting or significantly revising different parts of the research:

- Prof. Paul Colditz,
- Prof. Boualem Boashash,
- Dr. Ghasem Azemi,
- Dr. Shiying Dong,
- Dr. Samir Ouelha.

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

All thanks and gratitude are due to the Almighty—the first and foremost.

It would not have been possible for me to complete this thesis without the help, advice, encouragement and support of others and I thank them from the bottom of my heart.

I would like to extend my deepest appreciation to my advisors, Professors Paul Colditz and Boualem Boashash, for providing me with the opportunity of undertaking this research at the Perinatal Research Centre (PRC), University of Queensland Centre for Clinical Research (UQCCR). I bothered you a lot in different circumstances throughout the PhD process; thank you for tolerating me. I owe a debt of gratitude for your tutelage, your clinical and signal processing insights and, above all, your tireless efforts in editing different manuscripts and this thesis.

To Professor Paul—thanks again for giving me a formal exposure to clinical research, providing me with a positive learning environment and introducing me to your kind and helpful colleagues: Dr Shabeed Chelakkadan, Dr Jim Pelekanos, Dr Michael Navakatikyan, Associate Professor Rod Hunt, Samantha Francis-Pester and many more. I am overwhelmed by their invaluable support and help. You make my PhD journey much easier.

To Professor Boualem—I am eternally grateful to you for introducing me to state-of-the-art signal processing research applied to solve clinical problems. Your long-term experience and philosophy toward science and life will continue to influence me for the rest of my life. Thanks also for organising the lab visit to Qatar University and financial support. Thank you!

I sincerely thank Drs: Ghasem Azemi, Shiying Dong, Samir Ouelha and M.S. Khlif for your encouragement and your insightful comments during my research project. Professor Andrew Bradley, Dr Tracey Bjorkman, Dr Susan Sullivan and Dr Julie Wixey played an invaluable role during my three PhD milestones and provided valuable feedbacks. I am also grateful to Sheikh Shanawaz Mostafa, Dr David Green, Dr Tobias Werther and Dr Robert Ware for their support in high-performance computational and statistical analysis. I acknowledge the assistance of the clinical staffs collecting neonatal EEG data. My deepest condolences go to the families of those babies who died or suffered major handicap.

In addition, I acknowledge the financial support of The University of Queensland International (UQI) scholarship and the research assistantship from the Qatar National Research Fund under its National Priorities Research Program award number NPRP 6-885-2-364 and NPRP 4-1303-2-517.

Thanks to all the members of UQCCR: Associate Professor Dr Judith Greer, Dr Barbara Lingwood, Dr Simon Finnigan, Dr Kirat Chand, Tim, Annice, Abi, Mellissa, Steph, Emma, Zhouwei Xu, Ashwag, Aakanksha—to name but a few. I would like to express my deepest gratitude to Dr Janet Hammill—I call her *'My Aussie Mom'*. You have a very beautiful mind! Outside the research environment, I would like to express my deepest gratitude to all the juniors, friends and seniors who made my stay in Australia memorable. My home university, Khulna University, Bangladesh, is acknowledged for the logistic support it provided.

Last but not least, my warmest appreciation goes to my lovely wife, Tamanna, for her understanding and mental support, especially during time of stress associated with the thesis. I wholeheartedly thank to my mother, brother and sister for their endless love and support. I always remember my late father. Baba, I miss you a lot!

Keywords

Neonatal EEG background patterns, hypoxic ischaemic encephalopathy (HIE), systemic review, meta-analysis, Bayley Scales of Infant and Toddler Development, time-frequency distribution, hybrid genetic algorithm (HGA), locally optimised spectrogram (LOS), hybrid feature selection (HFS), support vector machine (SVM), neurodevelopmental outcome prediction.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 090399, Biomedical Engineering not elsewhere classified, 50%

ANZSRC code: 090609, Signal Processing, 50%

Fields of Research (FoR) Classification

FoR code: 0903, Biomedical Engineering, 50%

FoR code: 0906, Electrical and Electronic Engineering, 50%

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADTFD	Adaptive directional time-frequency distribution
BEES	Babies EEG study
BSID-III	Bayley Scales of Infant and Toddler Development–Third Edition
CD	Cognitive delay
CP	Cerebral palsy
CNV	Continuous normal voltage
CLV	Continuous low voltage
DSS	Decision support system
DW-MRI	Diffusion weighted MRI
ED	Early death
ECS	Electro cerebral silence
FT	Flat trace
GLCM	Grey-level cooccurrence matrix
HFS	Hybrid feature selection
HSROC	Hierarchical summary receiver operating characteristics
ISPR	Instantaneous spectral power ratio
LOSO	Leave-one-subject-out cross-validation
mRMR	Minimum redundancy maximum relevance
NEST	Neonatal electrographic seizure trial
OAR	One-against-rest
PMR	Psychomotor retardation
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAR	Robust artifact removal
SSC	Signal strength combination
SQ	Spastic quadriplegia

Chapter 1 Introduction

The human brain is the main control system of the body and its most complex structure, with billions of neurons and many dynamic interactions within the brain's networks. Brain injury is a leading cause of morbidity and mortality around the globe; in the United States alone, an estimated 1.7 million people suffer brain injury each year, and the estimated annual cost of brain injury is over US$60 billion [1]. For number of reasons, the case of neonatal brain injury is even more complex. For example, neonates are unable to verbally communicate with clinicians, making it difficult to identify signs and symptoms of injury. Additionally, the brain requires a continuous supply of oxygen and nutrients as the main fuel of the control system. There are risks of inadequate supply of oxygen and nutrients before, during and after delivery, arising from cord entanglement, placental abruption, antepartum maternal hypotension or trauma, intrapartum uterine hyperstimulation and postnatal cardiopulmonary abnormalities [2]. A global insufficiency of oxygen may result in the clinical condition of Hypoxic Ischaemic Encephalopathy (HIE); brain cells begin dying within minutes without oxygen and nutrients, and this leads to short- and long-term damage, with clinical manifestations including mental retardation, seizures, cerebral palsy, visual and auditory impairments and even death, depending on the area of the brain affected by lack of oxygen [2].

HIE is a common medical term describing brain injury due to lack of oxygen supply. 'Hypoxic' means a decreased amount of oxygen, 'ischaemic' a decrease in blood flow and 'encephalopathy' the injury to the brain; together, 'Hypoxic Ischaemic Encephalopathy' means a decreased amount of oxygen caused by a decrease in blood flow resulting in injury to the brain. HIE can be defined as the presence of any two of three features in the presence of a known perinatal event [3] (cord prolapse or rupture, pathological fetal heart rate tracing, placental abruption, placenta praevia, uterine rupture and shoulder dystocia etc.)—need for respiratory support shortly after birth, an Apgar score at 5 min of < 5, or evidence of acidosis ($pH < 7$) [3].

HIE is a major cause of mortality and morbidity in newborns throughout the world. The World Health Organization (WHO) reports that neonatal deaths account for 37% of mortality in children under five years of age and 23% of these deaths are associated with asphyxia [4] and it causes one million neonatal deaths globally per year [5, 6]. The occurrence of neonatal HIE is up to 8 per 1000 newborns in developed countries, increasing to 26 per 1000 newborns in developing countries [2]. In Europe, HIE is the third most common cause of neonatal mortality, resulting in 21% of term infant deaths [7]. The mortality rate is dependent on the level of HIE severity: 10% for moderate HIE and 60% for severe HIE. About 30% of survivors with moderate HIE, and close to 100% with severe HIE, develop permanent neurological disability [8-10]. The authoritative statistics released by the

Queensland Government in Australia in 2015 reported that, from 2007 to 2012, the incidence of intrauterine hypoxia and birth asphyxia was 4–6 per 1000 newborns [11]. The consequences among the survivors are significant, with morbidity comprising cognitive delay (45%), cerebral palsy (29%), visual loss (26%) and sensorineural deafness (9%) [7]. The consequences of neonatal HIE have a major impact not only on the affected person but also on their family as well of the costs to society. These facts indicate that, despite the improvements in medical care, asphyxia is still a major threat to the survival and health of newborns.

1.1 EEG as a prognostic and diagnostic tool

Different parameters are used for the diagnosis and monitoring of these critically ill newborns. They include fetal heart rate abnormality, clinical features, such as those recorded in the Apgar score, and blood biochemistry, such as blood pH, lactate and oxygen saturation. However, these complementary methods of assessment cannot differentiate different clinical grades of HIE with high sensitivity and specificity [12]. These clinical grades are scored mainly by clinical criteria developed by Sarnat and Sarnat in 1976: level of consciousness, tone, neonatal reflexes, autonomic function and seizures. Another grading system is based on alertness, feeding, tone, respiratory status, reflexes and seizure activity (range: 0–6) [13]. However, these grading systems are not very sensitive to dynamic changes that may occur during treatment.

As the brain is the main target, brain monitoring is ultimately needed to monitor structural changes and spatiotemporal functional changes. Structural changes are best monitored by magnetic resonance imaging (MRI), whilst functional changes are best monitored by electroencephalogram (EEG). The imaging technique (MRI scan) is expensive, of limited time resolution and requires the baby to be moved to the scanner. Moreover, long-term brain function monitoring is required for continuous assessment of dynamic changes for diagnosis and prognosis. Importantly, MRI does not help in the immediate postnatal period, as it is unable to reliably ascertain the presence and extent of HI injury until several days after the baby's birth [14]. In contrast, conventional EEG (cEEG) and amplitude integrated EEG (aEEG), undertaken in the early postnatal period, have been shown to be highly predictive of long-term outcome. Another motivation for using EEG compared to MRI is the relatively low cost of EEG systems. This argument can be crucial for the deployment of EEG technology in developing countries. These potentials of EEG to aid in the early diagnosis and classification of HIE severity has led to renewed interest in the field of neonatal neurophysiology. The benefits of EEG are that it can easily be implemented at the cot-side soon after birth and that it provides real-time measurement of cerebral function.

A normal background neonatal EEG is highly predictive of a normal outcome. Various abnormal EEG background patterns have been consistently associated with neurological abnormalities or death. These normal and abnormal EEG backgrounds are defined in terms of different amplitudes, phases and frequencies. Generally, continuous normal voltage (CNV) and discontinuous normal voltage (DNV) are treated as normal EEG background patterns, whereas symmetry, asynchrony, burst suppression (BS), continuous low voltage (CLV) and flat trace (FT) are considered to be abnormal backgrounds. EEG background patterns have been correlated with different neurodevelopmental outcomes including cognitive delay (CD), mental retardation (MR), early death (ED), the need for extracorporeal membrane oxygenation (ECMO), later cerebral palsy (CP), psychomotor retardation (PMR), dystonic quadriplegia (DQ), spastic quadriplegia (SQ) etc. in term neonates with HIE from within 6 months to 2 years [15-17].

Shany *et al.* investigated 39 newborns with burst-suppression (BS) and found that 8 died within 3 years [15]. Roij *et al.* [16] showed that 60 out of 160 newborns with flat trace (FT) or continuous low voltage (CLV) died or had a major handicap. Horst *et al.* [17] studied 30 newborns with CLV patterns and found that 6 died or had a major handicap within 24 months, while Douglass *et. al.* [18] assessed 22 newborns with only BS and found that 16 died or had a major handicap. These abnormal outcomes are broadly supported by other studies (see Chapter 2) which also report that various abnormal EEG patterns are associated with adverse neurodevelopmental outcomes. However, it is still not clear which EEG background pattern or patterns are most predictive of the neurodevelopmental outcome in term neonates with HIE.

The starting point of this thesis is to identify these specific EEG background patterns that best predict neurodevelopmental outcome. This will enable neurophysiologists and clinicians to undertake early management and treatment decisions[1] [19]; it will also provide opportunities for engineers to characterise the signal and obtain time-varying temporal and spectral information that is ultimately needed for an automated detection and prediction method. Automated detection will be helpful for clinicians and will reduce workload, training and subjectivity. This PhD project addresses the automated detection and classification of the EEG background patterns that best predict neurodevelopmental outcomes, using different signal processing and machine learning techniques. These approaches will be validated using multichannel neonatal EEG. Finally, a set of EEG signal

[1] The available treatments in the immediate post-natal period and during the two years prior to neurodevelopmental outcome tests include (i) resuscitation and supportive care e.g., normothermia, normoglycemia, permissive mild hypercapnia, adequate cerebral perfusion, room air versus 100% oxygen, resuscitation training, fluid management, electrolyte balance, (ii) therapeutic hypothermia by selective head or whole-body cooling administered within six hours of birth, (iii) general anaesthesia and medications to control seizures, and (iv) treatments to assist the baby's heart function and control blood pressure. Details can be found in [19].

features that best predict neurodevelopmental outcome in a large cohort of term neonates with HIE will be identified. This project will make engineering contributions through the design and optimisation of novel signal processing and machine learning techniques for the characterisation of EEG signals, detection, and classification of abnormal EEG patterns and prediction of neurodevelopmental outcome.

1.2 Definition of clinical terms

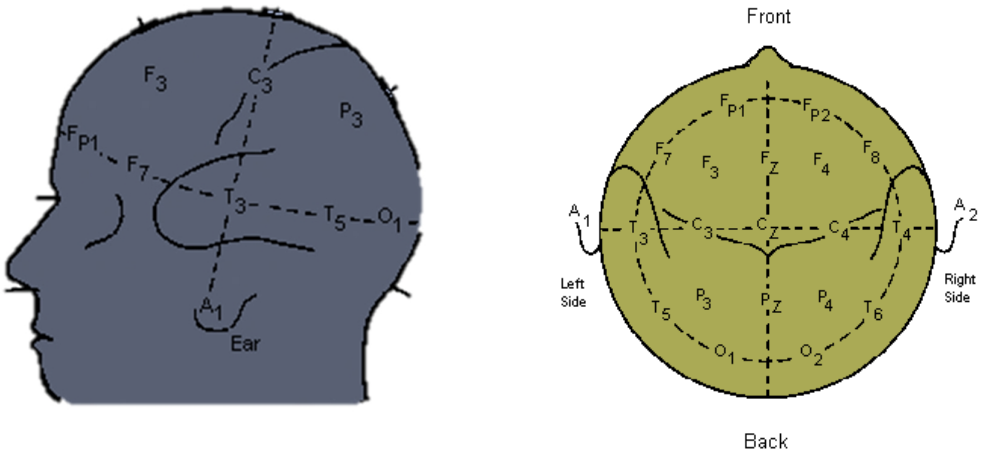
For the sake of clarity, a brief description of some general clinical terms is given here, while the description of more specific terms is provided within the relevant chapters.

Gestational age: (GA – also known as post-menstrual age), usually expressed in weeks + days: the time difference between the first day of the last normal menstrual period and the day of delivery.

Term neonate: an infant born at > 36 wks. GA. Infants with HIE eligible for hypothermia treatment are > 34 wks GA.

Apgar score: The Apgar score was developed by the US anaesthesiologist Virginia Apgar in 1953 [20] and has been routinely used for more than half a century to assess neonatal health within 1–10 min after birth. The Apgar score comprises five key components: heart rate, respiratory effort, muscle tone, reflex irritability and colour. Each component is scored from 0 to 2, giving at total Apgar score range of from 0 to 10 [21]. These components may not be of equal importance and heart rate and respiratory effort can be difficult to assess due to medical interventions [21].

International 10-20 EEG recording system: a standard EEG recording system which defines the arrangement procedure of EEG electrode placement over the scalp (see Figure 1.1). It is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The '10' and '20' refer to the 10% or 20% inter-electrode distance [22]. Each point in Figure 1.1 indicates a possible electrode position. Each site has a letter (to identify the lobe) and a number or another letter to identify the hemisphere location. The letters F, T, C, P, and O stand for frontal, Temporal, Central, Parietal and Occipital. (Note that there is no 'central lobe'; this is used for identification purposes). Even numbers (2,4,6,8) refer to the right hemisphere and odd numbers (1,3,5,7) refer to the left hemisphere. The z refers to an electrode placed on the midline. The smaller the number, the closer the position is to the midline. Nasion is the point between the forehead and nose. Inion is the bump of the occipital bone at the back of the skull.



(a) (b)

Figure 1.1: The 10–20 EEG system [22].

Burst-suppression: EEG burst suppression pattern (also called 'paroxysmal' [23] and 'spontaneous activity transients' (SATs) [24, 25]) is a collection of alternating bursts of high voltage with duration 1–10 s and composed of various patterns (delta and theta range frequencies with superimposed and inter-mixed spikes, sharp waves and faster activity) followed by periods of marked background attenuation known as suppression [26, 27] (see Figure 1.2).

Low voltage: Low voltage EEG activity can be defined as the disappearance of spindles and low voltage background activity [28].

Flat trace: Flat trace or isoelectric pattern is the extreme condition of cerebral inactivity which is also called 'electro cerebral silence' (ECS) [28].

Seizures: Seizures are paroxysmal or convulsive events caused by a hypersynchronous discharge of groups of neurons (see Figure 1.2). The morphology of neonatal EEG seizures varies, but characteristically consists of monophasic repetitive discharge of sharp or slow wave activity.

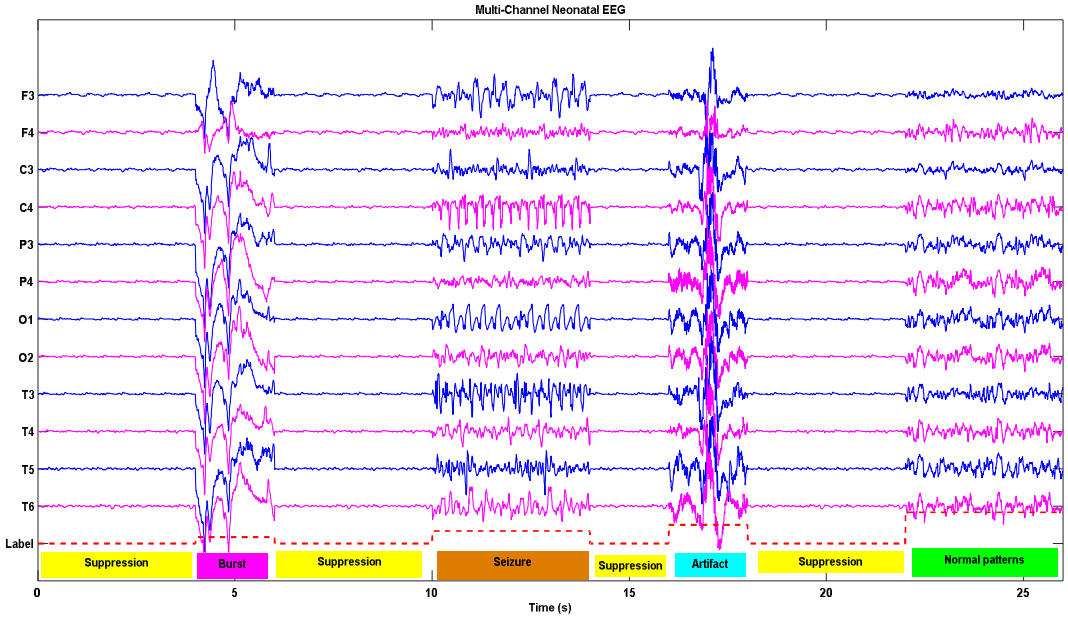


Figure 1.2: Multichannel neonatal EEG signals. Signal amplitude is in microvolts (μV). The bottom line shows the neurologist classification (see Chapter 4 for details).

Artifacts: Artifacts are unwanted signals produced by the random movement of the subjects, electrode contact, electrocardiogram, respiration, eye blinking and electrical interference in the NICUs [29]. These different types of artifacts are generalised and defined as an 'artifact' class in this project (see Figure 1.2).

Neurodevelopmental outcome: 'Neurodevelopment' represents the progressive, systematic transformation of behaviour and activities which are observed as neonates become older. Their physical and mental capacity and understanding of the world around them are expected to increase and mature as time advances. 'Neurodevelopmental outcome' represents this maturity. 'Poor neurodevelopmental outcome' includes the neurodevelopmental disorders cerebral palsy, intellectual disability, hearing loss and visual impairment etc. [30].

BSID-III: Bayley Scales of Infant and Toddler Development (version III), developed by psychologist Nancy Bayley. This is one of the most comprehensive and internationally recognised tools for neurodevelopmental assessment of infants in the five key developmental domains of cognition, language, social-emotional, motor and adaptive behaviour, and can be performed in the age range 1–42 months [31].

EEG frequency band: Frequency of EEG, in *Hz*, is the number of repetitions of a pattern within a second [32]. There are four main EEG frequency bands as described Table 1.1.

Table 1.1 EEG frequency band.

Activity	Frequency (Hz)
Beta, β	> 13
Alpha, α	8 – 13
Theta, θ	4 – 8
Delta, δ	0 – 4

1.3 EEG signal analysis

Neonatal EEG is commonly used as a tool to assess brain function and is suitable for long-term monitoring. The recorded signals provide important information on cerebral activities. In addition, neonatal EEG has been shown to be a robust predictor of neurodevelopmental outcome. In this project, EEG has been analysed in order to characterise, classify and predict the neurodevelopmental outcome. Joint time–frequency analysis has been considered as a signal processing approach in this project for signal characterisation and analysis. These considerations will inform the new signal processing methods.

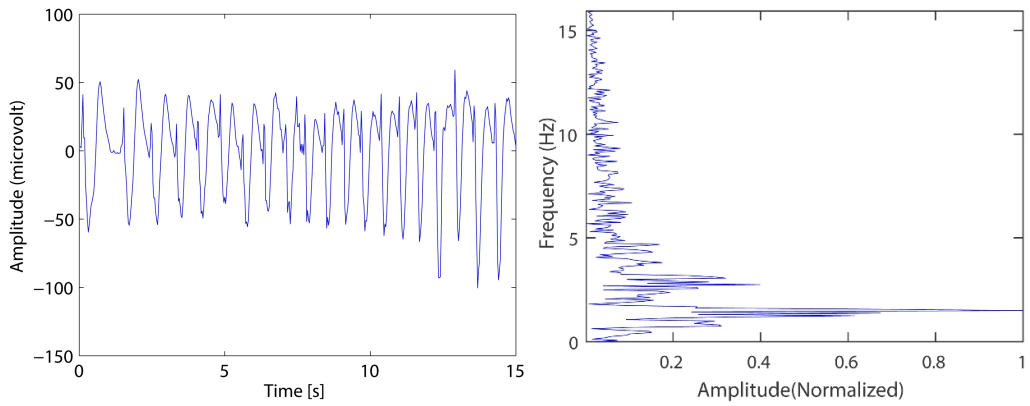
1.3.1 Signal characterisation

Signals are some form of measured quantity, normally recorded on measurement devices over a period of time. One of the most important aims of signal processing is to characterise signals and extract valuable information from the signals in order to describe signals and their properties. Signals are usually represented as functions of time, i.e. $x(t)$. Figure 1.3 shows an example of an EEG signal as a function of time. Analysis of the signal in the time domain is known as *time* (t) *domain analysis*.

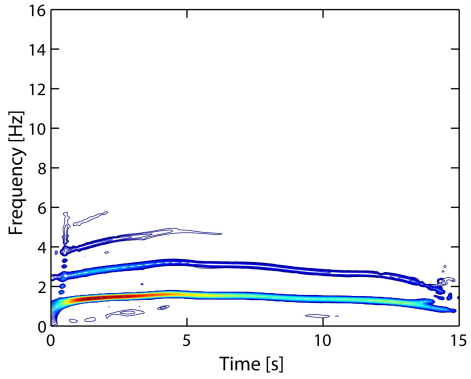
Figure 1.3(a) is pseudo-periodic, i.e. it has certain waveforms which are repeating. One can extract frequency information using *frequency* (f) *domain analysis* (see Figure 1.3 (b)) by Fourier transform of $x(t)$

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft}\, df \tag{1.1}$$

where $X\,(f)$ represents the f-domain signal.



(a) (b)



(c)

Figure 1.3: (a) t-domain plot of newborn EEG seizure signal and corresponding (b) f-domain plot and time-frequency (t, f) representation.

Signals can be categorised into stationary signals and nonstationary signals. Stationary signals have constant frequency or constant statistical proprieties over time, whereas nonstationary signals have frequency content that changes over time. Most real-world signals, including EEG signals, are nonstationary. The spectral contents of EEG vary in response to internal and external stimuli. The t-domain analysis and f-domain analysis can only provide limited information and cannot describe information that changes over time alone.

1.3.2 Need for time-frequency signal processing

Time-frequency representations (TFRs) are capable of fully characterising the nature of non-stationary signals over classical t-domain or f-domain representations. This is because in both t- and f-domain representations, the variables t and f are treated as mutually exclusive: to represent one variable, the other variable is 'integrated out'. Consequently, signals are non-localised with respect to the excluded variable. In contrast, joint time-frequency representation using time-frequency distribution (TFD) represents the signal energy distribution as a function of both t and f together

where t and f are not mutually exclusive. For example, Figure 1.3(c) shows the (t, f) representation of EEG and shows how the frequency content of this signal changes over time. This property motivates the researchers to process and analyse non-stationary signals such as biomedical signals and speech signals using TFDs in order to unveil the signal characteristics accurately. A more detailed definition of TFD can be found in Chapter 3.

1.3.3 Representing a signal in the time-frequency domain

There are many ways of representing a signal in the time-frequency domain. One of the simplest is the short-time Fourier transform (STFT) [33]. First, it divides the whole signal into short-time window (segments) and then applies Fourier transform to these segments. The squared magnitude of the STFT is a real-valued representation called the spectrogram. The standard spectrogram is easy to interpret. However, its performance depends on the analysis window that has been set heuristically and provides poor time or frequency resolution [34].[2]

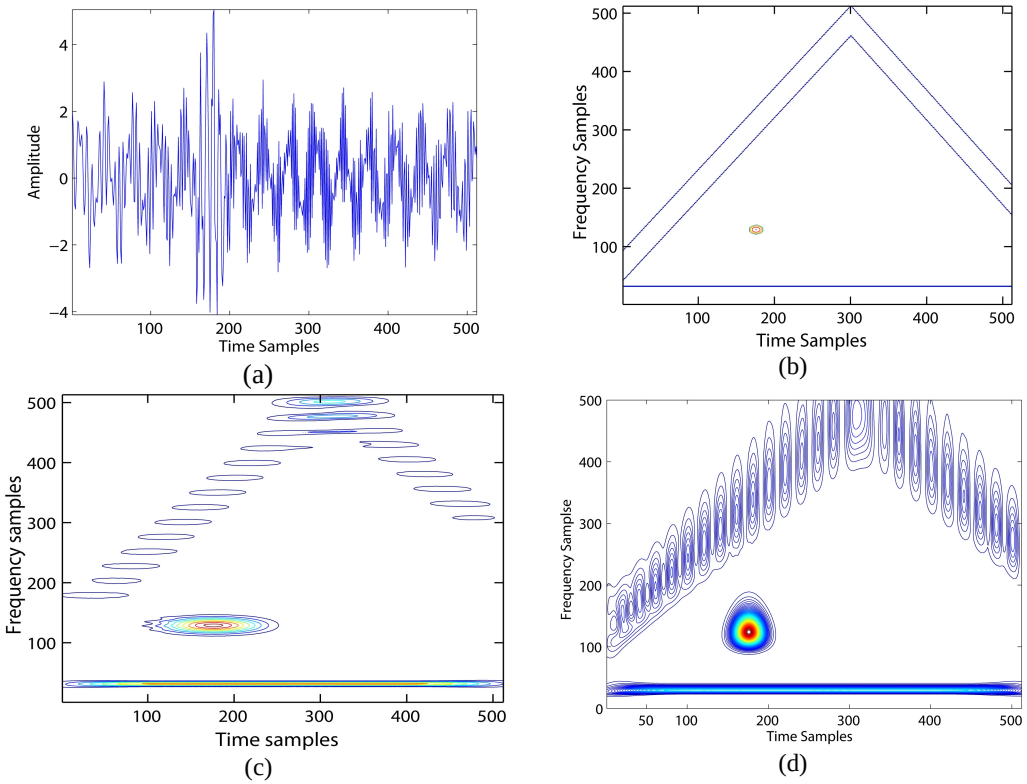


Figure 1.4: TFD representation of a simulated signal: (a) t-domain signal and (b) its corresponding ideal(t, f), (c) spectrogram and (d) scalogram representation.

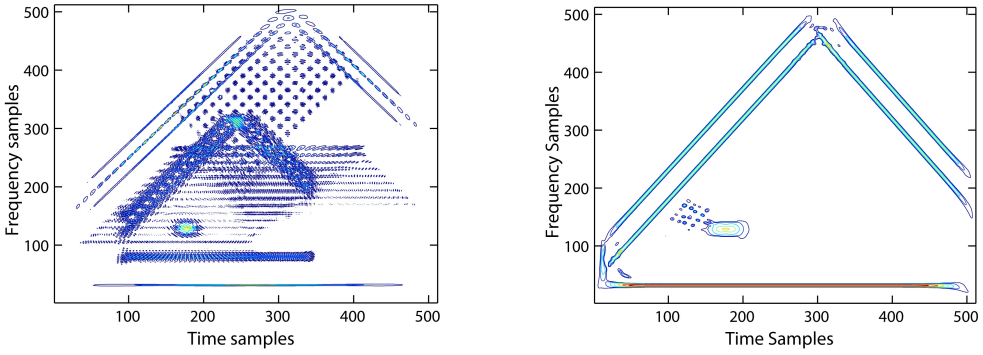
[2] In signal processing, resolution is used as a performance measure. It defines how well a method or approach separates/distinguishes different signal components in the time or frequency or time-frequency axis.

For example, Figure 1.4(c) plots the standard spectrogram of a multicomponent signal (Figure 1.4(a)). It provides poor resolution as the signal components are not separated (or visualised) and it is far from the ideal time-frequency distribution (Figure 1.4(b)).

Another method for generating time-frequency representations is wavelet transforms [3], which decompose the signal into a set of basis functions in terms of time and scale. As scale is inversely proportional to frequency, the wavelet transform can generate a time-frequency representation called the 'scalogram'. The wavelet transform has many successful applications, including in de-noising and image compression. A limitation of the scalogram is that it does not provide good resolution throughout the time-frequency representation, thus making interpretation of the representation difficult, as shown for example in Figure 1.4(d).

A quadratic time-frequency distribution (QTFD) is a quadratic transformation of the t-domain signal into the (t, f) domain. Although there are other classes of TFDs, QTFDs are probably the most useful due to their simple interpretation, high resolution and widespread use [33]; they have become a standard tool in many disciplines and will be used as the main signal processing approach in this project.

Wigner-Ville distribution (WVD) is the fundamental member of the QTFD class and provides high resolution (t, f) representation; however, it suffers from cross-terms (see Figure 1.5(a)). The cross-terms can be suppressed by smoothing the WVD with a properly adjusted kernel (see Chapter 3) function and this results in a high resolution and cross-term free (t, f) signature (see Figure 1.5(b)). One of the fundamental goals of this project is to design and optimise high resolution QTFDs and show 'how the optimised QTFDs offer better performance in terms of EEG signal characterisation and classification'.



(a) (b)

Figure 1.5: TFD representation using WVD and optimised adaptive directional TFD (ADTFD). See Chapter 3 for a detailed definition.

1.3.4 Tilings representation

Tilings are another way to quantify the effectiveness of QTFDs. Figure 1.6 shows the tilings for t-domain, f-domain and (t, f) -domain using rectangular tiles. The t-domain and f-domain analysis cannot provide frequency and time information, respectively. Fig 1.6(c) shows the tiling for STFT, which is obtained by using a fixed analysis window for all time. Figure 1.6(d) and (e) show the wavelet transform (WT) and wavelet packet (WP) transform. These transforms firstly divide the spectral contents into high and low frequencies, and then analyse the signal. For example, WT has narrow time windows, that yield better time resolution at higher frequencies, and wider time windows, that yield better frequency resolution at low frequencies. Figure 1.6(f) indicates that the generalised (t, f) method adapts the signal analysis window in time as well as frequency; from $(0, T_1)$ the tiling of generalised (t, f) is as in Figure 1.6 (c), as denoted by the grey line. From (T_1, T_2) the tiling is as in Figure 1.6 (e) and is denoted by the dark grey line, while from (T_2, T_3) the tiling is as in Figure 1.6 (d), as denoted by the light grey line. These facts suggest that STFT, WT and WP are the particular cases of the generalised (t, f) representation ([33], p. 98). This generalised (t, f) representation can be obtained from QTFDs and provides high resolution (t, f) representation, as seen in Figure 1.5. These rationales also underlie the use of a QTFD signal processing approach in this project.

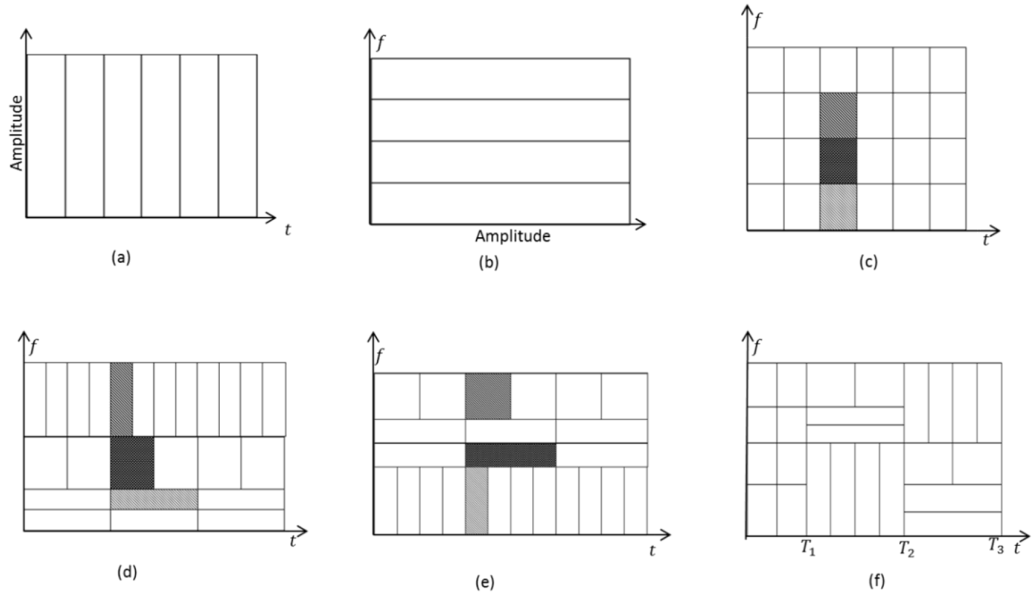


Figure 1.6: Tilings of the time-frequency plane: (a) time-domain tiling, (b) frequency-domain tiling, (c) STFT tiling, (d) wavelet tiling, (e) wavelet packet tiling and (f) generalised (t, f) tiling.

1.4 Significance and motivation

This project advances the field of newborn EEG analysis using time-frequency signal processing and machine learning techniques. The significance of this research falls within four major categories:

- Significance of the identification of specific EEG background patterns that best predict neurodevelopmental outcome

– HIE occurs following a peripartum hypoxic-ischaemic insult to the neonatal brain. EEG is a useful method of assessment in asphyxiated infants and various background patterns of EEG are the robust predictor of neurodevelopmental outcome. It is necessary to identify which specific background pattern(s) are the most predictive of neurodevelopmental outcome to allow clinicians to identify those neonates most appropriate to enrol in the studies of neuroprotective or neurorestorative therapies and neurointensive care as well as guide therapy.

- Significance of EEG abnormality characterisation

'Signal characterisation' is a broad term. As with other signals, EEG characterisation is important as it will allow visualisation and description of the signal in various domains such as t-domain, f-domain and (t, f) domain.

EEG is a non-stationary signal and TFD is well suited for the analysis of such a non-stationary signal over the t- and f-domains. It provides an optimal graphical way to visualise and characterise time-varying frequency content of the signal components. This can help in understanding the properties of the EEG signal patterns in order to determine the best approach (such as change detection or pattern recognition) to further analysing and processing the signal. The core idea is that different EEG background patterns show different signatures in the (t, f)-domain. Therefore, various features extracted from these domains offer discriminative information required for automated detection and recognition. To achieve a higher accuracy in detection or pattern recognition, these features need to be extracted from (t, f) representations of EEG signals using TFDs with high-resolution and minimal cross-term interferences.

Spectrogram and other QTFDs are widely used due to their high resolution and easy interpretation. The spectrogram is sensitive to analysis window, whereas the other QTFDs are sensitive to kernel parameter(s). Proper choice of analysis window and window parameter, as well as the kernel parameter(s), provides high resolution TFDs, minimises cross-term interferences and significantly improves signal characterisation and classification.

- Significance of EEG background classification/ Significance of newborn scalp EEG monitoring

Monitoring of scalp EEG in the newborn is useful for predicting long-term neurological consequences and in reducing mortality/morbidity [45]. The current dominant approach to newborn EEG assessment is based on visual inspection of the EEG by an expert. This process can be subjective,

requires high levels of expertise and usually takes considerable time, especially for long EEG recordings.

Currently, visual inspection is treated as the 'gold standard', whereas automated EEG analysis is a relatively new research area and is primarily a research tool. Much research focuses on specific patterns: for example, neonatal seizure detection and localisation [29, 35-37], and detection of the sleep-wake cycle [38]. Classification of a broader variety of EEG background patterns will increase the system complexity and the need for specific EEG data, especially in long-term multichannel EEG recordings. Restricted numbers of detectable patterns constrain the use of automatic algorithms in clinical settings. An automated detection and classification can aid in diagnosis and assist in subsequent clinical management.

- Significance of early prediction of neurodevelopmental outcome

Early prediction of neurodevelopmental outcomes is still a major challenge in neonates, as neonatal brain injury leads to a variable probability of developing various serious long-term neurodevelopmental disorders in childhood. There is an increased in risk of poor outcomes when longer durations of abnormal EEG results are observed in neonates [39, 40]. Different features can be extracted from EEG data to seek associations with the neurodevelopmental outcome.

A normal research procedure to predict the outcome in neonates using EEG is to collect EEG data at the very early stage of life, then assess the babies at the age of one or two years using some outcome measurement test such as BSID test and correlate the EEG features with the outcome. This feature or a set of features can then be considered as neurophysiological biomarker(s).

Once the neurophysiological biomarker(s) are identified, a model can be built to predict the outcome at a very early stage. This could benefit the baby by guiding new neuro-protective or neurorestorative therapies and neurointensive care to reduce morbidity and mortality. The final motivation of this research is to save the life of asphyxiated newborns by identifying those features that could be used as accurate neurophysiologic biomarker(s).

1.5 Aims and objectives

This project has the following aims and objectives.

Aim 1: Identify the EEG background features or patterns that best predict outcome in term neonates with HIE.

Aim 2: Design and optimise time-frequency distributions for EEG signal analysis.

Objective 1: Automatic fast optimisation of quadratic TFDs using the hybrid genetic algorithm.

Objective 2: Robust time-frequency representation based on fractional Fourier transforms.

Aim 3: Classify multichannel neonatal EEG background patterns.

Objective 1: Method-1 (Classification using single feature subset)

Objective 2: Method-2 (Classification using class-specific feature subset)

Aim 4: Test the identified EEG signal features in the prediction of neurodevelopmental outcome in term neonates with HIE.

1.6 Proposed approaches and methods

The following approaches and methods will be used to address the aims and objectives stated in the previous section. Figure 1.7 shows the concept map of this thesis.

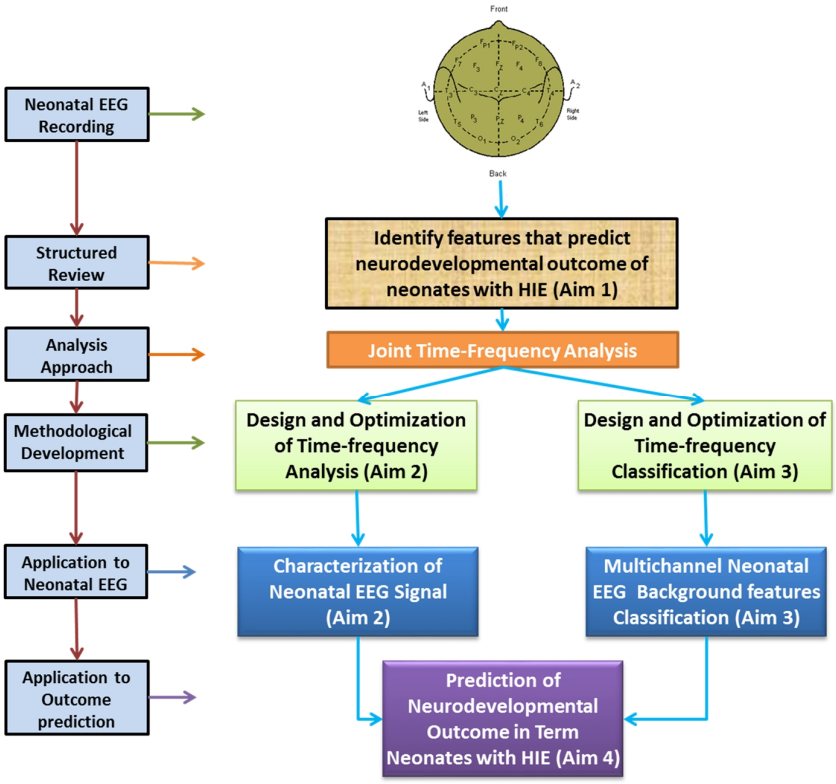


Figure 1.7: Concept map of the thesis.

Firstly, this thesis identifies the EEG background patterns; it then develops methods to characterise and classify these patterns; finally, it uses the developed methods to test prediction of the neurodevelopmental outcome.

Aim 1: Identify the EEG background features or patterns that predict outcome in term neonates with HIE

Attention has been drawn, in Section 1.1, to the fact that it is still unclear which background features (or patterns) of the EEG, in term neonates with HIE, best predict outcome. The published literature is very inconsistent in terms of feature definitions and there has been no systematic approach. The

starting point of this project is to identify from the literature the EEG background features or patterns that best predict the neurodevelopmental outcomes. A systematic review is conducted, and meta-analysis of the literature is performed to achieve this aim. The analysis categorises the background features (or patterns) and identifies the background patterns that predict neurodevelopmental outcome. See Chapter 2 for details.

Aim 2: Design and optimise time-frequency distributions for EEG signal analysis

Two optimisation methods – global optimisation (objective 2.1) and local optimisation (objective 2.2) – have been proposed to obtain high resolution TFDs. The (t, f) energy concentration measure (ECM) has been used to automatically select the optimal parameters in both algorithms. These methods are expected to increase the readability of the (t, f) signatures and reduce misinterpretation of the (t, f) signature by minimising the cross-terms for multicomponent and non-stationary signals without *a priori* signal information.

To address objective 2.1, a novel framework is developed that automatically adjusts the QTFD kernel parameters by using a hybrid genetic algorithm (HGA). To address objective 2.2, a novel spectrogram called the 'Locally Optimised Spectrogram' (LOS) is designed to obtain a high-resolution time-frequency representation based on the short-time fractional Fourier transforms (STFrFT). The HGA algorithm optimises the TFD globally, whereas LOS divides the TFD into several grids and performs local optimisation. See Chapter 3 for details.

Aim 3: Classify multichannel neonatal EEG background patterns

Two classification methods (method-1 and method-2) are designed and optimised to classify EEG background patterns including burst, suppression, normal and seizure in the presence of artifacts. These patterns are the predictors of neurodevelopmental outcomes. Classification method-1 uses a single feature subset for all classes, whereas classification method-2 uses class-specific features. Method-2 is expected to increase the classification performance as the selected feature subset is class-specific.

Various time domain, frequency domain and joint time-frequency domain features are first extracted and then concatenated to produce a long feature set. A hybrid feature selection (HFS) algorithm is proposed to simultaneously select the prominent feature subset as well as the classifier parameters. A state-of-the-art machine learning algorithm called Support Vector Machine (SVM) has been optimised and used to classify different neonatal EEG patterns using the feature subset selected by the hybrid feature selection (HFS) algorithm. This HFS algorithm uses a genetic algorithm that not only selects the prominent feature subset but also simultaneously selects the SVM tuning parameters

to ensure the optimal use of the SVM classifier. Three fusion techniques - channel fusion, feature fusion and decision fusion - are used for the multichannel classification. See Chapter 4 for details.

Aim 4: Testing of EEG signal features that best predict neurodevelopmental outcome in term neonates with HIE

Various time domain, frequency domain and joint time-frequency domain features are extracted from EEG in term neonates with HIE to seek association of these features with the neurodevelopmental outcome as assessed by BSID-III. BSID-III provides developmental scores for cognitive, language, motor and combined score. The hybrid feature selection (HFS) algorithm, developed in Aim 3, is used to select a prominent feature subset. EEG signal-based features are investigated with nonlinear support vector machines (SVMs) for their ability to predict neurodevelopmental outcome. See Chapter 5 for details.

The following applications are introduced for the proposed methods:

- Application of the optimised TFD to characterise newborn EEG
- Application of optimised classifier for multichannel EEG background patterns classification
- Application of EEG signal features to predict neurodevelopmental outcome.

1.7 Contributions

A number of original contributions, specifically in the fields of *signal processing* and *machine learning*, are made for the characterisation, detection, classification of biomedical signals especially neonatal EEG signal and, finally prediction of neurodevelopmental outcome. These contributions are summarised below aim-by-aim:

Aim 1:

- The systemic review and meta-analysis identified that burst suppression, low voltage and flat trace, trace alternant, asynchrony and asymmetry are the most predictive of poor neurodevelopmental outcome. These patterns can be potentially used as diagnostic markers.

Aim 2:

Objective 1:

- A novel framework for a fully automatic optimisation of Quadratic Time-frequency Distributions (QTFDs) has been proposed.
- The problem of optimisation of Reduced Interference Distributions (RIDs) (also called QTFDs due to their quadratic nature) has been formulated as a cost function of a modified energy concentration measure index to be minimised for the non-stationary signal without *a*

priori signal information, and a hybrid genetic algorithm (HGA) proposed to find the optimal and fine-tuning QTFD parameters.

- The proposed approach overcomes the need for human intervention, which is time-consuming.
- A fast and memory efficient optimisation has also been implemented in the optimised use of TFDs, especially for long recordings or multisensor data as for multichannel EEG recordings. This will significantly reduce the processing and consequent decision-making time.

Objective 2:

- A novel TFD, 'Locally Optimised Spectrogram' (LOS), is designed from the short-time fractional Fourier transform by locally optimising both window length and chirp rate. It automatically determines the locally optimal window parameters and fractional order (angle) for all signal components, leading to a high resolution and cross-term free time-frequency representation.
- The proposed method demonstrates superior performance by quantitative comparison with other state-of-the-art QTFDs. Different simulated multicomponent signals of varied amplitude and wide-ranging (t, f) characteristics are used to demonstrate the efficiency of the proposed method.
- A relationship between LOS and other TFDs has been derived.
- Finally, a qualitative assessment of global and local optimisation has been provided.

Aim 3:

- An improved characterisation of neonatal EEG background patterns is developed by using state-of-the-art optimised time-frequency distribution. Different EEG backgrounds show different (t, f) signature on the (t, f) plane which indicates the prominent use of TFD in the classification.
- Different established t, f and (t, f) domain-based feature have been extracted to characterise and classify the EEG background patterns and HFS algorithm is used to select prominent feature subset.
- Both classification methods (method-1 and method-2) have been applied to multichannel neonatal EEG background patterns classification. Different fusion techniques have been applied to solve this complex multiclass and multichannel problem and the classification performance of different approaches has been compared.
- A decision support system is also discussed as a potential application of the proposed approach.

Aim 4:

- Signal processing features are identified that correlate with neurodevelopmental outcome in a large study cohort of babies with HIE, with neurodevelopmental outcome defined by BSID-III.
- Various time domain, frequency domain and (t, f) domain features are extracted from the neonatal EEG recorded within the first 24 hours after birth to seek associations with the neurodevelopmental outcome.
- A probabilistic decision support system is derived as a potential application of the proposed system.
- The significance of this aim is that it will define those EEG features, both previously described and new, that most accurately predict neurodevelopmental outcome at 2 years. These features can be used as neurophysiological biomarker(s) to reduce morbidity and mortality in newborns.

1.8 Thesis organisation

The thesis is organised as follows:

Chapter 1 contains the background, rationale, aims and objectives of the thesis. It also briefly describes newborn HIE and associated EEG abnormalities. A relevant literature review has been provided within each chapter due to the multidimensional nature of this research.

Chapter 2 determines which specific background features of the EEG in term neonates with HIE best predict outcome. A structured literature review and meta-analysis of the literature is performed in order to determine the pooled sensitivity and specificity of the various background patterns.

Chapter 3 presents the design and optimisation of time-frequency distribution for the non-stationary signal characterisation of bio signals such as the EEG signal. There are two objectives: (i) to develop a state-of-the-art optimisation technique to fully optimise the kernel parameters of the existing QTFDs and compare this with other optimisation methods, and (ii) to design a novel TFD ('Locally Optimised Spectrogram' LOS), and compare it with other TFDs.

Chapter 4 presents the design and optimisation of a classifier for multichannel neonatal EEG background pattern classification. First it describes the background problem, including relevant literature review, then it describes methodology to extract features in order to classify EEG background patterns. The HFS algorithm is proposed for selecting prominent features and nonlinear SVM parameters simultaneously.

Chapter 5 tests a set of EEG signal features that best predict neurodevelopmental outcome in a large study cohort of term infants with HIE.

Chapter 6 provides concluding remarks and suggestions for future work.

Chapter 2 EEG background features that predict outcome in term neonates with Hypoxic Ischaemic Encephalopathy

2.1 Introduction

Hypoxic ischaemic encephalo-pathy (HIE) occurs following a peripartum hypoxic-ischaemic insult to the brain. A secondary energy failure follows, with build-up of reactive metabolites and injury ensues [32]. HIE is a major cause of morbidity and mortality in newborns [5, 41]. The spectrum of long-term morbidity in survivors ranges from mild motor and cognitive deficits to cerebral palsy and severe cognitive deficits [42]. The outcomes of HIE or 'birth asphyxia' can be devastating and permanent, making it a major burden for the patient, the family and society [43]. The ability to accurately predict outcomes in this population is important in identifying those most appropriate to enrol in studies of neuroprotective or neuro-restorative therapies.

What is already known on this topic?

Different EEG background patterns can predict neurodevelopmental outcomes in term neonates with HIE.

Published literature may use the same background pattern term but use different amplitude level, phase, frequency and duration in the definition.

What is the contribution of this study?

- ✓ A systemic review was undertaken to determine the specific EEG background patterns that best predict neurodevelopmental outcomes in term neonates with HIE.
- ✓ A meta-analysis has been performed to establish the prognostic value of the identified EEG background patterns and the pooled sensitivity and specificity have been calculated.
- ✓ Agreement on definitions is necessary for the effective implementation and use of EEG in NICUs and some definitions have been suggested.

EEG is a useful method of assessment in asphyxiated infants. Conventional and amplitude integrated EEG (aEEG) both perform well in predicting outcome [44] and in the early diagnosis and classification of HIE severity [12]. The benefits of EEG are that it can easily be implemented at the cot-side soon after birth and it provides a real-time measure of cerebral function [12]. Detecting seizures [29, 45, 46] and assessing response to anticonvulsants [47] are other useful functions of EEG recording in the clinical setting, particularly because many seizures in the newborn have no detectable clinical manifestations. A normal EEG is

highly predictive of a normal outcome, whereas various abnormal EEG features have been consistently associated with neurological abnormalities or death [47]. A systematic review and meta-analysis of the literature was undertaken to determine which specific background features of the EEG, in term neonates with HIE, have been described that best predict outcome.

2.2 Methods

2.2.1 Search strategy and inclusion and exclusion criteria

A literature search was conducted of the PubMed, EMBASE and CINAHL databases for studies published between January 1960 and April 2014. The search terms HIE or asphyxia, EEG, term newborn infants, prognosis or outcomes; and their derivatives were used to locate relevant studies. For PubMed, the specific search was: (((((("Asphyxia Neonatorum"[Mesh] OR asphyx*[tiab]) OR Hypoxic Ischemic Encephalopathy OR sarnat[tiab])) AND ((((Term* OR Neonate OR newborn OR neonat*) AND (Electroencephalo* OR EEG))), accessed 1 May 2014. Titles and abstracts were screened for relevant materials. Only studies published in English were considered. Two reviewers searched articles to determine which studies reported EEG background features and a neurodevelopmental outcome at $\geq$ 12 months age of term infants affected by HIE. Studies that identified the specific abnormal EEG background features were included e.g., burst suppression, continuous low voltage or low voltage and flat trace or isoelectric. Details are discussed in the discussion section.

2.2.2 Data extraction

Two reviewers (A. Awal and M. Lai) collected the following data from the included studies; number of infants, HIE Sarnat stage, type of EEG used, EEG background features identified, follow up period and neurodevelopmental assessments and outcome data. Where there was ambiguity in the description, for example in the EEG background feature classification or outcome level of disability, a third reviewer (P. Colditz) was consulted and a majority view prevailed. Data were extracted from published studies; no attempts were made to contact authors to clarify data or retrieve incomplete data.

2.2.3 Statistical analysis

The prognostic accuracy of each EEG background feature was assessed using the 2×2 matrix, shown in Table 2.1. Different terms of this matrix are defined as follows:

- Feature Present $(FP+)$: Presence of an abnormal EEG background feature (e.g. burst suppression, low voltage, and flat trace) is considered as $FP+$.

- Feature Absent $(FA-)$: Absence of an abnormal EEG background feature (e.g. CNV (continuous normal voltage), DNV (discontinuous normal voltage)) is considered as $FA-$.
- True Positive (TP): An abnormal EEG background feature or pattern associated with an abnormal neurodevelopmental outcome.
- False Positive (FP): An abnormal EEG background feature or pattern that falsely predicts a normal outcome.
- False Negative (FN): A normal EEG background pattern, which falsely predicts an abnormal outcome.
- True Negative (TN): A normal EEG background pattern associated with a normal outcome.

Table 2.1: 2 × 2 confusion matrix used in the analysis.

	Abnormal outcome	Normal outcome
Feature Present $(FP+)$	True Positive (TP)	False Positive (FP)
Feature Absent $(FA-)$	False Negative (FN)	True Negative (TN)

Sensitivity and specificity were calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2.1)$$
$$Specificity = \frac{TN}{TN + FP}$$

We constructed forest plots to show the sensitivity and specificity of each study graphically [48]. Forest plots were calculated by using Review Manager (RevMan) version 5.2 software [49].

Because of the study heterogeneity summary statistics cannot be provided using RevMan. We, however, estimated the 'summary point' for sensitivity and specificity and receiver operating characteristic (ROC) curve using STATA [50]. The '*metandi*' command was used for the meta-analysis which provides sensitivity, specificity with confidence interval using a random effects model from multiple studies to estimate the pooled sensitivity and specificity and best fit regression analysis parameters for Hierarchical Summary Receiver Operating Characteristics (HSROC) model to determine the summary ROC curve [50].

2.3 Results

Of the 860 articles generated by the initial search strategy, 566 studies remained when duplicates were removed. In the first stage, only 48 studies remained after hand-searching for the presence of reporting of EEG background features along with neurodevelopmental outcome in term neonates with HIE. A further 4 more were identified through hand-searching the references contained in the 48 studies. A total of 52 studies were finally identified. A further 21 studies were excluded because they

did not report on specific features of the background EEG; justification for the exclusion of these 21 studies is provided in the discussion section. Finally, 31 studies were included in the meta-analysis that which can distinguish among different abnormal background features. A flowchart of the search strategy is shown in Figure 2.1.

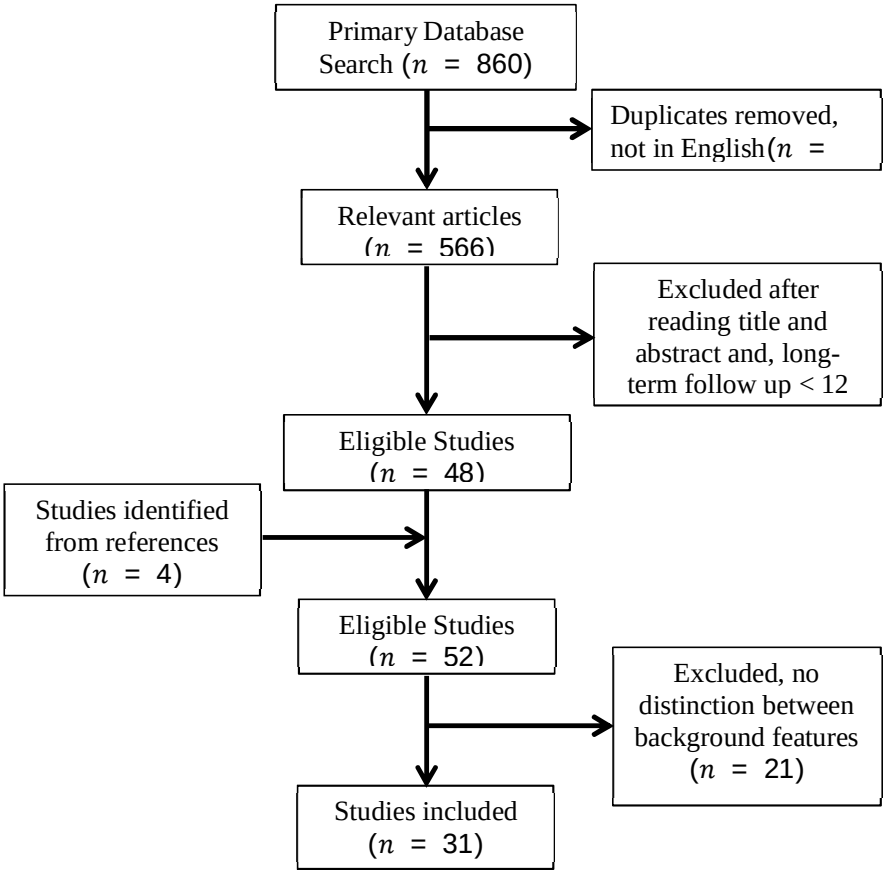


Figure 2.1: Flowchart of the search and study selection process.

Table 2.2: Summary of the included studies

Authors Year of Publication	#Asphyxia or HIE Stage I/II/III	Type of EEG	Timing of EEG	No. of infants, M:F ratio	Follow up age	Outcome measure	EEG and outcome*
al Naqeeb et al. 1999	14/17/7	aEEG	Within 12 hours of birth	20 $NR	18 – 24 months	Neurological examination Estimation of the Griffith's General Quotients	9 normal aEEG – all normal outcome 2 moderate aEEEG-1 died, 1 survived 9 BS- all died, severe to moderate outcome
Ancora et al. 2011	-/8/4	aEEG	Within 4 hours	12 all cooled	Mean age 17.3 months	Griffiths' mental Development Scales	2 normal aEEG- normal outcome 1 moderate aEEG-abnormal outcome

				NR			9 BS aEEG- 6 normal and 3 abnormal outcome
Azzopardi et al. 1999	12/5/4	EEG	Within 12 hours of birth	22 NR	12 – 24 months	A standardised neurological assessment, when possible by Griffiths developmental assessment	13 normal EEG- normal outcome 3 LV- abnormal outcome 6 BS- 4 normal and 2 abnormal outcome
Azzopardi et al. 2000	-	aEEG/ EEG	NR	16 NR	3 – 18 months	Griffiths general quotient	6 normal EEG-normal outcome 3 LV-all abnormal outcome 5 BS- 4 normal 1 abnormal outcome
Biagioni et al. 2001	8/16/1	EEG	Within the first 72 hours	25 13:12	2 years	Standardised protocol of neurological examination Griffiths' developmental scales	9 normal EEG- 8 normal and 1 abnormal 14 BS- 1 normal and 13 abnormal 2 LV- all abnormal
Bourez-Swart et al. 2009	n/a	aEEG/ cEEG	NR	12 NR	12 months	NR	4 normal- 3 normal and 1 abnormal 2 BS- 1 normal and 1 abnormal 6 discontinuous- 2 normal and 4 abnormal
Csekő et al. 2013	NR	aEEG	6 - 72 hours	70, 41:19 all cooled	18 to 24 months	Bayley Scales of Infant Development II (BSID II) Mental Developmental Index (MDI) or Psychomotor Developmental Index (PDI)	41 normal aEEG- 38 normal and 3 abnormal 10 BS- 3 normal and 7 abnormal 6 LV- all abnormal 1 FT- 1 abnormal
El-Ayouty et al. 2007	3/21/10	EEG	Within 72 hours	34 NR	18 months	Denver Developmental Screening Test II	7 normal EEG- all normal outcome 13 BS- all abnormal 5 LV- all abnormal
Eken et al. 1995	10/7/14	&CFM	within 24 hours	31 NR	3-24 months	Griffiths' developmental scales	12 normal EEG- 1 normal and 11 abnormal 11 BS- 2 normal and 9 abnormal 3 LV- all abnormal 5 FT- all abnormal
Finer et al. 1983	7/41/1	EEG	Within first week	49 26:23	27 months	Neurological examination	16 normal EEG- 14 normal and 2 abnormal 3 LV- all normal 26 BS- 19 normal and 7 abnormal
Gucuyener et al.	-/6/4	aEEG	Within 6 hours	10 6:4	8-18 months	Bayley Scales of Infant	Before Cooling: 3 normal aEEG- all normal 5 BS- all normal 2 FT-all abnormal

2012				All cooled		Development II (BSID II)	After cooling 8 normal aEEG- all normal 1 BS- abnormal 1 FT-all abnormal
Hallberg et al. 2010	3/16/4	aEEG	Within 6 hours	23 15:8 all cooled	4 and 12 months	Neurological examinations Alberta Infant Motor Scale	6 normal aEEG- normal outcome 15 BS- 10 normal, 5 abnormal
Hamelin et al. 2011	NR	EEG	1st EEG within 17 hours 2nd EEG within 72 hours	Cooled 4:12 Non-Cooled 9:14	12 months – 7 years	Neurological examination	Cooled 6 normal EEG- all normal 10 LV- 6 normal and 4 abnormal Non-Cooled 3 normal EEG- all normal 20 LV- all abnormal
Holmes et al. 1982	NR	EEG	Within 2 weeks	38 NR	24 months	Neurological and developmental examinations	15 normal EEG- all normal 11 BS- all abnormal 12 LV- 1 normal and 10 abnormal
Jose et al. 2013	9/13/9	EEG	Within 7 days	31 24:7	12 months	Denver Developmental Screening Test II Neurological examination	17 normal EEG- 15 normal and 2 abnormal 13 BS- all abnormal
van Lieshout et al. 1995	7/7/9	EEG	14 before 3rd day 9 before 10th day	23 NR	18 months – 7 years	WHO disability scoring Neurological examination	13 normal EEG- all normal 4 BS- 2 normal and 2 abnormal 3 LV- all abnormal 3 FT- all abnormal
Polat et al. 2012	11/10/3	EEG	NR	25 NR	44-48 months	Denver Neurological examination	13 normal EEG- all normal 6 BS- all normal 6 LV- all abnormal
van Rooij et al. 2005	NR	aEEG	Within 6 hours	160 NR	At least 24 months to 10 years	Griffiths mental developmental scale	70 normal aEEG- 64 normal and 6 abnormal 25 BS- 6 normal and 19 abnormal 65 FT- 5 normal and 60 abnormal
Selton et al. 1997	5/16/17	EEG	Between 2 - 7 days	38 14:24	12 months to 8 years	Amiel-Tison neurological evaluation	14 normal aEEG- 13 normal and 1 abnormal 10 BS- 2 normal and 8 abnormal 4 LV- 3 normal and 1 abnormal
Shankaran et al. 2011	-/71/37	aEEG	Within 9 hours	108 62:57 57 cooled 51 non-cooled	18 to 22 months	Neurological and developmental evaluations, Gross Motor Function Classification System (GMFCS), Bayley Scales of Infant Development II	24 normal aEEG- 18 normal and 6 abnormal 22 BS- 11 normal and 11 abnormal 26 LV- 13 normal and 13 abnormal 36 FT- 13 normal and 23 abnormal

Shany et al. 2006	14/14/9	aEEG	Within 3-6 hours	39 NR	3.6 years (mean)	Amiel-Tison and Gosselin neurological evaluation	23 normal aEEG- 21 normal and 2 abnormal 3 discontinuous- 2 normal and 1 abnormal 14 BS- 4 normal and 10 abnormal
Takeuchi et al. 1989	NR	EEG	Within 3 weeks	173 NR	12 months - 15 yrs, mean 3.2 yrs	Neurological examination	31 normal EEG- 29 normal and 2 normal 18 BS- all abnormal 20 LV- 10 normal and 10 abnormal 13 FT- all abnormal
Ter Horst et al. 2004	4/18/5	aEEG	60 min – 22 hrs	30 NR	24 months	Paediatric and neurological exam based on Touwen	19 normal- all normal 5 LV- all abnormal 1 FT- abnormal
Thoresen et al. 2010	NR	aEEG	Soon after birth until after cooling	74 31 non-cooled 43 cooled	18 months	Bayley Scales of Infant Developmental II	Normothermic: 12 normal aEEG- 8 normal and 4 abnormal 13 BS- 3 normal and 10 abnormal 1 LV- abnormal 5 FT- all abnormal Hypothermic : 16 normal aEEG- all normal 18 BS- 9 normal and 9 abnormal 1 LV- abnormal 8 FT- 1 normal and 7 abnormal
Thornberg et al. 1994	NR	CFM	Soon after admission	38 14:24	18-30 months	NR	20 normal- 17 normal and 3 abnormal 11 BS- all abnormal
Toet et al. 1999	-/17/20	aEEG	After admission	73 NR	12 months to 6 years	Griffiths mental developmental scale Items from Amiel-Tison and Renier evaluation and Touwens test Alberta Infant Motor Scale in children < 18mths	32 normal EEG- 27 normal and 5 abnormal 29 BS- 8 normal and 21 abnormal 6 LV- all abnormal 1 FT- abnormal
Toet et al. 2002	NR	CFM/ EEG	At 3 and 6 hours after birth	36 NR	3-24 Months	Griffiths mental developmental scale Items from Amiel-Tison and Renier evaluation and Touwens test Alberta Infant Motor Scale in children < 18mths	20 normal- 8 normal and 12 abnormal 7 BS-2 normal and 5 abnormal 5 FT- all abnormal
Watanabe et al. 1980	NR	EEG	Within 2 days	132 NR	NR	NR	23 normal EEG- all normal 31 BS- 30 normal and 1 abnormal

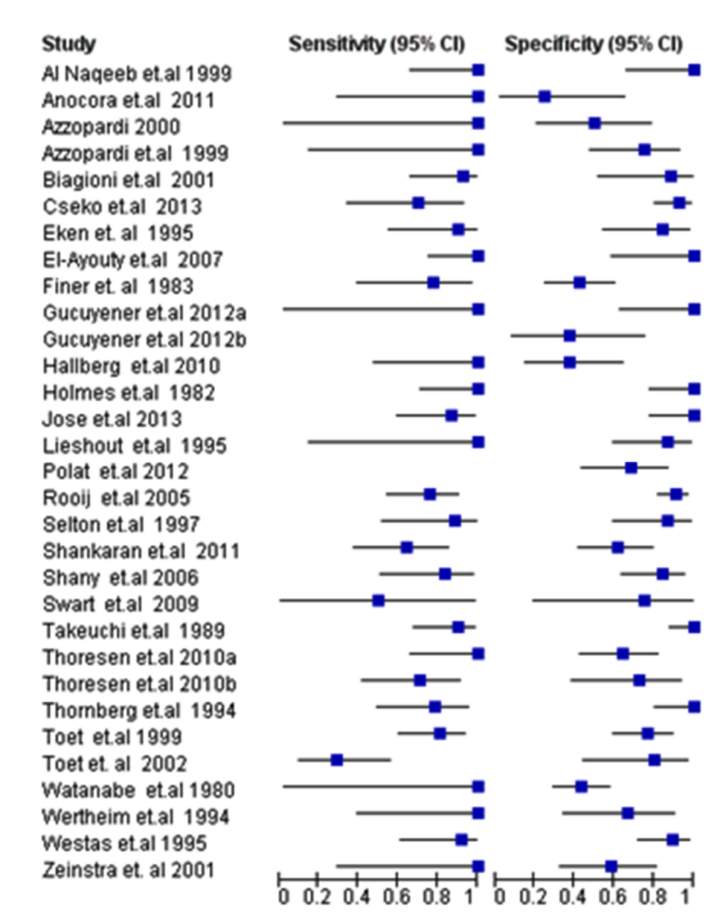
							74 LV- 9 normal and 65 abnormal
Wertheim et al. 1994	10/22/5	EEG	Soon after admission	37 19:18	12, 18 months	Griffith developmental scales	8 normal EEG- all normal 8 BS- 4 normal and 4 abnormal 7 FT- all abnormal
Hellström-Westas et al. 1994	9/5/24 (0/1/2-3)	aEEG	Within 6 hours	47 NR	12-18 months	NR	26 normal- 25 normal and 1 abnormal 14 BS- 3 normal and 11 abnormal 2 LV- all abnormal 5 FT- all abnormal
Zeinstra et al. 2001	-/27/9	EEG	1st: between 12 and 36 hours 2nd: between 7-9 days	36 NR	6 and 12 months	Examined by neonatologist and paediatric neurologist, WHO disability scoring	10 normal EEG- all normal 10 BS- 7 normal and 3 abnormal Other patterns are moderate EEG, Epileptic EEG, Asymmetry-Positive temporal wave.

2.3.1 Statistical analysis

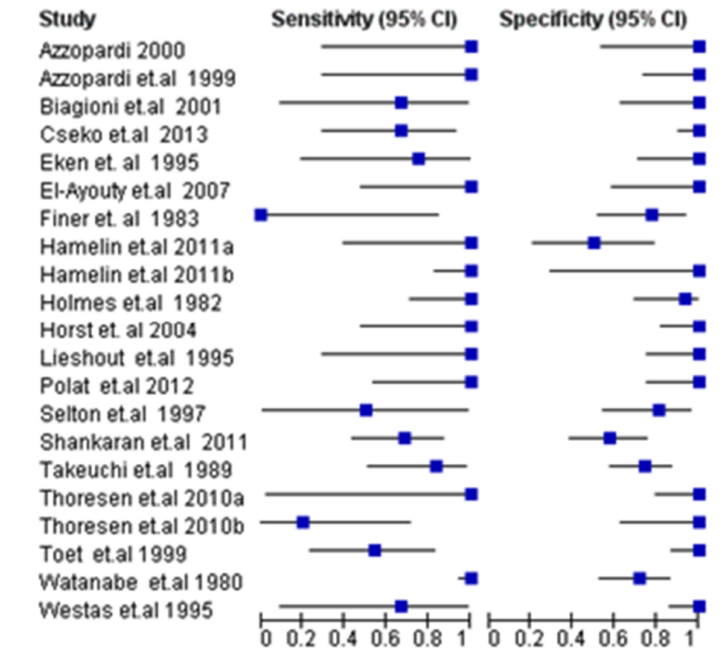
Total of 1948 term neonates with HIE with follow up available at 12 months of age or older were investigated in the 31 studies. Table 2.2 provides a summary of the numbers of infants and HIE severity. The score or severity of HIE was clearly described in 20 (64.5%) studies and hypothermia (HT) was used in 7 (22.5%) studies. In 4 (13%) studies it was necessary to extract data to construct the 2-way table.

The included studies used single, dual, and multichannel EEG; 14 (45%) studies used aEEG or single or dual channel EEG, 14 (45%) studies used international 10-20 EEG or continuous multichannel EEG, and 3 (10%) studies used both. The age at developmental assessment ranged from 12 months to 9 years.

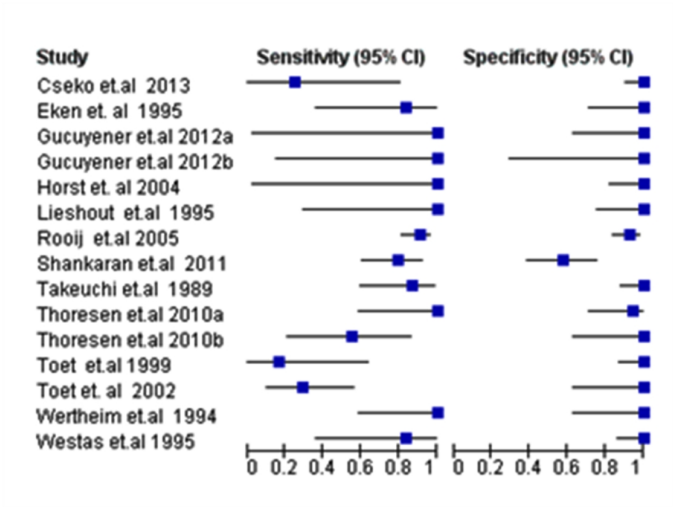
Twenty-nine of the 31 studies, involving a total of 914 term neonates reported the relationship between outcome and burst suppression pattern of EEG. The forest plot of sensitivity and specificity with 95% CI is given in Figure 2.2(a). 19 studies involving 567 neonates reported outcomes for low voltage EEG. A forest plot of sensitivity and specificity with 95% CI is given in Figure 2.2(b). 13 studies involving 493 babies reported outcomes for flat trace. Forest plot of sensitivity and specificity with 95% CI is shown in Figure 2.2(c).


Study
Sensitivity (95% CI)
Specificity (95% CI)
Al Naqeeb et.al 1999
Anocora et.al 2011
Azzopardi 2000
Azzopardi et.al 1999
Biagioni et.al 2001
Cseko et.al 2013
Eken et. al 1995
El-Ayouty et.al 2007
Finer et. al 1983
Gucuyener et.al 2012a
Gucuyener et.al 2012b
Hallberg et.al 2010
Holmes et.al 1982
Jose et.al 2013
Lieshout et.al 1995
Polat et.al 2012
Rooij et.al 2005
Selton et.al 1997
Shankaran et.al 2011
Shany et.al 2006
Swart et.al 2009
Takeuchi et.al 1989
Thoresen et.al 2010a
Thoresen et.al 2010b
Thornberg et.al 1994
Toet et.al 1999
Toet et. al 2002
Watanabe et.al 1980
Wertheim et.al 1994
Westas et.al 1995
Zeinstra et. al 2001
0 0.2 0.4 0.6 0.8 1
0 0.2 0.4 0.6 0.8 1

(a)


Study
Sensitivity (95% CI)
Specificity (95% CI)
Azzopardi 2000
Azzopardi et.al 1999
Biagioni et.al 2001
Cseko et.al 2013
Eken et. al 1995
El-Ayouty et.al 2007
Finer et. al 1983
Hamelin et.al 2011a
Hamelin et.al 2011b
Holmes et.al 1982
Horst et. al 2004
Lieshout et.al 1995
Polat et.al 2012
Selton et.al 1997
Shankaran et.al 2011
Takeuchi et.al 1989
Thoresen et.al 2010a
Thoresen et.al 2010b
Toet et.al 1999
Watanabe et.al 1980
Westas et.al 1995
0 0.2 0.4 0.6 0.8 1
0 0.2 0.4 0.6 0.8 1

(b)

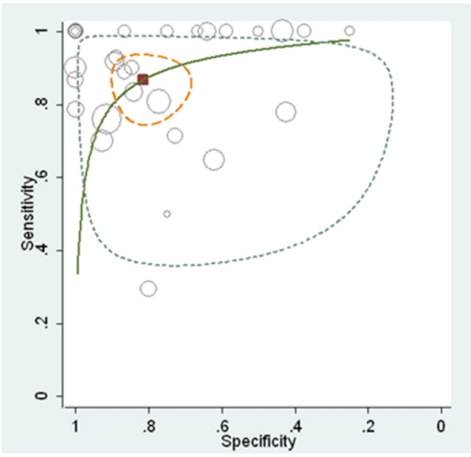


(c)

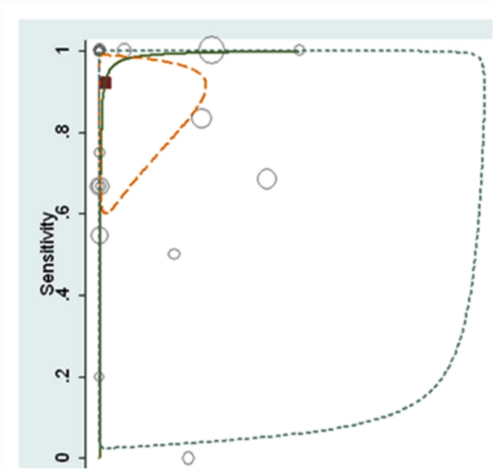
Figure 2.2: Forest plot of sensitivity and specificity for (a) burst suppression (b) low voltage (c) flat trace with outcome. Some studies [47, 51-57] used hypothermia in their HIE management. Letter 'b' and 'a' after the year of publication in the forest plots indicate the measure before and after cooling respectively of the same study.

2.3.2 Meta-analysis

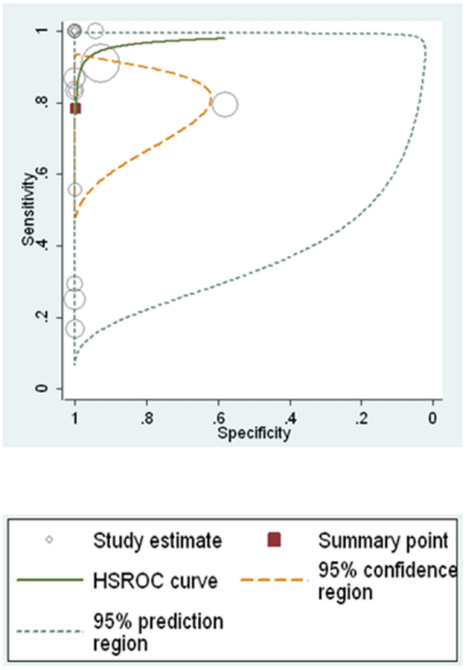
To generate a clear overview of the operating characteristics of different EEG background features, the ROC curves were estimated using HSROC model as recommended in the Cochrane handbook [49, 58]. Figure 2.3(a-c) shows the ROC curve using the HSROC model for burst suppression, low voltage and flat trace respectively. The pooled sensitivity and pooled specificity and overall summary of this review are given in Table 2.3.



(a)



(b)



(c)

Figure 2.3: Meta-analysis using HSROC model for: (a) burst suppression (b) low voltage (c) flat trace EEG related to outcome. The size of the circles is a visual representation of the number of subjects in each study. The summary point is the summary value for sensitivity and specificity confined by the 95% confidence region. The 95% prediction region gives a visual representation of the forecast of sensitivity and specificity in future studies (Stata, 2009).

Table 2.3: Pooled sensitivity and specificity with confidence interval for different EEG background patterns

EEG Background patterns	No. of Studies	No. of Neonates	Pooled Sensitivity		Pooled Specificity	
			Point Estimate	95% CI	Point Estimate	95% CI
Burst suppression	29	914	0.87	0.78 – 0.92	0.82	0.72 – 0.88
Low voltage	19	566	0.92	0.72 – 0.98	0.99	0.87 – 1.0
Flat trace	13	493	0.78	0.58 – 0.91	0.99	0.88 – 1.0

2.4 Discussion

This meta-analysis has established the prognostic value of different EEG background patterns in term neonates with HIE. To our knowledge this is the first review to analyse systematically different EEG techniques (CFM, aEEG, cEEG, international 10–20 EEG) and the background activity as a predictor of neurodevelopmental outcome. The meta-analysis clearly indicates that some features of EEG predict outcome accurately in term neonates with HIE.

One channel (CFM) or two channel (aEEG) EEG are currently widely used in NICUs due to ease of use and ease of interpretation. However several studies have shown that CFM or aEEG interpretation is user dependent [59, 60] and variable [61]. A recent meta-analysis has reported that sensitivity and specificity of aEEG and continuous EEG for predicting neurodevelopmental outcome are similar [44].

There are clearly trade-offs in using single or dual channel EEG or multichannel EEG [12, 44, 59, 60]. We therefore chose to review background features derived from studies using all 3 EEG data collection system. There are two critical issues that relate to the findings reported in this study: (a) issues relating to the definitions of abnormal patterns and (b) issues relating to the selection criteria for the studies included.

2.4.1 Issues relating to the definitions of abnormal patterns

A significant problem in the classification of different abnormal EEG features is that there is no universal definition of different abnormalities in terms of voltage level, phase or frequency. For example, voltage levels used to define an abnormality such as low voltage often differ between studies. Although not the primary goal of this review and meta-analysis, we note that in the case of isoelectric or flat trace EEG, different voltage amplitude levels are used in different studies: : $< 10\ \mu V$ was used in [62-65] , $< 5\ \mu V$ in [17, 66-75] and $< 3\ \mu V$ in [76]. Other studies did not provide an exact definition of the flat trace [77, 78] .

Similarly there was variation in the definition of 'low voltage': $< 5\ \mu V$ was used in [16, 17, 52, 73], $5 - 10\ \mu V$ by [74, 75], $5 - 15\ \mu V$ by [51, 69, 71], $5 - 20\ \mu V$ by [66, 68], $< 20\ \mu V$ was in [54, 79, 80], $5 - 25\ \mu V$ in [81-83], $< 25\ \mu V$ in [76], $< 30\ \mu V$ by [62] and $10 - 50\ \mu V$ in [64, 65]. Some studies did not provide a definition of low voltage [72, 78].

As with flat trace and low voltage, variations of the voltage amplitude used in defining burst suppression were different. To define suppression, $< 5\ \mu V$ was used by [16, 17, 70, 71, 73-75, 81], $0 - 2\ \mu V$ by [52], $5 - 10\ \mu V$ by [84, 85], $< 10\ \mu V$ by [62]. To define burst pattern: $\geq 25\ \mu V$ was used by [15, 74, 75] . $> 45\ \mu V$ was used by [79, 80], $> 50\ \mu V$ by [86], $75 - 250\ \mu V$ by [87] $100 - 1000\ \mu V$ by [88]. Unlike the voltage amplitude variation for defining bust suppression, in the frequency domain, most of the studies agreed that burst activity is dominant in the delta and theta bands.

The time and duration of EEG recordings has been variable. Most studies report using early EEG data – typically at 3-72 hrs but some have used recordings up to two weeks after birth to correlate with neurodevelopmental outcome as shown in Table 2.2. Some have contrasted times of recording: for example, Toet *et al*. showed that sensitivity and specificity of burst suppression, together with flat trace and low voltage, for poor outcome at 3 hrs was 85% and 77% respectively and at 6 hrs was 91% and 86% respectively [82]. However, not all reported better test characteristics at six hours than earlier

with Shany *et al*. (2006) reporting the sensitivity of burst suppression at 3 hrs to be 83% and at 6 hrs to be 75%.

2.4.2 Issues relating to the selection criteria for the studies included

Some studies described the relationship between individual background patterns and neurodevelopmental outcome without reporting on outcome of the babies who had been diagnosed with HIE and did not have the pattern: For example, burst suppression was reported in [18, 79, 89, 90] without providing information on the outcome of EEG without this feature. Thus we were unable to construct a 2-way table for the meta-analysis.

Some studies did not define the specific characteristics of EEG background patterns including low voltage, flat trace, and burst suppression but described more general classes such as minimal, mild, moderate, marked, and maximally depressed EEG background. In the studies of [64, 66, 67, 70, 71, 77, 85, 91, 92], we were able to extract information, but in 21 of 52 studies, it was not possible to extract individual EEG background information.

Some other background patterns including trace alternant, asynchrony and, asymmetry have also been reported but for only a very small number of babies. Murray *et al.* and al Naqeeb *et al.* reported that, 5 out of 7 newborns and 1 out of 40 newborns respectively, who showed some degree of asymmetry on EEG had a poor neurodevelopmental outcome whereas Zeinstra et al. reported 1 out of 36 newborns showing this feature had normal development [62, 64, 84]. Mariani *et al.* found that 75% of the newborns who had trace alternant had poor neurodevelopmental outcome [68]. Although these two background features have been reported to predict outcome moderately well in a relatively small number of infants, the ability of burst suppression, flat trace and low voltage EEG to predict outcome is superior in a much larger number of babies (see Table 3). A number of studies collapse these features, in variable and sometimes ill-defined ways and classify the EEG as ‘moderate’ in abnormality [32, 62, 63, 70, 85, 93, 94]. Although ‘moderate’ EEG predicts neurodevelopmental outcome, we chose to find the specific background patterns that best predict neurodevelopmental outcome and hence, we excluded these studies. Some studies [52, 84, 89, 95] include seizure as a background pattern but we did not classify seizure as a background feature and so excluded this feature from our analysis.

2.4.3 Assessment of neurodevelopmental outcome

Neurodevelopmental outcomes were measured using different assessment methods, criteria and at different ages from 12 months to 9 years. Studies used structured examinations [96-101] and/or general neurological examinations [102-106] to define 'normal' or 'abnormal' neurodevelopmental outcome as shown in Table 2.2. A few studies used their own follow up index to define 'normal' or 'abnormal' outcome [47, 69, 71]. Some studies sub-classified outcome as normal, minor or mildly abnormal, moderate and severely abnormal or death. During review of the relevant studies, it was agreed to collate normal, minor or mildly abnormal outcome as a 'normal outcome' and moderate or severely abnormal or death as an 'abnormal outcome'. This simplified the data to allow 2-way table calculations.

2.4.4 Strengths and weaknesses

This review is based on the relationship between EEG background patterns and outcome. The meta-analysis was performed taking the PRISMA statement [107] into account. Analysis using the random effects model and HSROC model was used to account for heterogeneity among numerous studies over the last 50 years with long-term neurodevelopmental follow-up ranging from 12 months to 9 years which added to the reliability of this structured review.

A meta-analysis published by Spitzmiller *et al.* (2007) was focused only on aEEG and included 8 studies [108] all of which were captured by our search criteria and included in our meta-analysis. This study does not report specific background patterns but reporting as severe abnormal EEG patterns with neurodevelopmental outcome.

Like other systematic reviews and meta-analysis, the presence of bias is possible. Selection bias may have resulted from the approach to selecting the different voltage amplitudes, phases, and frequencies used to define different abnormal EEG patterns as well as differences in the timing of the EEG. We introduced language bias by restricting to English language sources. Publication bias is also possible by selecting only journal articles and excluding conference paper, case studies, and review articles.

Like EEG background features, the presence of clinical and/or subclinical seizures can also predict neurodevelopmental outcome. A recent review covering literature from 1954 to 2013 suggests that 17.9% of neonates with seizures develop epilepsy and in 80.7%, epilepsy was associated with other neurological impairments [109]. Given the differing ability of CFM, aEEG and full EEG to diagnose seizures, we support the recent call for evidence-based studies of seizure recognition and management [110]. A more detailed focus on neonatal seizures is beyond the scope of this project.

2.4.5 Validity of result in newborns treated with hypothermia

Therapeutic hypothermia has been introduced into standard practice in the past decade in babies with HIE as it reduces rates of death, severe disability and cerebral palsy [111, 112]. EEG appears not to be significantly affected by HT [113, 114] and so it is reasonable to include studies with babies treated with HT as we have done for [47, 51-55].

2.4.6 Suggestions for type of EEG recording in NICUs

We have shown the potential for variation in predicting outcome to be caused by different EEG recording techniques, the timing of EEG recordings and the specific definitions used for defining background features. We make the following suggestions to assist in optimising the use and interpretation of EEG in NICUs:

Timing of EEG: Attempts should be made to report findings at 6 hours after birth because several studies have shown that features at this time have the best predictive characteristics [62, 82].

Suggested EEG techniques: There are trade-offs among EEG techniques. For longer term monitoring including for the detection of seizures, we suggest aEEG ideally with intermittent full array EEG as described in [115] to confirm features and seek evidence of seizures not seen in 2 channel EEG. aEEG electrode position should be at P_3 and P_4 location because it is over the cerebrovascular watershed region at high risk for acquired injury [116] and more prone to seizure [117].

Finally, continuous computer analysis allows robust measurement of features and use of frequency and time-dependent analytic strategies may allow features not detectable to visual analysis of EEG to be detected [115, 118]. Suggested definitions of relevant EEG background features are given in Table 2.4.

Table 2.4: Recommended definitions of EEG background features:

Patterns	Amplitude, duration or characteristics of EEG activity*
CNV[&]	Continuous background activity with voltage $\pm 10-25\ \mu V$ but without sleep stages.
DNV	Discontinuous trace, with voltage predominantly $> 5\ \mu V$.
Burst suppression[#]	High voltage $(> 30 \mu V)$ delta $(0.4-4 Hz)$ and theta $(4-8 Hz)$ activities lasting $1-10 sec$ with suppressed activity of $< 5\ \mu V$ lasting $> 2\ s$.
Modified burst suppression	Burst suppression with suppressed activity of $> 5\ \mu V$.
Asymmetry	$> 25\%$ consistent asymmetry between homologous brain regions can be treated as *abnormal* asymmetry. Asymmetry should be present in all states [23].
Asynchrony	$> 50\%$ of EEG burst that occurs asynchronously between hemispheres and the time difference of the onset of burst between hemispheres is $> 1.5\ s$ can be treated as *abnormal* asynchrony. To measure asynchrony, 5 consecutive minutes should be used [23].

Low voltage	Continuous background patterns $(5 \pm 3)\ \mu V$ throughout the record.
Flat trace	Mainly inactive (isoelectric tracing), consistently $< 5\ \mu V$.

*Age related issues need to consider while defining abnormal background feature. This is because pattern which is normal at preterm and may be treated as abnormal in term neonates [75].

& CNV with the band of aEEG activity altering in width, indicating cycling of sleep stages can be treated as CNV with cycling of sleep stages (CNV-S)[119].

#According to reactivity to external stimuli burst suppression can be classified as reactive and non-reactive burst suppression and both have prognostic value [18]. According to density of occurrence of burst per hours, burst suppression is also classified as burst suppression (+) and burst suppression (-) [burst suppression (+) denotes burst density ≥100 burst/h, and burst suppression (-) means burst density <100 burst/h [75]]. The time interval between two consecutive bursts is called inter-burst interval (IBI). IBI can be classified as type 1 and type 2 according to the cerebral activity [Type 1 denotes IBI ≥2 s with no cerebral activity in any channel and type 2 means rare transient or low voltage activity [120]]. Sometimes prolonged suppression periods with bursts are termed Permanent Discontinuous Activity (PDA) [23].

2.5 Overall summary

This chapter established the background features of EEG that accurately predict long-term neurodevelopmental outcome in term neonates with HIE. These results have been generated in studies of term neonates. Further studies are needed to determine the usefulness of EEG background patterns for prognostication of outcome in preterm infants.

Differences in definitions of and lack of an accepted classification system for EEG background patterns are a major impediment to synthesis of available information. Published material may use the same background pattern term but use different amplitude level, phase, frequency and duration in the definition. Agreement on definitions is necessary for the effective implementation and use of EEG in NICUs and we have suggested some definitions.

The studies reviewed here are from different decades covering the change from analogue to digital EEG equipment. However, we are interested in the ability of EEG background patterns in term infants to predict neurodevelopmental outcome rather than precision of this clinical tool. Despite these issues (Sensitivity 0.87 [95% CI (0.78-0.92)]; Specificity 0.82 [95% CI (0.72-0.88)], low voltage (Sensitivity 0.92 [95% CI (0.72-0.97)]; Specificity 0.99 [95% CI (0.88-1.0)], and flat trace (Sensitivity 0.78 [95% CI (0.58-0.91)]; Specificity 0.99 [95% CI (0.88-1.0)]) predict neurodevelopmental outcome with a high sensitivity and specificity. This contribution may motivate the neurophysiologists and the clinicians to take early management process and also opens the door for engineers to further processing the signals such as signal characterisation, automatic detection and classification. Most of the patterns are marked by visual inspection. Visual inspection of the EEG signal is routine but visual interpretation and classification is laborious and time-consuming, especially in the case of long recordings. It also requires experienced interpreters, who are not always available. To overcome these limitations, an automated detection and classification approach would

provide a decision support to the clinical staff. State-of-the-art signal processing and machine learning approach will be designed, optimised and applied for the analysis and automatic classification of these prognostic EEG background features in the next three chapters.

Chapter 3 Design and optimisation of time-frequency distribution and application to EEG[3]

3.1 Introduction

Accurate characterisation of EEG, including its background patterns in the (t, f) domain, helps understand the characteristics of EEG signals and assists in optimal representation and in extracting the key features from EEG for automatic detection and classification.

What is already known on this topic?

QTFD is very useful for the characterisation and analysis of nonstationary signals e.g. EEG signal.

Proper selection of QTFD kernel parameters, analysis window and window parameter(s) offers a high-resolution (t, f) signature and improves characterisation.

What is the contribution of this study?

- ✓ Two optimisation methods: (i) global optimisation (using Hybrid Genetic Algorithm (HGA)) and (ii) local optimisation (using Locally Optimised Spectrogram (LOS)) have been proposed for a fully automatic optimisation of QTFDs.
- ✓ Suitable for the non-specialist users and provides a useful tool for the analysis of non-stationary signals.
- ✓ Different measures have been used and compared to justify and validate the efficiency of the proposed methods.

In Chapter 1, it was shown that EEG is a non-stationary signal, i.e. the statistical properties of the frequency and amplitude content change over time, resulting from the relatively random firing of neurons. It has been shown that joint (t, f) representation, i.e. TFD, is well-adapted for the characterisation and analysis of this non-stationary signal. In addition, various dynamic features extracted from TFD can increase detection and classification accuracy [33]. However, a TFD with high resolution and high concentrations with no cross-terms is desirable in order to provide optimal or ideal representation of a signal and extract features for detection and classification.

Over the last decades, a large number of TFDs have been proposed by researchers to obtain the desirable properties mentioned above [33]. Generally, TFDs have some parameters; these are used to adjust the TFD kernels and improve the

[3] This chapter is an extended version of the following publications:
(i) M. A. Awal and B. Boashash, "An automatic fast optimization of Quadratic Time-frequency Distribution using the hybrid genetic algorithm," *Signal Processing*, vol. 131, pp. 134–142, Feb 2017.
(ii) M. A. Awal, S. Ouelha, S. Dong and B. Boashash, "A robust time–frequency representation based on the local optimization of the Short-Time Fractional Fourier Transform, *Digital Signal Processing*, vol. 70, pp. 125-144, Aug 2017.

characterisation of various non-stationary signals. The selection of optimal parameters depends on the nature of signal: the parameters that are optimal for one signal are different for another signal. The optimal selection of these parameters offers a highly energy concentrated (t, f) representation with signal-dependant resolution [34]. But, *how can we select these parameters for a multicomponent signal without a priori signal information*? This is an important problem as the solution of this problem will provide an automatic procedure to obtain a high-resolution TFD, which in turn can significantly improve signal characterisation and classification.

The most popular TFDs are the spectrogram at one hand and the Wigner-Ville distribution (WVD) at the other hand. The performance of the spectrogram is heavily dependent on the analysis window, which is set heuristically ([33], pp. 78–79) [121]. This fixed window approach is not suitable for multicomponent signals comprising both short and long duration overlapping components. This is a major limitation of the use of the spectrogram.

WVD is free from analysis window considerations and provides superior component concentration; however, it suffers from cross-terms, potentially leading to confusion and misinterpretation of the signal components (due to the quadratic nature of the transform) ([33], Chapter 3). A well-known technique is to smooth or filter the WVD by a kernel function to reduce the cross-terms, resulting in Reduced Interference Distributions (RIDs). Consequently, RIDs have been extensively used due to their high resolution and easy interpretation in the (t, f) −plane. However, the performance of RIDs depends on the choice of kernel function and each kernel function has one or more kernel parameters balancing concentration, resolution and cross-terms in RIDs.

The analysis window, window parameters and kernel parameters are often selected manually based on visual inspection (([33], pp.93–94, 297–298), [34]). This is time consuming and not feasible in certain applications – for example, in automatic pattern recognition and classification [29]. Additionally, the specialist knowledge of concentration and resolution that is required for correct interpretation precludes non-specialist operators. In this chapter, we address and mitigate this complexity by providing two optimisation methods (global optimisation and local optimisation) to fully automatic optimise the TFDs. Section 3.2 presents a global optimisation method that uses a hybrid genetic algorithm. Section 3.3 presents a local optimisation method based on the Locally Optimised Spectrogram. Section 3.4 and Section 3.5 provide a qualitative assessment of these two methods and their potential applications, respectively.

3.2 Method-1: global optimisation method

The hybrid genetic algorithm is proposed for a fully automatic optimisation of TFDs. First, we will define the problem mathematically and then describe the detailed methodology. The key topics covered by this section are as follows:

- The problem of optimisation of RIDs (also called QTFDs due to the quadratic nature) has been formulated as minimising the cost function of a modified energy concentration measure for non-stationary signals without *a priori* signal information (Section 3.2.2).
- A hybrid genetic algorithm (HGA) has been proposed to find the optimal and fine-tuning QTFD parameters (Section 3.2.3).
- Different state-of-the-art QTFDs like SPWVD, EMBD, CKD and ADTFD have been used to demonstrate the proposed approach (Section 3.2.4) and compared with other methods (Section 3.2.7).
- The robustness of the proposed method has been demonstrated under different signal-to-noise ratio (SNR) conditions in the application of IF estimation (Section 3.2.5).
- A fast and memory efficient optimisation has also been implemented in Section 3.2.6 for the optimised use of TFDs, especially for long recordings or multi-sensor data, for example, in multichannel EEG recordings.

Appendix 3.A describes the computer code used in this section.

3.2.1 Background Problem

For a given analytic signal[4] $z(t)$ associated with a real signal $x(t)$, the signal kernel $Kz(t,\tau)$ is defined as

$$K_z(t,\tau) = z\left(t+\frac{\tau}{2}\right) z^*\left(t+\frac{\tau}{2}\right) \tag{3.1}$$

[4] Analytic signals are used when producing their TFDs because the total bandwidth of the signals can be halved by removing the negative frequencies, allowing a minimal sampling frequency at half of the usual Nyquist rate. Also, by using analytic signals, the interference terms due to the interaction between positive and negative components can be avoided. Given a real signal $x(t)$, its analytic associate can be defined as: $z(t) = x(t) + j\,\mathcal{H}\{x(t)\}$; where $\mathcal{H}\{x(t)\}$ is the Hilbert transform of $x(t)$ and can be obtained by

$\mathcal{H}\{x(t)\} = \mathcal{F}^{-1}_{t\leftarrow f}\{[-j.sgn(f)\,\mathcal{F}_{t\rightarrow f}\{x(t)\}]\}$, where $sgn(f)\begin{cases}-1, & if\ f<0\\ 0, & if\ f=0\\ +1, & if\ f>0\end{cases}$

The Matlab function $hilbert(x)$ can be used to compute analytic signal.

where τ and $*$ indicate the temporal lag and complex conjugation. The Fourier transform (FT) of $K_z(t,\tau)$ with respect to τ results in the WVD of the signal, i.e.:

$$W_z(t,f) = \mathcal{F}_{\tau\to f}\{K_z(t,\tau)\} \tag{3.2}$$

WVD provides superior component concentration; however, it suffers from cross-terms, potentially leading to confusion and misinterpretation of the signal components (due to the quadratic nature of the transform) ([33], Chapter 3). A well-known technique is to smooth or filter the WVD by a kernel function to reduce the cross-terms, resulting in Reduced Interference Distributions (RIDs), $\rho_z(t,f)$. It can be represented in the continuous (t,f) domain as:

$$\rho_z(t,f) = W_z(t,f) \underset{t}{*}\underset{f}{*} \gamma(t,f) \tag{3.3}$$

Where $\underset{t}{*}$ and $\underset{f}{*}$ denotes the convolutions over time and frequency respectively. It can also be defined in the discrete domain [[33], pp. 364]:

$$\rho_z[n,k] = \underset{m\to k}{DFT}\left\{G[n,m] \underset{n}{*} (z[n+m]\, z^*[n-m])\right\} \tag{3.4}$$

where $G[n,m]$ is the discrete time-lag kernel, $z[n]$ that defines the analytic associate of the original signal, $y[n]$, $\underset{n}{*}$ denotes the convolution in the discrete time domain, and $(.)^*$ represents the complex conjugate. Table 3.1 provides the definition of different QTFDs, their time-lag kernel functions, and the controlling parameter(s) used in this study. For brevity, this study confines the discussion to state-of-the-art fixed and adaptive QTFDs like smoothed pseudo Wigner-Ville distribution (SPWVD), extended modified B distribution (EMBD), extended compact support kernel or compact kernel distribution (CKD) and adaptive directional TFD (ADTFD) [33]. These TFDs have proven popular, as they offer better resolution and concentration; this is because they can independently adjust smoothing kernel parameters along the time and frequency axis (they are also called 'separable kernel TFDs'), ensuring the signal-dependent kernel results in a high resolution TFD [122]. In this study, the adjustment of kernel parameters is formulated as an optimisation problem that provides the most compact and high-resolution time-frequency representation.

Table 3.1: QTFDs used in this study ([33], p. 341).

QTFDs	$G[n,m]$	Controlling parameters
WVD	$\delta[n]$	-
SPWVD	$w[n]\,w[m]$	$w[.]$ = Kaiser window; $\mathrm{w}_{\mathrm{L_t}}, \beta_t, w_{L_f}, \beta_f$
EMBD	$\frac{cosh^{-2\beta}[n]}{\sum_n cosh^{-2\beta}[n]}\frac{cosh^{-2\alpha}[m]}{\sum_n cosh^{-2\alpha}[m]}$	α, β; $0 \le \alpha \le 1, 0 \le \beta \le 1$
CKD [29]	$DFT_{l \to n}\left(\frac{e^{\frac{cD^2}{\left(\frac{l}{N}\right)^2}}}{-D^2}\right) e^{2c}\left(\frac{e^{\frac{cE^2}{\left(\frac{m}{N}\right)^2}}}{-E^2}\right)$	c, D, E
ADTFD	See Ref. [[33], pp. 299-305]	a, b

3.2.2 Formulation of the optimisation problem

An optimal QTFD maximises the $\rho_z[n,k]$ concentration and resolution, as well as minimising cross-terms for all signal components. To achieve this, a measure needs to be used as an objective or cost function (J). TFDs' information theoretic measures, the ratio of norms-based measure, the normalised instantaneous resolution measure and the energy concentration measure have been reported as TFD measurement indexes ([33], Chapter 7). Each index has strengths and weaknesses; for example, the norm-based concentration measure must be normalised to ensure accurate performance when cross-terms exists. It also discriminates the low-concentrated components relative to the highly concentrated ones within the same TFD; a detailed review can be found in [123]. A comprehensive resolution measure has been proposed in [124] and modified in ([33], pp. 438–44). This is an in-depth measure based on the morphology of a TFD and calculated by the weighted sum of concentration, resolution and cross-term interference [124]. However, the use of this method is computationally costly because the position of cross-terms can be defined only after an iterative procedure for a real-life multicomponent signal without *a priori* signal information [125]. On the other hand, the energy concentration measure (ECM) is efficient, simple to implement and does not suffer from the problem of low energy concentration for weak signal components [123]. Due to its simplicity, this index is used as an objective function for the proposed optimisation and can be defined for a normalised TFD $(\mathrm{i.e.}\sum_{n=1}^{N}\sum_{k=1}^{M}\rho_z[n,k] = 1)$ as:

$$\mathrm{J}(\rho_z[n,k]) = \frac{1}{2N}\left|\sum_{n=1}^{N}\sum_{k=1}^{M}|\rho_z[n,k]|^{1/2}\right|^2 \qquad (3.5).$$

The term $\left(\frac{1}{2N}\right)$ multiplies the original ECM index (hence, called modified ECM) to scale down the concentration as a convenience for the computation of some algorithms, e.g. gradient descent. The rationale is that the ECM index, i.e. Equation (3.5). is a square function and different function minimisers (e.g. see Equation (3.13) are based on the derivative of the cost function. The derivative of a square function produces 2 and $\frac{1}{2}$ will cancel out the 2. Thus, multiplying the original ECM index by $\frac{1}{2}$ will remove the doubling effect and scale down the ECM index. In addition, $\frac{1}{N}$ is also multiplied with the ECM index to make the cost function less dependent on the signal length (i.e. number of samples) and produces a normalisation effect. This justifies the use of $\frac{1}{2N}$ as the ECM index scaling factor. A lower value of $J(\rho_z[n,k])$ represents a more compact TFD representation. This value is signal dependent, as it is affected by the number and closeness of signal components, signal duration (i.e. signal length) and signal complexities. The objective function for the proposed method can be formulated as

$$\min_{kernel\ parameters} J\left((\rho_z[n,k]);\ kernel\ parameters\right) \tag{3.6}$$

This general framework for the QTFD optimisation problem can be interpreted as 'minimise the cost $J\left((\rho_z[n,k]), \text{kernel parameters}\right)$ as a function of QTFD kernel parameters'.

To illustrate this optimisation problem, the smoothed pseudo-WVD (SPWVD) is considered first. It belongs to a subclass of QTFDs known as separable kernel distributions [125]. Window length and window parameters are the kernel parameters in the case of SPWVD. A Kaiser-Kaiser window has been chosen in this study, as the controlling parameter (β) spans the fundamental window trade-off between main-lobe width and side-lobe level. The proper adjustment of the window length and controlling parameter can be done using a data-adaptive kernel. So, the controlling parameters of Kaiser-Kaiser SPWVD are $w_{L_t}, \beta_t, w_{L_f}, \beta_f$ which represent the window lengths and window parameter in the time and frequency axis, respectively. The optimisation problem in the case of the Kaiser-Kaiser SPWVD can then be formulated as:

$$\min_{w_{L_t},\beta_t,w_{L_f},\beta_f} J\left((\rho_{z_{SPWVD}}[n,k]);\ w_{L_t},\beta_t,w_{L_f},\beta_f\right) \tag{3.7}$$

which is interpreted as '*minimise the cost* $J\left((\rho_{z_{SPWVD}}[n,k]);\ w_{L_t},\beta_t,w_{L_f},\beta_f\right)$ *as a function of* $w_{L_t},\beta_t,w_{L_f},\beta_f$'. Similarly, for EMBD, CKD and ADTFD the optimisation problem can be formulated as

$$\min_{\alpha,\beta} J\left((\rho_{z_{EMBD}}[n,k]);\ \alpha,\beta\right) \tag{3.8}$$

$$\min_{c,D,E} J\left(\left(\rho_{z_{CKD}}[n,k]\right);\ c,D,E\right) \tag{3.9}$$

$$\min_{a,b} J\left(\left(\rho_{z_{ADTFD}}[n,k]\right);a,b\right) \tag{3.10}$$

3.2.3 Proposed optimisation: hybrid genetic algorithm

The optimisation of QTFDs is a complex problem. Classical gradient-based optimisation techniques for optimising QTFDs (i) are highly dependent on the initial values of the kernel parameters, (ii) generate only a single point at each iteration resulting in many iterations, and (iii) do not guarantee reaching a global minimum. These constraints motivate us to use the genetic algorithm (GA), as it is derivative-free, produces a population of TFD kernel parameters at each iteration and the best parameters in the population converge to an optimal QTFD [126]. Thus, the GA is a useful method of solving such a complex problem, and it can be defined as eight-component tuples, i.e. $GA = (E, F, P_0, P, \Phi, \Gamma, \Psi, T)$, where $E, F, P_0, P, \Phi, \Gamma, \Psi, T$ represent an encoding method of individual, fitness (or objective function), initial population, population size, selection, crossover and mutation operator and termination condition respectively. It is based on the 'natural selection' process, imitating the principles of biological evolution, i.e. repetitively adapt a population of individual solutions from the current population and treat them as parents to produce the offspring for the next generation. Over consecutive generations, the population approaches an optimum solution. An overview of GA can be found in [126], while the terminology and uses of 'GA' can be found in [127].

The GA can approach the optimal solution relatively fast, but to achieve convergence it requires the evaluation of many functions. The proposed method to optimise RIDs is to run the GA for a small number of generations. It will produce sub-optimal parameters for QTFDs that solves the initial value problem (unlike classical optimisation techniques), therefore making the technique fully automated. These QTFD parameters are then used as initial points for another optimiser that is faster and efficient for a derivative-free local search, based on the Nelder–Mead algorithm [128]. This combination is denoted as hybrid genetic algorithm (HGA) and shown in Figure 3.1. The HGA setup parameters and functions are presented in Table 3.2 (See Section 3.2.8 for details).

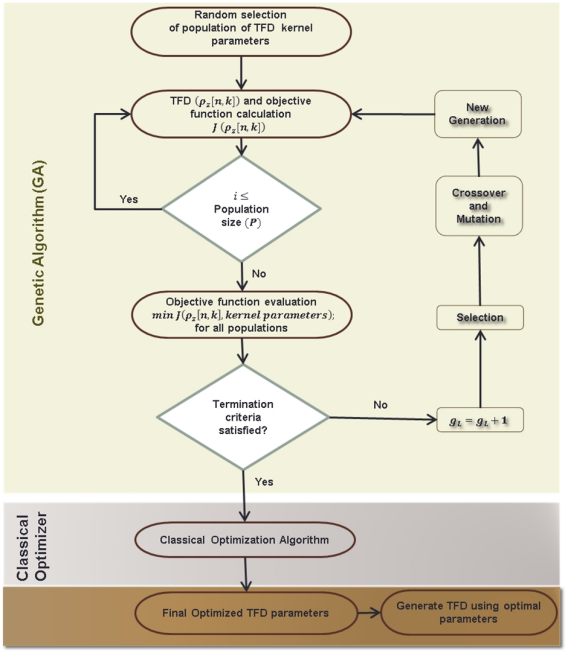


Figure 3.1: Hybrid genetic algorithm (HGA) for TFD parameters optimisation; $i = 1, \dots, P$ and $g_L = 1, \dots, G_L$ represent the population size and generation limit respectively. Firstly, a population $i = 1, \dots, P$ of TFD parameters are randomly selected within the bounds (see Table 3.3) and then evaluated and applies genetic operations (selection, cross-over and mutation) to generate another population of TFD parameters for the next generation. It iterates until the termination criteria are satisfied and finally GA provides the best fitted TFD parameters. These parameters are further refined by a Nelder-Mead algorithm and finally the optimal TFD parameters are selected.

Table 3.2: HGA setup.

Max. number of generations	See Table 3.3
Stall generations	50
Fitness function	Eqn. (3.6).
Population	
size	20
type	Double Vector
Selection function	Stochastic uniform
Elite count	2
Crossover	
type	Scattered
fraction	80%
Mutation	
Type	Gaussian
init. variance	1
Migration	
direction	forward
interval	20
fraction	20%
Scaling function	Fit scaling rank
Classical Optimiser	Nelder–Mead Algorithm

The GA can diverge from the feasible space if the creation of the initial populations is unbounded. To deal with that, populations are bounded by the known lower and upper bounds of the parameters. These bounds are recommended in previous studies ([33], pp. 125–28, 299–305), [125]) and shown in Table 3.3.

Table 3.3: Number of generations and QTFD parameter set up.

TFD Name / Bounds	SPWVD (Kaiser-Kaiser)				EMBD		CKD			ADTFD	
Parameters	$w_{L_t}^{\&}$	β_t	$w_{L_f}^{\&}$	β_f	α	β	c	D	E	a	b
Lower Bounds	5	0.01	5	0.01	0	0	0.01	0.01	0.01	2	4
Upper Bounds	$\lfloor \frac{N}{3} \rfloor$	20	$\lfloor \frac{N}{3} \rfloor$	20	1	1	10	1	1	3	30
# of Generations	20				20		30			20	

$\lfloor . \rfloor$ is the $floor$ function; $^{\&}$ should be integer.

3.2.4 Results on method-1

Two signals were used to illustrate the performance of the proposed method: (a) a multicomponent simulated signal $(s_a(t))$ with different orientations, and (b) a real-world bat signal $(s_b(t))$. For convenience, all (t, f) representations were given with the same MATLAB colormap, in order to show results in a fair and comparable manner. Different QTFDs with their optimal kernel parameters are shown to exhibit the resulting high resolution and high concentration TFD. A better trade-off was observed between the interferences and the (t, f) resolution when the parameters are optimised, as detailed below.

The simulated multicomponent signal $s_a(t)$ consists of linear FMs, constant FM and frequency-modulated Gaussian (see Section 3.3.4.1.4 for signal model). Figure 3.2 illustrates the optimal (t, f) representation of this signal using the SPWVD, EMBD, CKD and ADTFD with HGA selected parameters. For comparative purposes, the ideal TFD and the WVD have also been represented. The optimised EMBD provided a sharp localisation but suffers from cross-terms whereas the SPWVD failed to separate closely spaced LFM components; the CKD provided the best compromise among the fixed TFDs. The ADTFD showed the best performance and very close to ideal TFD as it is adapted to (t, f) directions. All fixed and adaptive TFDs showed a significant improvement in resolution when compared to the WVD.

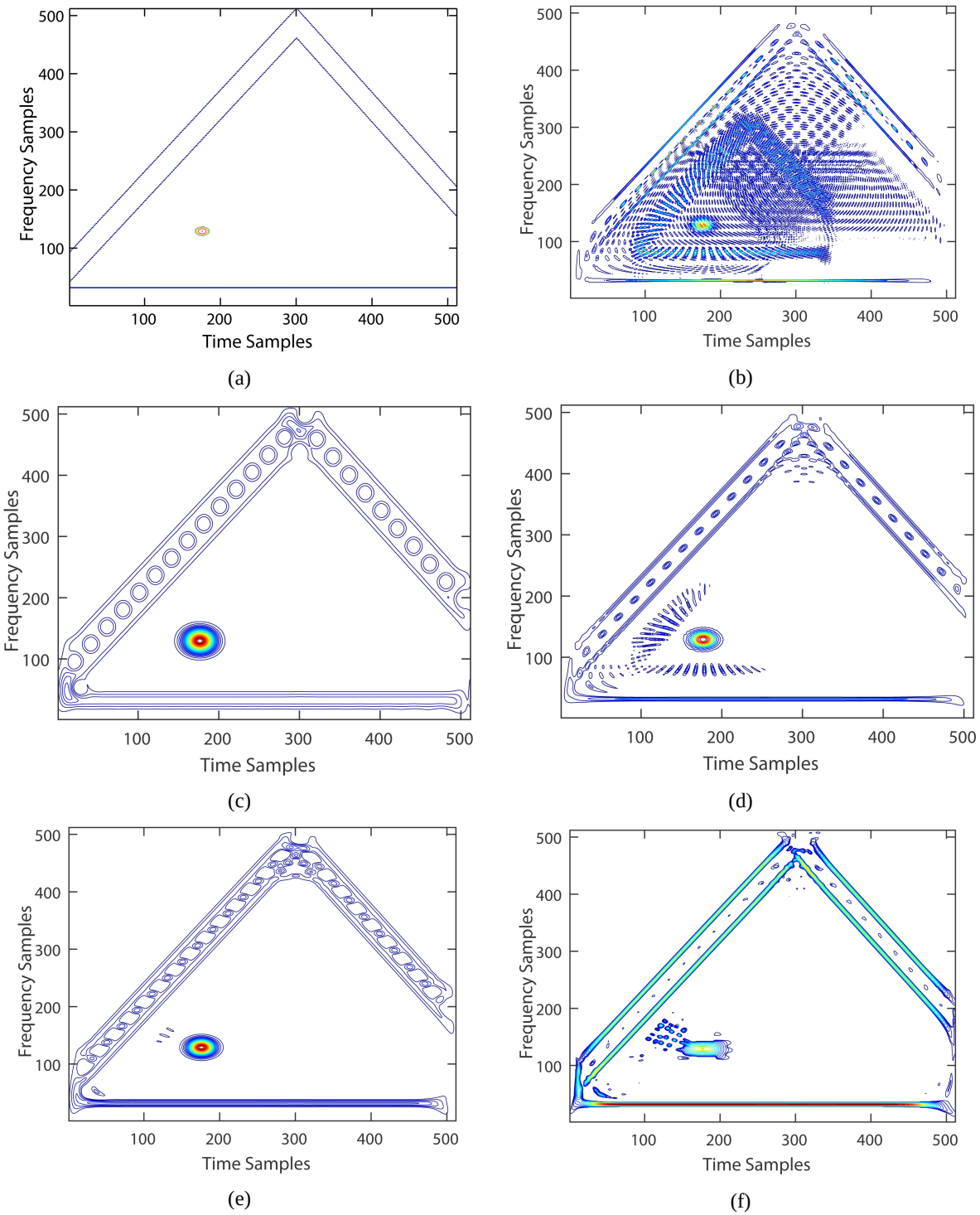


Figure 3.2: (a) Ideal TFD representation and (b) WVD. The optimisation of QTFDs using HGA method: (c) Kaiser-Kaiser SPWVD, $(w_{L_t}, \beta_t, w_{L_f}, \beta_f) = (97, 12.3, 143, 13.6)$, (d) EMBD, $(\alpha, \beta) = (0.3, 0.3)$ (e) CKD, $(c, D, E) = (7, 0.2, 0.2)$; and (f) ADTFD, $(a, b) = (3, 13.4)$. Optimised parameters are shown in brackets, ().

The real-world bat signal $s_b(t)$ was also used to show the efficiency of the proposed method. The optimised SPWVD and CKD provided cross-term free signatures, whereas the EMBD provided a sharp localisation but suffers from cross-terms. The optimised ADTFD offered the sharpest

localisation of all the components due to its adaptive procedure (see Figure 3.3). Note that an ideal TFD is not shown as it is not known precisely given that it is a real-world signal, and the WVD is also not shown as it is known to suffer from heavy cross-terms.

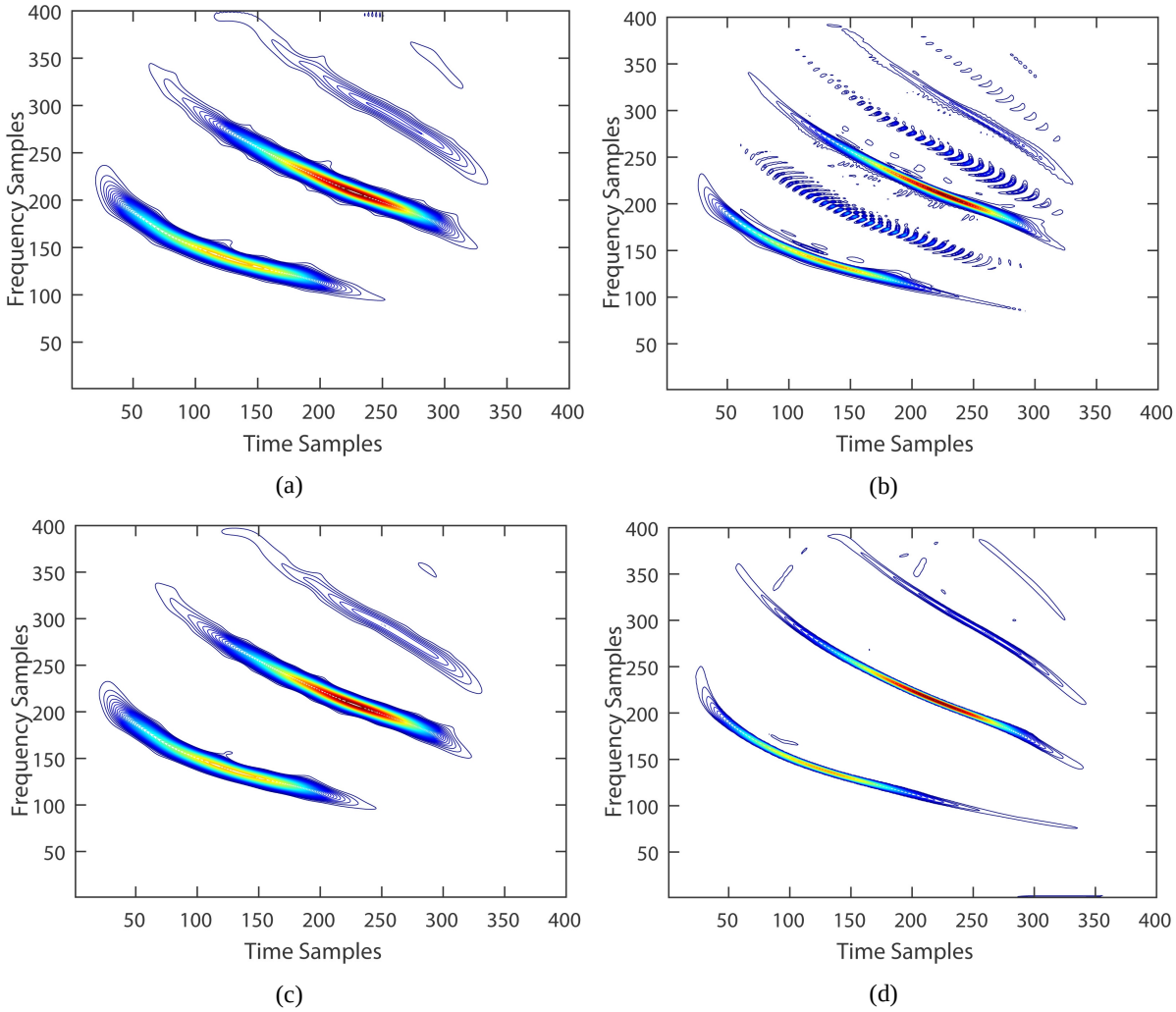


Figure 3.3: Optimisation of QTFDs using HGA method: (a) Kaiser-Kaiser SPWVD, $(w_{L_t}, \beta_t, w_{L_f}, \beta_f) = (59, 7.8, 160, 6.7)$, (b) EMBD, $(\alpha, \beta) = (0.2, 0.5)$ (c) CKD, $(c, D, E) = (3.5, 0.12, 0.1)$ and (d) ADTFD, $(a, b) = (3, 15.8)$, Optimised parameters are shown in brackets, ().

3.2.5 Instantaneous frequency (IF) estimation

One of the unique capabilities of (t, f) methods is the estimation of the signal IF ([33], pp. 611–19, 620–26). This IF estimation capability is often used as a performance evaluation criterion for TFDs, including its robustness in the presence of noise. In this study, the IF was estimated by the component extraction method presented in ([33], pp. 611–19; [129], see Appendix 3.C). The signal model is defined as:

$$x(t) = \cos(18.85t^3 + 125.664t) + \cos(18.85t^3 + 94.248t) + n(t); \quad (3.11)$$

where $n(t)$ is additive white Gaussian noise. Figure 3.4 shows the logarithmic mean square error (MSE) between the actual and the estimated IF under different SNRs by performing 100 simulations at each SNR level. For this signal, different optimised TFDs showed better performance than the WVD. In addition, the locally optimised TFD also performed better than the globally adaptive TFDs (e.g. global and local adaptive EMBD). As expected, the ADTFD with optimal parameters outperformed the other TFDs for IF estimation. These results show that an accurate estimation of the IF of the signal components is directly related to the auto-term resolution property and cross-term suppression property of (t, f) methods.

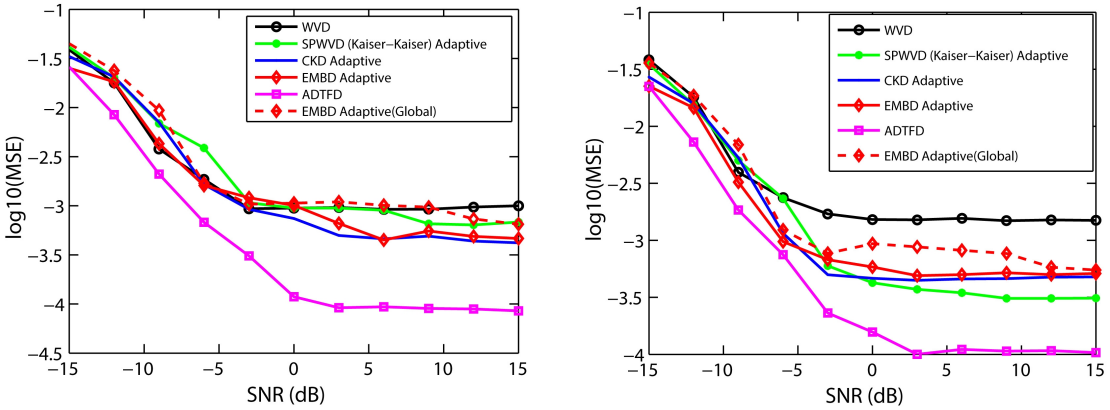


Figure 3.4: IF estimation of a multicomponent signals using different optimised QTFDs by the HGA method under different SNRs ranging from -15dB to $+15dB$; (a) logarithmic MSE for $x_1(t)$ signal component and (b) for $x_2(t)$ signal component. EMBD global adaptive represents that the same parameters (hence, called global) were used in EMBD in each iteration under all SNR conditions whereas EMBD local adaptive (as well as, in all other QTFD cases), the QTFD parameters were determined by the HGA in each iteration under all SNR conditions.

3.2.6 Fast and memory efficient implementations of QTFDs

In the era of 'big data', the issues of TFD computation and optimisation present a significant challenge in the case of long recordings or multisensor data such as multichannel EEG signals. This is because TFDs are two dimensional (2D) functions and require $N^2 \log_2 N$ numerical operations (a basic FFT element requires $N \log_2 N$ operations). An increase in the number of numerical operations will also result in an increase in computational time; programs often stop due to 'out of memory' errors for long data. Therefore, designing memory efficient and optimised TFD implementations is very important ([33], Section 6.6).

The separable-kernel TFD, where $G[n, m] = g[n].g[m]$, is usually oversampled. TFD computations and memory requirements can be significantly reduced by eliminating this oversampling. In addition, modern computers possess multicore processors and parallel computing can be used to further speed

up the HGA optimisation. Following the methodology presented in [33] (pp. 374–83) for generating fast TFDs and parallel computing, the efficient optimisation of TFDs was also implemented in this study. This process not only optimises TFDs in terms of optimising the TFD parameters, but also in terms of computational time. Table 3.4 compares the computational time of the optimal EMBD using a standard and memory efficient implementation. Further memory reduction can be done by generating under-sampled TFDs computing only a subset of the TFDs (called decimated TFD method).

Table 3.4: Computation time of standard and fast implementation of EMBD*

Signals (N)	Standard implementation (s)	Fast implementation (s)
Simulated signal($N = 512$)	104.6 ± 1.9	3.97 ± 0.2
Bat signal ($N = 400$)	53.1 ± 1.2	3.59 ± 0.4
Newborn EEG Seizure ($N = 1024$)	878.7 ± 43.1	11.5 ± 0.7

*This experiment was run in the Windows-7 on an Intel Core i-7 platforms having 8GB RAM and was connected to 4 'local workers' using Matlab's Parallel Computing Toolbox. Each signal was run 50 times and the mean±standard deviation of the computation time is presented here.

3.2.7 Comparison with other techniques

The TFD optimisation method can be implemented in a straightforward way by minimising the cost function using an exhaustive search. However, this process is computationally inefficient, especially in the case of multiple sensitive parameters. One of the state-of-the-art methods for optimising TFDs is gradient descent [123] and it is selected here as one of the fast and simplest methods. The main equation for optimising TFDs using gradient descent can be expressed as

$$\theta_{i+1} = \theta_i - \mu \frac{\partial J[\rho(n,k;\theta_i)]}{\partial \theta_i} \tag{3.12}$$

where θ is the set of controlling parameter(s) (e.g. for EMBD $\theta \in \alpha, \beta$) and μ is the step size selected by the user. The step size should not be too small or too large, to avoid slow convergence and divergence respectively. However, as stated before, the calculation of the differential part of Equation (3.12) is a complex task; it can be approximated based on $J[\rho(n,k;\theta_i)]$ calculated with θ_i and its previous value θ_{i-1} and repeated simultaneously to converge to the solution.

$$\theta_{i+1} = \theta_i - \mu \frac{J[\rho(n,k;\theta_i)] - J[\rho(n,k;\theta_{i-1})]}{\theta_i - \theta_{i-1}} \tag{3.13}$$

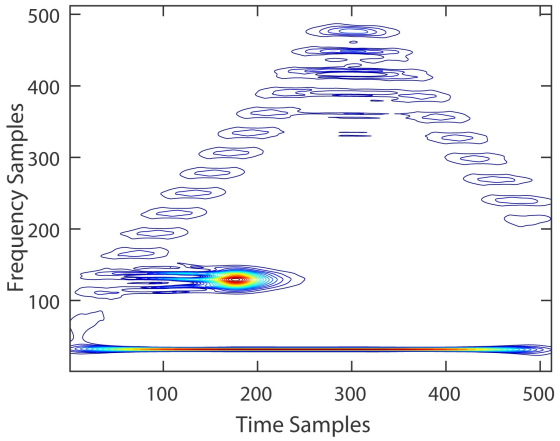
This method can be applied to optimise any QTFDs, e.g. EMBD. The EMBD, $\rho_{EMBD}(n,k;\alpha,\beta)$, has two parameters, α and β, and the optimisation of α and β using this method can be expressed as

$$\alpha_{i+1} = \alpha_i - \mu \frac{\mathbb{J}\left[\rho_{z_{EMBD}}(n,k;\alpha_i,\beta_i)\right] - \mathbb{J}\left[\rho_{z_{EMBD}}(n,k;\alpha_{i-1},\beta_i)\right]}{\alpha_i - \alpha_{i-1}} \tag{3.14}$$

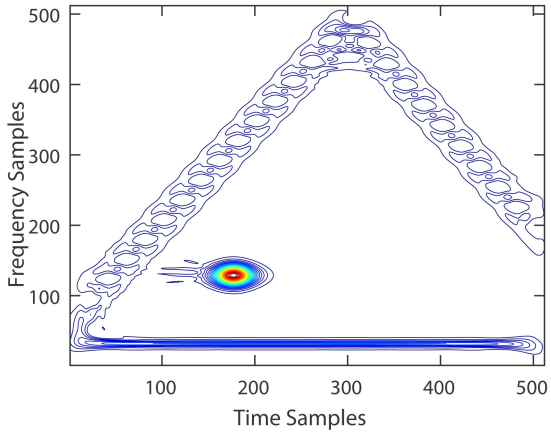
$$\beta_{i+1} = \beta_i - \mu \frac{\mathbb{J}\left[\rho_{z_{EMBD}}(n,k;\alpha_i,\beta_i)\right] - \mathbb{J}\left[\rho_{z_{EMBD}}(n,k;\alpha_i,\beta_{i-1})\right]}{\beta_i - \beta_{i-1}} \tag{3.15}$$

Equations (3.14) and (3.15) are iterated simultaneously to converge to the solution. The main method with the application to optimise EMBD has been discussed here. Other QTFDs can be optimised in the same way.

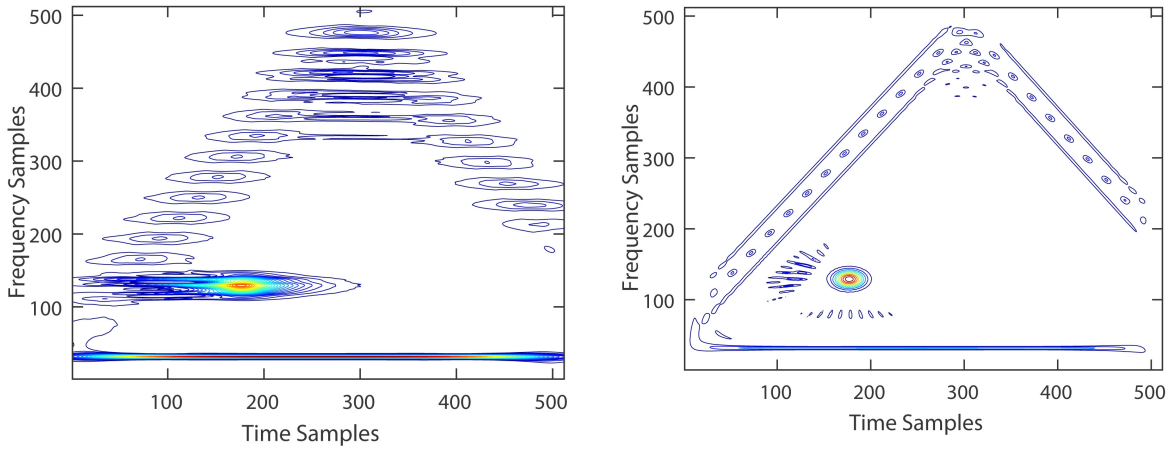
Figure 3.5 shows the optimised representation of the EMBD and the effect of choosing initial parameters and step size. This happens due to different choice of initial parameters and step size; the algorithm converges to different local minima. Another well-known optimisation algorithm is the Nelder–Mead Algorithm [128] which can also be used to find the minimum or maximum of an objective function in a multidimensional space. This algorithm was also applied to optimising TFDs, as the complex derivative is not to be known, unlike the gradient descent method [128]. However, like the gradient descent method, users need to supply initial parameters. Figure 3.5(c–d) illustrates the optimised EMBD representation using this classical optimisation algorithm.



(a)



(b)



(c) (d)

Figure 3.5: EMBD parameter optimisation using gradient descent: (a) The final optimised parameters calculated by this method are $\alpha = 0.02, \beta = 0.60$ [ECM index 90.5×10^3] when the initial parameters are $\alpha_0 = 0.01, \beta_0 = 0.6$; $\alpha_1 = 0.1, \beta_1 = 0.85$ and $\mu = 1/50$ and (b) $\alpha = 0.16, \beta = 0.39$ [ECM index 66.10×10^3] when $\alpha_0 = 0.16, \beta_0 = 0.4$; $\alpha_1 = 0.25, \beta_1 = 0.5$ and $\mu = 1/100$. The final optimised EMBD parameters calculated by using Nelder–Mead algorithm: (c) $\alpha = 0.01, \beta = 0.5$ [ECM index 112.1×10^3] when initial parameters are $\alpha_0 = 0.5, \beta_0 = 0.9$ and (d) $\alpha = 0.06, \beta = 0.12$ [ECM index 101.7×10^3] when initial parameters are $\alpha_0 = 0.01, \beta_0 = 0.01$. Optimal ECM index are shown in brackets [.].

3.2.8 Discussion of method-1

A fully automatic optimisation of QTFDs using the HGA method is useful for characterising non-stationary signals without *a priori* signal information. The main improvement of HGA is to improve the accuracy of getting precise, accurate, signal-dependent TFD kernel parameters with significantly reduced search space with respect to a direct search. In this way, HGA is able to improve the TFD representation of non-stationary signals in a reasonable time and makes the process fully automated. In addition, this study provides a general framework for optimising the parameters of any QTFD. This method can be used to optimise the parameters of any other TFDs. Reference [130] uses evolutionary programming to optimise only Multiform Tiltable Exponential TFD (MTED) based on the distribution norms. However, as mentioned in Section 3.2.2, this method of optimising TFD fails to perform in the expected way when the interferences appear, especially in the case of multicomponent signals comprising high and low amplitude components [123].

Another issue is the selection of HGA setup parameters and functions shown in Table 3.2 as it plays a vital role on the optimisation. This selection was based on several experiments performed on different signals and by observing the convergence curve defined as 'objective function $J(\rho_z[n,k]$ versus number of generation plot'. For example, Figure 3.6 shows the convergence curve when the CKD was applied to the simulated signal represented in Figure 3.2. It can be seen that, after 20 'generations', the solution $(J(\rho_z[n,k]))$ is almost saturated and, therefore, $G_L = 30$ was chosen with confidence for the CKD. In addition, the Nelder-Mead algorithm was applied to further refine these

parameters and, finally, the optimal parameters were chosen. Similar experiments have been done for other TFDs as well.

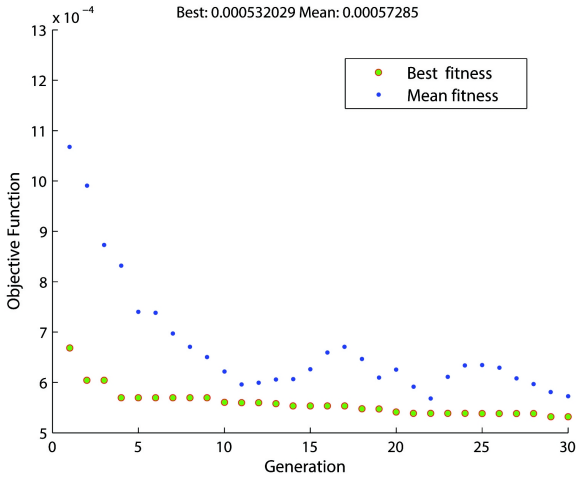


Figure 3.6: Convergence curve for CKD parameter optimisation.

To summarise:

- The Gradient descent method is faster than HGA, but it is highly dependent on the choice of initial value of the kernel parameters and step size, μ. Traditionally, users need to supply the initial kernel parameters and step size. This requires specialist knowledge of TFDs and the signal under analysis; users often use trial and error (initial guess) by supplying different initial parameters and step size. The convergence of the HGA method is independent of the initial parameters and is suitable for fully automated optimisation of QTFDs for most non-stationary signals.
- The classical optimisation Nelder–Mead algorithm is derivative-free and also converges faster than HGA. However, this method is also dependent on the correct choice of the initial kernel parameters and it suffers the same problems as the gradient descent.

Due to the advantages of the HGA mentioned above, various QTFDs have been optimised using this method. Note that the performance of different QTFDs is different depending on the nature of non-stationary signals and the shape of the kernel. This study proposes a method for optimising various QTFDs and users need only to choose the optimal TFD for the signal they are going to analyse. One can also generate multiple 'optimised' TFDs for their signal and select the most optimal TFD, either by inspection or by selecting the TFD that offers the minimum cost, $J(\rho_z[n,k])$. This fact indicates that, though this study automatically optimises the TFDs in terms of balancing concentration, resolution, cross-terms and computation, it still requires specialist knowledge in terms of correct interpretation of the results. That can be addressed in future work by combining AI (Artificial

Intelligence) techniques with TFD representation and implementation methods. Further improvement of the proposed method can also be achieved by applying post-processing steps which are now under consideration ([33] pp. 418–23).

3.3 Method-2: Local optimisation method

A global optimisation method was presented in Section 3.2. This method selects single global parameters to optimise the TFD. However, when a signal is composed of closely spaced chirp signals, or of a mixture of short and long duration components which overlap within the (t, f) domain, the global optimisation method cannot significantly improve the (t, f) resolution or concentration. This justifies the need for a local optimisation of TFD. To overcome this issue, we propose an improved (t, f) representation, named Locally Optimal Spectrogram (LOS), based on the fractional Fourier transform (FrFT) to locally enhance the resolution. This method is well-suited for the analysis of multicomponent, non-stationary signals when we do not have *a priori* signal information, a common situation when one deals with real-life signals, e.g. electroencephalograph (EEG) signals. The results demonstrate the effectiveness of the LOS using different simulated signals and a real-life application using clinical EEG signals. The key topics covered by this section are as follows:

- A simple and efficient optimisation procedure is proposed to enhance the resolution of the spectrogram for non-stationary and multicomponent signals by taking the window length and chirp rate into account. This process ensures the compact representation of the local signal behaviour both in time and frequency (Section 3.3.2).
- Different simulated multicomponent signals of varied amplitude and wide-ranging (t, f) characteristics were used to demonstrate the efficiency of the proposed method. The LOS ws compared with other state-of-the-art fixed and adaptive time-frequency methods in Section 3.3.4.
- Various time-frequency measure indexes were used to ensure (1) a thorough and rigorous quantitative evaluation, and (2) a fair comparison of the LOS with other state-of-the-art methods in Section 3.3.4.2.
- The robustness of the LOS was evaluated for instantaneous frequency (IF) estimation purposes. In Section 3.3.5, a simulation was run 100 times under different SNR conditions (Monte-Carlo approach) comparing the effectiveness and robustness of the LOS with other TFDs.
- Section 3.3.6 discusses the relationship between LOS and other TFDs; this relationship generalises the proposed approach.

Note that we have focused only on the analysis of QTFDs due to their simple interpretation, high resolution and widespread use. The computer programs used in this study are described in appendix 3.A. Appendix 3.B provides a pseudocode for generating the fractional S-Method.

3.3.1 Proposed method – Locally Optimal Spectrogram

The proposed LOS is derived from fractional Fourier transform and, therefore, it is important to briefly describe the fractional Fourier transform prior to discussing the proposed method.

3.3.1.1 Fractional Fourier Transform (FrFT)

The FrFT is the generalisation of the classical Fourier transform (FT). It can be regarded as a rotation by an arbitrary angle α in the (t, f) plane [131, 132]. The classical Fourier transform (FT) corresponds to a rotation over an angle $\alpha = \pi/2$ in the (t, f) plane. The FrFT is defined as [133]:

$$X_\alpha(u) = \int_{-\infty}^{\infty} x(t) K_\alpha(u, t)\, dt \tag{3.16}$$

where $\alpha = p(\pi/2)\,; p \in \mathbb{R}$ and the kernel K_α is defined by: $K_\alpha(u, t) = \sqrt{\frac{1 - j\cot\alpha}{2\pi}}\, e^{j\left(u^2 \cot\alpha/2\right)} e^{j\left(t^2 \cot\alpha/2\right)} e^{-jut/\sin\alpha}, \alpha \neq p\pi$ (3.17)

Special cases of the FrFT are $X_0(u) = x(u), X_\pi(u) = x(-u)$ and $X_{\pi/2}(u)$ corresponding to the classical FT. The orthogonal pair (u, v) characterises a new physical quantity in the fractional Fourier domain and is related to (t, f) as [134]:

$$\begin{pmatrix} t \\ f \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \tag{3.18}$$

Therefore, the (u, v) plane is only the rotation of the (t, f) plane in the fractional domain by an angle α.

3.3.1.2 Rationale for using the FrFT

Figure 3.7 interprets the concept of FrFT in the (t, f) plane. By applying the classical FT, $\mathcal{F}(x(t))$, a time domain signal $x(t)$ is changed to its frequency domain counterpart, $X(f)$, which rotates the signal over an angle $\pi/2$ counter-clockwise. By again applying FT i.e. $\mathcal{F}^2(x(t)) = x(-t)$ rotates $2\pi/2$ angle and similarly $\mathcal{F}^3(x(t)) = X(-f)$ rotates $3\pi/2$ and $\mathcal{F}^4(x(t)) = x(t)$ rotates $4\pi/2$ [133].

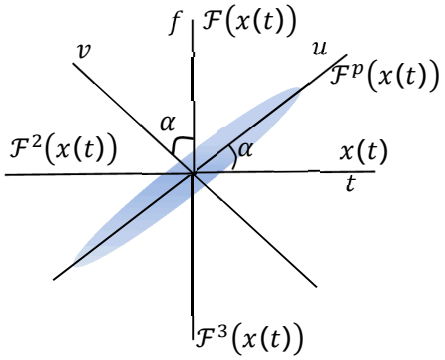


Figure 3.7: Illustration of classical FT and fractional FT and the representation of a signal whose principal axis corresponds to fractional time-frequency axis. α is the angle ($\alpha = p(\pi/2)$) and p is the transform order ranging from 0 to 4.

The FrFT $\mathcal{F}^p(x(t))$ provides a generalisation of the classical FT and offers improved flexibility when designing high resolution time-frequency signatures, as the signal chirp rate can be adapted by this approach. In the case of the signal whose principal axis does not correspond to the time or the frequency plane, as in Figure 3.7, the FrFT applies an affine transformation in the phase plane leading to an optimum signal representation [134]. This can be done by properly adjusting the FrFT transform order (angle). This justifies the analysis of a signal in the fractional Fourier domain.

3.3.1.3 Short-Time Fractional Fourier Transform (STFrFT)

By generalising the STFT in the same manner as the FT, the short-time fractional Fourier transform (STFrFT) can be defined as ([135], pp. 135–36):

$$\mathrm{STFrFT}_\alpha(u,v) = \int_{-\infty}^{\infty} X_\alpha(u+\tau)w(\tau)e^{-j2\pi v\tau}\,d\tau \tag{3.19}$$

$$\mathrm{STFrFT}_\alpha(u,v) = \int_{-\infty}^{\infty} x(t+\tau)w(\tau)K_\alpha(u,\tau)\,d\tau \tag{3.20}$$

These formulations indicate that the lag truncation can be applied prior to or after signal rotation with the same results ([135], pp. 135–36). The fractional spectrogram (FrSpec) is calculated by squaring the magnitude of STFrFT, i.e.

$$\mathrm{FrSpec}(u,v) = |\mathrm{STFrFT}_\alpha(u,v)|^2 \tag{3.21}.$$

Equations (3.19)(3.21) indicate that the performance of the STFrFT and FrSpec is determined by the rotation angle α, the shape and the length of the analysis window. The transform order determines the rotation angle by $\alpha = p\,(\pi/2)$. Regarding the window shape, the confined Gaussian window (CGW) can be selected as it has been shown that it gives a better performance than traditional windows (e.g. Gaussian window, hamming window, hanning window) under minimal root mean square time-bandwidth product [136]. However, it requires tedious minimisation procedures and hence an approximation of the CGW called 'Approximate Confined Gaussian' window (ACGW) [136] is used in this study.

The requirements for producing high resolution $FrSpec$ are:

(a) the optimisation of the transform order (angle) and window parameters, and

(b) all components of a multicomponent and nonstationary signal must be clearly resolved without *a priori* signal information.

This study proposes an improved local optimisation technique designed to meet these requirements.

3.3.2 Proposed optimisation

The proposed algorithm uses a set of transform angles ($\alpha(i)$) and analysis window lengths ($L_w(i)$). This is because, for an arbitrary multicomponent and nonstationary signal, without *a priori* information the selection of appropriate values for (α, L_w) is not trivial [137]. Therefore, in order to find the most compact representation for all signal components, an estimate of α_{opt} and $L_{w_{opt}}$ is required. To find this estimate, a search algorithm spanning the space of α, L_w chooses the optimal window length and transform angle for the input signal. Then $STFrFT_{\alpha_{opt}, L_{w_{opt}}}(u, v)$ is chosen by maximising or minimising a specific objective function in a small (t, f) region, such that only one component lies in the region. In this way, the most compact (t, f) representation of all signal components can be achieved, resulting in a high resolution (t, f) signature. The proposed optimisation requires two major elements that must be selected prior to calculation:

(1) settings of window length(L_w), and fractional angle(α), and

(2) proper objective function or optimisation criterion.

The following strategy is used to meet the requirements.

3.3.2.1 Settings of window length (L_w), and fractional angle (α)

The window length plays a key role in the LOS. Different studies used $L_w = L_s/4$, where L_s is the signal length, as an analysis window for spectrogram ([33], pp.729–30). In this study, we define a set of L_w using the following equation:

$$L_w(i) = \lfloor (i \times 10) + (L_s/40) + 1 \rfloor; i = 0,1, \dots, (D-1), \quad (3.22)$$

where $D \in \mathbb{N}$ is a positive integer and $\lfloor . \rfloor$ is a $floor$ function.

It can be seen from Equation (3.22) that, $L_w(i)$ ranges from short to wide windows that can take care of both short and long duration signal components in the optimisation technique. It is also adaptive, in the sense that, $L_w(i)$ is changed according to signal length L_s.

Like the window length, the fractional angle $\alpha = p(\pi/2), p \in \mathbb{R}$ also plays a key role in the optimisation. In this study, p is derived from $E = 31$ evenly distributed values between 0 and 4 to cover the whole 2π rotation. This corresponds to only $0.209\, rad \approx 12°$ increment. This is a trade-off between precision and computation time. Mathematically,

$$\alpha(j) = p_j(\pi/2), j = 0, 1, \dots, (E-1); \text{ where } E \in \mathbb{N} \text{ is a positive integer.} \quad (3.23)$$

The algorithm also requires a suitable objective function which is discussed in the next section.

3.3.2.2 Objective function of the algorithm

An optimal search method for the fractional order (angle) in the case of a multicomponent signal requires a local measure that considers TFD in a small region [137]. There are several (t, f) measurement criteria, for example, information theoretic measure [138], the ratio of norms-based measure ([33], pp. 401–08), normalised instantaneous resolution measure [124], and energy concentration measure (ECM) [123]. Each method has pros and cons, as discussed in Section 3.2.1. On the other hand, the ECM index is very simple, efficient to implement [123] and overcomes the limitations of the previous methods. Due to its simplicity, this index is useful for selecting the parameters α_{opt} and $L_{w_{opt}}$ of the LOS that give the most compact representation of all signal components. Unlike the norm-based concentration measures, this measure is not restricted by low energy concentration for weak signal components. It is used in this section for optimisation and expressed as

$$J[\mathrm{FrSpec}(u,v)] = \left(\sum_u \sum_v |\mathrm{FrSpec}(u,v)|^{1/2}\right)^2 \tag{3.24}$$

with $\sum_u \sum_v |\mathrm{FrSpec}(u,v))| = 1$ being the normalised unbiased energy constraint. The optimal choice according to this criterion is the one that produces the minimal value of $J[\mathrm{FrSpec}(u,v)]$ [123]. This ECM index permits us to determine the most appropriate fractional order in small regions of the (t,f) plane.

3.3.2.3 Proposed LOS

The proposed optimisation algorithm has the following three main steps: (1) calculation of FrSpecs, (2) calculation of sub-optimal FrSpecs, and (3) calculation of the final LOS.

Step 1: Calculate F number of FrSpecs (for use in step 2). To do this, firstly initialise $D \in \mathbb{N}$ and $E \in \mathbb{N}$ and calculate a set of $(L_w(i))$ and $\alpha(j)$ using Eqns. (3.22) and (3.23) respectively. After that, calculate F number of fractional spectrograms, where $F = D \times E$, by iterating window length $(L_w(i))$ from 0 to $D-1$ and fractional angle $\alpha(j)$ from 0 to $E-1$. The flowchart shown in Figure 3.8 demonstrates the step-by-step process for calculating F number of FrSpecs. Figure 3.9(a) also shows the $(F = \{jk | jk = 0,1,2,\dots,F-1\})$ numbers of FrSpec.

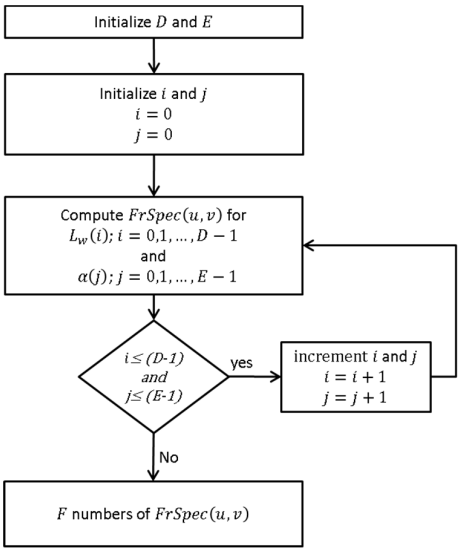
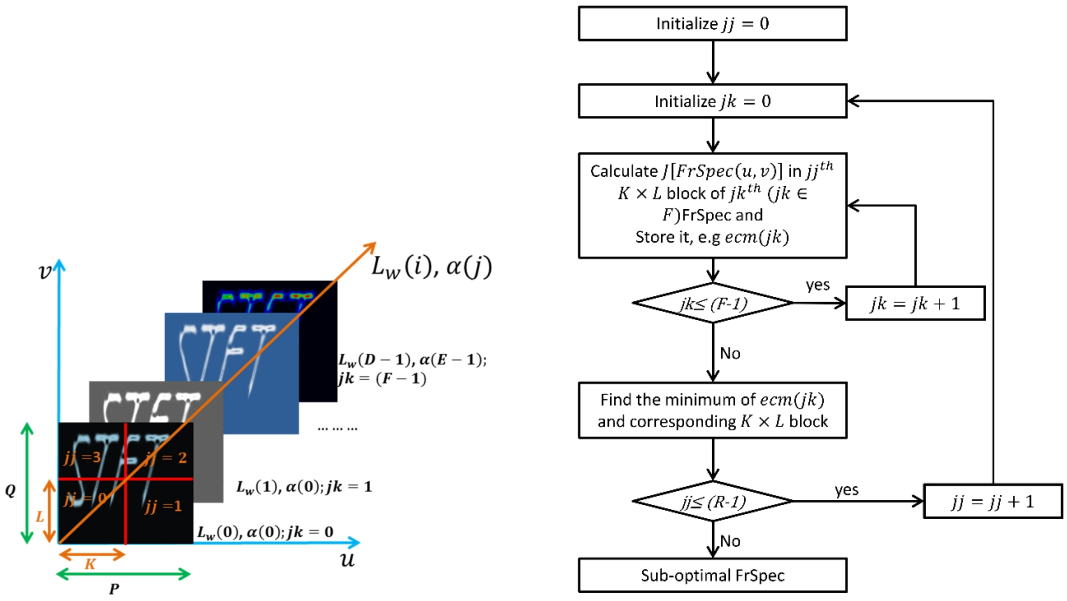


Figure 3.8: Calculation of F number of FrSpecs; i and j represent the index for the analysis window length and transform order respectively.

Step 2: A near-optimal FrSpec is determined from the calculated FrSpecs. To do that, each FrSpec (size e.g., $P \times Q$) is divided into $R = \{jj | jj = 0,1,\dots,R-1\}$ small blocks of size $K \times L$; i.e. $(P \times Q) = R \times (K \times L)$ and compute the ECM index for each block. Finally, a near-optimal

FrSpec is chosen that minimises the $J[FrSpec(u,v)]$ in each $K \times L$ block from the F number of FrSpecs.

Figure 3.9 shows an example of the step-by-step process for the calculation of a sub-optimal FrSpec. To illustrate, consider $R = 4$. FrSpecs of size $P \times Q$ are divided into 4 small blocks, i.e. $R = 4$; $(jj = 0,1,2,3)$ of size $K \times L$. Calculate the ECM index and then iterate from $jk = 0$ to $F - 1$ to obtain the particular $K \times L$ block that has the minimum $J[FrSpec(u,v)]$ value. This process is repeated for all other $(K \times L)$ blocks to obtain an optimal FrSpec for a particular R which we denote 'sub-optimal FrSpec'. In the same way, several sub-optimal FrSpecs can be generated for different values of R, e.g. $R = 2, 16$ (see Figure 3.10). The higher value of R divides the FrSpec into smaller blocks and vice versa.



(a) (b)

Figure 3.9(a-b): Procedures for the calculation of sub-optimal FrSpec; 'jj' represents the index for each $K \times L$block whereas 'jk' represents the index for each FrSpec.

Step 3: In the experiment, it has been found that the dividing parameter R plays a key role due to the number and orientation of signal components of a multicomponent signal in the TFD plane. Therefore, it is decided to use several $R \in \mathbb{N}$ and repeat step 2. Each particular value of R provides a full sub-optimal FrSpec matrix. For instance, 4 values of R generate 4 (t,f) signatures, i.e. R_0, R_1, R_2, R_3. The final LOS can be generated by dividing each R_i signature into S small blocks and finding the local maximum energy across R_i as in step 2. Note that the size of S is same for all R_i.

To illustrate, consider a particular S block of R_0. The energy of that block in R_0 is compared with the corresponding S block of all other R_i signatures and picks the particular S block that has the maximum energy. This process is repeated for all other S blocks to obtain the final and full LOS signature. Figure 3.10 depicts the process used to obtain the final LOS.

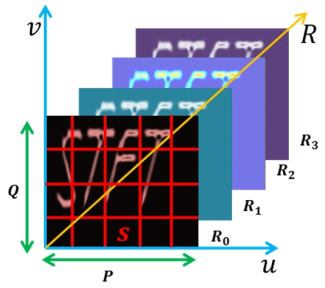


Figure 3.10: Methodology to obtain LOS from R_i (t, f) signature. For instance, 4 values of R generate 4 (t, f) signatures i.e. R_0, R_1, R_2, R_3. Like step 2, find the local maximum energy of a small block S and repeat the process to obtain the whole LOS.

3.3.3 (t, f) measurement criteria

The (t, f) measurement criteria including TFD information theoretic measures, the ratio of norm-based measure, the normalised instantaneous resolution measure and the ECM index are used to evaluate the performance of TFDs quantitatively. Note that the discrete version of the TFD is used in this section for the sake of better understanding and clarity, i.e. $TFD[n, k]$ is the discrete version of continuous $TFD(t, f)$, where n, k represent discrete time and frequency respectively.

3.3.3.1 TFD information theoretic measures

TFDs represent a pseudo-energy density in the (t, f) domain [138]; hence entropy measures can be used to quantify the complexity of a TFD ([33],pp 401–04). Conceptually, a highly concentrated TFD with a lower number of signal components is less complex, and thus has lower entropy, than a signal with a large number of components with cross-terms. The following entropy measures are applied in this study.

3.3.3.2 (t, f) Renyi entropy (TFRE)

The (t, f) Renyi entropy, defined as in Equation (3.25), is an improved entropy method based on the (t, f) Shannon entropy by applying the log operator [139]. A normalisation using a distribution volume is applied so that it detects zero mean cross-terms when $\alpha = 3$ [29, 33].

$$\text{TFNRE} = -\frac{1}{2} log_2 \left(\sum_{n=1}^{N} \sum_{k=1}^{M} \left[\frac{TFD[n,k]}{\sum_{n=1}^{N} \sum_{k=1}^{M} |TFD[n,k]|} \right]^{\alpha} \right); \alpha > 2 \quad (3.25)$$

3.3.3.3 Ratio of norm-based measure

The ratio of norm-based measure can be used to discriminate a low resolution TFD from a high resolution TFD, which is defined as [29, 123]:

$$\mathrm{RN} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{M} TFD^4[n,k]}{(\sum_{n=1}^{N}\sum_{k=1}^{M} TFD^2[n,k])^2} \tag{3.26}$$

3.3.3.4 Energy concentration measure

Another approach to evaluate the performance of a TFD is to measure its region of support. A highly concentrated TFD has a smaller region of support compared to a blurred TFD [140], as defined in Equation(3.24). A modified version of the energy concentration (MEC) can be defined as [140]

$$\mathrm{MEC} = \sum_{n=1}^{N}\left[\sum_{k=1}^{M} |TFD[n,k]|^{1/2}\right]^2 \tag{3.27}$$

3.3.3.5 Hoyer measure

The best TFD should represent the signal components as sparsely as possible in the (t,f) domain, which means that the components can be represented sharply in the (t,f) plane. Beside this conventional TFD measures, the Hoyer measure can be used to measure the sparsity as it satisfies most of the desirable properties of sparsity [141]. As TFD meets the requirements of compressed sensing, i.e., sparsity complexity [142], this study also uses this measure in the context of (t,f) performance evaluation [141].

The Hoyer measure (HM) can be defined in the (t,f) context as [141]

$$\mathrm{HM} = \left(\sqrt{H} - \frac{\sum_{n=1}^{N}\sum_{K=1}^{M}|TFD[n,k]|}{\sqrt{\sum_{n=1}^{N}\sum_{K=1}^{M}|TFD[n,k]|^2}}\right)\left(\sqrt{H}-1\right)^{-1} \tag{3.28}$$

where H is the total size of the TFD, i.e. $H = N \times M$.

In all the described measures except the RN and HM, a smaller value represents a more compact and therefore desirable representation. For the RN and HM, a higher value gives a better representation.

3.3.4 Results of method-2

The LOS was calculated by Equation (3.21) using the ACG window [136]. The ACG window length L_w and fractional angle α were optimised using the proposed techniques. The values $D = 16$ and $E = 31$ were used in this study (see Section 3.3.7 for further explanation). The dividing parameter R was set to $2, 4$ and 16 after several experiments on different signals which represent three (R_0, R_1, R_2)

fractional spectrograms; see Section 3.3.7 for further discussion. This explores the (t, f) signature from higher to lower blocks respectively and corresponds to long and short duration signal components. To obtain the final LOS a flexible value of $S = \left\lfloor \frac{L_s}{16} \right\rfloor$, where $\lfloor . \rfloor$ is the $floor$ function, is adapted to the signal length, L_s.

3.3.4.1 Simulations study

To illustrate the performance of the LOS method, four simulated multicomponent signals of varied amplitude and wide-ranging (t, f) characteristics were used and the signal model is $y(t) = \sum_i A_i \, x_i(t)$ [121]. Since the study is focused on the QTFDs, we include various state-of-the-art (t, f) methods of this class for comparison: spectrogram, S-transform (ST), S-method (SM), Wigner-Ville distribution (WVD) and different smoothed versions of the WVD which have shown good performance ([33], pp. 274–83). The WVD renders better (t, f) resolution but suffers from cross-terms. The smoothed version of WVD has been introduced to reduce these interferences (hence called reduced interference distribution, or RID) – for example, smoothed WVD (using Hamming window), Choi-Williams distribution (CWD), modified B distribution (MBD), extended modified B distribution (EMBD) and compact-kernel distribution (CKD). Table 3.1 provides definitions of the RIDs, and a detailed introduction is available in [33] (pp. 274–83). RIDs have one or more parameter(s) to control the smoothing along the lag and Doppler axes in the ambiguity domain and to reach a compromise between the resolution and cross-term effects. The parameters are tuned so that the best representations are achieved according to the visual inspection of the resulting (t, f) plots by an experimental search as in [29] and measuring the ECM criterion. Therefore, these optimised TFDs give a fair comparison with the proposed LOS. In addition, this study also compares the LOS with an adaptive directional TFD (ADTFD) which is adapted to direction of local components ([33], pp. 299–307). Although this study confined the study to the QTFD framework, wavelet-based (t, f) is also used for a complete comparison. Continuous wavelet transform (CWT) is a linear time-scale transform where scale is a reciprocal of frequency ([33], pp. 78–79). In fact, the scalogram (which is obtained by squaring the magnitude of the CWT) can be considered as a special case of QTFD ([33], pp. 92–98). The Morlet wavelet was used to implement the scalogram analysis.

All (t, f) representations provided in this study are given with the same MATLAB colormap in order to show results in a fair and comparable manner. Contour plots were used to visualise all the signal components and their interferences. The ideal (t, f) is represented in order to compare with different TFDs. The sampling frequency of the simulated signals is 1 Hz and the signal duration is $T = 512$ s.

3.3.4.1.1 Signal type 1: Frequency-overlapping signal terms

The signal comprises (i) two linear frequency-modulated (FM) Gaussians ($x_1(t)$ and $x_2(t)$), (ii) a constant FM Gaussian, $x_3(t)$, and (iii) a narrow-band sinusoid $x_4(t)$. Figure 3.11 shows the corresponding signal model and its (t, f) representations.

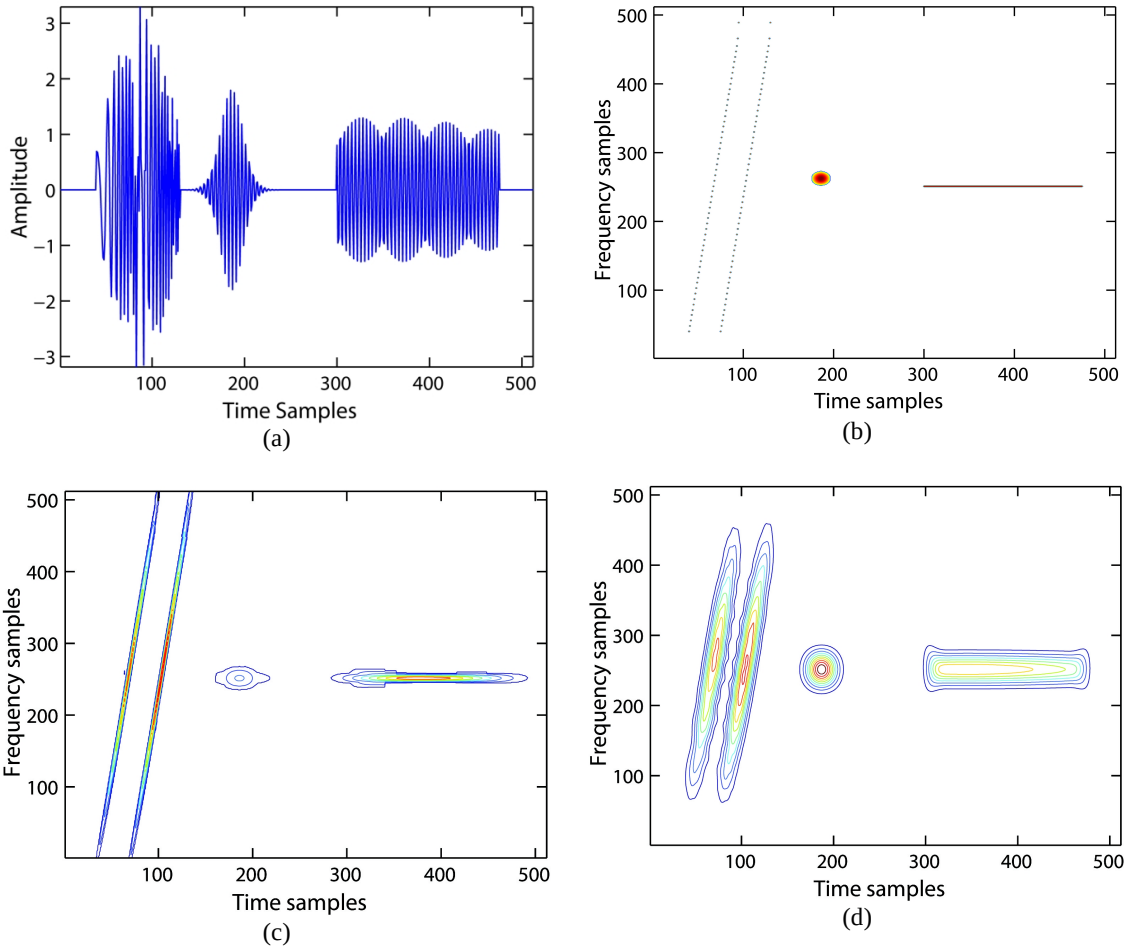
$$x_1(t) = \begin{cases} 2.5e^{-\frac{9(t-103)^2}{6272}+j\frac{2\pi 1000(t-70)^2}{511^2}} & 40 \le t \le 95 \\ 0 & else \end{cases}$$

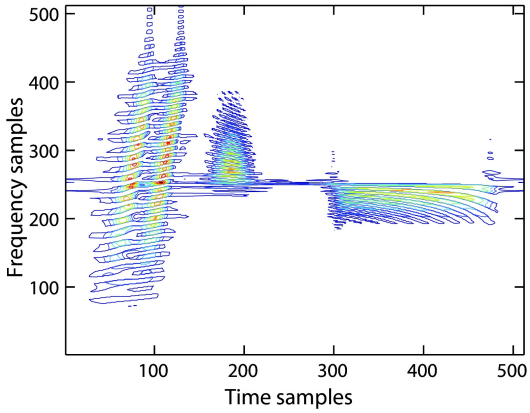
$$x_2(t) = \begin{cases} 2.7e^{-\frac{9(t-103)^2}{6272}+j\frac{2\pi 1000(t-70)^2}{511^2}} & 75 \le t \le 130 \\ 0 & else \end{cases}$$

$$x_3(t) = 1.8e^{-\frac{(t-185)^2}{338}+j\frac{2\pi 125t}{511}} \qquad 0 \le t \le 511,$$

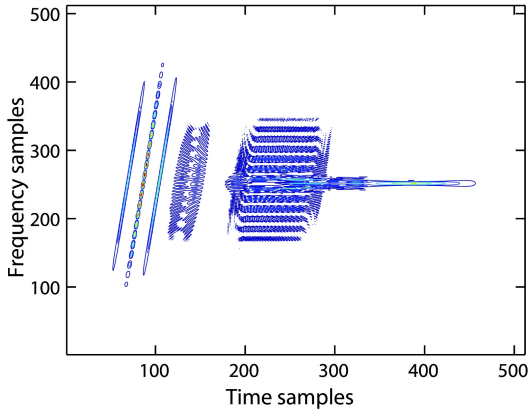
$$x_4(t) = \begin{cases} 1.3e^{-\frac{(t-344)^2}{80000}+j\frac{2\pi 125t}{511}} & 300 \le \le 475 \\ 0 & else \end{cases},$$

$$y(t) = x_1(t) + x_2(t) + x_3(t) + x_4(t) \tag{3.29}$$

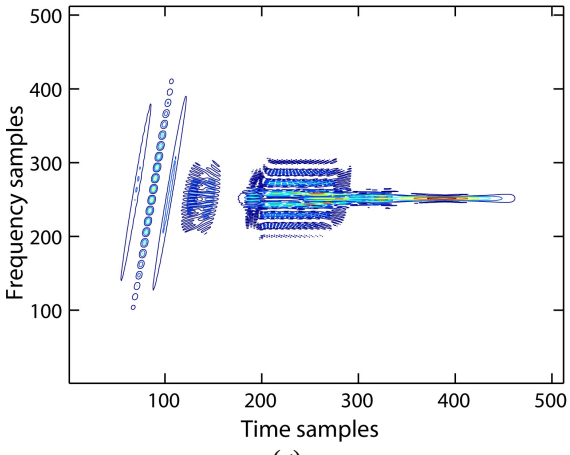



Frequency samples
Time samples
100
200
300
400
500

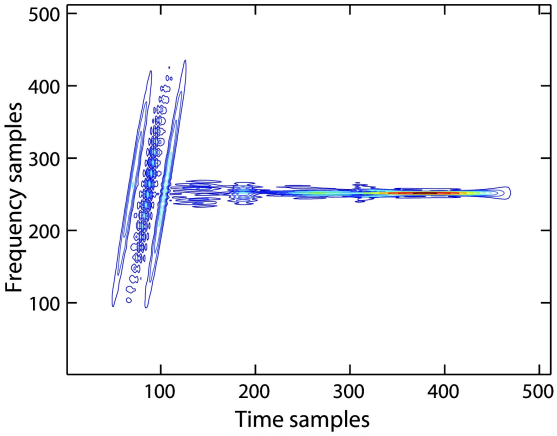
(e)


Frequency samples
Time samples
100
200
300
400
500

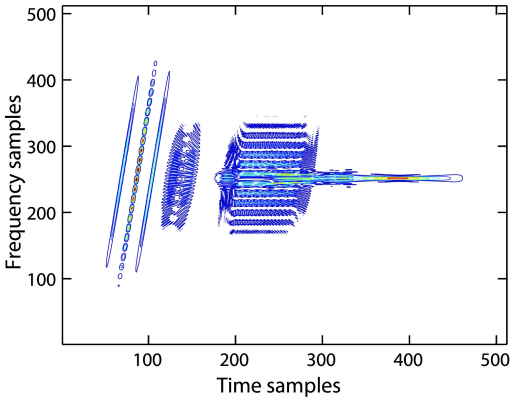
(f)


Frequency samples
Time samples
100
200
300
400
500

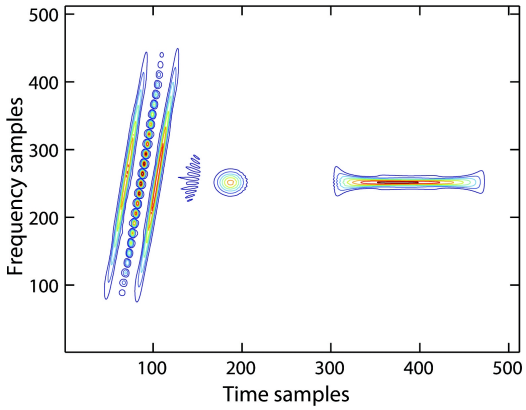
(g)


Frequency samples
Time samples
100
200
300
400
500

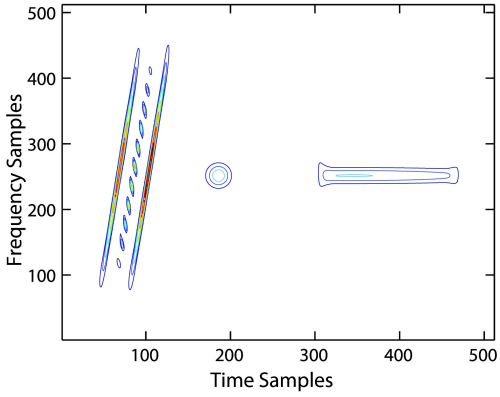
(h)


Frequency samples
Time samples
100
200
300
400
500

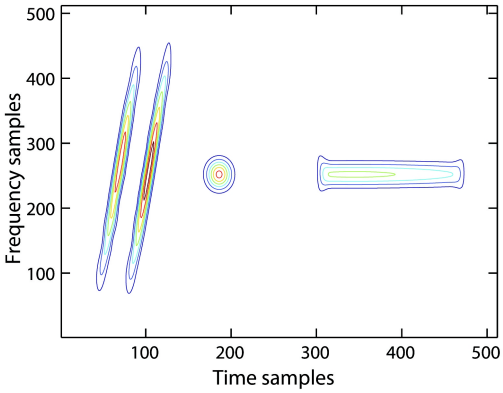
(i)


Frequency samples
Time samples
100
200
300
400
500

(j)


Frequency Samples
Time Samples
100
200
300
400
500

(k)


Frequency samples
Time samples
100
200
300
400
500

(l)

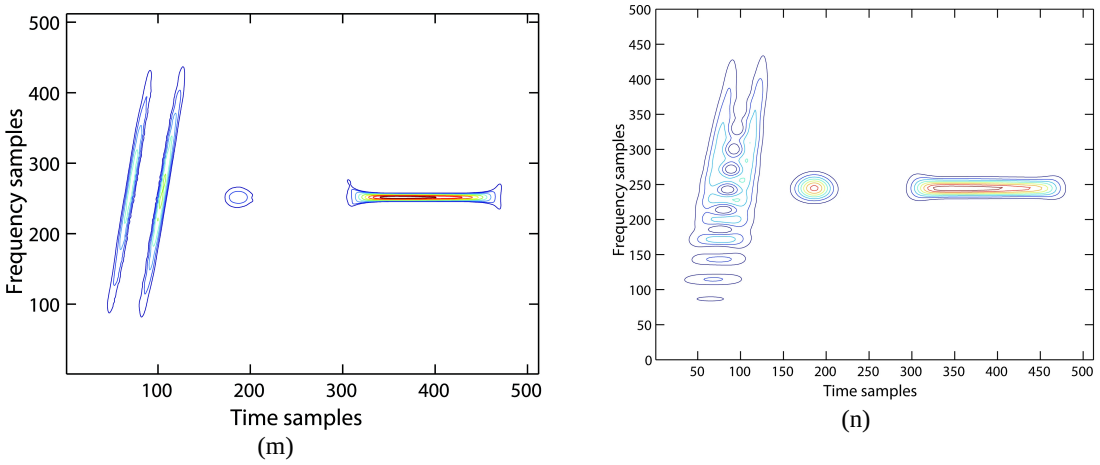


Figure 3.11: (a) signal type 1 and its (t, f) representation: (b) ideal time-frequency representation, (c) LOS, (d) spectrogram (Hamming window length 41 samples), (e) ST, (f) WVD, (g) SWVD (Hamming window length = 11 samples) (h) CW($\sigma = 20$), (i) MBD ($\beta = 0.9$), (j) EMBD ($\alpha = 0.25, \beta = 0.12$), (k) CKD ($c = 8.5, D = 0.08, E = 0.3$), (l) SM (Hamming window length 41 samples and correction terms (L=14)), (m) ADTFD ($a = 3, b = 5$), and (n) scalogram.

The LOS showed good concentration, resolution and was close to the ideal (t, f) representation, whereas the spectrogram exhibits poor resolution and the ST renders poor resolution and cannot separate all the components. Other QTFDs offer lower resolution due to the presence of cross-terms. Conversely, the ADTFD provided cross-term free signatures, but the energy concentration of the LFM components $(x_1(t), x_2(t))$ are lower than the LOS by 5×10^3 (as shown in Figure 3.11 and Table 3.5). In addition, scalogram renders poor resolution as it cannot separate out the close LFM components.

3.3.4.1.2 Signal type 2: Time-overlapping signal terms

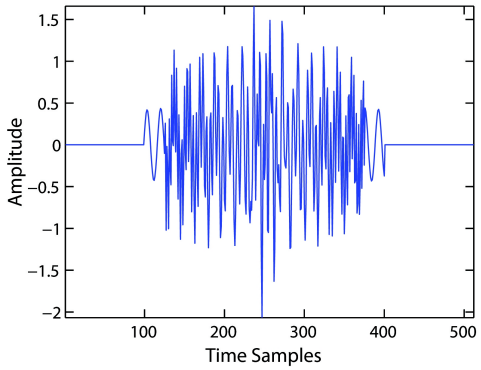
The signal comprises (i) a sinusoid $x_1(t)$ (ii) a quadratic chirp $x_2(t)$, and (iii) a constant frequency-modulated Gaussian $x_3(t)$. The signal and its (t, f) representations are shown in Figure 3.12.

$$x_1(t) = \begin{cases} 0.5e^{-\frac{(t-250)^2}{125000}+j\frac{2\pi 30t}{511}} & 100 \le t \le 400 \\ 0 & else \end{cases}$$

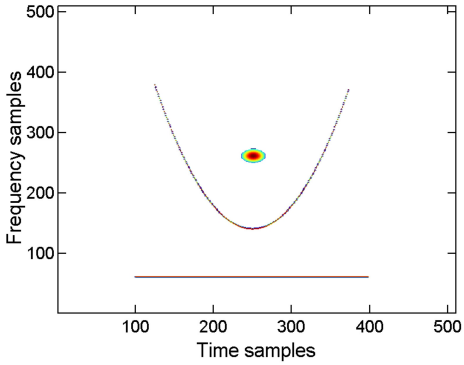
$$x_2(t) = \begin{cases} 0.8e^{-\frac{9(t-250)^2}{781250}+j\frac{\pi 4000(t-250)^3}{3\times 511^3}+j\frac{2\pi 70(t-250)}{511}} & 125 \le t \le 375 \\ 0 & else \end{cases}$$

$$x_3(t) = \begin{cases} e^{-\frac{(-250)^2}{450}+j\frac{2\pi 130t}{511}}, & 200 \le \le 300 \\ 0, & else \end{cases}$$

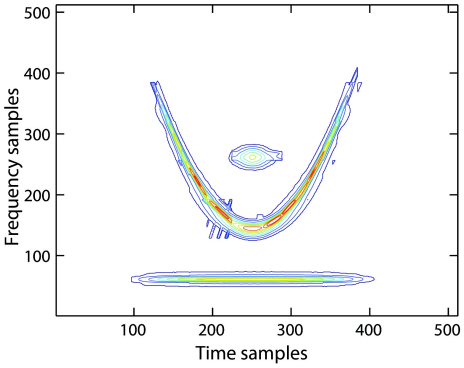
$$y(t) = x_1(t) + x_2(t) + x_3(t) \qquad (3.30)$$


Amplitude
Time Samples

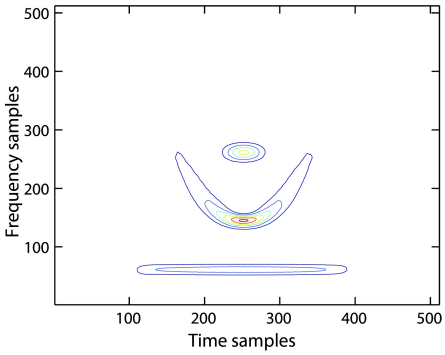
(a)


Frequency samples
Time samples

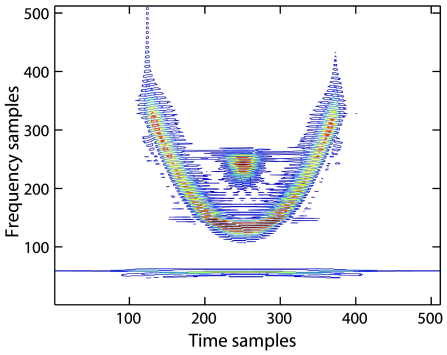
(b)


Frequency samples
Time samples

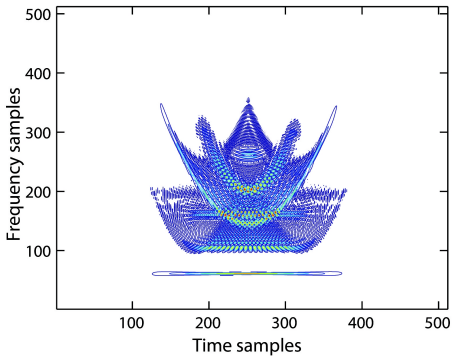
(c)


Frequency samples
Time samples

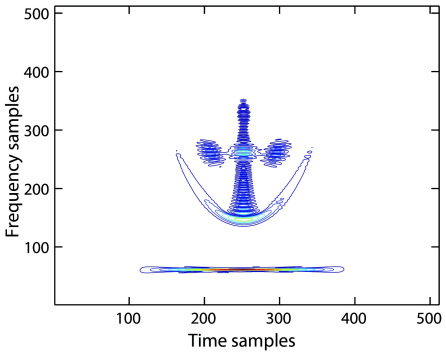
(d)


Frequency samples
Time samples

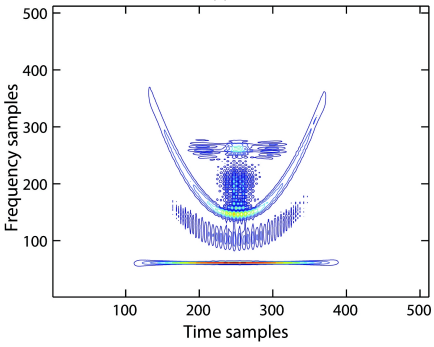
(e)


Frequency samples
Time samples

(f)


Frequency samples
Time samples

(g)


Frequency samples
Time samples

(h)

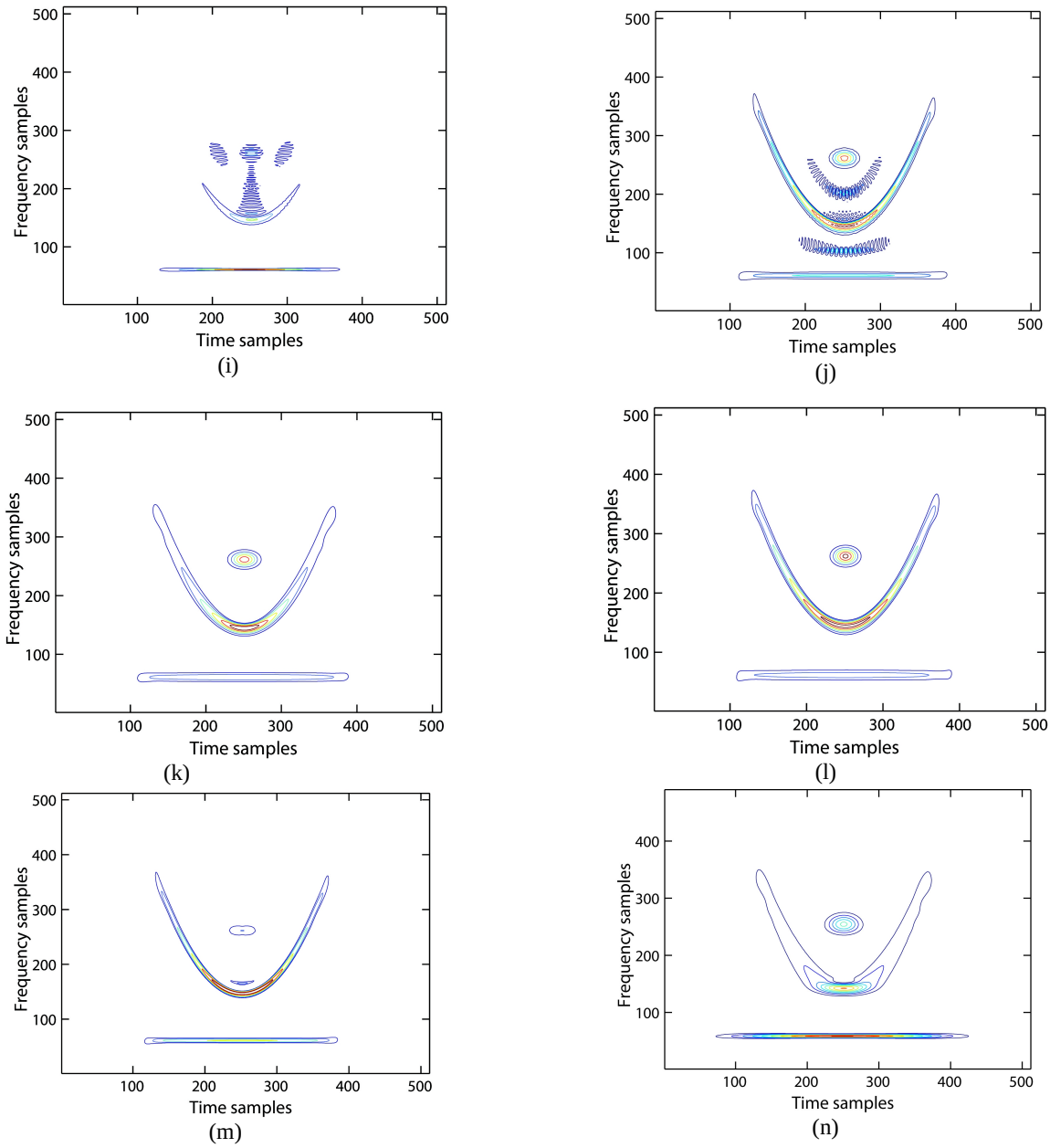


Figure 3.12: (a) signal type 2 and its (t, f) representations: (b) ideal time-frequency representation, (c) LOS, (d) spectrogram (Hamming window length 85 samples), (e) ST, (f) WVD, (g) SWVD (Hamming window length = 35 samples), (h) CW$(\sigma = 20)$,(i) MBD $(\beta = 0.1)$, (j) EMBD $(\alpha = 0.25, \beta = 0.12)$), (k) CKD $(c = 5.9, D = 0.12, E = 0.11)$, (l) SM (Hamming window length 85 samples and correction terms (L = 15)), (m) ADTFD ($a = 3, b = 7$), and (n) scalogram.

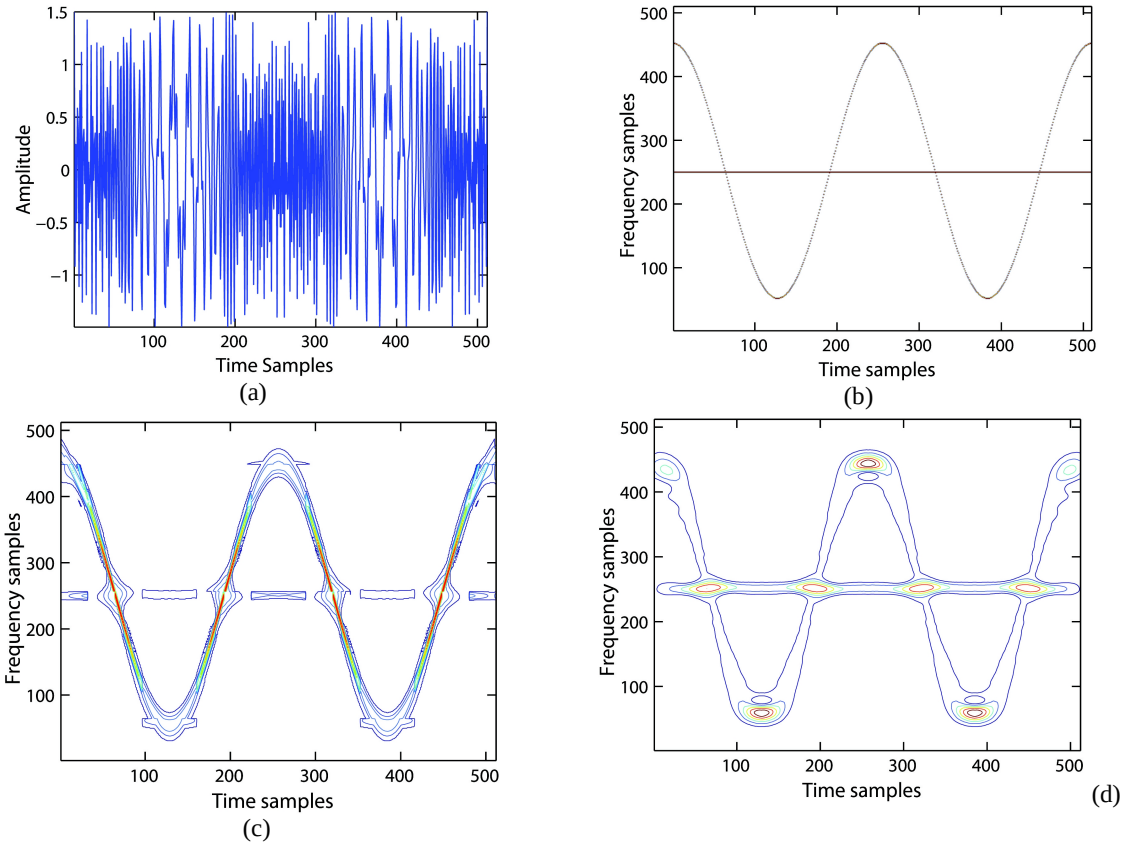
The LOS method exhibited better resolution and energy concentration as it separates all the components and offers a cross-term free signature, whereas the standard spectrogram separated all the components but provides poor resolution. Other quadratic methods provided better concentration but suffer heavily from interferences, except CKD and SM which offered better resolution but lower energy concentration. On the other hand, the ADTFD provided sharp localisation of all the components with significantly lower cross-terms, except for the constant FM Gaussian $(x_3(t))$

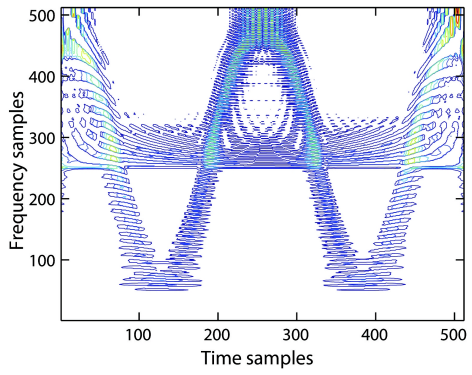
component where the energy concentration was lower than other TFDs. Furthermore, the scalogram concentration of this signature is lower compared to most of the QTFDs.

3.3.4.1.3 Signal type 3: Intersecting time-overlapping signal terms

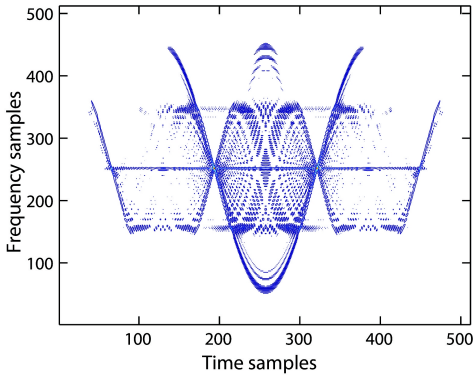
The signal comprises a nonlinear (sinusoidal) frequency-modulated component intersecting with a sinusoid at multiple points in the (t, f) plane. The corresponding signal and its (t, f) representations are shown in Figure 3.13.

$$y(t) = e^{j16\pi \sin\left(\frac{4\pi t}{511}\right) + j\frac{2\pi 125 t}{511}} + 0.5e^{j\frac{250\pi t}{511}} \tag{3.31}$$

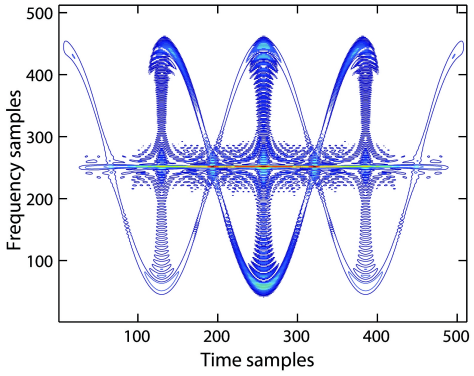



Frequency samples
Time samples
100
200
300
400
500

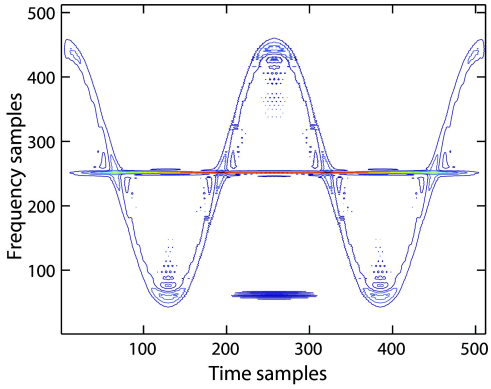
(e)


Frequency samples
Time samples
100
200
300
400
500

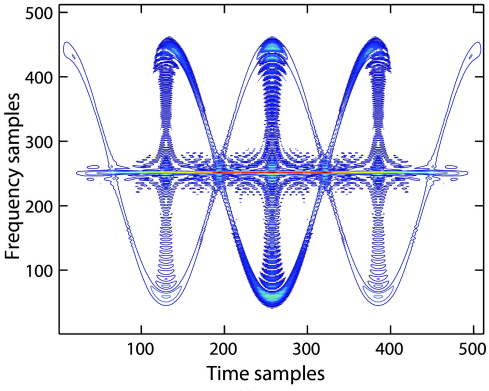
(f)


Frequency samples
Time samples
100
200
300
400
500

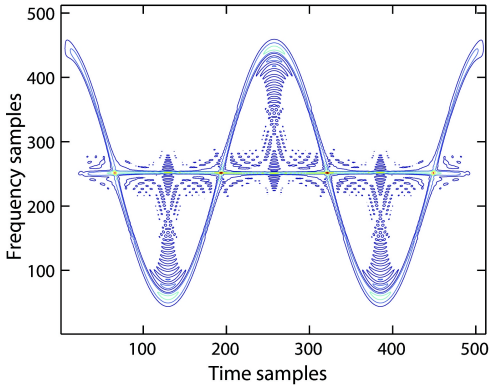
(g)


Frequency samples
Time samples
100
200
300
400
500

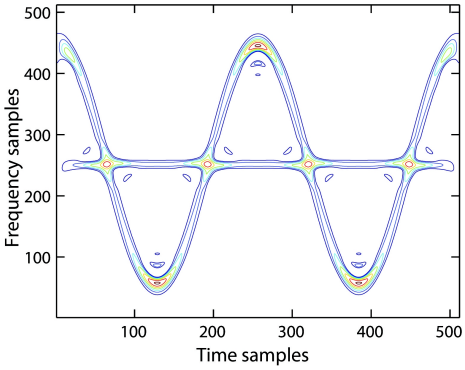
(h)


Frequency samples
Time samples
100
200
300
400
500

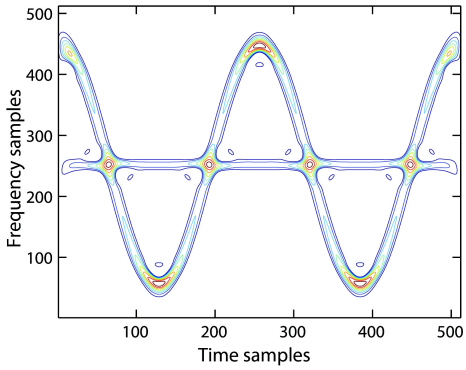
(i)


Frequency samples
Time samples
100
200
300
400
500

(j)


Frequency samples
Time samples
100
200
300
400
500

(k)


Frequency samples
Time samples
100
200
300
400
500

(l)

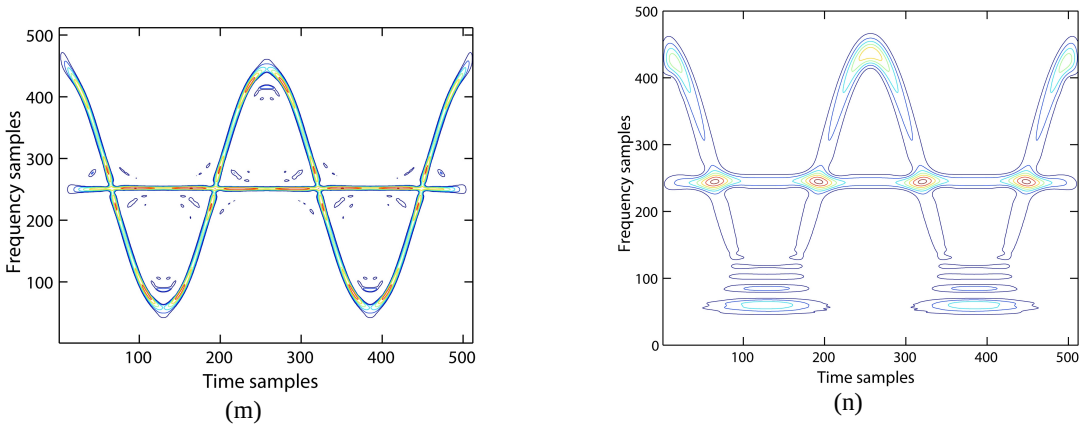


Figure 3.13: (a) signal type 3 and its (t, f) representations: (b) ideal time-frequency representation, (c) LOS, (d) spectrogram (Hamming window length 85 samples),(e) ST, (f) WVD, (g) SWVD (Hamming window length = 15 samples), (h) CW($\sigma = 10$), (i)MBD ($\beta = 0.15$), (j) EMBD ($\alpha = 0.3, \beta = 0.4$)), (k) CKD ($c = 2.5, D = 0., E = 0.1$), (l) SM (Hamming window length 85 samples and correction terms (L = 11)), (m) ADTFD ($a = 3, b = 11$), and (n) scalogram.

The LOS method offered better (t, f) representation, whilst standard spectrogram provided blurred representation. All other QTFDs provided superior (t, f) concentration but suffer from severe interferences. On the other hand, the CKD, SM and ADTFD provided a sharp localisation of each signal component while maintaining lower interferences. Additionally, the resolution and concentration of the scalogram of this signal is lower than most of the QTFDs and generates spectral leakage for the sinusoidal signal component.

3.3.4.1.4 Signal type 4: Signal terms with different orientations

This signal consists of four signal components with different orientations consisting of linear FM signals $(x_1(t), x_2(t))$, constant FM $(x_3(t))$ and frequency-modulated Gaussian $(x_4(t))$; see equations (3.32)–(3.34).This type of signal is difficult to represent without severe compromise between (t, f) concentration and resolution in the (t, f) domain. The signal components are formed as:

$$C_i(t^i) = \sum_{i=1}^{5} e^{j2\pi\left[f_0^i(t^i - t_0) + \frac{f_1^i - f_0^i}{2T^i}(t^i - t_0)\right]} \tag{3.32}$$

where (f_0^i, f_1^i) denotes the initial and final frequency of the i^{th} component respectively. The chosen parameters are:

$$(f_0^i, f_1^i) = (0.09,0.5), (0.04,0.45), (0.5,0.2), (0.45,0.15), (0.03,0.03)$$

for $i = 1,2, \ldots, 5$; $t_0 = 50$ and when $i = 1,2$; t^i is $0 \leq t^i < 300$ and $T^i = 300$; when $i = 3,4$; $300 \leq t^i < 512$, $T^i = 212$ and, when $i = 5$; $0 \leq t^i \leq 511$, $T^i = 512$. In the implementation,

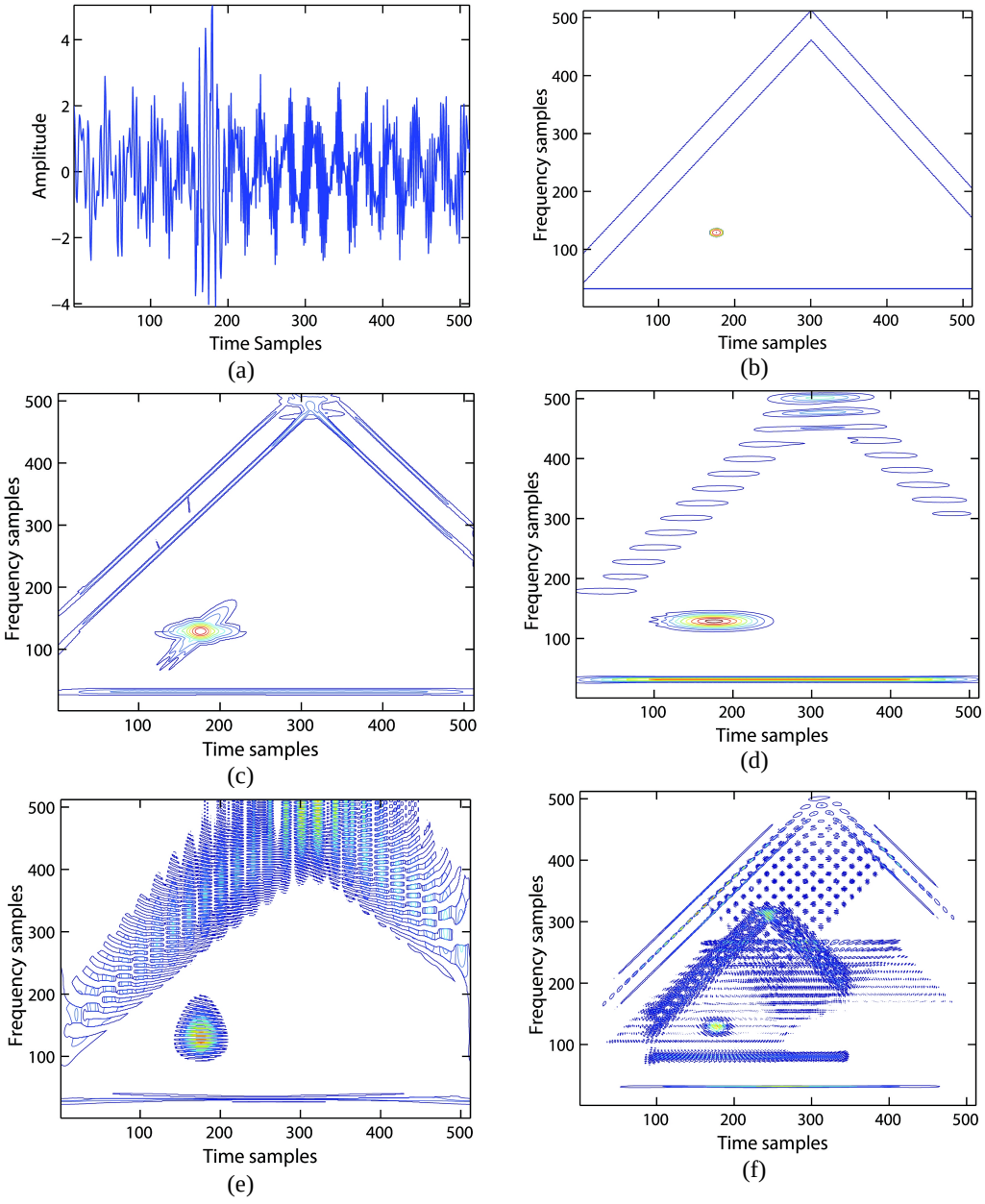
$C_1(t)$ and $C_3(t)$ as well as $C_2(t)$ and $C_4(t)$ sub-components are concatenated to maintain the signal length of 512 expressed in Equation (3.33). Another component, $x_4(t)$, is a frequency-modulated Gaussian expressed in (3.34) and the final multicomponent signal model is expressed in (3.35).

$$x_1(t) = [C_1(t)\ C_3(t)];\ x_2 = [C_2(t)\ C_4(t)];x_3(t) = C_5(t) \tag{3.33}$$

$$x_4(t) = 4e^{-0.05(t-175)^2}\ sin(2\pi 0.125t)\ ;\quad 0 \le t\ \le 511 \tag{3.34}$$

$$y(t) = x_1(t) + x_2(t) + x_3(t) + x_4(t\ ) \tag{3.35}$$

The corresponding signal and its (t, f) representations are given in Figure 3.14.



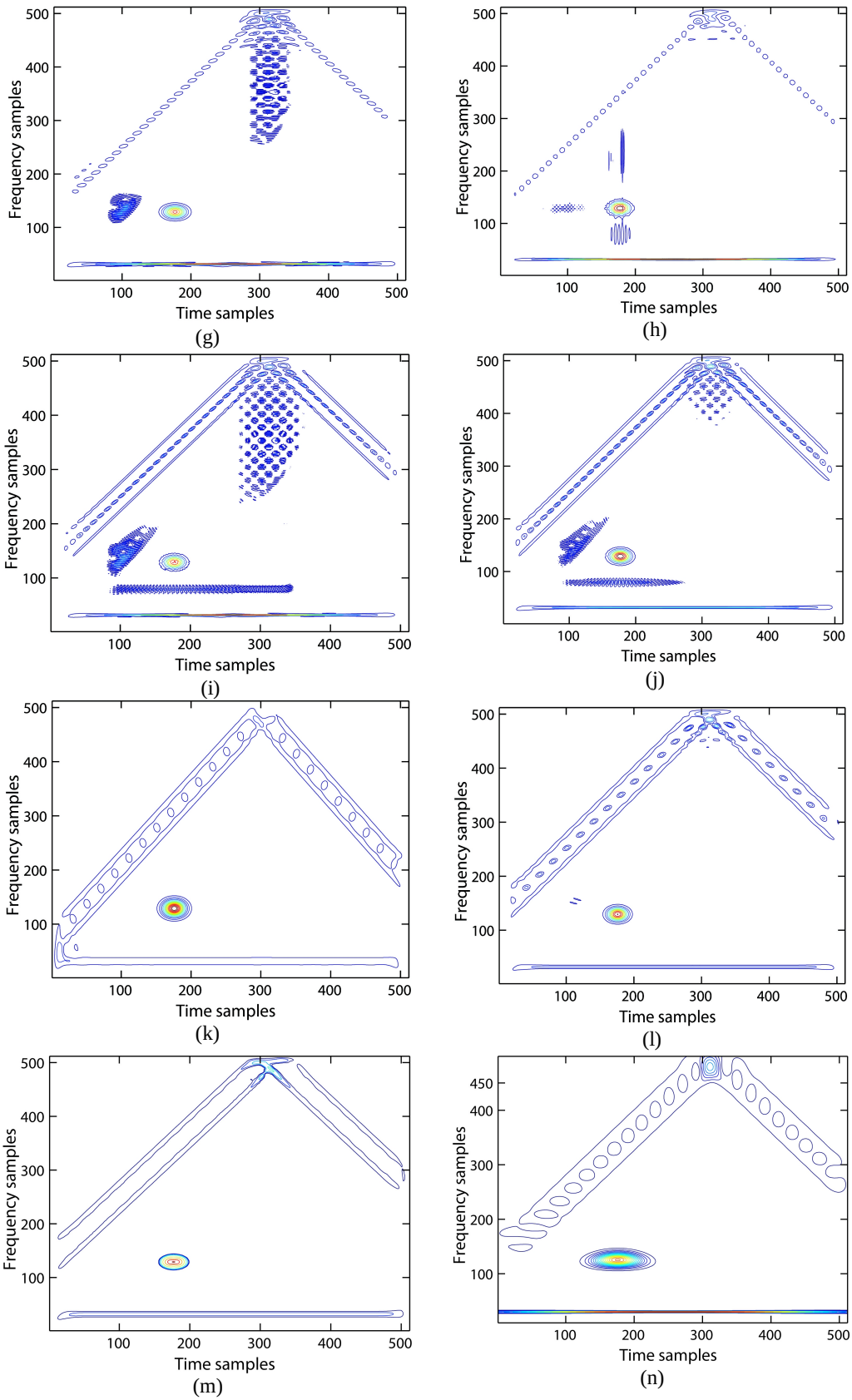


Figure 3.14: (a) signal type 4 and its (t, f) representations, (b) ideal time-frequency representation, (c) LOS, (d) spectrogram (Hamming window length 225 samples), (e) ST, (f) WVD, (g) SWVD (Hamming window length= 25 samples), (h) CW($\sigma = 15$), (i) MBD ($\beta = 0.25$), (j) EMBD ($\alpha = 0.25, \beta = 0.6$)), (k) CKD ($c = 9.5, D =$

$0.12, E = 0.16$), (l) SM (Hamming window length 225 samples and correction terms (L = 16)), (m) ADTFD ($a = 3, b = 5$), and (n) scalogram.

The LOS method handled this signal sufficiently well as it is locally optimised by the optimal window length and transform order (angle), whereas the standard spectrogram failed to characterise all the signal components. Other QTFD methods offered sharp localisation but with the expense of severe interference. CKD and SM provided less interference whilst ADTFD generated few interference (t, f) terms, but the energy concentration of the signal components were lower than the LOS. Additionally, scalogram of this signal cannot separate the $x_1(t)$ and $x_2(t)$ signal components.

From the above (t, f) representations, it can be noted that none of the methods can provide excellent resolution, sharp localisation and cross-term free (t, f) signature for all signal components. However, the LOS and the ADTFD methods offer a good compromise, given that they are adaptive. However, although ADTFD is adaptive it still requires selecting optimal ADTFD parameters manually. Conversely, LOS is fully automatic, as it selects the optimal parameters to provide high-resolution (t, f) signature.

3.3.4.2 Quantitative assessment of LOS

The analysis of Table 3.5 is based on the type 1 signal used in Section 3.3.4.1.1. For the ECM and the MEC, the LOS performs better than all other methods. The reason is that these measures are used as the objective function of the LOS. In other cases, for example in the case of TFNRE and RN, the ADTFD is the best performing among the different TFD methods. However, as stated above, these measures are not suitable for the signal composed of lower to higher amplitudes [123]. Moreover, the TFNRE and RN measures of the LOS are very close to the best performing ADTFD. Therefore, considering all the measures, the LOS method performs better due to its local adaptive property.

Table 3.5: (t, f) quantitative assessments

TFD / $(\mathbf{t}, \mathbf{f})$ measurement	LOS	WVD	SWVD	CWD	MBD	Spec	EMBD	CKD	SM	ADTFD
TFNRE	12.7	15.2	14.7	13.7	15.1	14.5	13.8	13.2	14.0	12.5
RN$(\times 10^{-3})$	0.25	0.14	0.15	0.33	0.12	0.08	0.16	0.25	0.11	0.38
EC$(\times 10^{3})$	9.06	480.6	236.4	181.3	385.9	51.4	62.7	36.3	32.6	14.5
MEC$(\times 10^{3})$	0.04	1.27	0.66	0.56	1.03	0.22	0.25	0.16	0.14	0.07
HM	0.84	0.94	0.92	0.88	0.93	0.68	0.81	0.78	0.74	0.84

3.3.5 Instantaneous frequency estimation

One of the desired (t, f) features is the instantaneous frequency (IF) [143], so the IF estimation capability is often used as a performance evaluation criterion for TFDs, including robustness in the presence of noise. The IF estimation performance of the LOS under different SNRs is estimated using the component extraction method [129] (see Appendix 3.C). This method also requires the number of components before estimation. The signal model is defined as:

$$x(t) = \cos(18.85t^3 + 125.664t) + \cos(18.85t^3 + 94.248t) + n(t); \quad (3.36)$$

where $n(t)$ is additive white Gaussian noise. Figure 3.15 shows the logarithmic mean square error (MSE) between the actual and estimated IF under different SNRs by performing 100 simulations. The LOS performed better than the fixed TFDs under all SNR conditions. The performance is comparable to the adaptive technique, the ADTFD. For example, for an SNR of $-6dB$ considering the first component (Figure 3.15(a)), the LOS achieved a logarithmic MSE of -3.4 dB, while different fixed and adaptive TFDs (standard spectrogram, SM, CKD, ADTFD) achieved $-1.61, -2.50, -2.73, -3.22$ respectively. When estimating IF for the second component (Figure 3.15(b)) with the same SNR$(-6dB)$, the LOS achieved a logarithmic MSE of -3.40, while different fixed and adaptive TFDs (standard spectrogram, SM, CKD, ADTFD) achieved $-1.75, -2.66, -2.89, -3.25$ respectively. These findings indicate that an accurate estimation of the IF of the signal components is directly related to the auto-term resolution property and cross-term suppression property of (t, f) method and, hence, indicates the high resolution of the LOS.

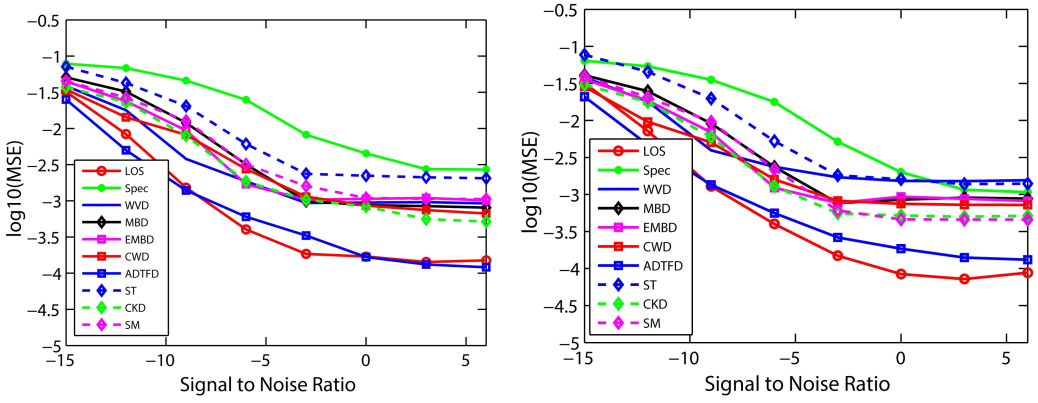


Figure 3.15: IF estimation of a multicomponent signals using different TFDs under different SNRs ranging from $-15dB$ to $+6dB$; (a) logarithmic MSE for the first signal component and (b) for the second signal componment.

3.3.6 Relationship between the LOS and the other QTFDs

The LOS derived from the STFrFT is related to the other TFDs by adding L correction terms, i.e. $2Re[\mathrm{STFrFT}(n,k+i)\mathrm{STFrFT}^{*}(n,k-i)], i = 1,2,\dots,L$ in the recursive formulation of discrete fractional S-method (FrSM) [144]. As the LOS is derived from the fractional domain, we denote it as FrSM instead of S-method (SM) and it can be written in the same ways as the SM [144]:

$$\mathrm{FrSM}_{L}(n,k) = \mathrm{FrSM}_{L-1}(n,k) + 2Re[\mathrm{STFrFT}(n,k+L)\mathrm{STFrFT}^{*}(n,k-L)] \qquad (3.37)$$

with initial $(L = 0)$ distribution $\mathrm{FrSM}_{0}(n,k) = LOS = |\mathrm{STFrFT}(n,k)|^{2}$. The optimal angle and window derived from the LOS methodology have been used to compute FrSM.

Equation (3.37) can be used to relate the LOS to the other QTFDs. There are three different cases: (i) taking no correction terms $(L = 0)$, Equation (3.37) becomes non-negative and has no cross-terms, i.e. LOS; (ii) taking $(0 < L < \frac{L_s}{2})$ correction terms, it relates to FrSM and enhances the concentration with the expense of interferences (cross-terms), and (iii) by gradually increasing L to $L \geq L_s/2$ correction terms, it becomes the WVD, which is highly concentrated but suffers from cross-terms. The optimal correction terms (L) can be obtained by measuring the MEC index, for which this relationship gives a minimum MEC index value.

These adjustments of L result in a good compromise between cross-term interference and the desired properties. Normally a small number of correction terms are sufficient to produce a high concentration TFD and the objective function of LOS can be used to determine optimal correction terms. Figure 3.16 illustrates the relationship of the LOS with other TFDs. In this way, one can improve the auto-terms concentration, control the cross-terms and combine the good properties of the LOS and the highly concentrated WVD.

Once the WVD can be calculated from the LOS, the other QTFDs can also be generated easily (see [33], Chapters 2 and 3). Thus, this generalisation allows the LOS to be converted to other QTFDs. The pseudocode of this generalisation is described in Appendix 3.B [144]. Figure 3.17 simulates an example to show the relationship between the LOS and the other QTFDs of the proposed method [145].

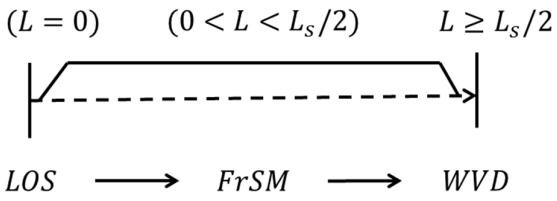


Figure 3.16: Generalisation of the proposed method: from LOS to WVD. L_s is the length of the signal.

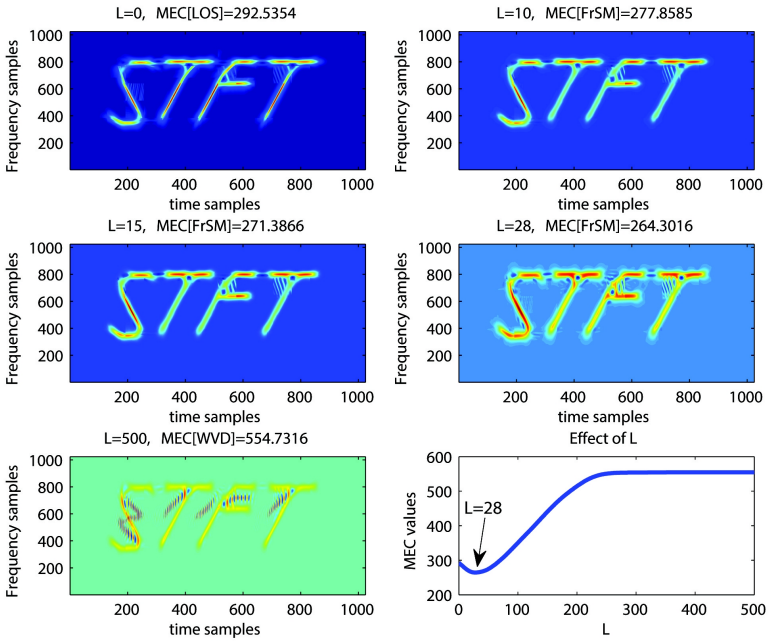


Figure 3.17: Illustration of the generalisation of the LOS. LOS renders no cross-terms and after adding few correction terms it enhances the energy concentration: the fourth one from the top shows the highest compromise between concentration and cross-terms whereas the fifth one, essentially a WVD, provides the highest concentration but suffers from cross-terms. In the analysis $\mathrm{MEC}(\times 10^3)$ was used to measure the concentration and the best choice according to this measure is $\mathrm{L} = 28$.

3.3.7 Discussion of method-2

An improved methodology to optimise the spectrogram has been described in this study. It automatically determines the optimal parameters from a set of the window length L_w ($L_w = \{i | i = 0,1,2,\dots,D-1\}$) and the fractional angle α ($\alpha = \{j | j = 0,1,2,\dots,E-1\}$) for all signal components using local optimisation in order to obtain a high resolution (t, f) signature.

Implementation of the LOS requires iterative calculation to determine the optimal parameters; therefore, a reasonable concern is the running time in applications. We tested the number of D in relation to the optimisation. Figure 3.18 demonstrates the LOS characterisation quality (MEC) and code run-time against D using the type 1 simulated signal of length $L_s = 512$. Observation shows that the MEC converges after the 5^{th} iteration and the optimisation performance is saturated thereafter.

However, considering the run-time and energy concentration, $D = 16$ is safety zone and is therefore used in this study. The run-time is less than $20\,\mathrm{s}$, which is acceptable in most applications. This simulation was run on a Windows-7, Intel Core i-7, 8GB RAM computer. The amount of run-time also depends on the signal length; the larger the signal length, the higher the computational cost, i.e. run-time. In addition, modern computers possess multi-core processors and parallel computing can be used to further speed up the processing time. This is beyond the scope of this study.

The dividing parameter R also plays a role, due to the number and orientation of signal components. We report that $R = 2, 4, 16$ are found to provide minimum value of $J[\mathrm{FrSpec}(u, v)]$ for a wide-range of time and frequency varying signals. A more robust measure to determine R will be considered in further study of the method. Further improvement of the LOS can be achieved by adopting pre-processing and post-processing steps as in [33, 140].

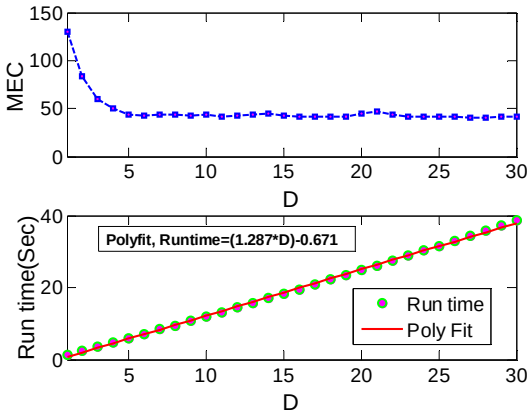


Figure 3.18: Convergence of proposed method (upper plot) and run-time measurement (lower plot).

To summarise:

- The standard spectrogram offers cross-term free (t, f) signature but it is too sensitive to the analysis window selection.
- Different state-of-the-art fixed QTFDs offer sharp localisation but suffer from cross-terms. One can reduce the cross-terms by adjusting the control parameters, but this task requires human intervention which is not suitable for an automated system. This justifies the approach of parameter adaptation and optimisation.
- Among the proposed adaptive methods, the ADTFD adapts its kernel using only the direction information and provides a good (t, f) resolution with reduced cross-terms, but it still requires human intervention to adjust the controlling parameters ([33], pp.299–307).

- The LOS is locally optimised for both the analysis window and angle and does not require any human intervention or a *priori* signal information. It provides a cross-term free, high resolution and concentration of (t, f) auto-terms.
- The LOS can be related to other TFDs by adding *L* correction terms.

3.4 Comparison between method-1 and method-2

Method-1 and method-2 can be useful tools for signal synthesis and processing, especially for non-stationary signals. Both methods are fully automatic in the sense that users only need the signal they are going to analyse. The algorithms automatically choose optimal parameters for representation with improved resolution. A difference between the two methods is that the global optimisation method uses single global parameters, whereas local optimisation method divides the TFD into small grids and performs local optimisation. Moreover, the global optimisation method cannot significantly improve the (t, f) resolution and concentration for signals composed of closely spaced chirp signals, or a mixture of short and long duration components which overlap within the (t, f) domain. In addition, method-1 uses a complex optimisation algorithm called HGA. Therefore, a thorough understanding of genetic algorithm is also required. This complex optimisation increases system complexity and slows down the running program. On the other hand, a simplified local optimisation procedure is proposed in method-2. This section presents a comparison between the two proposed methods in terms of their computational time and performance. The HGA-based method uses the same optimisation criterion as LOS for a fair comparison. From Table 3.6, it can be seen that the average run-time for the LOS is 19.4 s, which is 5.70, 9.38 and 53.4 times lower than the standard implementation of HGA-optimised EMBD, CKD, and ADTFD respectively for a signal length of 512.

Table 3.6: Run-time comparison*

Run-time (in s)	HGA-Optimised TFDs			LOS
	EMBD	CKD	ADTFD	
Mean time	110.5	181.9	1035.9	**19.4**
Minimum time	103.6	102.5	1004.8	**19.1**
Maximum time	119.0	203.3	1103.4	**19.7**
Standard Deviation	±3.1	±32.8	±30.3	**±0.1**

*These experiments were run in the Windows-7 on an Intel Core i-7 platform having 8 GB RAM. The experiment was run 50 times and the mean, minimum, maximum and standard deviation of the computation time are presented here.

A more qualitative comparison between two methods is given in Table 3.7 in terms of assumption, visualisation, automatic selection, robust to noise, high resolution, low cross-terms and low computational time. For example, in standard implementation the QTFD parameters are assumed or manually selected (Assumption) whereas in HGA optimised QTFD or LOS, there is no assumption. The other properties can be described in a similar way.

Table 3.7: Qualitative comparison of the QTFD optimisation

	QTFDs	Assumption	Visualisation	Automatic selection	Robust to noise	High Resolution	Low Cross-terms	Low Computational time
Standard Implementation	Spectrogram	+++	+	x	+	+	+	+++
	WVD	x	+	NA	+	+++	+	+++
	EMBD	+++	++	x	+	+++	++	+++
	CKD	+++	++	x	++	+++	++	+++
	ADTFD	+++	+++	x	++	+++	+++	++
Global (HGA) Optimised	HGA-SPWVD	x	++	+++	++	++	++	+
	HGA-EMBD	x	++	+++	++	+++	++	++
	HGA-CKD	x	++	+++	++	+++	++	+
	HGA-ADTFD	x	+++	+++	+++	+++	+++	+
LO*	LOS	x	+++	+++	+++	+++	+++	++

+++, ++, +, x represent strongly satisfied, moderately satisfied, weakly satisfied and not satisfied, respectively. NA: not applicable. LO* represents locally optimised QTFD.

3.5 Application to neonatal EEG

EEG signals possess time-varying spectral contents, and the variations of EEG amplitude and frequency contents are representative of certain neurological disorders [146]. Therefore, (t, f) characterisation and processing of EEG signals can provide improved information over traditional visual interpretation by representing an optimal (t, f) signature of the time-varying spectral EEG contents. In this section, three neonatal EEG patterns were considered for (t, f) characterisation: namely, seizure, suppression and burst. The presence of these patterns in newborn EEG is associated with adverse neurodevelopmental outcome (see Chapter 2). Early characterisation and detection of these patterns can aid in diagnosis and clinical management.

Given the non-stationary nature of EEG signals, the optimised TFD can provide a useful tool for precise characterisation of these patterns. Figure 3.19 illustrates the (t, f) characterisation of seizure, suppression and burst patterns. Note that, only the $0-16\,Hz$ frequency range is considered in the

(t, f) representation, as dominant neonatal EEG components lie within this frequency range. It can be seen from Figure 3.19 that WVD cannot provide precise characterisation of these patterns. In fact, it provided significant interferences that decreases the (t, f) readability and can result in misinterpretation. On the other hand, optimised TFDs (e.g. CKD and LOS) can provide a precise characterisation of these patterns. The (t, f) representations using optimised CKD and LOS for these patterns show how the frequency contents of varied amplitudes, i.e. instantaneous frequency and instantaneous bandwidth, change over time. Furthermore, the local signal components are better captured by LOS than HGA-CKD due to the local optimisation. It is also revealed that most of the signal components lie within the delta ($0-4$ Hz) and theta ($4-8$ Hz) bands.

Another important finding is that the (t, f) representations or characterizations of these patterns are different, and thus highly discriminative. Consequently, different dynamic and non-stationary features can be extracted from these optimised and discriminative TFDs and can improve the signal detection and classification. This is one of the core motivations of designing and optimising TFDs, as these findings will be used in the next two chapters.

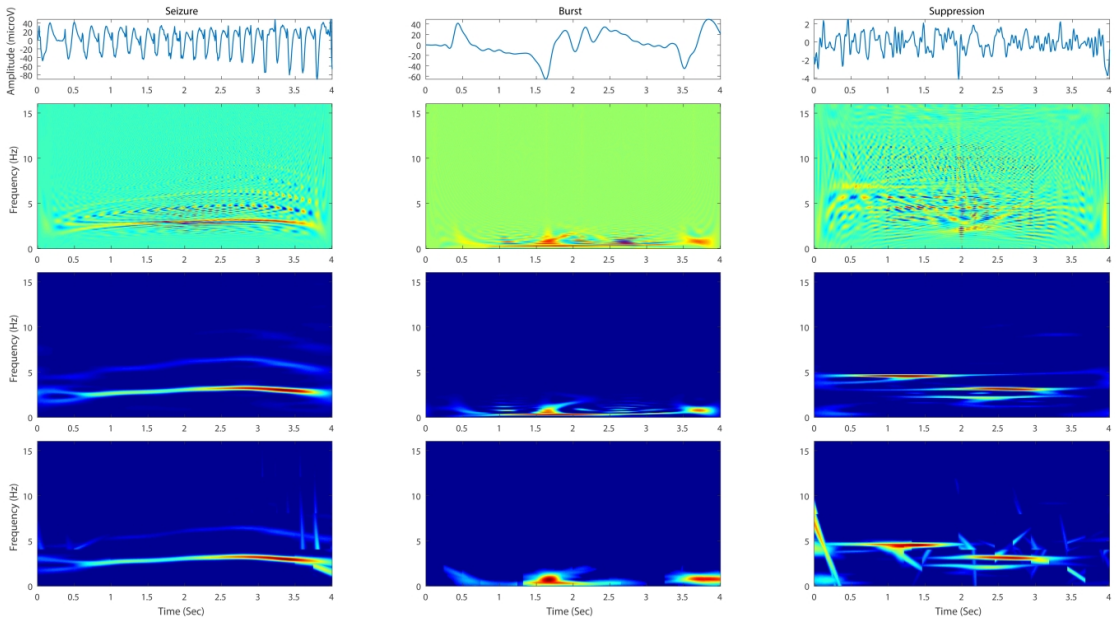


Figure 3.19: (t, f) characterisation of EEG signal. First row presents neonatal EEG signals: seizure, burst, suppression. Second row represents the WVD of these patterns whereas third and fourth row represent HGA-CKD and LOS of these patterns.

To illustrate the application of the developed methodology, this study uses a database including 12 -channel $(Ch = 12)$ EEG burst-suppression (B-S) data recorded from newborns admitted to the Royal

Brisbane and Women's Hospital in Brisbane, Australia. A flowchart of the methodology is presented in Figure 3.20.

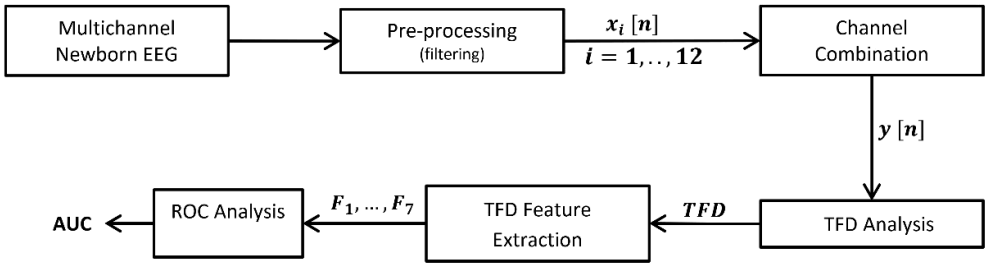


Figure 3.20: Methodology for detecting multichannel neonatal EEG burst and suppression. AUC was calculated by thresholding.

Table 3.8: AUC analysis for EEG burst suppression classification.

Features / TFDs	Max SVD	SVD entropy	ECM	Deviation IF	$(\mathbf{t},\mathbf{f})$ Renyi entropy	$(\mathbf{t},\mathbf{f})$ mean	$(\mathbf{t},\mathbf{f})$ variance	Average
LOS	0.99	0.71	0.98	0.80	0.79	0.97	0.98	**0.88**
EMBD	0.95	0.68	0.95	0.53	0.60	0.96	0.94	0.80
CKD	0.98	0.53	0.98	0.83	0.62	0.97	0.98	0.84
ADTFD	0.98	0.53	0.98	0.83	0.64	0.96	0.98	0.84

EEG burst and suppression patterns are marked by a paediatric neurologist. The signals were inspected visually to remove episodes of major artifacts. There are 115 epochs of burst and 115 epochs of suppression extracted from artifact-free EEG signals from 3 newborns; further details can be found in Chapter 4. The sampling frequency is $\mathrm{fs} = 256\ \mathrm{Hz}$. After pre-processing and channel combination, B-S data were analysed using different TFDs. Seven TFD-based features (listed in Table 3.8, Appendix 3.D) were extracted from the B-S data. The receiver operating characteristic (ROC) was applied to evaluate the performance of the (t, f) features in discriminating B-S patterns, and the area under ROC curve (AUC) is calculated. The AUC is often used to measure the quality of a feature or a classifier for the task of discriminating between two patterns (burst/suppression in this application) ([33], pp. 931–35). The AUC value lies between 0.5 and 1. A classifier with an AUC value of 0.5 corresponds to a random-guessing classifier, whereas an AUC value of 1 is a perfect one. Table 3.8 shows that the LOS performed better than other global optimised QTFDs as the averaged AUC value (0.88) of all the features is the highest. The effectiveness of the detection features is highly dependent on an accurate (t, f) representation that (i) eliminates erroneous information resulting from the cross-terms and (ii) strengthens useful information contributing to the features [147]. The LOS

superior characterisation performance is attributed to the local adaptation to each B-S signal, and this consequently leads to superior discriminating power of the features. This finding confirms that the local adaptation and optimisation of the LOS is also useful for signal detection and classification. Note that this study uses a simple case of binary (burst/suppression) class problem, but will be applied to a complex multiclass EEG classification problem in the next chapter.

3.6 Overall summary

Over the last few decades, TFDs have gained in popularity and become a standard tool in many disciplines. However, despite all improvements it is still difficult to adjust the kernel parameters and window parameters for the optimal use of TFDs. This chapter addresses this gap by providing two fully automatic procedures, namely global optimisation (Section 3.2) and local optimisation (Section 3.3). These methods contribute a data-adaptive kernel and spectrogram analysis window which can significantly reduce the presence of cross-terms and enhance the energy concentration and resolution. Experiments on simulated signals, and characterisation of real world bat and EEG signals using different performance measures, have shown that the TFDs optimised by the proposed methods can achieve good (t, f) resolution, optimise the local signal components and outperform the other state-of-the-art methods in various situations. The robustness of these methods has been tested under various SNR values by comparing the accuracy of IF estimation. These optimised TFDs are not only suitable for the characterisation of non-stationary signals but also suitable to improve signal detection and classification. The next two chapters will apply the optimised TFDs to extract dynamic features for EEG background patterns classification (Chapter 4) and neurodevelopmental outcome prediction (Chapter 5).

Chapter 4 Multichannel EEG background patterns classification in term neonates with HIE

4.1 Introduction

Perinatal HIE follows a peripartum hypoxic–ischaemic insult to the brain and is a major cause of morbidity and mortality; [148, 149]. In survivors, the common sequelae include cerebral palsy, cognitive and motor impairments, neuro-developmental delay and epilepsy. Pathological studies also show a pattern of injury, predominantly cortical or of relative cortical sparing but with deep grey matter injury particularly involving hippocampi, lateral geniculate nuclei, putamen, ventrolateral thalami and dorsal mesencephalon [150]. In NICUs, it is necessary for clinicians to obtain prompt and accurate diagnosis of HIE, assessment of its severity and guidance in the treatment plan. Clinical scoring systems such as the Sarnat score are routinely used in diagnosis but are not very sensitive to dynamic changes that may occur during treatment [151]. Complementary methods of assessment such as cord blood, pH, Apgar score and base deficit individually also lack sensitivity to different grades of HIE severity injury and are not reliably detected by conventional MR imaging in the first days post insult, precluding them from use in early diagnosis [12]. EEG is a low-cost and non-invasive tool which can easily be implemented at the cot-side soon after birth to monitor dynamic changes in cerebral function [12]. EEG is also used for detection of seizures and assessment of response to anticonvulsants in the clinical

What is already known on this topic?

Visual inspection and classification of various EEG background patterns are time consuming and require experienced interpreters.

Features extracted from (t, f) domain can improve the classification performance.

Feature selection and SVM parameter optimisation are necessary.

What is the contribution of this study?

- ✓ Different EEG background patterns have been automatically classified.
- ✓ Different features are extracted from t, f and (t, f) domain for the classification using RBF-SVM.
- ✓ A hybrid feature selection (HFS) algorithm comprised of mRMR and genetic algorithm has been proposed for the selection of prominent features and SVM parameters.
- ✓ Two classification methods have been developed for this multichannel and multiclass EEG background patterns classification problem.

setting [47], and is particularly useful because many seizures in the newborn have no detectable clinical manifestations.

Visual inspection of the EEG signal is routine but visual interpretation and classification is laborious and time-consuming, especially in the case of long recordings. It also requires experienced interpreters, who are not always available. To overcome these limitations, an automated detection and classification approach would provide a decision support to the clinical staff. The system needs to monitor and assess recovery of background EEG activity, as a normal EEG is highly predictive of a normal outcome, whereas various abnormal EEG features have been consistently associated with neurological abnormalities or death [47].

Currently visual inspection is the 'gold standard'; automated EEG analysis is a relatively new research area and still primarily a research tool. Most research has focused on specific patterns: for example, seizure detection and localisation [29, 35-37], burst-suppression classification [24, 152-154] and detection of the sleep-wake cycle [38]. Classification of a broader variety of EEG background patterns will increase the system complexity.

Another important problem is the selection of the most useful features present in neonatal EEG background. Studies mainly use different t-domain, f-domain features. EEG is a non-stationary signal, i.e. the statistical properties of the frequency and amplitude content change over time, resulting from the random firing of neurons; joint time-frequency (t, f) distribution is well adapted in this setting as it takes the signal non-stationarities into account. Time-frequency distribution has not been widely used in neonatal EEG background classification. Previous studies suggest that dynamic features extracted from TFD show good performance in detecting newborn seizures [29] and automatic grading of EEG background patterns [155]. The extraction of prominent features plays a vital role in the classification, but it is also necessary to optimise feature selection and classifier parameters as the classification accuracy depends on the optimal use of both features and classifier parameters.

To overcome the above limitations, we have developed a new methodology for the appropriate use of these features and classifier and (t, f) -features are introduced in order to classify multichannel neonatal EEG background patterns. The major aspects and key contributions are:

- Characterisation of neonatal EEG background patterns using a state-of-the-art optimised TFDs. Different EEG backgrounds show different (t, f) signatures (Section 4.4.2.1.2).
- The extraction of different t, f and (t, f) -domain features to characterise and classify the EEG background patterns (Section 4.4.3).
- Proposal of a hybrid feature selection (HFS) algorithm. A state-of-the-art classifier called Support Vector Machine (SVM) has been optimised and used to classify different neonatal

EEG patterns using the feature subset selected by the HFS algorithm. This HFS algorithm uses a genetic algorithm that not only selects the prominent feature subset but also simultaneously selects the SVM tuning parameters to ensure the optimal use of the SVM classifier ('Classification using single feature subset', Section 4.4.6).

- Classification using single feature subset selects a single feature subset for all classes which has been extended to select class-specific features. This method ('Classification using class-specific feature subset') increases the classification performance as the selected feature subset is *class-specific* (Section 4.4.7.1).
- The application of both classification methods (classification using single and class-specific feature subsets) to multichannel neonatal EEG background pattern classification. Different fusion techniques have been applied to solve this complex multiclass problem and classification performance of different approaches is compared in Section 4.5.
- A decision support system as a potential application of the proposed approach is discussed in Section 4.8.

4.2 Clinical rationale of HIE and the EEG background patterns

This study has focused on EEG background patterns that best predict the neurodevelopmental outcome. Meta-analysis shows that burst and suppression patterns predict neurodevelopmental outcome (see Chapter 2) and that they are present in 50–60% of babies with HIE.

During the period of suppression, about 95% of the cortical cells are electrically silenced. The cerebral cortex is unable to participate efficiently in information processing and cognitive processes. These phenomena are modulated, as a function of voltage amplitude and frequency, by the general state of the central nervous system and, therefore, the seriousness of the encephalopathy [156]. Depending on the severity and evolution of brain injury, cortical neurons can exhibit two phenomena: (i) if the brain is severely injured, voltage amplitude and frequency gradually decrease, the burst period becomes shorter, the suppression period longer, and eventually a continuous isoelectric EEG (flat trace) results [32]; (ii) as EEG activity increases, a repetitive hypersynchronous discharge of a population of cortical neurons, seizure, may occur [157]. Normalisation of EEG background may occur with either discontinuity or full recovery of continuous voltage [158, 159]. Figure 4.1 illustrates background EEG changes during severe HIE.

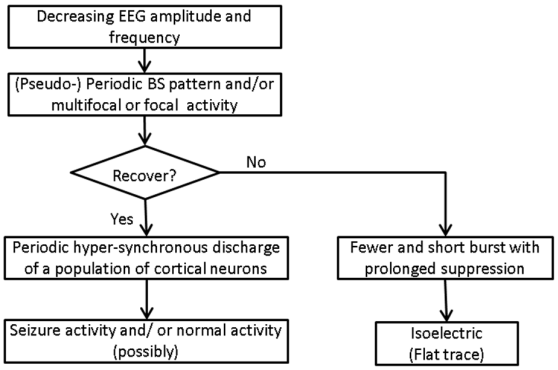


Figure 4.1: Evaluation of EEG changes during severe HIE.

Different types of artifacts are present within EEG recordings and include electrode and wire interference, movement, and transient muscle artifact, to name a few. These need to be identified and removed to facilitate accurate analysis of the signal. Seizure and artifact are not considered a part of the EEG background. However, it is necessary to distinguish seizure and artifact, in addition to burst and suppression, to have utility in the clinical setting. Thus, the following five classes will be used for classification: burst, suppression, seizure, normal and artifact.

4.3 Materials

12-channel $(Ch = 12)$ clinical EEG signals were used from 40 newborns admitted to the NICU, Royal Brisbane and Women's Hospital, Brisbane, Australia. Neonates ≥ 35 weeks gestation were enrolled for EEG monitoring if they depressed at birth and required extensive resuscitation after birth; exhibited clinical evidence of encephalopathy; or seizures developed within 72 hours of birth. This study was conducted with approval from the Human Research Ethics Committees of the Royal Brisbane and Women's Hospital, Brisbane, Australia. Written, informed consent was obtained from at least one parent of each neonate who participated in this study.

EEG signals were collected using a MEDELEC Profile System (Medelec, Oxford Instruments, UK) after skin preparation using 12 Ag/AgCl electrodes placed according to the international 10–20 standard. Monopolar recordings were collected from F3, F4, C3, C4, P3, P4, O1, O2, T3, T4, T5 and T6 positions. A1 and A2 are the reference electrodes. The five patterns were marked by paediatric neurologists, and their classification is used as the 'ground truth'. Figure 4.2 plots a multichannel newborn EEG containing all five patterns. The marking of different EEG background patterns is shown at the bottom of the plot. The classifications denote generalised patterns; focal (localised) activities have been ignored. The EEG from each channel is first down-sampled from 256 Hz to 64 Hz and then segmented into 4-s epochs.

This database contains artifacts produced by the random movement of the subjects, electrode contact, electrocardiogram, respiration, eye blinking and different interference in the NICUs. The different

types of artifacts are generalised and regarded as an 'artifact' class in this study. A total of 35,542 segments i.e. 39.5 h of EEG data comprising the above five patterns were finally extracted and used in this study. The mean duration of EEG per neonate was about 1 h.

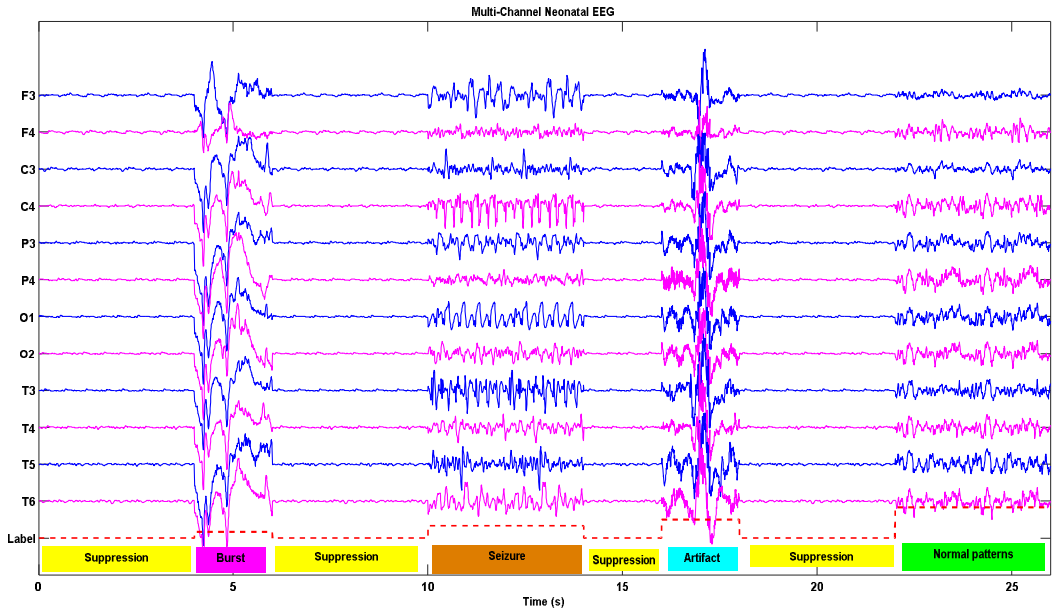


Figure 4.2: Neonatal EEG showing different EEG background patterns. Signal amplitude is in microvolts (μV). The bottom line shows the neurologist classification.

4.4 Methods

Figure 4.3 shows the general methodology used in newborn EEG background classification. The analyses have been done in high performance computing system at the Research Computing Centre of the University of Queensland.

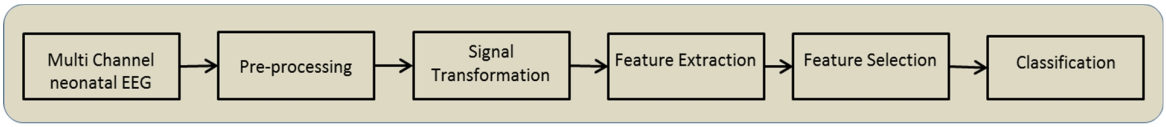


Figure 4.3: General pipeline of the multichannel EEG background classification.

4.4.1 Pre-processing

The DC component is removed from the EEG signals, which are then band pass filtered to 0.5–25 Hz, a range that contains the vast majority of the energy of neonatal EEG signals.

4.4.2 Signal transformation

The t-domain multichannel EEG signal is transformed into f-domain using Fourier transform, and also into the (t, f) domain using the methods presented in the next section. This is because some

features extracted from these domains show more discriminative properties than those in the t-domain.

4.4.2.1 Time-Frequency analysis

A time domain signal can be represented in (t, f) domain by translating, modulating and scaling a basis function with a certain time and frequency localisation. Matching pursuit (MP) is one of the important classes in this category.

4.4.2.1.1 Time-Frequency Matching Pursuit (TFMP)

TFMP is an iterative atomic decomposition technique that represents a signal $x[n]$ using an over-complete dictionary D [33] (See Appendix 4.A).

The dictionary atoms are formulated from discrete Gabor atoms, which can be written as [160]:

$$g_{(s,u,\omega,\theta)}[n] = \frac{K_{(s,u,\omega,\theta)}}{\sqrt{s}} e^{-\pi(n-u)^2/s^2} \cos[2\pi\omega(n-u) + \theta] \tag{4.1}$$

where $s \in \mathbb{R}^+; u, \omega \in \mathbb{R}; \theta \in [0, 2\pi]$. $K_{(s,u,\omega,\theta)}$ is a normalisation factor such that $\left\|g_{(s,u,\omega,\theta)}\right\|^2 = 1$ and s, u, ω, θ correspond to an atom's scale, translation, modulation and phase respectively. The features used are the number of atoms, the mean and variance of scale, translation, modulation and phase. Each EEG segment is decomposed by the Gabor dictionary using a 5% residual energy as stopping criterion and different features are extracted and presented in Table 4.2.

4.4.2.1.2 Time-Frequency Distribution (TFD)

A t-domain signal can be represented by the signal's energy distribution simultaneously as a function of both time and frequency called time-frequency distribution (TFD). The quadratic time-frequency distribution (QTFD) is one of the most useful classes of TFD and has shown promising results for non-stationary signal analysis and classification [29, 33]. The QTFD of a signal $x[n]$ can be calculated using [33] (p. 341):

$$\rho[n, k] = \underset{m \to k}{\overset{DFT}{}}\{G[n, m] *_n^* (z[n + m] z^*[n - m])\} \tag{4.2}$$

where $G[n, m]$is the discrete time-lag kernel of the QTFD, $z[n]$ is the analytic associate of $x[n]$, $*_n^*$ represents discrete time domain convolution, and * denotes the complex conjugate. For an N-point signal $y[n]$, $\rho[n, k]$ is represented by a $N \times M$ matrix where M is the number of FFT points used for calculating the TFD. In this chapter, extended modified B distribution (EMBD) and more recent CKD are chosen due to their high resolution and having shown promising results in classifying neonatal EEG seizure signal classification [29]. Table 4.1 shows their time-lag kernel functions and their

control parameter. These control parameters are optimised for each EEG segment by the method presented in Chapter 3.

Table 4.1: QTFDs including kernel function and control parameters used in this study ([33],pp. 341).

QTFDs	$G[n,m]$	Control parameters
EMBD	$\frac{cosh^{-2\beta}[n]}{\sum_n cosh^{-2\beta}[n]}\frac{cosh^{-2\alpha}[m]}{\sum_n cosh^{-2\alpha}[m]}$	$\alpha,\beta;$ $0\le\alpha\le 1, 0\le\beta\le 1$
CKD	$DFT_{l\to n}\left(\frac{cD^2}{e^{\left(\frac{l}{N}\right)^2-D^2}}\right)e^{2c}\left(\frac{cE^2}{e^{\left(\frac{m}{N}\right)^2-E^2}}\right)$	c, D, E

In addition, the LOS (see Chapter 3) and the S-Method [135] have been used as they are prominent state-of-the-art TFD's. Figure 4.4 shows the optimised CKD representation of the five-target EEG background patterns as an example.

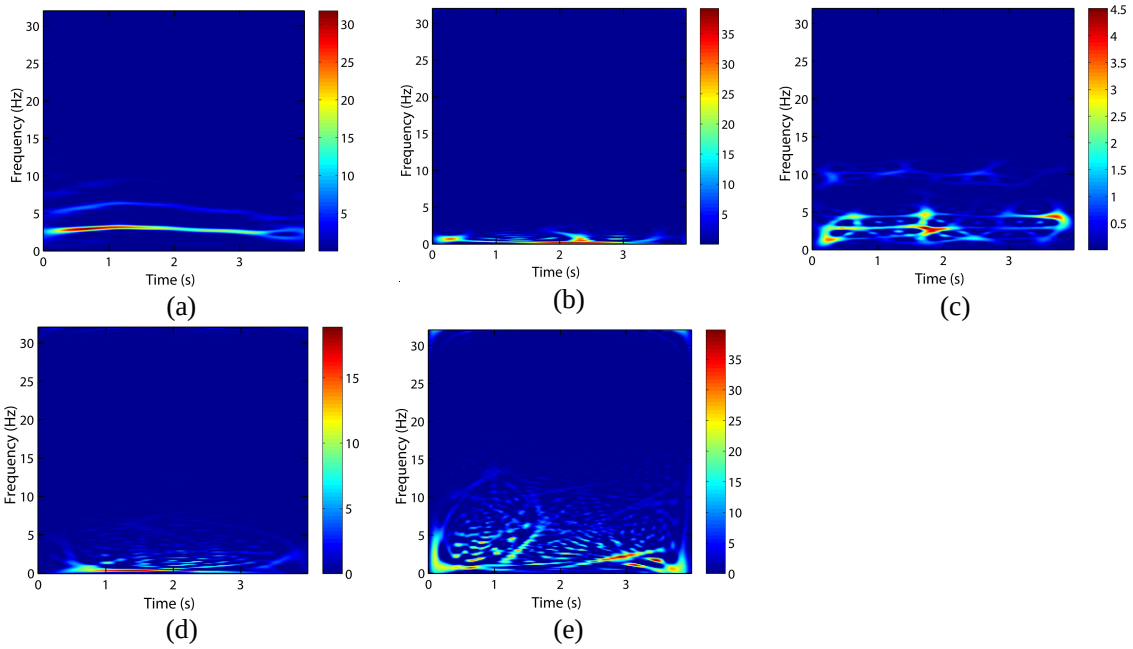


Figure 4.4: (t,f) of different EEG background patterns using optimised CKD: (a) seizure, (b) burst, (c) suppression, (d) normal and (e) artifact.

4.4.3 Feature extraction

Feature extraction is a key part of automatic classification, as features represent a particular pattern from objects. Limited prior knowledge is available on whether and what features would perform well for the considered multiclass problem. A large set of features is extracted from the three domains: t-domain, f-domain and (t,f) domain. These features are previously used for EEG signal classification [35, 161, 162] and describe the EEG from different perspectives. t-domain and f-domain features represent temporal and spectral characteristics, whereas (t,f) domain features represent non-

stationary characteristics of a signal. Table 4.2 lists the extracted features and provides references with information on their implementation. These features provide temporal, spectral, entropy information and take the EEG non-stationarity into account; hence they are included in this study.

4.4.3.1 t-domain and f-domain feature extraction

Different t-domain and f-domain features have been used to describe the neonatal EEG signal for classification [148] and those used in this study are summarised in Table 4.2 . These features are used in neonatal seizure detection and EEG background classification. Therefore, we explore those features for the current classification problem.

Table 4.2: EEG feature extraction from different domains.

Domain	Features	Features Name
Time	Statistical features	First four (mean, variance, skewness, kurtosis) statistical moments [35] Coefficient of variation, RMS power, Max, Min[35] Mean of the lower and upper envelope ([33],pp 944-945)
	Entropy based features	Hjorth moments [148] Higuchi fractal dimension [163] Renyi, Shannon and Tsaillis Entropy [164, 165] Approximate and Sample Entropy [165-167] Hurst exponent[168]
Frequency	Spectral and relative power based features	Spectral Flatness, Flux and Entropy [35] Spectral Edge frequency at 80, 90 and 95% [169] Relative power in delta, theta and alpha bands[170]
Time-frequency (t, f)	(t, f) signal based features	Matching Pursuit features [171] First four statistical moments of the TFD [29] TF Entropy, Entropy Flatness and Flux [29] Instantaneous Frequency mean and range [29] Maximum and Entropy of the singular values Gini Index [172], ECM
	(t, f) band specific features [in Delta, Theta, Alpha, Beta bands]	First four statistical moments of the $ISPR$ [161] Coefficient of Variation of the $ISPR$ [161]
	(t, f) image based features	Convex Hull, Perimeter and Compactness [29] Centred region, Rectangularity and Aspect ratio from the image moments [29] 20 GLCM features extracted using [173, 174] 16 SFTA features extracted using [175]

4.4.3.2 (t, f) based feature extraction

TFDs are rich in information, but the full (t, f) -matrix, in the present context a matrix of dimension 256×256, cannot be directly used as a set of features for the classification as this would significantly

increase the dimensionality of the problem. To avoid this, a small representative set of features describing the relevant information for the signal classification is extracted as described below.

4.4.3.2.1 (t, f) statistical features

Studies [162, 176] have extracted different prominent TFD-features to characterise neonatal EEG signals. In [29], the authors used statistical (t, f) features to classify neonatal seizure. These features could be useful as they take the EEG non-stationarities into account. In addition, TFD can be considered as a bi-dimensional density function and TFDs of different EEG background patterns show different distributions [29]. Therefore, features extracted from TFDs offer significant and discriminative information, justifying their inclusion in the feature set (see Table 4.2).

4.4.3.2.2 Band specific (t, f) features

The (t, f) features of specific EEG bands, typically delta $(0 - 4\ Hz)$, theta $(4 - 8\ Hz)$, alpha $(8 - 12\ Hz)$ and beta $(12 - 30\ Hz)$, may be informative. The instantaneous spectral ratio (ISPR) has been calculated separately for each band. The ISPR has been used in electromyogram (EMG) analysis and we have adapted it for use in EEG. It is calculated for a TFD at each time instant as [161]:

$$ISPR[n] = \frac{\sum_{k=f_l}^{f_u} |\rho[n,k]|}{\sum_{k=1}^{f_H} |\rho[n,k]|} \tag{4.3}$$

where f_l and f_u represent the lower and upper frequency bins associated with the frequency band and f_H represents the highest frequency bin. Once the ISPR of a band is calculated, the mean, variance, coefficient of variance, skewness and kurtosis from these bands are extracted (see Table 4.2).

4.4.3.2.3 TFD image related features

TFD is a 2D matrix which can be transformed to a binary intensity image; different geometric features can be extracted from this image. By following the methodology presented in [29], different image related features are extracted, as in Table 4.2. In addition, Grey-level co-occurrence matrix (GLCM) and segmentation-based fractal texture analysis (STFA) feature extraction have been extensively used in image processing, especially in texture classification [174]. These feature extraction methods can be applied to a TFD. One study used image-based features to detect adult seizures [177]. These features could also be useful for classifying the TFD in terms of observed textural patterns.

GLCM can be seen as a directional pattern counter with a specific distance d and angle θ between neighbouring image pixel pairs for grey-scale images. It reflects the comprehensive information on the direction, adjacent interval and amplitude variations for image grey-level. For a given TFD-image

$\rho'(n,k)$ (which is calculated from TFD, $\rho[n,k]$ [29]), the corresponding GLCM can be calculated as [174]:

$$P[i,j,d,\theta] = \sum_{t=0}^{n}\sum_{f=0}^{m}\begin{cases}1, if\ \rho'[(n,k)] = i\ and\ \rho'[n + dcos\theta, k + dsin\theta] = j \\ 0, \quad otherwise\end{cases} \tag{4.4}$$

where $P[(i,j,d,\theta)]$ in GLCM describes the relative frequencies or occurrences with which two pixels separated by a particular displacement distance d and a specified angle θ occur on the image, one with grey-level i and the other with grey-level j. Several GLCM features are extracted by the method presented in [173, 174].

SFTA uses fractal dimension to describe the fractal structure of the surface indicating the grey levels of an image. This is one of the recent approaches, and the authors of [178] have shown that the SFTA-based features outperformed other texture feature descriptors and, hence, added in the feature vectors (see Table 4.2).

4.4.3.3 Feature vector formulation

After extracting the features from t-domain, f-domain and (t,f) domain, all features are concatenated to create the final feature, vector (FV_{all}), expressed as:

$$FV_{all} = [FV_t\ \ FV_f\ \ FV_{(t,f)}] \tag{4.5}$$

where $FV_t, FV_f, FV_{(t,f)}$ are the t, f and (t,f) domain features respectively. $FV_{(t,f)}$ also contains features extracted from TFMP and the selected TFD.

4.4.4 Multiclass Support Vector Machine (SVM)

After feature extraction, prominent features need to be selected. In this study, features are selected according to a fitness function calculated from multiclass SVM.

SVM is a binary classifier that can be extended to a multiclass problem. The SVM classifier has been widely used in many applications including EEG classification as it generalizes well and is computationally stable [179]. One of the popular methods is one-against-rest (OAR). There are also two other methods for solving multiclass problem: one-against-one (OAO) and directed acyclic graph-SVM (DAG-SVM). The classification accuracy across these methods (OAR, OAO and DAG-SVM) is statistically insignificant [180]. With a 5-class problem OAR is the most efficient in terms of computation time and is the chosen approach for this chapter.

The OAR method constructs Q SVM models, where Q is the total number of classes. The p^{th} SVM is trained with all examples in the p^{th} class with positive labels, and all the other examples with negative labels as illustrated in

Figure 4.5. Thus given l training data $(x_1, y_1), \dots, (x_l, y_l)$, where $x_i \in R^n, i = 1, \dots, l$ and $y \in [1, \dots, Q] = [\text{seizure}, \text{burst}, \text{suppression}, \text{artifact}, \text{normal}]$ is the class of x_i, the p^{th} SVM solves the following problem [180]:

$$\min_{\boldsymbol{w}^p, b^p, \xi^p} \frac{1}{2}\|\boldsymbol{w}^p\|^2 + C\sum_{p=1}^{l}\xi_j^p$$
$$s.t.\begin{cases} \left((\boldsymbol{w}^p)^T\phi(\boldsymbol{x}_j) + b^p\right) \geq 1 - \xi_j^p, & \text{if } y = p \\ \left((\boldsymbol{w}^p)^T\phi(\boldsymbol{x}_j) + b^p\right) \geq -1 + \xi_j^p, & \text{if } y \neq p \\ \xi_j^p \geq 0, & j = 1, \dots, l \end{cases} \tag{4.6}$$

where the training data x_i is mapped to a higher-dimensional space by function ϕ and C is the regularisation parameter used to tune the classifier. When data are not linearly separable, there is a penalty term $C\sum_{p=1}^{l}\xi_j^p$ which can reduce the number of training errors. The basic concept behind SVM is to search for a balance between the regularisation term $\frac{1}{2}\|\boldsymbol{w}^p\|^2$ and the training error (the second term in Eqn. (4.6).

At the classification phase, a sample vector $\boldsymbol{v}$ is classified in the class i for which the decision function produces the largest value [180].

$$\text{Class of } i = \arg\max_{p=1,\dots,l}((\boldsymbol{w}^p)^T\phi(\boldsymbol{v}) + b^p) \tag{4.7}$$

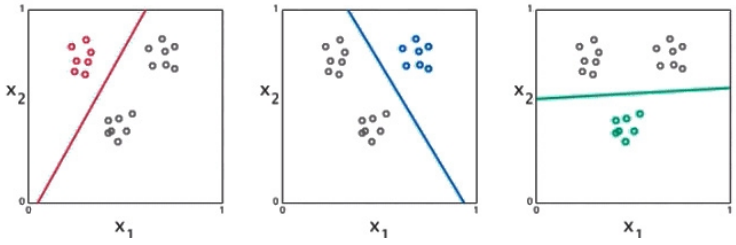


Figure 4.5: Multiclass Classification using the OAR method. A 3-class classification problem is used as an example. A 3-class problem provides 3 binary class problems using the OAR method.

In practice, the dual problem of Equation (4.6) is solved using quadratic programming. SVM can also be used in nonlinear classification tasks with the application of kernel functions. Any function that satisfies Mercer's theorem [181] can be used as a specific kernel function to compute a dot product in feature space. There are different kernel functions used in SVMs, such as linear, polynomial and

RBF (radial basis function). In this chapter, the RBF kernel is used due to its widespread application. In addition, the number of parameters that need to be optimised is small compared to other kernels (e.g. polynomial kernels). The values of $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ are between 0 and 1 and it can avoid numerical issues such as values of different polarities i.e., positive and negative values, higher numerical differences between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ [181]. It can be defined as

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right); \gamma > 0 \tag{4.8}$$

where $\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2$ is the squared Euclidean distance between the two feature vectors and $\gamma = \frac{1}{2\sigma^2}$ is a variance parameter that controls the kernel shape through the adjustment of the variance value. The selection of the appropriate kernel functions as well as kernel parameter(s) is very important, since the kernel defines the feature space in which the training examples will be classified. In the current context, (C, γ) parameters need to optimise for the optimal use of the RBF-SVM along with the feature selection.

4.4.4.1 Probabilistic estimation

In addition to binary decision making, posterior probabilities derived from the SVM distances provided by each of the classifiers in the multiclass SVM are also evaluated. Probabilistic outputs are generally more intuitive for clinical staff and have been utilised in previous EEG analysis systems [148]. The probabilistic information is useful in both assigning a class to an individual EEG segment and plotting a continuous probabilistic output of the EEG recordings for use in the clinical environment.

The output of the SVM classifier is computed for each epoch and this output is then transformed to posterior probabilities using a sigmoid function:

$$Pr(y = 1 | d_{hyp}) = \frac{1}{1 + exp(Ad_{hyp} + B)} \tag{4.9}$$

where d_{hyp} is the distance to the separating hyperplane [148], i.e. the output of the SVM classifier, A and B are the parameters of the sigmoid function estimated on the training dataset using the method presented in [182].

4.4.5 Classification performance measures

In the segment-based approach, classification performance is computed by comparing the detector outcome for a given segment to the neurologist marking of the same segment. The detector performance using segment-based classifications is evaluated in terms of:

$$Recall = Sensitivity = \frac{TP}{TP + FN} \times 100\%$$
$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (4.10)$$
$$Precision = PPV = \frac{TP}{TP + FP} \times 100\%$$
$$ACC = \frac{TP + TN}{(TP + FN + TN + FP)} \times 100\%$$

where TP = true positive, TN = true negative, FP = false positive and FN = false negative [183].

In the diagnostic accuracy test, generally background data (i.e. normal voltage) is more prevalent (negative class (N)) than other patterns such as burst and seizure (i.e. positive class (P)) and hence the dataset is unbalanced [184]. In this situation, specificity has a tendency to take values close to 1 [184]. With such an asymmetry, TN will generally take large values because of the size of negative class, i.e. $TP \leq P \ll N$. In this context, it is expected that $TN \gg FP$ and, thus, $Specificity = \frac{TN}{TN+FP} \rightarrow 1$ should more or less always be the case. Thus, specificity has little importance compared to $Precision$ [184], and hence is not included in this study. $Precision$ and $Recall$ should take precedence over specificity when evaluating the performance of a classifier on an unbalanced dataset. One common measure to quantify the classifier performance is $Fmeasure$ which can be defined as

$$Fmeasure, F_\beta = \frac{(1 + \beta^2)(Precision \times Recall)}{\beta^2.(Precision + Recall)} \quad (4.11)$$

where β is the weight parameter that controls weight between $Precision$ and $Recall$, is usually set to 1 in the $Fmeasure$ calculation to provide same importance to $Precision$ and $Recall$ [185] and hence used in the project. β = 2 and 0.5 can also be used. When β = 2, $Fmeasure$ weighs recall higher than precision (by placing more emphasis on false negatives); when β = 0.5, $F_{0.5}$ weighs recall lower than precision (by attenuating the influence of false negatives) [183]. For this reason, this measure is used in the fitness function. The advantage of $Fmeasure$ is that it takes both the $Precision$ and $Recall$ classifiers into account [184].

Furthermore, there is an equivalent relationship between statistical significant value such as p-value with diagnostic test statistics - sensitivity, specificity, disease prevalence and positive predictive value (see Appendix 4.C) [186]. Pharoah states that, "Test sensitivity is equivalent to statistical power. Test specificity is equivalent to the P-value. Disease prevalence in the population being tested is equivalent to the prior that the null hypothesis is false. What the clinician wants to know is the probability that the subject has the disease given that the test is positive. This is the positive predictive value and, in hypothesis testing, is equivalent to the probability that the alternative hypothesis is true given the

data. We are all comfortable with the fact that the positive predictive value depends on sensitivity, specificity and disease prevalence" [186].

Considering the above rationale, recall, precision, Fmeasure and accuracy have been used in this thesis. Of the total dataset, 50% has been used for testing and 50% for training. The justification behind this is to show the strength of the proposed algorithm in terms of classification accuracy. In addition, t-test (when comparing 2 groups) and analysis of variance (ANOVA) (when comparing more than 2 groups) have been performed to determine whether there is a statistically significant between results for the different methods/approaches. To accomplish this, as suggested by [187-189], 10-fold cross-validation test is performed first. 10-fold cross-validation test is performed first. It reduces the variation related to data selection and allows the results to be averaged to yield a robust calculation of the performance of the SVM [188]. In 10-fold cross-validation, the total feature set is divided into 10 mutually exclusive subsets or folds (i.e. folder size is 10% of the total feature set). The first fold is used for testing and the other nine folds (i.e. 90% of the total feature set) are used for the training process and repeated until each fold has been tested. The advantage of this method is that each data point appears exactly once in the test set (i.e. mutually exclusive). Therefore, it minimizes the bias associated with the random sampling of the training and test set. However, the cross-validation accuracy would be depended on the random assignment of data into a specific fold [189]. To solve this problem, a common practice is to stratify the folds themselves. Experimental studies show that, stratified cross-validation tends to generate comparison results with lower bias and lower variance when compared to regular k-fold cross-validation [189]. In stratified k-fold ($k = 10$ in this study) cross-validation, each fold is created in such a way that they contain approximately the same proportion of class/labels as the original dataset [187, 189]. The stratified 10-fold cross-validation (CV) has been used this chapter. The 10-fold stratified CV provides 10 results for each method/approach and, t-test and ANOVA have been performed on the stratified 10-fold cross-validation results to determine whether there is a statistically significant difference or not in different methods/approaches. Finally, a multi-comparison of mean accuracy of different methods/approaches is shown graphically to visualize the comparison.

4.4.6 Feature selection

Learning from such a high dimensional feature set leads to high computational load and decreases the classification accuracy due to irrelevant and redundant features. Manual selection of the most relevant features is time consuming and has difficulty determining any nonlinear relationship among the features. Furthermore, in machine learning and artificial intelligence it is known that a feature that

is completely irrelevant by itself can be very relevant when taken with other features and improves the classification performance [190]. Therefore, it is necessary to use a prominent feature selection method to (a) deal with the 'curse of dimensionality'[191], (b) improve model performance, (c) provide a faster and more cost-effective model [192], (d) gain a deeper insight into the underlying processes, and (e) select optimal tuning parameters of the classifiers simultaneously.

Feature selection methods can be structured into three clusters: filter-based, wrapper-based and embedded [193]. All have some major drawbacks and we therefore propose a hybrid feature selection (HFS) algorithm. First, we briefly describe the background problems and then formulate an objective function for the HFS algorithm.

4.4.6.1 Background problems

Filter-based approaches, such as t-test or correlation-based feature selection (CFS), maximum relevance and minimum redundancy, show lower computational complexity and computational time than wrapper methods but ignore interaction with the classifier, i.e. they do not take the classifier performance into account. Wrapper-based methods are search-based methods and can be sub-categorised into (i) deterministic search and (ii) randomised methods. Deterministic wrappers such as sequential selection and its variants, sequential forward selection (SFS), add or remove features sequentially. They are computationally intensive ('greedy search') and have a risk of overfitting [193]. On the other hand, randomised methods such as simulated annealing, genetic algorithm and differential evaluation incorporate randomness into their search procedure to escape local minima but still tend to suffer from overfitting [193]. The embedded approach is less computationally intensive than wrappers but provides a classifier-dependent selection. However, filter-based methods and deterministic wrapper-based methods cannot select the classifier tuning parameters at the same time as the feature selection. Simultaneous selection of an optimal feature set and optimal classifier parameters is important to increase the classification performance [193] and ensure the optimal use of a classifier.

Although this simultaneous selection is not related to feature selection method, in the present context simultaneous selection of an optimal feature set and optimal classifier parameters is ultimately needed as we have used RBF-SVM classifier which has two tuning parameters (C, γ) that need to be optimised simultaneously with the feature selection process. To solve this problem, we propose a hybrid feature selection algorithm which combines both filter and wrapper methods. It meets the general requirements of a good feature selector, improving the classification rate by (i) avoiding the

overfitting problem, (ii) simultaneously selecting an optimal feature set and classifier parameters, and (iii) having less computational cost than a genetic algorithm searching the full feature space.

4.4.6.2 Hybrid Feature Selection (HFS) algorithm

The HFS algorithm is comprised of maximum relevance and minimum redundancy (mRMR) [194] and a genetic algorithm. The mRMR method is computationally fast. Peng *et al*. demonstrated that mRMR method is a useful feature selection algorithm for wide-range of feature selection problems [194]. In addition, classification accuracy also increases when mRMR is applied prior to wrapper-based methods [195]. In the HFS algorithm, maximum relevance and minimum redundancy (mRMR) [194] is used to initially reduce the original features (let's say, the number of original features (FSo)), down to a more manageable z most discriminative features, and a genetic algorithm is then used to select the most prominent feature set from this reduced set. It also selects the optimal (C, γ) parameters simultaneously.

The first z best features, ranked by mRMR ($z < FSo$), are selected first. The mRMR ranks the features using mutual information; see Appendix 4.B for details. This process eliminates redundant features, reducing the risk of overfitting at the GA stage. The GA then simultaneously selects the features and the SVM parameters according to a fitness (or objective) function. As this feature selection and parameter optimisation process combines mRMR and GA, we denote this proposed approach the hybrid feature selection (HFS) algorithm. The flow diagram of the HFS is shown in Figure 4.6. Unlike the GA-based approach, the salient properties of HFS are that it (a) converges quickly as a smaller number of generations is needed than GA, and (b) reduces the risk of overfitting, as mRMR is used to filter the redundant features.

The performance of the classifier depends heavily on the fitness function. Here the fitness function maximises the classification rate, while weighting it against the feature set length. Mathematically the optimisation problem can be formulated as:

$$\max_{(SVM\ parameters, Feature\ subset)} (FitnessFun; SVM\ parameters, Feature\ subset) \qquad (4.12)$$

which can be interpreted as *'maximise the FitnessFun as a function of SVM parameters, Feature subset'*. The $Fmeasure$ is used as the fitness function i.e.

$$FitnessFun = Fmeasure \qquad (4.13)$$

The $Fmeasure$ is chosen as it shows more accurate classification performance indicator and widely used in different applications [196, 197]. In the present context, i.e. for RBF-SVM, the general equation of optimisation (Eq. (4.12)) can be written as:

$$\max_{(C,\gamma,Feature\ subset)} (FitnessFun; C, \gamma, Feature\ subset) \tag{4.14}$$

The following steps have been taken in order to precisely establish a HFS-based feature selection and parameter optimisation algorithm. Figure 4.6 represents the system architecture.

Step 1: Normalisation: The numerical range of one feature or attribute can be different from another feature and normalisation is used to avoid this dominance. It can be defined as:

$$F_{Normalized} = \frac{F - F_{\mu}}{F_{\sigma}} \tag{4.15}$$

where F and $F_{Normalized}$ are the original and normalised feature value and F_{μ}, F_{σ} are the mean and standard deviation of the feature value respectively. These F_{μ}, F_{σ} are computed from all the data available. This process will also reduce the equipment-related variations and physiological differences from one subject to another [198].

Step 2: Data partitioning: Data is split into training and testing sets (Figure 4.6 (a, b)). The main training set is further partitioned into a training set (50%) and a validation set (50%) by randomly splitting the main training set and repeating in every generation. This process reduces the bias towards a certain training set. The subdivided training set, i.e. second training set, is used to train the SVM and build a nonlinear model, while the validation set is used to evaluate the fitness function.

Step 3: Feature ranking: Feature ranking is done using the mRMR method. Ranking of the features is in descending order (from best to worst feature) and takes the first z higher ranked features.

Step 4: Genetic Algorithm: Selection, crossover and mutation are used to generate a population from (C, γ) and the highest ranked features from Step 3.

Step 5: Fitness evaluation: After genetic operation (C, γ) and features subset are determined and RBF-SVM model is built. This model is applied to classify the validation dataset to evaluate the fitness function.

Step 6: Termination criteria: Steps 4 and 5 are repeated, refining the model fitness. When the termination criteria, i.e. the number of generation, are satisfied the process ends.

Step 7: Classification: After acquiring optimal SVM parameters and an optimal feature set using the HFS algorithm, a SVM model is built and evaluated against the test set. Note that the testing set is not used in any part of the optimisation process, to ensure a fair judgement.

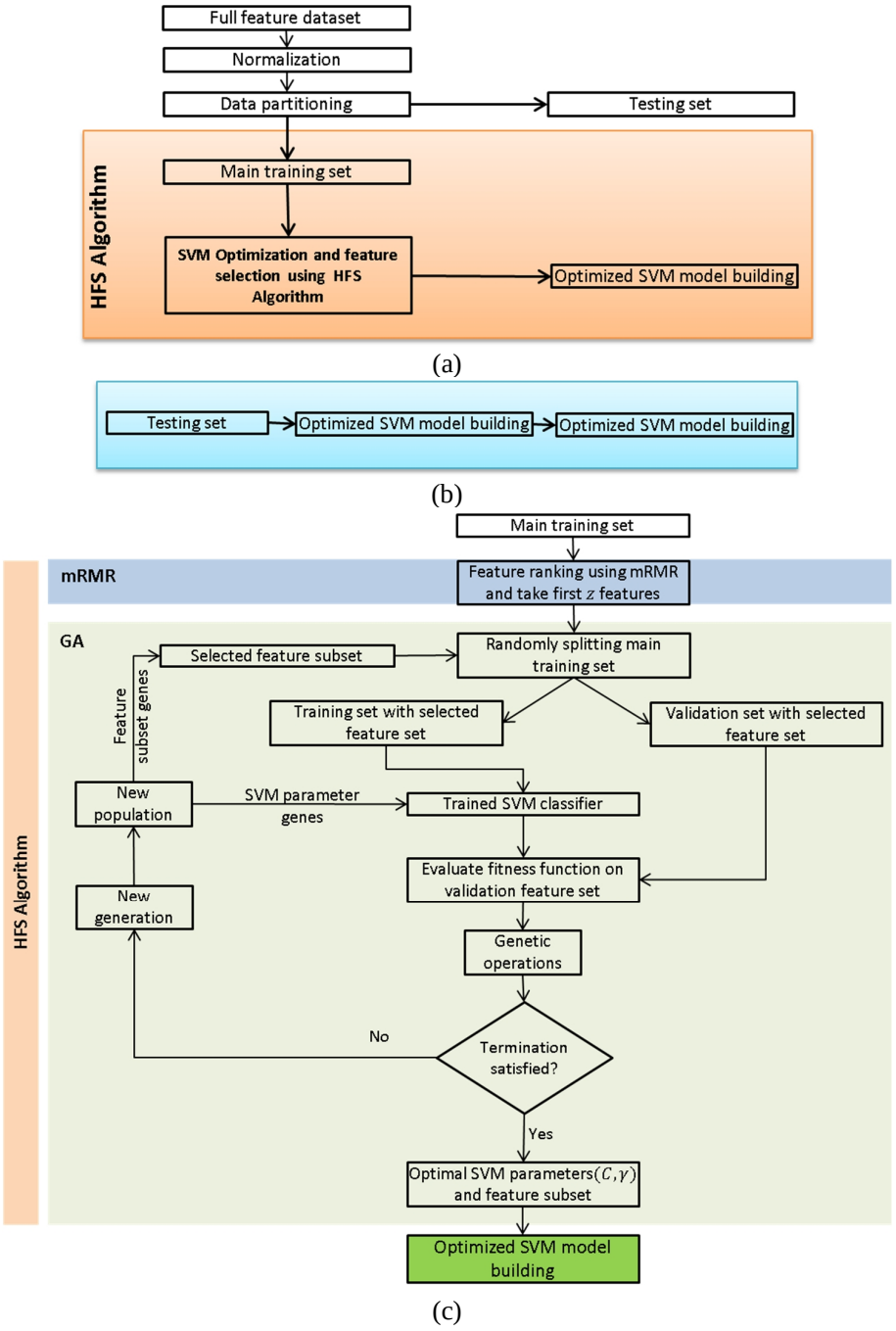


Figure 4.6: System architectures: (a) dataset and classification process, (b) testing of SVM model and (c) System architectures for the HFS algorithm.

4.4.7 Classification

As described above, the HFS algorithm generates a reduced subset of the feature vector, containing only the most discriminative features among the five classes. The most common approach for

classification is to design the classifier to select a single set of features. We will refer to this classification as classification using single feature subset. Pineda-Bautista *et al*. [199] have proposed extending this method by selecting a specific feature subset for each class. This is known as classification using *class-specific feature subsets;* see Figure 4.7.

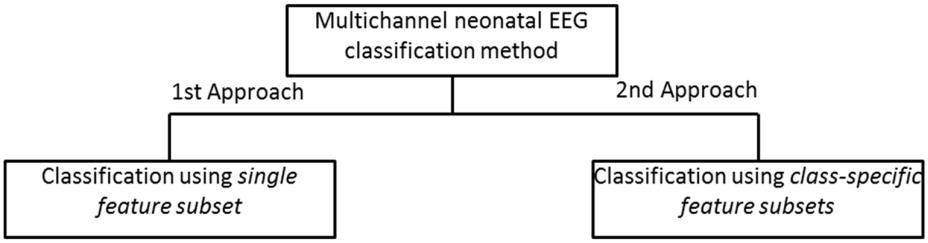


Figure 4.7: Neonatal EEG classification approaches.

4.4.7.1 Classification using class-specific feature subsets

Class-specific features have been shown to increase the classification performance in terms of accuracy, as the selected features are class specific [199]. Motivated by the general framework for class-dependent feature selection [199], this study introduces a class-specific feature selection procedure and parameter optimisation method for each class. Unlike [199], this method not only selects prominent feature sets specific to the individual classes, but also selects the optimal classifier (SVM) tuning parameters. However, as there are now unique classifiers for each class, a signal strength-based combining (SSC) algorithm is used to combine the classifiers and obtain final decision and posterior probabilities [148].

Classification using a class-specific feature subset consists of the following five steps:

Step 1: Class binarisation: Using OAR classification, a Q-class ($Cl_q;\ q = 1,2, \dots, Q$) classification problem is transformed into Q binary problems. For example, all instances of class 1 are used as positive class, whereas instances of rest of the classes $(Q - 1)$ are regarded as a negative class. This process is repeated to form Q binary problems.

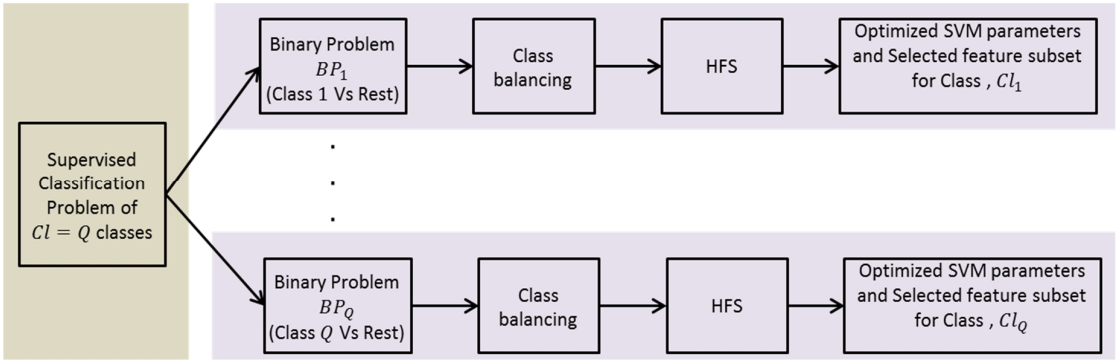
Step 2: Class Balancing: As OAR class binarisation strategy is used; the generated class could be imbalanced and consequently affect the feature selection and optimal SVM parameter selection. While there are several solutions [200], the ADASYN (ADAptive SYNthetic sampling) approach is used in this study due to its simple interpretation and widespread use [201]. It improves the machine learning with respect to the data distribution by (i) shifting the classification decision boundary adaptively toward the difficult instances, and (ii) reducing the bias introduced by the class imbalance. Note that, this balancing should only be applied to training and optimization process not in the testing set.

Step 3: Class-specific feature selection and parameter optimisation: After class balancing, the proposed HFS algorithm is used to select the prominent feature set specific to each individual class and optimise the SVM parameters; see Figure 4.8(a).

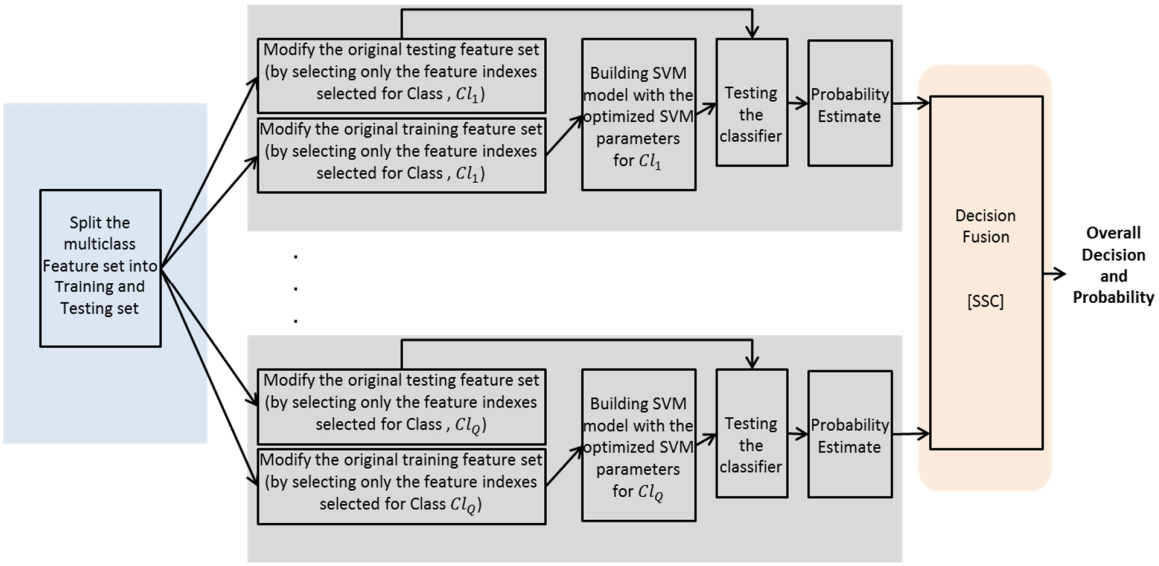
Step 4: Classification: In the classification stage, for each class $Cl_q; q = 1,2,\dots,Q$, a SVM classifier is trained for the original multiclass problem, using the optimal SVM parameters and feature indexes selected for *specific* class, Cl_q. In this way, a classifier ensemble $D = \{d_1, d_2, \dots, d_Q\}$ is formed. Finally, a decision fusion technique called SSC is used to obtain the final decision value; see Figure 4.8(b). The SSC algorithm also provides a final decision probability [202]. Section 5.3.1 provides more details about the implementation of final decision probability using SSC algorithms.

For example, the current classification problem is $Q = 5$-class problems, i.e. $Cl_q; q = 1,2,\dots,5 = [\text{seizure}, \text{burst}, \text{suppression}, \text{artifact}, \text{normal}]$. For each class, previously selected *class-specific* feature subset and SVM parameters (from Step-3 (Figure 4.8 (a)), are selected and generate a class-specific SVM model. Five class-specific SVM models provide 5 output decisions, i.e. $D = \{d_1, d_2, \dots, d_5\}$. Finally, the SSC algorithm is used to provide a single output from this decision vector, D.

Step 5: Testing: When a test set matrix T is provided, its original features must be reduced to the size of the feature's index vector selected for Cl_q used by the *class-specific* SVM; see Step 4. This algorithm not only gives a final decision value but also provides a final decision probability [202].



(a)



(b)

Figure 4.8: Class specific feature selection and multiclass classification. (a) Class binarisation, class balancing, class-specific feature selection and parameter optimisation; (b) classification.

4.5 Application to multichannel EEG background pattern classification

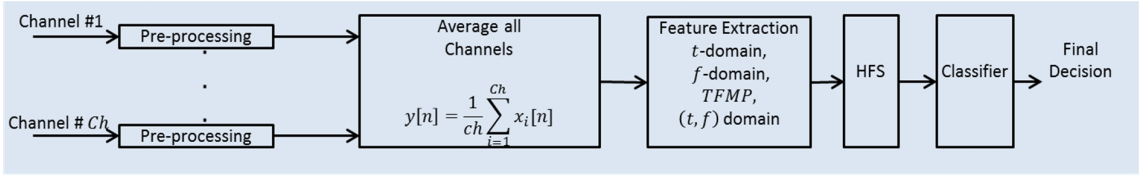
As the EEG recordings are multichannel, a sensible fusion of the channels is necessary [29]. There are three general stages where the channels may be fused with respect to multichannel EEG classification: (a) channel averaging or channel fusion, (b) feature pooling or feature fusion and (c) decision fusion. These three techniques will be utilised and compared in this study.

4.5.1 Channel fusion

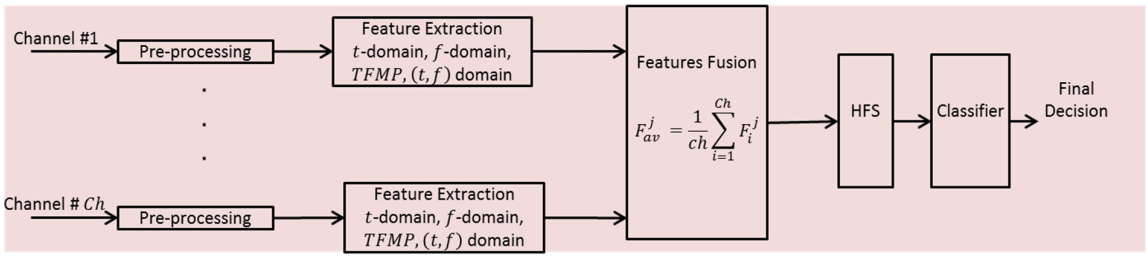
In channel fusion, multichannel EEG signals are averaged into a single average channel before extracting features (see Figure 4.9 (a)). System complexity and computational time are reduced as only a single average channel is used instead of all channels. Channel fusion can be expressed as:

$$y_{av}[n] = \frac{1}{ch}\sum_{i=1}^{Ch} x_i[n] \tag{4.16}$$

Where, $x_i[n]$ is a sample of the signal recording from the i^{th} channel, Ch is the total number channels and $y_{av}[n]$ is the averaged signal.



(a)



(b)

Figure 4.9: EEG fusion techniques: (a) channel Fusion and (b) feature Fusion.

4.5.2 Feature fusion

In feature fusion, the extracted features from different channels are fused. The simplest feature fusion technique is to average the individual channel features into single features. This feature fusion technique is used in this study and can be expressed as:

$$F_{av}^{j} = \frac{1}{ch}\sum_{i=1}^{Ch} F_i^j \tag{4.17}$$

where F_i^j denotes the j^{th} feature of the i^{th} channel and F_{av}^{j} is the j^{th}averaged feature across Ch channels; see Figure 4.9(b).

4.5.3 Decision fusion

Each channel in the multichannel EEG signal can be classified individually. This process produces a channel-level decision or probability vector. The decision vector from each channel can be computed uniquely and a decision fusion algorithm (SSC algorithm) used to compute an optimal final decision (see Figure 4.10).

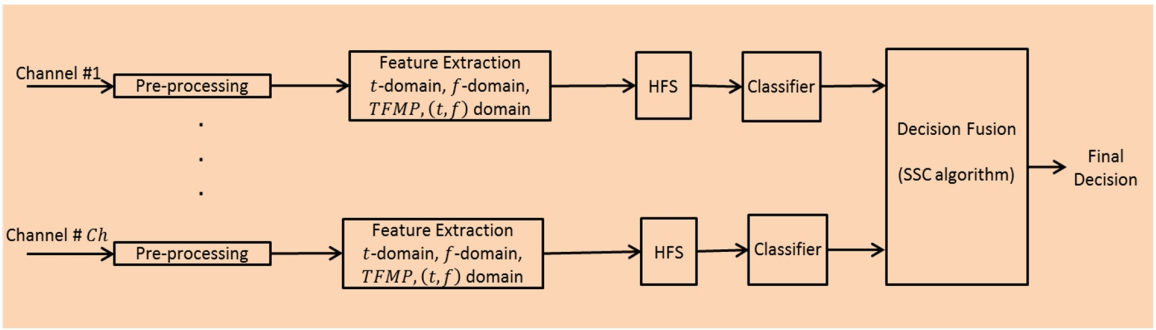


Figure 4.10: EEG fusion technique: decision fusion.

4.5.4 SSC algorithm

The SSC algorithm effectively integrates individual votes from all the unique classifiers computed for each EEG channel in an ensemble learning system.

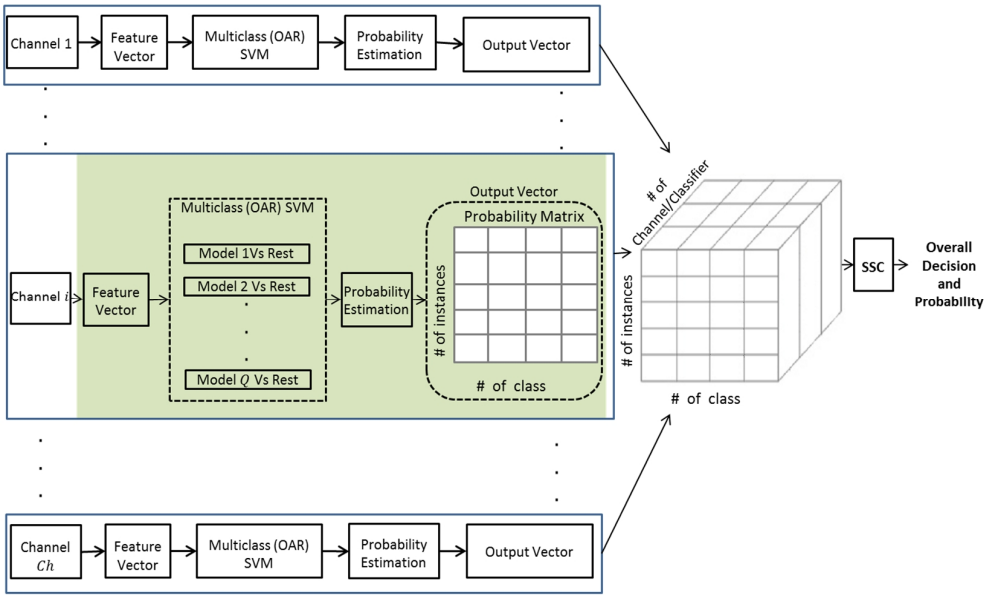


Figure 4.11: Multichannel EEG decision fusion using SSC algorithm. This process produces a probability matrix $U_i \in Pr^{T\times Q}, Pr \in [0,1]$, where U_i denotes the probability matrix of the i^{th} channel, T is the number of testing instances and Q is the number of classes or EEG patterns. A three-dimensional $(3D)$ decision probability matrix $Dp \in Pr^{T\times Q\times Ch}$ is formed by combing the probability matrix U_i of all channels. SSC algorithm is applied on this $3D$ matrix and overall decision and decision probability are found.

This algorithm takes the classifier's probabilistic output from different channels as input and provides the final combined output decision and corresponding final posterior probability. By following the methodology presented in [202] and Figure 4.11, the output probabilities of each classifier (or channel) are computed. Unlike the majority voting algorithm, this algorithm provides the output decision probability along with the overall decision label.

4.6 Results

The following settings were used in this experiment:

1. After pre-processing, the t-domain multichannel neonatal EEG segments were transformed in f-domain, (t, f) domain. A vector of $FSo = 134$ features was formed for all EEG channels, as well as for the averaged channel.
2. The total feature dataset was divided into two: 50% of the dataset was used as the main training dataset and the remaining 50% dataset was used as the testing dataset. The main training dataset was used for SVM model selection, parameter optimisation and optimal feature subset selection using HFS algorithm. The main training dataset was again randomly subdivided into training (50%) and validation (50%) sets and repeated in every generation. Note that the testing set was completely unseen by the classifier during model selection and

the same testing set was used everywhere for a fair comparison of different methods and fusion techniques.

3. Initially, $z = 20$ features were selected using mRMR approach and this sub-optimised feature set was then refined by the genetic algorithm in order to find the optimal SVM parameters as well as optimal feature subsets. This process eliminates the redundant and irrelevant features and cuts down the high dimensional original features, FSo, to a manageable feature set for GA. 20 generations with 25 populations in each generation are used in the GA. The other parameters are: crossover rate 0.7, mutation rate 0.02, two-point crossover, roulette wheel selection and elitism replacement.
4. GA can diverge from the feasible search space if the RBF-SVM tuning parameters (C, γ) are not bounded. The lower bound and upper bound for $C : 2^{-2}, 2^{12}$ and for $\gamma : 2^{-10}$ and 2^{4} respectively, were chosen. These larger ranges should be enough for the selection of optimal (C, γ) parameters [180].
5. The experiment is run with identical parameters and settings for both classification methods.

We present the performance of the classifier for the various configurations described above. First, we present the relative performance of the four TFDs: EMBD, LOS, CKD and SM (Section 4.6.1.1). We then analyse the performance gain provided by considering the TFD in combination with the t-domain an f-domain features (Section 4.6.1.2). This is performed on the simplest scenario, channel fusion and a single-feature subset.

Next, the performance of the single-feature subset classifier is compared with the use of class-specific feature subsets (Section 4.6.2). Then the performance of the more complex feature fusion and decision fusion techniques is presented (Section 4.6.2). Finally, a comparison with other classifiers will be presented (Section 4.6.3).

4.6.1 TFD performance

To analyse the TFD performance, we have performed two experiments to answer the following questions.

(a) What is the relative performance of the various TFDs?

(b) Are (t, f) features contributing to the classification accuracy when applied to neonatal EEG background features?

(c) How to visualise the relative classification performance in different domains classifying different patterns?

To answer the first question, the HFS algorithm was applied to the (t, f) domain features in isolation, and the relative performance of the four TFDs was evaluated. To answer the second question, three feature vectors (FV) were created:

$$\begin{aligned} FV_{t\ and\ f} &= [FV_t\ FV_f\ ] \\ FV_{(t,f)} &= [FV_{(t,f)}] \\ FV_{all} &= [FV_t\ FV_f\ FV_{(t,f)}\ ] \end{aligned} \tag{4.18}$$

where $FV_{t\ and\ f}$ consist of only t -and f -domain features, $FV_{(t,f)}$ only the features derived from the TFD's, and FV_{all} consists of all the t, f and (t, f) -features.

4.6.1.1 Choice of TFD

Table 4.3 presents the classification performance for the LOS TFD in terms of seizure, burst, suppression, normal and artifact. The last row of Table 4.3 shows the average performance measures of all classes (patterns). The other TFDs had been analysed in the same way, and Table 4.4 presents the average values for all four TFDs. Note that, *average classification performance measures*[5] are used in rest of the tables.

Table 4.3: Classification performance (in %) using single feature subset and channel fusion technique when LOS was used to extract (t, f) features.

Patterns	Precision	Recall	Fmeasure	Accuracy
Seizure	90.4	88.8	89.6	90.4
Burst	91.0	90.7	90.8	91.1
Suppression	89.9	89.5	89.7	89.9
Artifact	81.6	82.6	82.1	81.6
Normal	83.5	84.7	84.1	83.5
Average	**87.3**	**87.2**	**87.3**	**87.3**

[5] For example, average accuracy (e.g., see Table 4.4, last column) defines an average value of accuracy (in %) classifying all five patterns calculated by following Table 4.3. The other measures i.e. average precision, average recall, average Fmeasure can be defined in the same way. This is done to reduce the number of tables and to simplify the results.

Table 4.4: Overall average classification performance (%) using single feature subset and channel fusion technique when $FV = FV_{(t,f)_{TFD}}$.

HFS selected features when	Average precision	Average recall	Average Fmeasure	Average accuracy
$FV_{(t,f)_{EMBD}}$	81.7	80.4	80.8	81.0
$FV_{(t,f)_{LOS}}$	87.3	87.2	87.3	87.3
$FV_{(t,f)_{CKD}}$	87.3	86.9	87.1	87.1
$FV_{(t,f)_{SM}}$	84.5	83.7	84.0	84.2
Overall average	**85.2**	**84.6**	**84.8**	**84.9**

As discussed, the $Fmeasure$ is the most robust measure of performance for a classifier when it is operating on unbalanced data. LOS performed the best across all four measures, followed by CKD, then SM, and finally EMBD.

ANOVA test, applied on stratified 10-fold CV accuracy, was performed to detect any difference between four TFD (EMBD, LOS, CKD and SM)-based features. There was a significant difference ($p<0.001$). However, difference in classification accuracy using $FV_{(t,f)_{LOS}}$ and $FV_{(t,f)_{CKD}}$ did not provide a statistically significant result ($p=0.08$, t-test, 10-fold CV). Figure 4.12 shows how the accuracy varies among different TFD-based features.

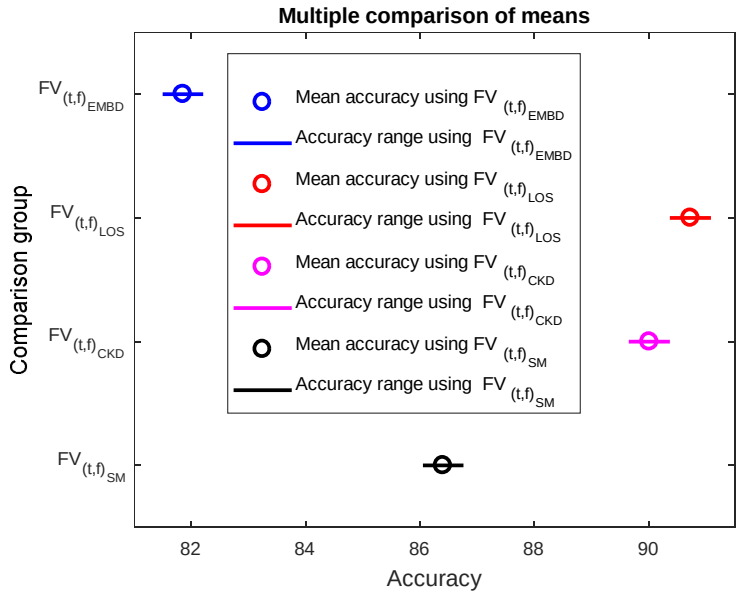


Figure 4.12: Multiple comparison of classification performance using four different TFD based features.

4.6.1.2 (t, f) domain vs t -and f-domain features

Table 4.5 presents the performance of the classifier when operating on $FV_{t\ and\ f}$, only on the $FV_{(t,f)}$, and FV_{all} features. The t- and f-domain feature vector operating in isolation obtains an average Fmeasure of 85.7% whereas (t,f) -domain feature vector provides an average Fmeasure of 84.8%. However, referring back to Table 4.4, the LOS TFD's average $Fmeasure$ is 87.3%, better than the result obtained by the t -and f -domain feature vector in isolation. More importantly, when considered together, the Fmeasure performance is improved to 91.6%, higher than the classification using only t- and f-domain feature vector ($FV_{t\ and\ f}$) and using only (t, f) domain feature vector ($FV_{(t,f)}$).

Table 4.5: Overall average performance (%) using single feature subset and Channel fusion techniques when HFS selects the features from $FV_{t\ and\ f}$, $FV_{(t,f)}$ and FV_{all}.

HFS selected features when	Average precision	Average recall	Average Fmeasure	Average accuracy
$FV_{t\ and\ f}$	86.4	85.4	85.7	86.0
$FV_{(t,f)_{TFD}}$	85.2	84.6	84.8	84.9
FV_{all}	91.8	91.4	91.6	91.7

ANOVA test, performed in stratified 10-fold CV accuracy, showed that the difference in classification accuracy among these three methods ($FV_{t\ and\ f}$, $FV_{(t,f)}$ and FV_{all}) was statistically significant (p <0.001). Difference in classification using $FV_{t\ and\ f}$ and FV_{all} as well as $FV_{(t,f)}$ and FV_{all} were also statistically different (p<0.001) but the classification using $FV_{t\ and\ f}$ and $FV_{(t,f)}$ was equivalent (p=0.50); see Figure 4.13.

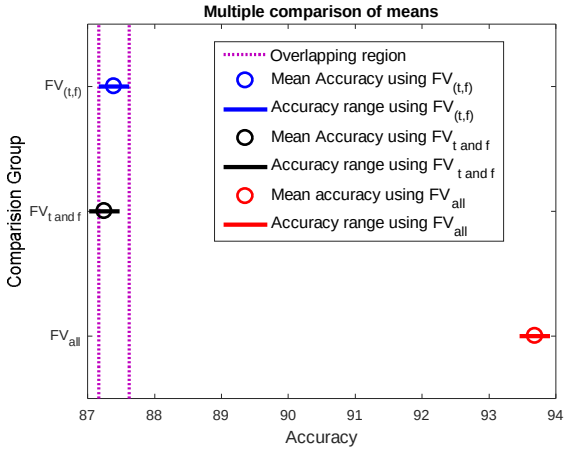


Figure 4.13: Multiple comparison of classification performance in three different feature vectors.

Table 4.6 presents the relative performance of the four TFDs. Interestingly, when combined with the t -and f -domain features, the worst performer in isolation, EMBD, obtained the highest average $Fmeasure$ of 92.6%. The last row represents the average classification performance for all TFDs. Although it is possible to show the individual classification performance with FV_{all} including each TFD, the 'overall average' result, comprised of FV_t , FV_f and all four TFDs, is calculated by following Table 4.6 and used in all other cases for a fair comparison. The overall average classification $Fmeasure$ and accuracy using channel fusion are 91.6% and 91.7% respectively.

Table 4.6: Overall average performance (%) using single feature subset and channel fusion techniques.

HFS selected features when	Average precision	Average recall	Average Fmeasure	Average accuracy
$FV = [FV_t\ FV_f\ \mathrm{FV}_{(t,f)_{\mathrm{EMBD}}}]$	92.6	92.4	92.5	92.6
$FV = [FV_t\ FV_f\ FV_{(t,f)_{\mathrm{LOS}}}]$	92.0	91.4	91.8	92.0
$FV = [FV_t\ FV_f\ FV_{(t,f)_{CKD}}]$	91.6	91.2	91.4	91.4
$FV = [FV_t\ FV_f\ FV_{(t,f)_{SM}}]$	91.0	90.6	90.8	90.9
Overall average	**91.8**	**91.4**	**91.6**	**91.7**

ANOVA test, performed in stratified 10-fold CV accuracy, showed that the difference in classification accuracy among these four groups was different ($p < 0.001$); see Figure 4.14. However, difference in classification using $FV = [FV_t\ FV_f\ \mathrm{FV}_{(t,f)_{\mathrm{LOS}}}]$ and $FV = [FV_t\ FV_f\ \mathrm{FV}_{(t,f)_{\mathrm{CKD}}}]$ was not statistically significant ($p=0.051$, t-test, 10-fold CV). Classification using $FV = [FV_t\ FV_f\ \mathrm{FV}_{(t,f)_{\mathrm{EMBD}}}]$ provided the highest classification accuracy among these four groups.

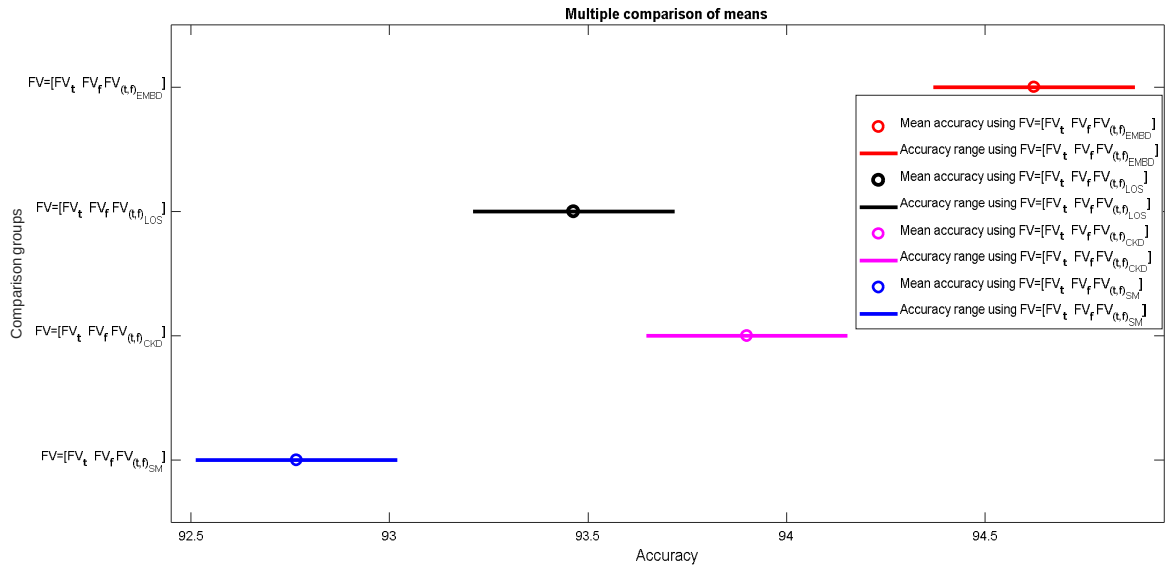


Figure 4.14: Multiple comparison of classification performance in four different feature vectors.

It is clear that the use of the full feature set (FV_{all}) consistently performs better than using either the t -and /or f-domain features ($FV_{t\ and\ f}$), or the (t, f) domain features ($FV_{(t,f)_{TFD}}$) in isolation. Same conclusion has also been found in recent studies, e.g., seizure detection and classification in newborns using (t, f) based features [29, 203]. They have shown that the (t, f) signal-based features generally yield improved performance when compared to the corresponding t-domain and f-domain features. Classification accuracy using (t, f) image-based features further improves the classification performance. Considering these, rest of the analysis is provided in terms of the FV_{all} result.

4.6.1.3 Confusion matrix plot

To visualise the classifier performance predicting different EEG patterns, the confusion matrix is also shown for different methods. The readers and the data analysers can easily visualise and interpret the overall performance of a classifier. The rows of the confusion matrix correspond to the predicted class (Output Class), and the columns corresponds the actual class (Target Class); see Figure 4.15. The diagonal elements show the correct classification information, whereas the non-diagonal elements show misclassification information and the bottom right element represents the overall accuracy. In this way, the confusion matrix for other fusion techniques can be plotted. It can be observed that the classification methods differentiate burst, suppression and seizure patterns better than normal and artifacts. This is because some artifacts look very similar to normal background and misclassification mainly occurs between these two classes.

To describe the confusion matrix plot, consider Figure 4.15(a). The first upper left diagonal cell shows the number and percentage of correctly classified seizure patterns. For example, 2971 EEG seizure segments are correctly classified as seizure, which corresponds to 16.7% of the testing dataset. 16, 11, 150 and 195 seizure segments are incorrectly classified as burst, suppression, artifact and normal segments respectively. Out of 3556 seizure predictions, 88.9% are correct and 11.1% are wrong. In this way, the performance of prediction for all other patterns can be interpreted. The last diagonal cell represents the overall performance: 86% of the predictions are correct and 14% are wrong classifications.

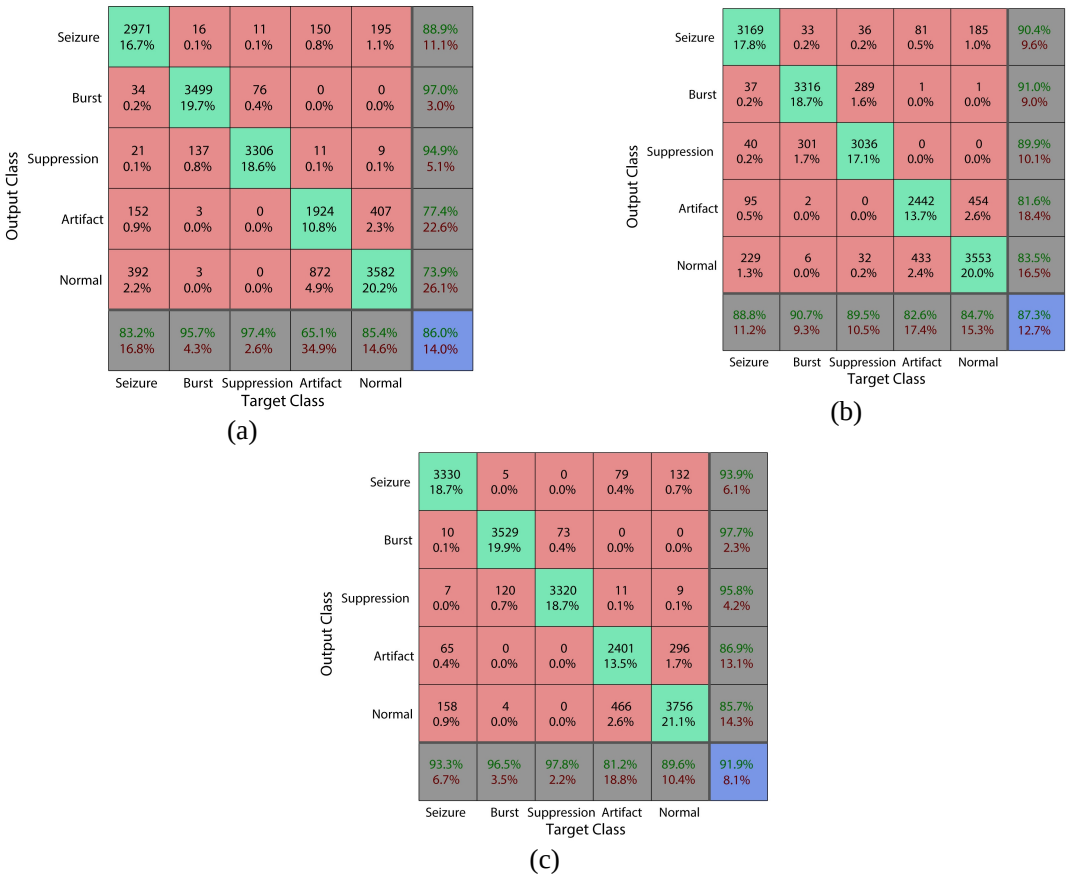


Figure 4.15: Confusion matrix plot when (a) the FV consists of t and f domain features (i.e. $FV = FV_{t\ and\ f}$); (b) when $FV = FV_{(t,f)}$ and (c) $FV = FV_{all}$ with LOS using single feature subset. In this figure, $\mathrm{FV}_{(t,f)}$ features are extracted from the LOS and channel fusion techniques are used. Confusion matrix for other TFDs can be plotted in the same way.

4.6.2 Classification using single vs class-specific feature subset

The performance of two classification methods – classification using (i) single feature subset and (ii) class-specific feature subset – using channel fusion are presented in Table 4.7. It can be seen that the average $Fmeasure$ of classification using single feature subset (91.6%) is slightly higher than classification using class-specific feature subset (90%). This is because channel fusion averages all the channels at the channel level which has significant additive, uncorrelated noise, and is assumed to be observing a singular source. This is not the case with EEG and, hence, decreases the performance.

Table 4.7: The overall average performance comparison (in %) between two methods.

Classification using	Fusion Technique	Average precision	Average recall	Average Fmeasure
Single feature subset	Channel Fusion	91.8	91.5	91.6
Class specific feature subset	Channel Fusion	90.7	89.6	90.0

Table 4.8 compares the overall average performance for both single feature subset and class-specific feature subset under different fusion techniques. Figure 4.16 provides a graphical representation of the performance of different fusion techniques. It can be observed that the overall average accuracy for channel fusion, feature fusion and decision fusion using single feature subset is 91.7%, 94.9% and 96.5% respectively and, using class-specific feature subset, 90.1%, 95.4% and 98.7% respectively (see Figure 4.16). Therefore, it can be concluded that classification using class-specific feature subset is more accurate than classification using single feature subset.

Table 4.8: The overall average performance comparison (in %) among the two methods and the three fusion approaches.

Feature selection method	Fusion Technique	Average precision	Average recall	Average Fmeasure
Single feature subset	Channel Fusion	91.8	91.5	91.6
	Feature Fusion	95.0	94.1	94.8
	Decision Fusion	96.8	96.2	96.4
Class specific feature subset	Channel Fusion	90.7	89.6	90.0
	Feature Fusion	95.7	95.1	95.3
	Decision Fusion	98.8	98.6	98.7

The improvement of feature fusion and decision fusion over channel fusion is likely due to the preservation of spatial information; channel fusion loses this information due to averaging. The best performance, with an average $Fmeasure$ of 98.7%, is obtained with decision fusion and class-specific feature subsets. This result is likely due to the added modelling power these two features provide the classifier.

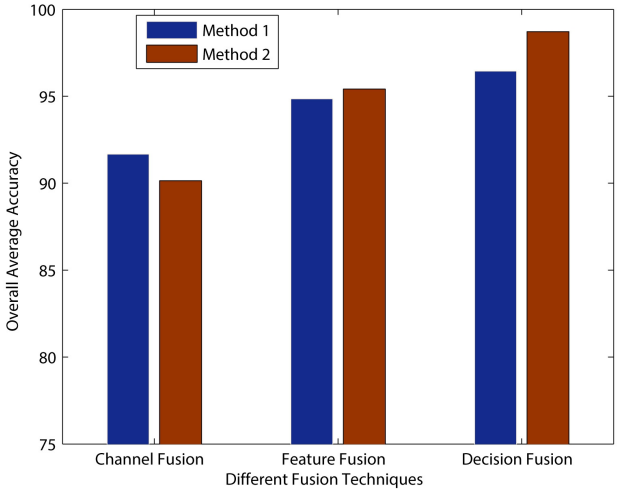


Figure 4.16: Effect of different fusion techniques.

4.6.3 10-fold cross-validation results

All results presented above are based on the fixed testing dataset. The testing dataset, comprising 17,771 epochs (50% of the total database), is independent, not being included in the training and validation process. This approach was used to show how strongly the optimised SVM classifier can classify the pattern with only 50% of the training data. However, the whole dataset cannot be utilised for testing by this approach. To address this, an N-fold cross-validation can be used to test the whole database. In the N-fold cross-validation, the i^{th} fold is used for testing and (N-1) folds (i.e., all except the i^{th} fold), are used for training and optimisation; this is repeated until all folds are tested. Finally, test results for each fold are averaged. This method 'reduces the variation related to data selection, and allows the results to be averaged to yield a robust calculation of the performance of the SVM' [188]. A 10-fold cross-validation[6] was also used, and the 10 results from 10 folds were then averaged to produce a single estimation. Table 4.9 shows the overall average 10-fold cross-validation performance among two feature selection methods and the two feature fusion approaches. This table was calculated by applying the same procedure producing Table 4.8 in 10-fold cross-validation case. As with previous results, it can be observed that the overall average 10-fold cross-validation accuracy for channel fusion and feature fusion using a single feature subset is 94.1% and 94.3% respectively and, using a class-specific feature subset, 95.0% and 97.1% respectively. In addition, compared to previous results as presented in Table 4.8, the 10-fold cross-validation performance is increased by more than 2%. This is due to the fact that a higher ratio (90% in this case) of the total data is used for training than testing (10%, in this case) in each fold. The method can classify difficult examples correctly and hence increases the classification rate.

Table 4.9: The overall average 10-fold cross-validation performance comparison (in %) among the two methods and the two fusion approaches.

Feature selection method	Fusion Technique	Average precision	Average recall	Average Fmeasure	Average Accuracy
Single feature subset	Channel Fusion	94.2	94.0	94.1	94.1
	Feature Fusion	94.5	94.1	94.2	94.3
Class specific feature subset	Channel Fusion	95.1	94.9	95.0	95.0
	Feature Fusion	97.2	97.1	97.1	97.1

[6] In 10-fold cross-validation, the total feature set is divided into 10 folds (i.e. folder size is 10% of the total feature set). The first fold is used for training and the other nine folds (i.e. 90% of the total feature set) are used for the training process and repeated until each fold has been tested.

4.6.4 Comparison with other classifiers

The results are compared with other well-known classifiers such as linear discriminant analysis (LDA)[7], naive Bayes, logistic regression, neural network, and Bayesian network. These classifiers are also used in neonatal EEG background patterns such as burst-suppression classification [162]. In all cases the overall average performance is generated following Table 4.6 and compared with other classifiers; see Table 4.10. The optimal parameters of these classifiers are chosen by an experimental search that provide the highest accuracy. The HFS selected features are still useful for these classifiers and used for a fair judgement. However, the classification performance of the other classifiers is lower than the RBF-SVM (see Table 4.5 and Table 4.10). These results suggest that the optimal RBF-SVM is more useful predicting different EEG background patterns. Therefore, this study uses optimal RBF-SVM in all other cases.

Table 4.10: Overall average performance (%) of different classifiers when $\mathrm{FV} = \mathrm{FV}_{\mathrm{all}}$ and fusion = channel fusion.

Classifier	Average precision	Average recall	Average Fmeasure	Average accuracy
Linear Discriminant Analysis (LDA)	74.2	72.9	71.3	72.2
Naive Bayes	66.3	67.8	62.5	65.7
Logistic Regression (Ridge 10^{-8})	74.1	73.1	73.3	73.8
Neural Network: (Learning rate 0.3, momentum 0.2) (Multilayer Perceptron)	82.8	79.0	79.2	80.7
Optimized SVM	91.8	91.5	91.6	92.6

4.7 Discussion

This study proposes a new methodology to select single and class-specific feature subsets and presents results from applying this methodology to multichannel neonatal EEG classification. This contribution involves the characterisation of neonatal EEG signal in t-domain, f-domain and (t, f) domain; simultaneous selection optimal features and classifier parameters using HFS algorithm.

This study also investigates various feature fusion techniques as well as single and class-specific feature selection. To the best of our knowledge, no study has attempted to discriminate patterns we have analysed. Reference [162] used burst-suppression patterns using spectral edge frequency (SEF95), 3 Hz power, median, variance and Shannon entropy and has shown an average area under the curve (AUC) of 94% using SVM on a database consisting of six term infants. Reference [176] used burst and normal patterns in the presence of artifact using wavelet-based statistics, f -domain

[7] LDA may not be a relevant or optimal classifier for a nonlinear problem. This classifier is chosen for to show that linear classifier like LDA is not suitable for the present complex nonlinear problem.

and nonlinear features and has shown 78% sensitivity on a database consisting of sixteen term newborns.

Several studies collapse the patterns and grade the patterns as mild, moderate or severe abnormal background trace using voltage amplitude definition. For example, Matic *et al*. [204] use a tensor-based classification method to grade background EEG abnormalities and achieve 89% accuracy on a database consisting of 34 term newborns. Ahmed *et al*. use a supervector approach and achieve 87% accuracy to grade EEG on a database consists of 54 term newborns [148].

The performance of the proposed methodology compares extremely well with contemporary techniques. The methodology achieves an accuracy of 98.7% for the second method (class-specific feature subset) when decision fusion is used, significantly improving on past works. We posit that part of this improvement is due to the inclusion of (t, f) features in classification, which is not present in the prior studies mentioned above. We show a significant improvement can be made, from approximately 85% to 90% accuracy, when (t, f)-features are combined with t-domain and f-domain features. Additionally, some studies do not use feature selection and the redundant features could be causing overfitting and degrading classification accuracy. The proposed HFS algorithm not only selects prominent features but also selects the optimal classifier parameters simultaneously. This provides a substantial advantage over alternative feature selection algorithms in terms of classification accuracy and computational complexity.

The prominent features selected by the various configurations of fusion techniques and classification methods are different. Describing every feature that appears would be excessive. The most consistent features that appear across the various configurations will be discussed. They are: Higuichi fractal dimension, Renyi entropy, approximate entropy, sample entropy, Hurst exponential, SVD entropy; number of atom and mean and variance of TFMP modulation; statistics of band specific (t, f) features, ECM, aspect ratio, and GLCM Information measure of correlation.

Neonatal EEG background, as well as seizure signal, can be characterised as a nonlinear dynamical system [163] and different nonlinear dynamics such as Higuichi fractal dimension, Renyi entropy, Hurst Exponential and SVD entropy measure the change of brain signal complexity. Burst, normal and seizure states often exhibit significantly more complex nonlinear patterns when compared to artifact and suppressed states, explaining the prominence of features quantifying these traits. These features have been consistently used in various studies [148, 205] to characterise neonatal EEG and prediction of neurodevelopmental outcome.

(t, f) features focused on a specific band are regularly included, particularly the delta band. It can be seen from Figure 4.4 that a large proportion of EEG power lies in this band, and is noted as a more dominant pattern in infants, reducing in prominence with maturity [205]. It follows that these features would be useful in discriminating normal and abnormal patterns. The (t, f)-based image processing features, such as aspect ratio and GLCM Information measure of correlation, are also seen to be good candidate features, possibly because they characterise the TFD signatures of different EEG patterns. TFMP also provides promising results in detecting seizure [171]. Finally, the statistics computed from the modulation of Gabor atoms and the number of atoms is regularly seen.

Selection of SVM tuning parameters is also important. In this study, we have shown how one can simultaneously select the optimal features as well as optimal classifier parameters, enhancing the classification accuracy using an HFS algorithm. Although RBF-SVM is optimised in this chapter, the general framework provided in Equation (4.12) can be used to optimise other SVM kernel parameters.

To summarise the content of this chapter:

- The channel fusion technique is computationally efficient when compared to other fusion techniques but decreases the classification performance, as useful information, particularly spatial information, is lost.
- Feature fusion is more computationally demanding than channel fusion. This is because the technique must extract features for every channel before averaging. However, this increases the classification accuracy.
- The decision fusion technique is significantly more computationally demanding than the other two techniques, as it not only extracts all the features but must search the much larger feature space occupied by all the independent channels. However, it also shows the highest classification accuracy, improving on prior works.
- Classification using class-specific feature subset is more accurate than classification using single feature subset. Interestingly, the computation time for the both methods is the same. The reason for this is that classification using single feature subset utilises all the EEG patterns simultaneously, whereas classification using class-specific feature subset utilises only two patterns at a time in a one-against-rest iterative fashion (see Figure 4.8).

Feature extraction techniques also play a key role, as a good candidate feature choice enhances classification accuracy. This study shows the usefulness of (t, f) features when they are combined with t -and f -domain features and it is therefore beneficial to consider (t, f) features when performing neonatal EEG classification.

One of the disadvantages of the proposed classification method is computational load. This is due to the multichannel nature of EEG and effort required to build an optimal SVM model. Once the model has been built the classification task is significantly faster. The number of features to be computed is significantly reduced, and the task of classification for an SVM is computationally efficient. On a modern PC, a 4-s segment of recording could be analysed in less than a second; consequently, real-time classification is possible once a model has been obtained. Further strengths and weaknesses have been discussed in section 4.7.1.

This proposed system classifies neonatal EEG into the five proposed patterns with improved performance over contemporary work. These patterns are frequently observed in term neonates with HIE and are also associated with a range of adverse neurodevelopmental outcomes. In addition, the duration and occurrence of certain patterns are also very important in clinical settings, as they are predictive of neurodevelopmental outcome: for example, burst density, defined as the number of min spent in the burst state per hour. Also, Menache *et al*. [86] observed that neonates exhibiting an inter-burst interval (IBI) (length of suppression between two burst patterns) duration > 30 s have 100% probability of experiencing severe neurologic disabilities or death and an 86% chance of developing epilepsy. The output from the classifier can be used to calculate both these indicators. However, direct prediction of neurodevelopmental outcome is beyond the scope of this chapter. To produce a robust classifier of abnormal vs normal patterns, the age of the subject must also be taken into account. Certain patterns that are quite normal in preterm neonates are considered abnormal in term neonates. For example, the maximum normal duration of IBIs allowable at 26–27 weeks gestational age (GA) is about 9 s, versus about 3 s in 38–40 weeks GA [23].

4.7.1 Strengths and weaknesses

- Features selected by GA-based feature selection could be overfitting in the case of high dimensional feature space. In contrast, HFS algorithm minimises the risk of overfitting because the HFS algorithm comprises maximum relevance and minimum redundancy (mRMR) and a genetic algorithm.
- Most of the studies select the features first and then optimise the SVM tuning parameters using grid search algorithm. However, this search algorithm performs poorly. In addition, extra time is needed as the feature selection and parameter optimisation are done separately. In contrast, the HFS algorithm simultaneously selects the SVM tuning parameters and optimal feature set. It reduces computation time and complexity compared to separate optimisation.

- The (t, f)-features used in this chapter show that HFS can improve the classification accuracy when combined t- and f-features. This demonstrates that (t, f)-features take the non-stationary property of real EEG signals and improve the classification accuracy.
- This chapter also discusses a wide range of options or strategies for the users classifying multi-channel signals, e.g., multi-channel EEG signal classification: from channel fusion to decision fusion, from single feature selection to class-specific feature selection. There is a trade-off between computational complexity and accuracy. Users can select the option, presented in this chapter, according to their need and available hardware resources, HPC facilities etc.

Weaknesses:

- Although the HFS algorithm requires less time than a GA-based approach, the computational cost is higher than filter-based approaches as it requires more time in the training and optimisation process, especially in the high-dimensional features space. However, once the optimised SVM is built, the testing process is less complex and faster as fewer features are generally selected than in the original feature set during the training and optimisation process.
- The results are based on 50% of the total data and on 10-fold cross-validation. Validation as well as calculation of p-value on a separate database from different babies would be useful. This limitation is largely due to data collection constraints, and especially the marking of data by a neurologist, which is a complex and time-consuming procedure.

4.8 Applications of the proposed method

A decision support system (DSS) can be built to support clinical staff. We have created a program that visually presents the probable patient state in terms of posterior probabilities. The RBF-SVM can provide output probabilities of different patterns which can be used to identify different patterns. For example, Figure 4.17 builds a probabilistic output of 60 epochs consisting of seizure, burst and suppression. The corresponding actual output marked by the neurologist is also shown in Figure 4.17 to visualise the classification probability. The classification probability > 0.5 indicates the presence of a background pattern.

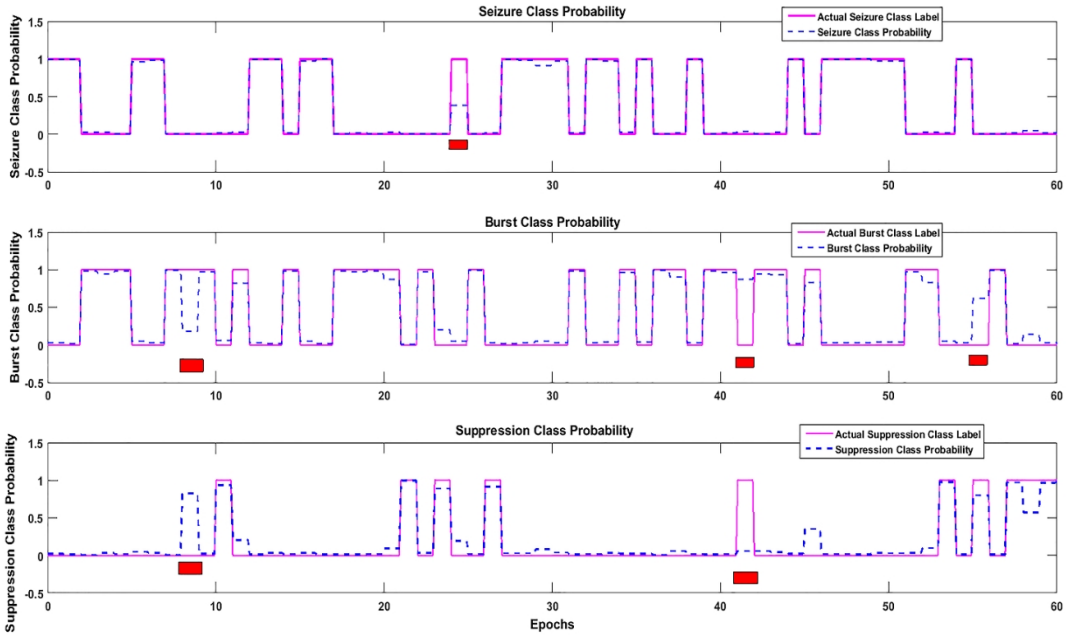


Figure 4.17: Probabilistic output for the DSS. The '1's in the solid line represent the presence of the actual patterns and '0's the absence of that pattern. The dotted line represents the probabilistic output between 0 and 1. The red marks represent the misclassification segments. Note the individual segments are randomly selected from the source dataset for illustration; they do not represent a continuous recording.

Like seizure probability plot, a probabilistic plot for artifact class can also be created. The probability of artifact close to ≈ 1 can be treated as highly contaminated (artifact) EEG, whereas probability ≈ 0 can be treated as artifact-free EEG. This probabilistic plot, i.e. 'EEG quality' plot, can be used for continuous quality monitoring of EEG (see Figure 4.18).

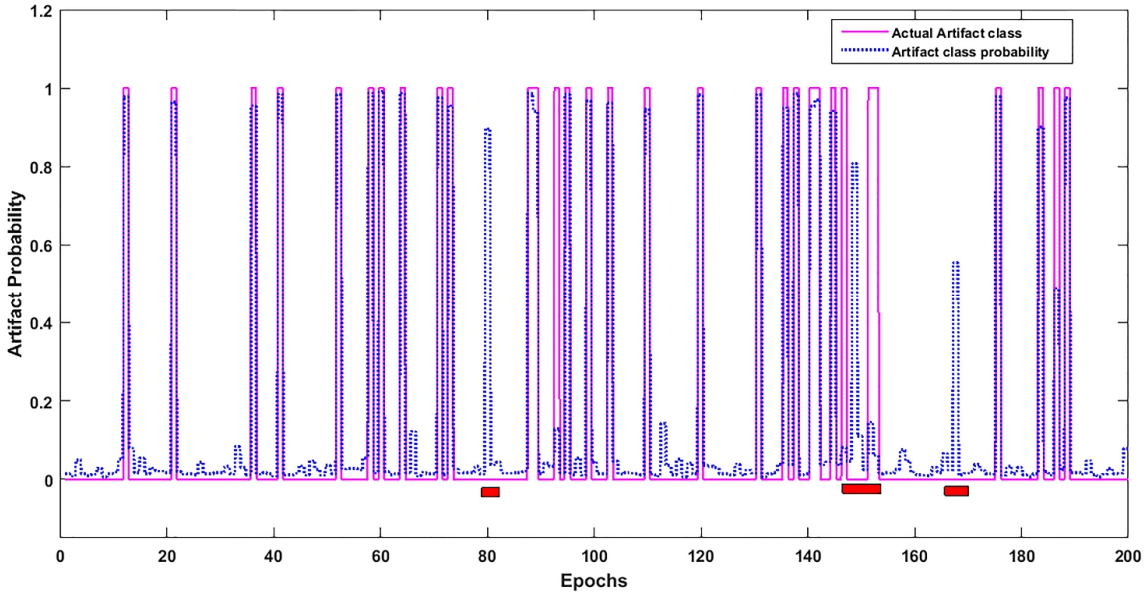


Figure 4.18:'EEG quality' plot.

4.9 Overall summary

In this chapter, multichannel neonatal EEG background patterns such as burst, suppression and normal patterns in the presence of seizure and artifact are classified using six different possible approaches; two classification methods (classification using single and class-specific feature subset) and three fusion techniques (channel fusion, feature fusion and decision fusion). We show how different approaches and their combinations enhance classification accuracy. In addition, we have shown the importance of (t, f)-features when they are combined with the t -domain and f -domain features. For example, a feature vector containing only t -domain and f -domain n features (i. e. $FV = FV_{t\ and\ f}$) using the channel fusion technique achieves an overall average accuracy of 86% in this study, which is increased to an overall average accuracy of 91.7% when features from all domains are taken into account.

Feature fusion and decision fusion can increase the classification accuracy to 94.8% and 96.4% respectively. Furthermore, classification using class-specific feature subset improves decision fusion's overall average accuracy to 98.7% when compared to a single-feature subset. The system could provide a significant reduction in workload for physicians or allow for monitoring in situations where a specialist is not available. Translation of these findings, especially the DSS or 'EEG quality' plot, into an EEG monitoring system would be very useful for end users. The increases in classification accuracy also increase the computational complexity, particularly in the learning stage. There is a trade-off between computational resources and accuracy among different approaches and users can choose the approach according to their needs and available computational resources. Furthermore, the HFS simultaneously selects the optimal feature subset as well as optimal classifier parameters that increase the classification accuracy. Finally, the feature extraction routine, HFS algorithm and the different fusion techniques and classification methods presented in this study are applicable to many other classification problems such as brain-computer interface, machine condition monitoring and fault detection.

Chapter 5 Testing of EEG signal features that best predict neurodevelopmental outcome in term neonates with HIE

5.1 Introduction

What is already known on this topic?

Different features extracted from EEG are associated with adverse neurodevelopmental outcomes in term neonates with HIE.

What is the contribution of this study?

- ✓ A prominent feature set, extracted from EEG close to 24h, has been selected by a HFS algorithm and validated using a LOSO cross-validation method.
- ✓ An optimised SVM model has been built to test the model on a test database.
- ✓ A decision support system has been created for continuous monitoring of EEG to predict neurodevelopmental outcome.

Neonatal HIE is a major cause of morbidity and mortality in newborns [148, 149]. With the advent of potential neuroprotective and neuro-restorative therapies, early and accurate diagnosis of HIE has become increasingly important. At the same time, reliable prognostic information is essential for counselling parents and caregivers [205].

To date, several different types of monitoring have been studied for outcome prediction. Laptook *et al.* [206] and Mirsa *et al.* [207] reported that Apgar score, based on simple clinical characteristics in the first 5 min after birth, provides useful prognostic information. However, the American Academy of Paediatrics and the American College of Obstetrics and Gynaecology have drawn attention to the fact that it should not be used alone to predict outcome [208]. The meta-analysis of biochemical markers of injury severity or of outcome, in serum, urine and/or cerebrospinal fluid in neonates with encephalopathy, by Ramaswamy *et al.* [209] concluded that no biomarker had yet been sufficiently validated to warrant routine clinical use.

Recently, Alderliesten *et al.* correlated MRI-apparent diffusion coefficient (ADC) of the basal ganglia and thalamus, lactate/N-acetylaspartate (LAC/NAA) and N-acetylaspartate/choline (NAA/Cho) ratios extracted from diffusion-weighted MRI (DW-MRI) and proton magnetic resonance spectroscopy (H-MRS) with neurodevelopmental outcome [210]. The optimal timing of an MRI examination for prognosis of outcome in newborns with HIE is the second week of life and therefore its use for early prognostication may be limited [205].

On the other hand, EEG has been shown in many studies to be a robust predictor of neurodevelopmental outcome. Conventional and amplitude-integrated EEG both perform well in the prediction of outcome [44] and in the early diagnosis and classification of HIE severity [12]. The benefits of EEG are that it can easily be implemented at the cot-side soon after birth, and it provides a real-time measure of cerebral function and its dynamic changes across time [12].

A normal EEG is highly predictive of a normal outcome, and various abnormal EEG features have been consistently associated with neurological adverse outcomes or death [47]. For evidence, Shany *et al.* investigated 39 newborns with burst-suppression (BS) and found that 8 died within 3 years [15]. Roij *et al.* [16] showed that 60 of 160 newborns with flat trace (FT) or continuous low voltage (CLV) died or had major handicap. Horst *et al.* [17] studied 30 newborns with CLV patterns and found that 6 died or had major handicap, while Douglass *et. al.* [18] assessed 22 newborns with only BS and found that 16 died or had major handicap. These abnormal outcomes are broadly supported by findings reported by others (see Chapter 2).

A few studies, not included in the meta-analysis because of insufficient papers, have related EEG frequency, power, coherence [84, 106, 211-213], complexity and entropy [214] to outcome. Scher *et al.* related power analysis and spectral correlation with neurodevelopmental outcome in both preterm and term infants [213], especially lower power spectral measures in the higher frequency band during sleep. Quentin *et al.* correlated higher approximate entropy (ApEn) with lower burst suppression ratio (BSR) and low ApEn correlates with poor outcome [214]. Iyer *et al.* used different burst dynamics such as area under burst, scaling slope, burst symmetry and burst sharpness in preterm infants [215]. Like burst dynamics, Dereymaeker *et al.* proposed 'suppression curve', a measure of discontinuity of preterm EEG [216]. However, in the case of term infants, not only burst-suppression patterns but other abnormal patterns exist. Hence, building a model using only those features may be very limited for term infants.

Most of the studies using the more sophisticated analytic methods mentioned above assume EEG is a stationary signal. However, EEG is a non-stationary signal, i.e. statistical properties of the frequency and amplitude content change over time, resulting from the relatively random firing of neurons. Joint time-frequency (t, f) techniques are well suited to analysis of non-stationary signals. To the best of our knowledge, JTF-based features have not been explored in relation to predicting neurodevelopmental outcome. Previous studies suggest that dynamic features extracted from JTF show outstanding performance in detecting newborn seizures [29] and automatic grading of EEG background patterns [155]. We propose that JTF-based dynamic features may also be useful in predicting outcome. In addition, most of the cited works were limited to linear methods that consider

a single feature at a time. A nonlinear and complex relationship between these predictors has not been explored, and in fact may improve accuracy considerably, with each parameter providing complementary information.

In this study, various temporal, spatial and (t, f) based dynamic features are extracted and selected to build a nonlinear support vector machine (SVM) model to predict the neurodevelopmental outcome using EEG collected during the first 24 h after birth. This proposed method provides a promising extension of the field of neonatal neurophysiology with this decision support system.

The major aspects and key contributions of this study are as follows:

- First, a novel pre-processing step is used to remove the unwanted noise and artifacts. This process uses a robust artifact removal (RAR) method (Section 5.3.1).
- Various time domain, frequency domain and (t, f) -domain features are extracted from the neonatal EEG recorded within the first 24 h after birth to seek associations with the neurodevelopmental outcome in term neonates with HIE (Section 5.3.2).
- A novel hybrid feature selection algorithm (HFS) is used to select a prominent feature set and optimise the SVM tuning parameters. This HFS algorithm is used in the statistically least-biased leave-one-subject-out (LOSO) cross-validation process to ascertain the most consistent feature set (Section 5.3.4).
- A nonlinear SVM model is built based on the most consistent feature subset found by the LOSO system, and a neonatal EEG database is used to test the predictive capability of the model (Section 5.4.2).
- Finally, a probabilistic decision support system is derived as a potential application of the proposed system (Section 5.5).

5.2 Materials

Two datasets of term infants with HIE have been used: (i) the BEES study dataset (collected at the Royal Brisbane and Women's Hospital (RBWH) in a prospective, longitudinal, observational cohort study, with both continuous 2-channel and shorter recordings of multichannel EEG), and (ii) the NEST study dataset (collected from RBWH and other NICUs in a multicentre RCT, with continuous 2-channel EEG).

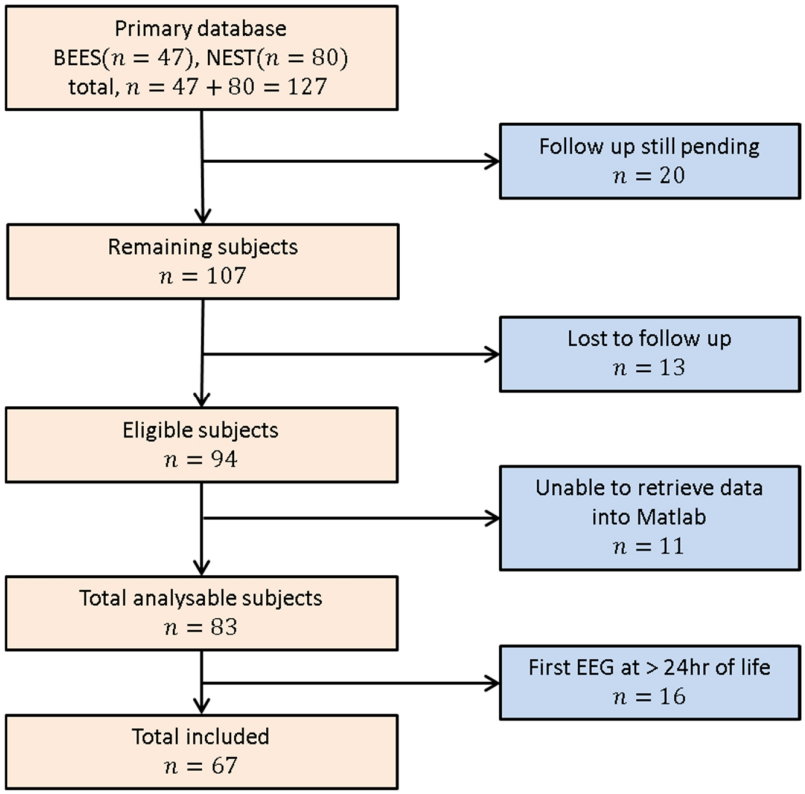


Figure 5.1: Diagram of subject numbers as included in the study.

Table 5.1: Demographics of the term data set with EEG and clinical outcome information

Information	Value	Range (Min−max)	Standard deviation
Clinical details			
Subjects, n	67		
Sex, male/female, n	38/29		
Gestational age, Weeks (mean)	39.3	34 – 42	1.89
5-min Apgar score (mean)	4	0 – 9	2.54
EEG information			
Recording duration (close to 24 hours after birth)	2 h		
Neurodevelopmental outcome (BSITD)			
Cognitive (composite score) (mean)	84	54 – 154	26.2
Language (Composite score) (mean)	81	47 – 150	25.9
Motor (Composite score) (mean)	82	46 – 127	25.1
CB-III (Composite score) (mean)	83	51 – 147	25.7
D1, good/poor, n	42/25		
D2, optimal/suboptimal, n	40/27		

Parental written informed consent was obtained for all subjects. The study had ethical approval from the human research ethics committees at the Royal Brisbane and Women's Hospital, Brisbane, Australia (BEES) and the Royal Children's Hospital, Melbourne, Australia (NEST). This trial is registered with the Australian New Zealand Clinical Trials Registry (ANZCTR), number ACTRN12611000327987. Subject names were encrypted before analysis.

As mentioned in Chapter 1, HIE is defined as the presence of two of the following three features in the presence of a known perinatal event [3]: need for respiratory support shortly after birth, an Apgar score at 5 min of < 5, or evidence of acidosis $(\mathrm{pH} < 7)$ [3]. A total of 127 term infants were available and 67 infants were included in the analysis (see Table 5.1). EEG data from 40 babies were randomly selected from the 67 infants for feature selection, parameter optimisation and finally for optimised model generation. The remaining 27 infants' EEGs were used to test and validate the model.

5.2.1 aEEG/ EEG monitoring

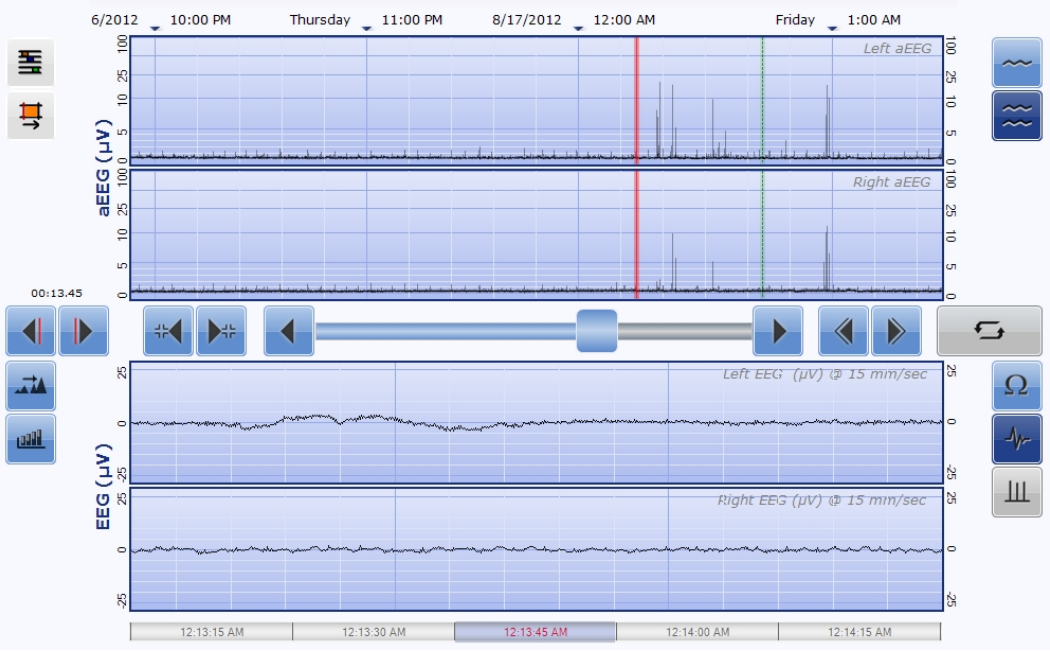


Figure 5.2: The 2-channel aEEG/EEG screen of a subject with abnormal outcome. The top 2 rows show the 3-h condensed amplitude integrated EEG (aEEG) background for the left and right channels. The bottom two rows are the 10-s windows of EEG tracing over the central parietal regions of the left and right hemispheres. This aEEG/EEG pattern shows a flat trace pattern over both hemispheres.

All infants were monitored with a 2-channel EEG ($C3-P3$ and $C4-P4$) using hydrogel electrodes with a BRM 2/3 or OBM monitor (BrainZ Instruments, Natus Medical, San Carlos, California, USA) [3, 39]. Figure 5.2 illustrates EEG recorded using an OBM monitor, which samples with a frequency of $200Hz$. EEG data were then exported to Matlab for pre-processing and all other analyses.

5.2.2 Neurodevelopmental outcome

A standardised neurodevelopmental test at 24 months corrected age with the Bayley scales of infant development, versions II and III (BSID-II and BSITD-III), was performed by an independent psychologist blind to the clinical history. BSID-II provides Mental Developmental Index (MDI) and Psychomotor Developmental Index (PDI) sub-scores, whereas BSITD-III splits the original MDI scores into cognitive, receptive language, and expressive language scores, and the PDI into fine and gross motor scores. BSITD-III also provides a social-emotional score and an adaptive behaviour score. BSID-II tests were performed on 9 babies and all other cases were tested using BSITD-III. BSID-II scores have been converted to BSID-III using a published recommended approach [217, 218] (see Table 5.2) to unify the analysis. In addition, a combined BSITD-III score (CB-III) was calculated by averaging BSITD-III's cognitive and Language scores [31]. Newborns unable to be tested with the BSITD-III, due to severe cerebral palsy (CP), severe developmental delay or death prior to testing, were assigned a score of 54 for the BSITD-III composite score [30].

Table 5.2: Conversion table for (a) BSID-II MDI and (b) PDI scores to BSID-III composite cognitive/language and motor scores, respectively.

BSID-II MDI	60	65	70	75	80	85	90	95	100	105	110
BSID-III Combined Cognitive /Language Scores **[217, 218]**	80	83	87	90	93	96	100	103	106	109	113

BSID-II PDI	60	65	70	75	80	85	90	95	100	105	110
BSID-III Combined Motor Scores **[217, 218]**	78	82	86	90	94	98	102	106	110	114	118

Two dichotomous outcome groupings of neurodevelopmental outcome have been used [215]:

I. Dichotomous-1 (D1): classifies the neurodevelopmental outcome as 'good' by grouping normal/mildly abnormal outcome and 'poor' by grouping Vs moderate/severe abnormality. Poor outcome was defined by a score <85 in all subscales or <70 in any subscale; otherwise the outcome was classified as 'good'.
II. Dichotomous-2 (D2): Newborns are considered as optimal or suboptimal based on a cut-off CB-III score of 85, 1 standard deviation below the mean standardised combined score of 100.

5.3 Methods

In Chapter 2, the EEG patterns that best predict the neurodevelopmental outcome were identified. In Chapter 4, various signal features have been extracted to classify these identified EEG patterns. The findings suggest a strong relationship between EEG signal features and neurodevelopmental outcome

(Figure 5.3). In this chapter, this relationship will be thoroughly explored using state-of-the-art signal processing and machine learning approaches.

The general approach is shown in Figure 5.4. Various features are extracted and selected from the neonatal EEG record soon after birth in term neonates with HIE to seek association with the neurodevelopmental outcome obtained at two years.

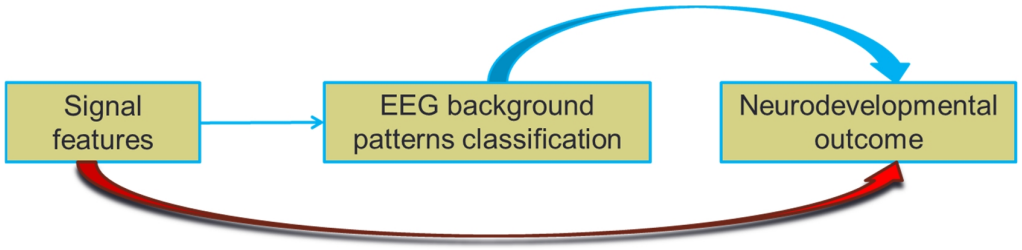


Figure 5.3: Relationship between EEG signal features and outcome.

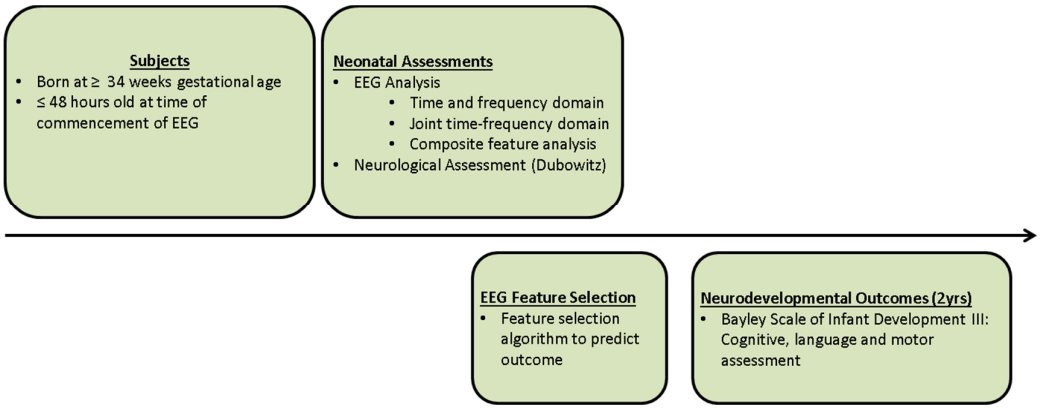


Figure 5.4: Study overview.

5.3.1 EEG pre-processing

EEG pre-processing is an important step before analysis as the inclusion of different types of artifact can produce misleading results. Recordings with an impedance $> 10\ k\Omega$ or EEG amplitude exceeding $+100\ \mu V$ or $-100\ \mu V$ were excluded first [39, 219]. EEG signals were then high-pass filtered with a cut-off frequency of 0.5 Hz, de-trended to remove linear trend using Matlab function, $detrend$, and finally a robust artifact removal (RAR) method was applied [220]. This method also obeys the recent EEG artifact removal guidelines [221], which recommend combining 'more than one algorithm to correct the signal using multiple processing stages' [221]. In the RAR method, independent component analysis (ICA) and wavelet denoising (WD) are combined to remove artifacts [220]. EEG is processed in multiple frames three times, each time with a different partitioning of the signal. The number of frames are calculated from the signal length [220]. Finally, an adaptive folding and low-

pass filter are used to obtain a clean EEG signal. This method has been tested on neonatal EEG signal [220] and shown to not significantly affect the signal, and so was used in this research. Figure 5.5 shows the block diagram of the RAR method and details of the RAR method can be found in [220]. A 2-h data segment was then chosen for each subject as close as possible, but prior, to 24 h after birth.

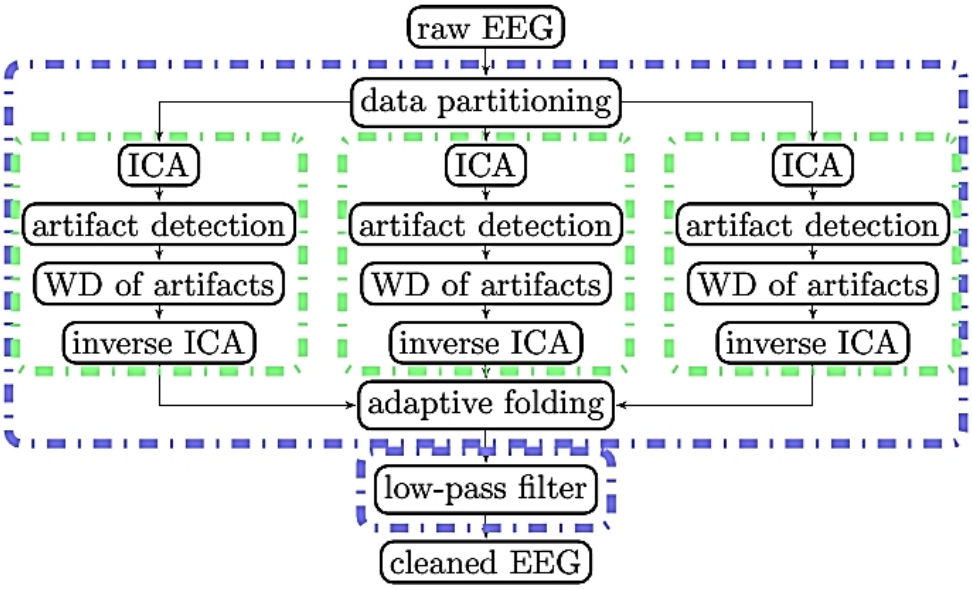


Figure 5.5: Different steps of the RAR method [220]. Firstly, artifacts with high amplitudes are removed (first blue block) and after that wavelet-based ICA (wICA) (green blocks) is used to remove short duration artifacts. This figure is reproduced from [220] with permission. [Permission was obtained from *Physiological Measurement (*http://iopscience.iop.org/journal/0967-3334*), an IOPScience journal]*.

5.3.2 EEG feature extraction

After pre-processing, all EEG datasets were down-sampled to 64 Hz and segmented into 60 s epochs with no overlapping. The segment length of 60 s was chosen arbitrarily but has been previously used in clinical studies predicting neurodevelopmental outcome [41, 205] and is a convenient sample length for use in JTF calculation.

Limited prior knowledge is available on what signal processing features most accurately predict outcome, so an exploratory large set of features was extracted in the time (t) domain, frequency (f) domain and joint (t, f) domain. We chose to describe the signal from as many perspectives as possible to provide a systematic, efficient and exhaustive characterisation of neonatal EEG background activity. All analyses were performed using the high-performance computing system at the Research Computing Centre at the University of Queensland. Table 5.3 lists the features used in this study. Describing all the features in detail would be trivial and we have therefore provided references (see Table 5.3) for their mathematical derivations and/or code implementations. Features were mainly extracted from 1-min epochs and these were denoted *local or short-term features*. Some

features are also extracted from the entire 2-h EEG dataset and these are denoted *global or long-term features.*

5.3.2.1 Local features extraction

5.3.2.1.1 Time domain features extraction

Different time-domain features such as mean, variance and mean inter-burst interval (IBI) provide the temporal evaluation of the EEG signal. For example, the variance corresponds to the energy of the signal, skewness is a measure of the symmetry of the amplitude distribution and kurtosis is a measure of the relative peakiness or flatness of the amplitude distribution [222]. During the pre-ictal period, a decrease in variance and an increase in kurtosis were observed [222]. Similarly, time between two burst periods provides a promising indicator of neurodevelopmental outcome [86]. In addition, EEG is the electrical activity of brain taking place at higher hierarchic levels of the central nervous system, which is a suitable area for entropy and nonlinear time series analysis techniques [223] and normal and abnormal EEG provides distinguishable entropy and nonlinear information. For this reason, different nonlinear features were also extracted from neonatal EEG signals. Table 5.3 includes all the time-domain features used in this study. These features have also been shown to be useful and have been consistently used for outcome prediction [205] and neonatal seizure detection [148].

5.3.2.1.2 Frequency domain features extraction

The frequency content in EEG is of great importance, as different physiological and cognitive processes are reflected in activity in different frequency ranges of EEG [224, 225]. Moreover, neonatal HIE is a low-frequency, high-impact condition that poses specific challenges [226]. An EEG with normal outcome may provide different information compared to an EEG with poor outcome in different frequency bands. Therefore, the power in different frequency bands such as delta$(0.5-4\ \mathrm{Hz})$, theta $(4-8\ \mathrm{Hz})$, alpha $(8-12\ \mathrm{Hz})$, beta $(12-30\ \mathrm{Hz})$ and their ratios, such as delta-to-theta power ratio (DTR) and delta-to-alpha power ratio (DAR) may also useful. In addition, the 'spectral edge frequency' is a quantification of the power distribution along the spectral range of a given signal [222]. These features have been frequently used in neonatal EEG [225, 227]. Table 5.3 lists the frequency domain features used in this study.

5.3.2.1.3 Time-frequency domain features extraction

JTF analysis takes the time-varying spectral contents of the EEG into account. The (t, f) can be represented by the signal's energy distribution simultaneously as a function of both time and frequency called time-frequency distribution (TFD). The QTFD is one of the most useful classes of

TFD and has shown promising results for non-stationary signal analysis and EEG classification [29, 33]. Therefore, the QTFD was chosen for this study. The Wigner-Ville distribution (WVD) is a highly concentrated and sharply localised QTFD, but the resolution is very poor due the quadratic nature of the transform and it produces cross-term interferences in a multicomponent signal such as EEG. These cross-terms can be reduced (or smoothed) by applying a data-dependent kernel. Chapter 3 provides a Hybrid Genetic Algorithm (HGA) to obtain a fully-optimised and high-resolution QTFD for a non-stationary signal like EEG.

The epoch size of the 60 s segment of EEG signal was 3840 (i.e. $60\,\text{s} \times 64\,\text{Hz} = 3840$ sample points). Currently, standard QTFD implementation algorithms cannot handle such a large EEG epoch using a standard standalone computer because TFDs are two dimensional (2D) functions and require $N^2 \log_2 N$ numerical operations (a basic FFT element requires $Nlog_2N$ operations; N is the signal length). To deal with this, a fast, memory-efficient and optimised QTFD is needed (see Chapter 3). We have implemented a fast and optimised QTFD (see previous chapter) called extended modified B distribution (Fast EMBD) (see Chapter 3). This TFD was used to extract (t, f) features. For example, instantaneous frequency (IF) and EEG band-specific instantaneous spectral power ratio (ISPR) are the unique features of a TFD and can be treated as dynamic features. Different statistical features such as mean, variance etc. are extracted from the IF and ISPR and used (see Table 5.3) with their derivation found in the previous chapter.

5.3.2.2 Composite features extraction

The inter-relationship between two hemispheres can provide useful information about functional connectivity and has been shown to be associated with the cognitive outcome [228]. This inter-hemispheric dependence can be measured by EEG coherence, phase-locking value (PLV) and brain symmetry index. Table 5.3 provides a list of composite features used.

Table 5.3: EEG feature extraction from different domains.

Features	Domain	Feature based on	Features Name
Local features	Time(t)	Statistical features ($FV_{t_{stat}}$)	First four (mean, variance, skewness, kurtosis) statistical moments [35] Coefficient of variation, RMS power, Max, Min [35] Mean of the lower and upper envelope ([33],pp 944-945) Number of zero-crossing in first and second derivatives[205] Mean and median of IBI (inter-burst interval) [229]
		Entropy based features (Nonlinear features) ($FV_{t_{entro}}$)	Hjorth moments, Hjorth complexity [148, 195] Higuchi fractal dimension, algorithmic entropy [163] Renyi, Shannon and Tsaillis entropy [164, 230] Approximate and Sample entropy [166] Hurst exponent [168, 230]

	Frequency (f)	Spectral and relative power-based features (FV_f)	Spectral Flatness, Flux and Entropy [35] Spectral Edge frequency at 80, 90 and 95% [169, 195, 222] Spectral Edge power at 80, 90 and 95% [169, 195, 222] Total power in delta, theta and alpha bands [222] Relative power in delta, theta and alpha bands w.r.t total EEG power [170, 222] Delta to theta power ratio (DTR) [231] Delta to alpha power ratio (DAR) [222, 232] Mean bout in delta, alpha and theta band [229] Bout percentage in delta, alpha and theta band[229]
	Time-frequency (t, f)	(t, f) -signal based features $\left(FV_{(t,f)_{sig}}\right)$	First four statistical moments of the TFD [29] TF Entropy, Entropy, Flatness and Flux [29] Instantaneous Frequency mean and range [29] Maximum and Entropy of the singular values Gini Index [172], ECM
		(t, f)-band specific features $\left(FV_{(t,f)_{band\_spe}}\right)$ [in Delta, Theta, Alpha, Beta bands]	First four statistical moments of the $ISPR$ [161] Coefficient of Variation of the $ISPR$ [161]
	t, f ,phase	Composite features Between two channels (FV_{comp})	Mean, variance and max of magnitude squared coherence Brain symmetry index (BSI) [205, 233] Mean, variance and max of phase-locking value (PLV) [234]
Global features	–	Global features (FV_{global})	Low activity duration and Low activity density High activity duration and High activity density Global mean (IBI) [229] Global median (IBI) [229]

5.3.2.3 Global features extraction

All the features extracted above are from a 1-min epoch. However, we hypothesise that not only features extracted from the 1-min EEG epoch, but also features extracted from long duration EEG signal may be useful in predicting neurodevelopmental outcome. Several features have been extracted from the whole 2-h EEG signal and these features are denoted *global features*. The longer-term evolution of the EEG may provide discriminative information that might be useful in predicting neurodevelopmental outcome. Low activity duration and low activity density, high activity duration and high activity density, global mean and median of (IBI) are proposed as global features in this study.

Low activity duration is defined as the amount of time (in min) the EEG amplitude lies between $+5\ \mu V$ and $-5\ \mu V$. The low activity duration density is defined as the low activity duration per hour. These features are the indication of low voltage and our structured review and meta-analysis (see Chapter 2) shows that low voltage is one of the robust predictors of neurodevelopmental outcome (see Chapter 2). Like low voltage duration, high activity duration is defined as the amount of time EEG amplitude exceeds $+60\ \mu V$ or $-60\ \mu V$. Similarly, high activity duration density is defined as the high

activity duration per hour. Global IBI is the IBI calculated from the whole 2-h EEG signal [229]. Like low voltage, IBI is also one of the predictors of outcome [86].

5.3.3 Feature vector formulation

All the features mentioned in Table 5.3 have been extracted from 2-channel EEG and concatenated to create the final feature vector (FV_{Final}). The following steps were used to create the final feature vector, FV_{Final}.

1. Firstly, all the features extracted from the left EEG channel are used to create the left EEG feature vector (FV_L) by concatenation. This can be written mathematically as:

$$FV_L = \left[\left(FV_{t_{stat_L}}\right)\ \left(FV_{t_{entro_L}}\right)\ \left(FV_{f_L}\right)\ \left(FV_{(t,f)_{sig_L}}\right)\ \left(FV_{(t,f)_{band\_spe_L}}\right)\left(FV_{global_L}\right)\right] \quad (5.1)$$

where subscript 'L' represents the left-channel EEG. Note that FV_{global_L} is a one-dimensional ($1D$) matrix as it is extracted from entire 2-h left channel-EEG signal whereas all other feature vectors are $2D$ matrix consisting of P rows: each row represents the features extracted from 1-min epoch. The FV_{global_L} is repeated P times (using Matlab $repmat$ function) to form a $2D$ matrix to meet the requirement of matrix concatenation.

2. The feature vector (FV_R) from the right EEG channel is also created by following step 1. This can also be expressed as:

$$FV_R = \left[\left(FV_{t_{stat_R}}\right)\ \left(FV_{t_{entro_R}}\right)\ \left(FV_{f_R}\right)\ \left(FV_{(t,f)_{sig_R}}\right)\ \left(FV_{(t,f)_{band\_spe_R}}\right)\left(FV_{global_R}\right)\right] \quad (5.2)$$

where subscript 'R' represents the right-channel EEG.

3. Both feature vectors extracted from the left and right-channel EEG i.e. FV_L and FV_R are fused together. The mean and absolute difference between FV_L and FV_R are used to form FV_{mean} and FV_{diff} vectors. These can be expressed as:

$$FV_{mean} = \left(\frac{FV_L + FV_R}{2}\right) \quad (5.3)$$

$$FV_{diff} = |FV_L - FV_R| \quad (5.4)$$

4. Finally FV_{mean}, FV_{diff} and FV_{comp} features vectors are concatenated to form the high dimensional final feature vector (FV_{Final}). This can be expressed as:

$$FV_{Final} = \left[(FV_{mean})\ \left(FV_{diff}\right)\ \left(FV_{comp}\right)\right] \quad (5.5)$$

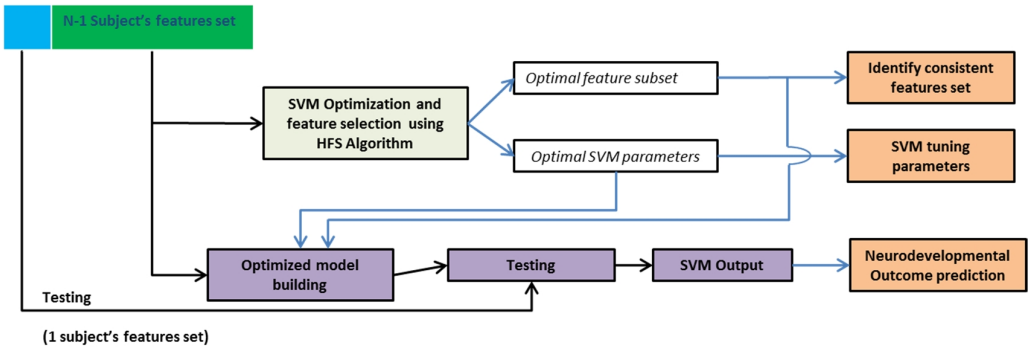
5. By following steps 1–4, FV_{Final} is created for all subjects. These final feature vectors are used for feature selection and neurodevelopmental outcome prediction.

5.3.4 Feature selection

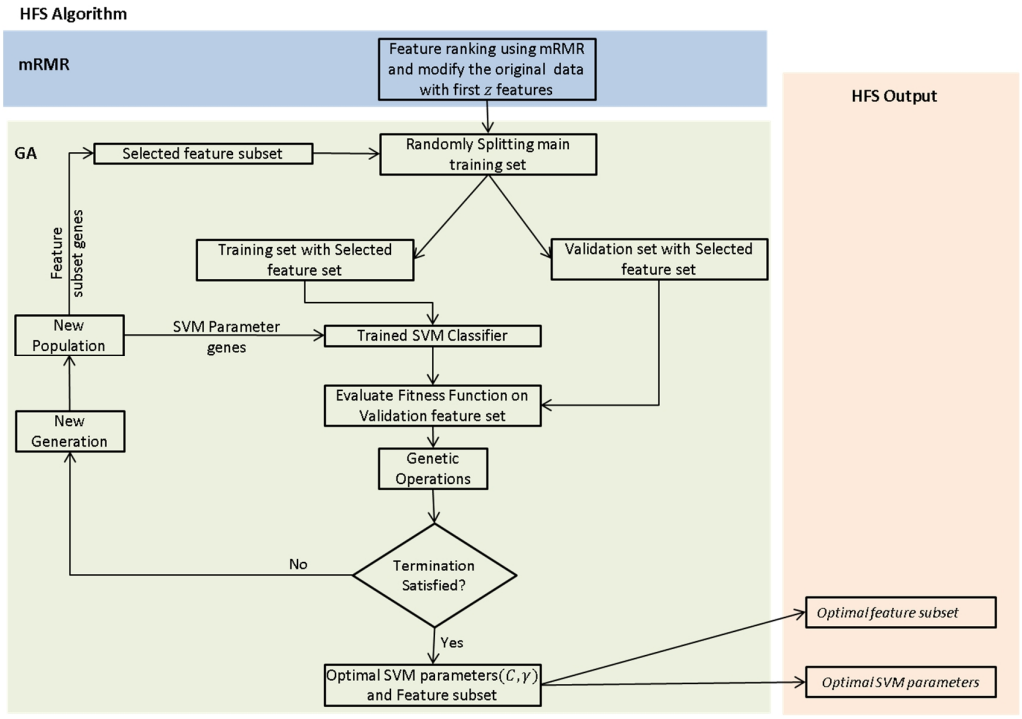
The FV_{Final} generates a high dimensional feature set. Learning from such a high dimensional feature set leads to high computational load and decreases the classification accuracy due to irrelevant and redundant features. Therefore, the most-discriminative and non-redundant feature subset needs to be selected. This is known as 'feature subset selection' or 'feature selection'.

The support vector machine (SVM) classifier is used as it is computationally stable and has been successfully applied to EEG signal classification [29]. The SVM classifier finds a hyperplane by maximising the margin between the classes using the data instances (support vectors) close to the hyperplane and it has a regularisation parameter (C). A detailed description of the SVM has been provided in the previous chapter. The Gaussian radial basis kernel (RBF) is used in this study as it requires only one parameter (γ) to be optimised [29]. Conversely, other popular kernels, e.g. polynomial kernel, require several parameters to be optimised.

In the current context, (C, γ) parameters need to optimise for the optimal use of the RBF-SVM along with the feature subset selection. In the previous chapter, a hybrid feature selection (HFS) algorithm was proposed to simultaneously select prominent feature subset and SVM tuning parameters (C, γ). The HFS algorithm comprises a maximum relevance and minimum redundancy (mRMR) [194] method and a genetic algorithm (see previous chapter for details).



(a)



(b)

Figure 5.6: (a) Workflow in the LOSO performance assessment, and (b) the HFS algorithm used in the LOSO cross-validation system.

Note that the feature is normalised before applying the HFS algorithm because the numerical range of one feature or attribute can be different from another feature and normalisation is used to avoid the dominance of the feature which has a higher numerical range over the feature which has the lower numerical range. Each feature is normalised by the following equation.

$$F_{Normalized} = \frac{F - F_{\mu}}{F_{\sigma}} \tag{5.6}$$

where F and $F_{Normalized}$ are the original and normalised feature value and F_{μ}, F_{σ} are the mean and standard deviation of the feature value respectively. These F_{μ}, F_{σ} are also stored to normalise the test instances.

The HFS algorithm is adapted to a leave-one-subject-out (LOSO) cross-validation system. Figure 5.6(a) shows a schematic workflow of the LOSO cross-validation system. Figure 5.6(b) shows the HFS algorithm used in the feature selection and parameter optimisation part of Figure 5.6(a). In the HFS algorithm, the maximum relevance and minimum redundancy (mRMR) method [194] is used to initially reduce the original features, (let's say, the number of original features (FS_o)), down to a more manageable z most discriminative features, and a genetic algorithm(GA) is then used to select the most prominent feature set, from this reduced set. It also selects the optimal (C, γ) parameters

simultaneously. Note that GA can diverge from the feasible search space if the RBF-SVM tuning parameters (C, γ) are not bounded. The lower bound and upper bound for $C: 2^{-2}, 2^{12}$ and for $\gamma: 2^{-10}$ and 2^{4} are chosen. These larger range should be enough for the selection of optimal (C, γ) parameters [180]. Details of the HFS algorithm can be found in the previous chapter.

In the LOSO system, the feature set extracted from one subject is used for testing while the features extracted from all other $(N-1)$ subjects are used for training and optimisation. A model is built using this optimal feature subset and optimal SVM parameters. This optimised and nonlinear model is used to predict neurodevelopmental outcome on the test data. Importantly, the testing data is unseen or unused in the optimisation process. This process is iterated until all subjects are tested. Note that the selected feature subset may not be consistent for all iterations or subjects. To identify features that appear to be consistently useful in most of the LOSO iterations, the top N features, after applying the HFS algorithm, are identified by using a majority voting system. The majority voting algorithm was performed by counting the number of times a feature appears in LOSO iterations and then sorting (higher to lower counts) to find the most consistent features. This process will offer better feature subset selection that correlates best with neurodevelopmental outcome in most subjects. These most consistent features will be used to build the final SVM model. An averaged value of the SVM tuning parameters across all LOSO iterations will be used in the final model. This final optimised SVM model will be used to test a completely unseen test database.

To determine the number of features that provides the highest accuracy in predicting good/poor and optimal/suboptimal outcome, we repeated the procedures with different number feature sets, i.e. $z = 10, 15$ and 20 features. The prediction performance is measured in terms of Sensitivity, PPV, $Fmeasure$ and accuracy as defined below.

5.3.5 Performance assessment matrices

The prediction performance can be assessed using the following standard quality measures:

$$\begin{aligned} Recall = Sensitivity &= \frac{TP}{TP + FN} \times 100\% \\ Specificity &= \frac{TN}{FP + TN} \times 100\% \\ Precision = PPV &= \frac{TP}{TP + FP} \times 100\% \\ ACC &= \frac{TP + TN}{(TP + FN + TN + FP)} \times 100\% \end{aligned} \tag{5.7}$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative, PPV= positive predictive value. Specificity is less important than Precision [184], and hence is not included

in this study (see previous chapter for a more detailed explanation). A more robust measure is $Fmeasure$ as it takes the classifier's Precision and Recall into account. $Fmeasure$ is defined as:

$$Fmeasure, F_{\beta} = \frac{(1 + \beta^2)(Precision \times Recall)}{\beta^2.(Precision + Recall)} \quad (5.8)$$

Note that β is usually set to 1 in the $Fmeasure$ calculation; see section 4.4.5 for detail justifications.

5.4 Results

Prediction results for good/poor outcome and then for optimal/suboptimal outcome will be described. The LOSO cross-validation was used in both cases.

For each subject, Sensitivity, PPV, $Fmeasure$ and accuracy were calculated along with an average value of these performance measures across all subjects (Table 5.4). A feature set comprising 15 features shows the highest average accuracy of 88.4% and average $Fmeasure$ of 90.8% compared to a feature set of 10 or 20 features.

For the optimal/sub-optimal outcome classification, as in Table 5.4, an average prediction (optimal/sub-optimal) result across all the subjects was calculated and is presented in Table 5.5. The prediction result using 10 features shows an average accuracy of 85.6% and the accuracy decreases when more features are added; for example, prediction results using 15 features shows a mean accuracy of 82.5%. Although a feature set comprising 20 features increases the average accuracy of 89.6%, a model using 20 features may overfit the system and decrease the classification performance for the new testing instances.

Table 5.4: The average performance (%) of the LOSO system predicting good and poor outcome.

Good and poor outcome				
Result using	**Sensitivity**	**PPV**	**Fmeasure**	**Accuracy**
10 Features	87.5	95.0	88.8	87.5
15 Features	88.4	95.0	90.8	88.4
20 Features	83.6	97.5	86.4	83.6

Table 5.5: The average performance (%) of the LOSO system predicting optimal and sub-optimal outcome.

Optimal and sub-optimal outcome				
Result using	**Sensitivity**	**PPV**	**Fmeasure**	**Accuracy**
10 Features	85.6	97.5	88.9	85.6
15 Features	82.5	95.0	86.6	82.5
20 Features	89.6	97.5	92.2	89.6

5.4.1 Selection of the most consistent features

Attention has already been drawn to the fact that the selected feature subset may not be consistent for all LOSO iterations or subjects. A majority voting algorithm was used to obtain the most consistent features.

Table 5.6: The most consistent 15 features for the prediction of good/poor outcome.

Selected 15 features
Bout percentage in Delta band
Bout percentage in Theta band
(t, f) Singular Values entropy
Mean bout in Alpha band
Low activity duration density
Low activity duration
(t, f) Shannon entropy
IBI Median
kurtosis of ISPR in Delta band
Global Mean IBI
(t, f) variance ISPR in Delta band
kurtosis
DAR
Mean of lower envelope
(t, f) Skewness ISPR in Alpha band

As can be seen from Table 5.4, a SVM model comprising 15 features provides the highest average cross-validation accuracy in predicting good/poor outcome and, therefore, the most consistent 15 features were selected from LOSO iterations. Table 5.6 lists the most consistent 15 features for the prediction of good and poor outcome.

Similarly, it can be seen from Table 5.5 that a model comprising 10 features provides higher cross-validation accuracy for predicting optimal/suboptimal outcome than a model comprising 15 features and, therefore, the most consistent 10 features were selected from LOSO iterations. Table 5.7 lists the most consistent 10 features for the prediction of optimal/suboptimal outcome.

Table 5.7: The most consistent 10 features for the prediction of optimal/suboptimal outcomes.

Selected 10 features
Bout percentage in Theta band
Low activity duration density
Low activity duration
Global median IBI
(t, f) ISPR in Theta band
Mean bout in Alpha band
Bout percentage in Delta band

(t, f) Shannon entropy
Difference in kurtosis
Skewness

5.4.2 Testing of the model on a test database

Although the LOSO cross-validation has been shown to be the most unbiased estimator, validation of the proposed model developed by the selected features would be useful to show the robustness of the model. To validate the model, we have used a completely separate database recorded from 27 term neonates with HIE. This database was not used in either the model or the LOSO cross-validation system. The performance of the proposed model is shown for individual subjects. Table 5.8 lists the prediction performance of good and poor neurodevelopmental outcome, while Table 5.9 lists the prediction performance of optimal and suboptimal outcome. Table 5.8 and Table 5.9 also provide gold standard binary neurodevelopmental outcome at two years, assessed by the neurologists to compare between gold standard neurodevelopmental outcomes and the results predicted by the model in terms of Fmeasure and accuracy. To illustrate, Subject 1 can be chosen. The gold standard binary neurodevelopmental outcome for Subject 1 is defined as '*good'* and '*optimal'*. The model predicted the outcome as *'good'* with Fmeasure and accuracy of 100% and 100% (see Table 5.8); and *'optimal'* with Fmeasure and accuracy of 97.4% and 95% (see Table 5.9), respectively. Other subjects stated in Table 5.8 and Table 5.9 can be interpreted in the same way.

The model lacked accuracy in predicting the neurodevelopmental outcome of subjects 11 and 17. The neurodevelopmental outcome for subject 11 is defined as 'good' and 'optimal' but the model predicts 'poor' and 'suboptimal', whereas for subject 17 the opposite phenomenon occurred. In the case of subject 11, a moderately abnormal EEG trace was observed at 24 h but improved after that. The 2-year BSID-III cognitive and language scores of this subject were 130 and 115 respectively, well above the average value, but the motor score was 82, which is below the average. Although this subject had a lower than average motor score, the neurodevelopmental outcome was thresholded to 'good' outcome (see Section 5.2.2). In the case of subject 17, a mildly abnormal EEG trace was observed at 24 h but the HI severity increased after that and the baby died prior to 2 years of age. The EEG features extracted at that time point predicted a 'good' outcome using the model, but the patient died which was classified as a bad outcome. Table 5.10 summarises the results of Table 5.8 and Table 5.9 by showing the mean value of the all subjects. Note that the $Fmeasure$ is derived from $\mathrm{Sensitivity}, \mathrm{PPV}$ and, therefore, this redundant information, i.e. $\mathrm{Sensitivity}, \mathrm{PPV}$ are not presented in Table 5.8 and Table 5.9.

Table 5.8: Prediction of good and poor outcome on the test dataset

Subject No	Neurologist 'Gold standard'*	Model prediction	
		Fmeasure	Accuracy
Subject 1	Good	100	100
Subject 2	Good	97.9	95.83
Subject 3	Good	97.0	94.17
Subject 4	Good	68.1	51.67
Subject 5	Good	99.6	99.17
Subject 6	Good	100	100
Subject 7	Good	100	100
Subject 8	Good	100	100
Subject 9	Good	100	100
Subject 10	Good	99.6	99.17
Subject 11	Good	0	0
Subject 12	Good	100	100
Subject 13	Good	91.9	85
Subject 14	Good	100	100
Subject 15	Good	100	100
Subject 16	Good	100	100
Subject 17	Poor	0	0
Subject 18	Poor	100	100
Subject 19	Poor	91.4	84.2
Subject 20	Poor	90.9	83.3
Subject 21	Poor	6.45	3.33
Subject 22	Poor	100	100
Subject 23	Poor	100	100
Subject 24	Poor	100	100
Subject 25	Poor	94.3	89.2
Subject 26	Good	100	100
Subject 27	Poor	85.7	75

*This is gold standard binary neurodevelopmental outcome at 2 years assessed by neurologists.

Table 5.9: Prediction of optimal and suboptimal outcome on the test dataset

Subject No	Neurologist 'Gold standard'*	Model prediction	
		Fmeasure	Accuracy
Subject 1	Optimal	97.4	95
Subject 2	Optimal	67.4	50.8
Subject 3	Optimal	66.7	50
Subject 4	Optimal	34.5	20.8
Subject 5	Optimal	87.3	77.5
Subject 6	Optimal	100	100
Subject 7	Optimal	100	100
Subject 8	Optimal	100	100
Subject 9	Optimal	45.2	29.2
Subject 10	Optimal	97	94.2
Subject 11	Optimal	0	0
Subject 12	Optimal	96.1	92.5
Subject 13	Optimal	100	100
Subject 14	Optimal	100	100
Subject 15	Optimal	100	100
Subject 16	Optimal	98.3	96.7
Subject 17	Sub-optimal	0	0
Subject 18	Sub-optimal	100	100
Subject 19	Sub-optimal	100	100
Subject 20	Sub-optimal	100	100
Subject 21	Sub-optimal	66.7	50
Subject 22	Sub-optimal	100	100
Subject 23	Sub-optimal	88.4	79.2
Subject 24	Sub-optimal	100	100
Subject 25	Sub-optimal	99.6	99.2
Subject 26	Optimal	15.4	8.33
Subject 27	Sub-optimal	100	100

*This is gold standard binary neurodevelopmental outcome at 2 years assessed by neurologists.

Table 5.10: The average performance (%) on the test database**.

Outcomes	Optimal Features	Fmeasure	Accuracy
Good/Poor	Using 15 Features	85.8	83.8
Optimal/Suboptimal	Using 10 Features	78.0	75.4

** This average performance was calculated from Table 5.8 and Table 5.9.

5.4.3 Comparison with other studies/ approaches

The proposed method was derived from wide range of features, ranging from t-domain to joint (t, f)-domain features, and from local to global features. Studies also used entropy-based features such as approximate entropy (ApEn) [235, 236] and sample entropy (SpEn) [237] with outcome. Approximate entropy and sample entropy with optimal parameters were applied in our database for a fair comparison with our results. Table 5.11 presents a quantitative comparison predicting good and poor outcome, while Table 5.12 presents a quantitative comparative study predicting optimal and suboptimal outcome for different studies/approaches. It can be seen that accuracy is about 69.3% and 67.5% predicting good and poor outcome using ApEn and SpEn based studies respectively, and about 67.4% and 67.1% respectively predicting optimal and suboptimal outcome. The accuracy is more than 12% lower than our proposed method. The classification result of our proposed method is significantly better (i.e. $p<0.001$, ANOVA test) than ApEn, and SpEn based studies in terms of predicting good/poor and optimal/suboptimal outcome. However, ApEn and SpEn-based studies are simpler than our proposed method, as only entropy-based features are used. Therefore, only ApEn and SpEn need be calculated, whereas in our proposed method 15 features must be calculated. The prediction accuracy is more important than system complexity, especially in the case of neurodevelopmental outcome prediction.

Table 5.11: The average performance (%) on the test database for different studies/approaches in predicting good and poor outcome.

Features	Fmeasure	Accuracy
Using ApEn	70.3	69.3
Using SpEn	68.8	67.5
Using HFS selected 15 t- and f- features	80.6	79.0
Results from this study i.e. using HFS selected 15 (t-, f- and (t, f)) features	85.8	83.8

Table 5.12: The average performance (%) on the test database for different studies/approaches in predicting optimal and suboptimal outcome.

Features	Fmeasure	Accuracy
Using ApEn	69.1	67.4
Using SpEn	68.4	67.1
Using HFS selected 10 t- and f- features	72.8	70.4
Results from this study i.e. using HFS selected 15 (t-, f- and joint (t, f)) features	78.0	75.4

From Table 5.6 and Table 5.7, it can be seen that several (t, f)-features have been selected by the HFS algorithm, which shows the effectiveness of (t, f) features in predicting neurodevelopmental outcome. To prove this, whole analysis was re-run independently, on the same test database, with only t- and f-based features. The accuracy in predicting good and poor outcome is 79%, whereas it is 70.4% in predicting optimal and sub-optimal outcome. Thus, (t, f)-features can improve the prediction of neurodevelopmental outcome when they combine with t- and f-domain features; in our test database, this improvement is about 5% and the result is statistically significant ($p<0.001$, t-test). The HFS selected t- and f-domain features have been provided in Appendices 5.A and 5.B.

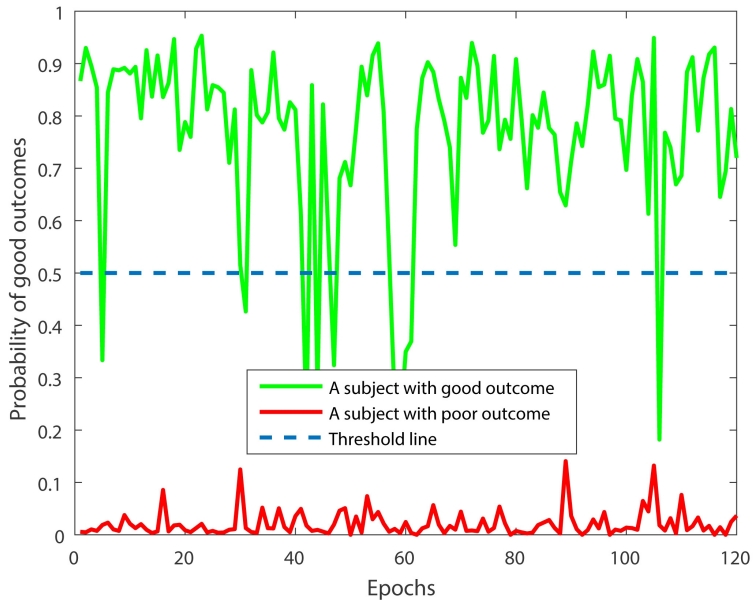
5.5 Application of the proposed model

It is possible to build a decision support system (DSS), a computer-based system to assist decision-making from the proposed model. We have created a program that visually represents the probable neurodevelopmental outcome in terms of posterior probabilities. The RBF-SVM can be employed to obtain output probabilities. These probabilistic outputs are generally more intuitive for clinical staff [148].

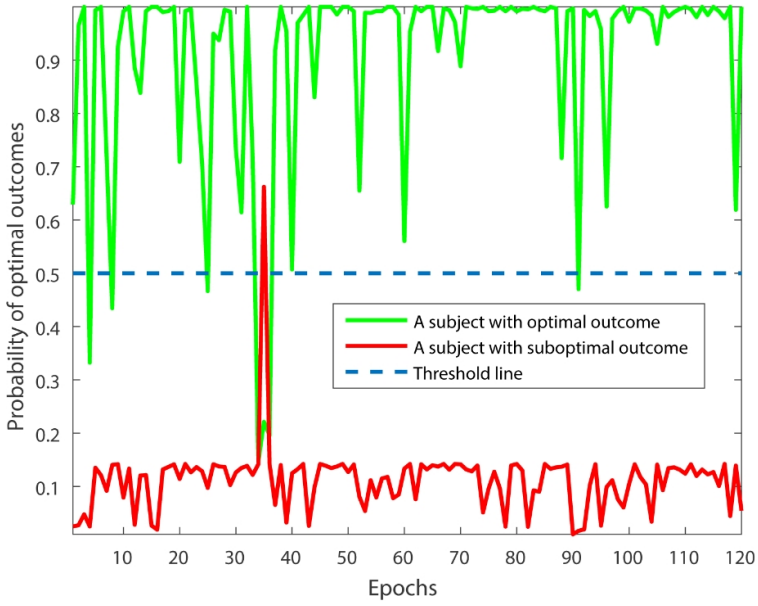
In the present context, the SVM model provides a binary decision output, i.e. good/poor or optimal/sub-optimal. These binary decisions can be converted into posterior probabilities using a sigmoid function. This function can be expressed as [148]:

$$Pr(y = 1|d_{hyp}) = \frac{1}{1 + exp(Ad_{hyp} + B)} \tag{5.9}$$

where d_{hyp} is the distance to the separating hyperplane [148], i.e. the output of the SVM classifier, and A and B are the parameters of the sigmoid function estimated on the training dataset using the method presented in [182].



(a)



(b)

Figure 5.7: An example of the DSS: (a) probabilistic trend for good poor outcome and, (b) optimal and suboptimal outcome.

Using this method, the probabilistic output for each EEG epoch can be produced and a series concatenation of probabilistic outputs calculated from all the EEG epochs provides a continuous trace of probable neurodevelopmental outcome. Figure 5.7(a) shows an example of the raw (unsmoothed) probabilistic trace for a newborn with good outcome and a newborn with poor outcome, whereas Figure 5.7(b) shows an example of the raw (unsmoothed) probabilistic trace for a newborn with

optimal outcome and a newborn with suboptimal outcome. Smoothing filters such as Savitzky–Golay smoothing filter [238] can be used to improve and smooth the probabilistic output.

5.6 Discussion

Early prediction of long-term neurodevelopmental outcome is still a major challenge in infants. EEG is a useful method of assessment in asphyxiated infants and has been shown to be a robust predictor of neurodevelopmental outcome (see Chapter 2). We have selected and extracted features from neonatal EEG to seek association with the cognitive and psychomotor, i.e. neurodevelopmental, outcome by applying state-of-the-art signal processing and machine learning approaches.

The key findings can be summarised as:

- Different features extracted from EEG predict neurodevelopmental outcome in term neonates with HIE.
- Both local and global features contribute to outcome prediction.
- Several (t, f) features (see Table 5.6 and Table 5.7) have been selected by the HFS algorithm suggesting that joint time-frequency analysis, i.e. optimised TFD, improves prediction of neurodevelopmental outcome.
- Joint optimisation of features and SVM classifier parameters is needed to build an optimised model for outcome prediction.
- The DSS developed would be very useful to clinicians to counsel parents and to guide early clinical management decisions including the efficacy of anti-epileptic drugs.

It is unlikely that only one feature will serve as a robust predictor of neurodevelopmental outcome. A set of features may generalise the process and one of the important aspects of this study was to establish the most prominent set of features that best predict neurodevelopmental outcome. The LOSO cross-validation has been applied to a high dimensional feature set extracted from 40 infants.

Looking at the features list, it can be seen that features in certain frequency bands may contain more information than others. For example, the proportion of oscillation in the delta and theta band (called 'bout percentage') provides significant and discriminative information. Lower values of delta-to-alpha power ratio (DAR) and higher values of the mean of the lower envelope of the signal are observed in more severely affected babies, suggesting that these measures are valuable in outcome prediction. Abnormal babies usually have low voltage activity. This is clearer from the global features: a higher value in low activity duration and low activity density predict abnormal

neurodevelopmental outcome. In addition, TFD-based features in certain frequency bands provide better prediction than others. This is because TFD can be considered as a bi-dimensional energy density function and TFD of normal and abnormal EEG generate different distributions. A normal EEG is highly predictive of a normal outcome, whereas various abnormal EEG features have been consistently associated with neurological abnormalities or death (see Chapter 2); therefore, features extracted from TFD offer significant and discriminative information to predict neurodevelopmental outcome. Moreover, ISPR, which is one of the salient features of TFD, describes the signal non-stationarities. Different statistics of this feature in different bands also provide significant and distinct information.

A nonlinear model is built-up by the selected features and, although it is possible to provide a sensible rationale for the individual features, they are only important when considered together. In machine learning and artificial intelligence it is known that '*a feature that is completely useless by itself can provide a significant performance improvement when taken with others'* [190]. Therefore, in statistical analyses using Spearman correlation tests, Pearson correlation tests etc. as commonly used in clinical studies, seeking relationships between features and outcomes, different conclusions may be reached. These approaches cannot take into account the influence of other features simultaneously nor determine any nonlinear relationship among the features.

Another important issue is the validity of the use of limited channel (i.e. 2-channel) EEG or amplitude integrated EEG (aEEG). A recent meta-analysis reported that the sensitivity and specificity of a number of simple features from limited channel EEG or aEEG are 93% and 90% in predicting neurodevelopmental outcome, whereas in the case of full-channel EEG – i.e. continuous EEG (cEEG) – it is 92% and 83% respectively [44]. These statistics suggest the prediction of neurodevelopmental outcome using limited channel EEG or aEEG is good. However, in other research applications such as detection and classification [29], source localisation [239], or focal activity detection, a cEEG or even dense electrode array EEG (64-channel EEG) [239] is required

The structured review and meta-analysis presented in Chapter 2 showed that different EEG background patterns can predict neurodevelopmental outcome. It is also possible to use seizure detection algorithms and correlate seizure burden with outcome. However, a recent study showed that aEEG background patterns exhibit superior prediction of poor neurodevelopmental outcome (sensitivity of 91.7%, PPV of 78.6%) compared to seizure burden detected using aEEG (sensitivity of 94%, PPV of 57%) [240].

5.6.1 Limitations of the study and proposed recommendations

The accuracy of this system falls short of 100%, at 83.8% in predicting good/poor outcome and 75.4% in predicting optimal/suboptimal outcome. The following approaches may be useful and should be explored to improve the prediction accuracy:

- Multimodal approaches, i.e. the addition of other modalities such as HRV (heart rate variability), and MRI might be very useful to increase the accuracy.
    - HRV provides a non-invasive assessment of both the parasympathetic and sympathetic control of heart rate (HR) [241, 242]. There is evidence of alternation of HR following HIE and HR is also associated with EEG grade of HIE and neurodevelopmental outcome [242]. Therefore, different features extracted from HR and HRV could be very useful along with EEG-based features.
    - The features extracted from diffusion-weighted MRI (DW-MRI) and proton magnetic resonance spectroscopy (H-MRS), such as apparent diffusion coefficient (ADC) of the basal ganglia and thalamus, lactate/N-acetylaspartate (LAC/NAA) and N-acetylaspartate/choline (NAA/Cho) [210], brain volume [243] and white and grey matter lesions [244], could be useful. However, the best time to undertake MRI is 7–10 days and earlier imaging may under- or over-estimate the degree of brain injury [205].
    - Clinical features including blood pH, base deficit and Apgar score are also associated with outcome after HIE [245]. It would be useful to explore these additional features along with EEG features.
- The use of ensembles of classifiers, i.e. a combination of different classifiers [246], may improve accuracy. However, this would increase the computational load significantly as well as increasing the complexity of ascertaining the process leading to a decision.

5.7 Overall summary

EEG allows a robust and early prediction of neurodevelopmental outcome and can be implemented safely and at relatively low cost at the cot-side soon after birth. In this study, we applied state-of-the-art signal processing techniques to extract a high dimensional feature set. A highly discriminative and

non-redundant feature subset as well as classifier (SVM) parameters have been selected from this high dimensional feature set by a HFS algorithm. This algorithm is applied to the statistically most unbiased LOSO cross-validation system and an optimised model is created using the most consistent feature subset. This model has been tested on a separate dataset and a DSS has been built as a potential application of the model to visualise the probable neurodevelopmental outcome in a continuous probabilistic fashion. This work could provide a promising step toward building an objective decision support tool in term neonates with HIE for the early prediction of neurodevelopmental outcome. The present methodology could also be adapted for the prediction of outcome of other high-risk groups, such as very preterm infants. Translation of these findings into an EEG monitoring system would be very useful for end users. Further development and testing is also necessary and some suggested research directions have been provided. Ultimately, an optimised system needs to be evaluated in further clinical trials.

Chapter 6 Conclusion and future directions

6.1 General summary

HIE is one of the major causes of morbidity and mortality in newborns and EEG is a useful diagnostic and prognostic tool for the analysis of these asphyxiated newborns. Different specific background patterns of EEG have been shown to be robust predictors of neurodevelopmental outcome. Characterisation of these patterns is very useful for exploring brain activity under different conditions, such as severity of HIE. Further, accurate and automatic detection of these EEG patterns is significant in practice, as EEG is usually monitored in babies with HIE in the NICU. EEG analysis can provide a useful tool to support early and appropriate clinical management for those newborns who are at risk of HI cerebral injury. Automated analysis has the potential to both reduce the workload of clinicians and provide a more uniform and accurate analysis. Prediction of neurodevelopmental outcome using EEG features at an early stage would be very useful in guiding therapy and discussing options with parents as well as providing opportunities for improved health outcomes.

Motivated by this, this thesis (i) identifies those specific EEG background patterns that have been described to best predict neurodevelopmental outcome in term neonates with HIE (Chapter 2); (ii) designs and optimises signal processing tools for the characterisation of these EEG patterns (Chapter 3), (iii) develops methods for the accurate and automatic detection and classification of these patterns (Chapter 4) and, finally (iv) tests a model using the identified features in a patient dataset to establish its test characteristics in predicting 2-year developmental outcome (Chapter 5).

Different EEG background patterns in multichannel neonatal EEG have been used for analysis (characterisation and classification) and prediction of neurodevelopmental prediction. To take the EEG non-stationarity into account and obtain time and frequency information simultaneously, high-resolution quadratic time-frequency distributions have been designed and optimised. Different features from the high resolution QTFDs and in the time and frequency domain have been selected using a hybrid feature selection method. State-of-the-art feature selection and classification method(s) have been designed and optimised for the automatic classification of EEG background patterns that predict neurodevelopmental outcome.

The research question posed was: '*How can optimised high-resolution QTFDs, and optimised classification method(s), be developed to provide better characterisation and classification for multichannel neonatal EEG signals for the purpose of accurately predicting neurodevelopmental outcome*?'

This thesis makes several novel contributions, comprising both theoretical and application-specific contributions, in addressing the research question. These are highlighted below on a chapter-by chapter-basis within the overall concept map described in Section 1.6.

In Chapter 2, a systematic review was conducted, and meta-analysis of the literature undertaken to determine which specific background features of the EEG, in term neonates with HIE, best predict outcome. A literature search was conducted using the PubMed, EMBASE and CINAHL databases from January 1960 to April 2014. Studies included in the review described recorded EEG background features and neurodevelopmental outcomes at a minimum age of 12 months and were published in English. Pooled sensitivities and specificities of EEG background features were calculated, and meta-analyses performed for each background feature.

In Chapter 3, two optimisation methods were presented for the optimal use of TFD. Section 3.2 presented a global optimisation method that uses a hybrid genetic algorithm (HGA). This 'black box' approach automatically adjusts the QTFD kernel parameters by HGA, which results in the optimal use of QTFDs for non-specialist users without requiring any additional input beyond the signal itself. This method is suitable for multicomponent and non-stationary signals without *a priori* signal information (such as, for example, the EEG signal).

Section 3.3 presented a local optimisation method based on LOS. LOS utilises short-time fractional Fourier transform (STFrFT). The key strength of the LOS is that it automatically determines the locally optimal window parameters and fractional order (angle) for all signal components, leading to a high resolution, cross-term free time-frequency representation. This method is also suitable for multicomponent and non-stationary signals without *a priori* signal information such as the EEG signal.

In Chapter 4, two classification methods were presented and validated for classifying multichannel EEG background patterns including: burst, suppression, normal, seizure and artifact. These patterns are the predictors of a range of neurodevelopmental outcomes. Classification method-1 uses a single-feature subset for all classes (patterns) whereas classification method-2 uses class-specific features. Method-2 is expected to increase the classification performance as the selected feature subset is class-specific.

Various time domain, frequency domain and joint time-frequency domain features have been first extracted and then concatenated to produce a long feature set. A hybrid feature selection (HFS) algorithm, comprising an mRMR and a genetic algorithm, is proposed to simultaneously select the prominent feature subset and the classifier parameters. A state-of-the-art machine learning algorithm,

the support vector machine (SVM), was optimised and used to classify different neonatal EEG patterns using the feature subset selected by the HFS algorithm. This HFS algorithm not only selects the prominent feature subset but also selects the SVM tuning parameters simultaneously to ensure the optimal use of the SVM classifier. Three fusion techniques, channel fusion, feature fusion and decision fusion, were used and compared for the multichannel classification. The decision fusion technique provided the best classification performance.

In Chapter 5, various time domain, frequency domain and (t, f) domain features were extracted from the neonatal EEG recorded within the first 24 h after birth to seek associations with the neurodevelopmental outcome in term neonates with HIE assessed using the industry-standard Bayley Scales of Infant Development, version III (BSID-III). The BSID-III provides developmental scores for cognitive, language and motor domains and a combined score. The hybrid feature selection (HFS) algorithm developed in Chapter 4 was used in the statistically least biased leave-one-subject-out (LOSO) cross-validation process to ascertain the most consistent feature set associated most strongly with outcome. A nonlinear SVM model was built based on the most consistent feature subset found by the LOSO system and a neonatal EEG database was used to test the predictive capability of the model. Finally, a probabilistic decision support system was derived as a potential application of the proposed system.

6.2 Key conclusions and perspectives

Based on the results presented in this thesis, the following key conclusions can be made and are presented here chapter by chapter.

From Chapter 2:

- *Different specific EEG background patterns in babies with HIE can predict neurodevelopmental outcome:* Different EEG background patterns such as burst suppression, low voltage, flat trace, trace alternant, asynchrony and asymmetry predict neurodevelopmental outcome in term neonates with HIE. Based on a systematic review of the literature, the pooled sensitivity and specificity of these EEG background features are: burst suppression (Sensitivity 0.87 [95% CI (0.78–0.92)]; Specificity 0.82 [95% CI (0.72–0.88)], low voltage (Sensitivity 0.92 [95% CI (0.72–0.97)]; Specificity 0.99 [95% CI (0.88–1.0)], and flat trace (Sensitivity 0.78 [95% CI (0.58–0.91)]; Specificity 0.99 [95% CI (0.88–1.0)]). Some other background patterns, including trace alternant, asynchrony and, asymmetry, have also been proposed as important in terms of prognosis but, because they are less prevalent, have been reported only in a small number of babies.

- Variability in the definitions of amplitude, phase, frequency and duration is present in the published literature; this highlights the need for agreement on the definitions.

- *A unique definition for EEG background patterns is needed:* An agreement on definitions for EEG background patterns is necessary for the effective implementation and use of EEG in NICUs, and some definitions have been suggested.

From Chapter 3:

- Automatic optimisation is important to characterise the non-stationary signal using QTFDs that provides a high resolution (t, f) signature: The statistical properties of non-stationary signals, including neonatal EEG, change over time. Therefore, analysis window and window size (for spectrogram) and kernel parameter(s) of QTFDs need to be chosen either by visual inspection or by automatic optimisation to obtain high resolution (t, f) signatures. The optimal window, window size and kernel parameter(s) for a non-stationary signal, even a segment (or epoch) of a long non-stationary signal, is signal dependant. Choosing optimal parameters by visual inspection is time consuming and not applicable in certain situations, such as automatic detection and classification. Therefore, automatic optimisation is necessary.

- The (t, f) optimisation problem needs a simple but effective objective function and the 'Energy Concentration Measure (ECM)' index can satisfy this requirement: By minimising (or maximising) an objective, a suitable algorithm can provide high resolution QTFDs. The ECM can be used for this purpose as it is a simple but effective objective function. Unlike norm-based concentration measures, this measure does not discriminate low concentrated components compared to the highly concentrated ones within the same TFD.

- The kernel parameter(s) of QTFDs need to be optimised in order to obtain a high resolution (t, f) signature and the proposed novel hybrid genetic algorithm (HGA) offers a fully automatic optimisation of QTFDs: The optimal kernel parameter(s) of QTFDs can enhance the (t, f) resolution by reducing cross-terms. An automatic optimisation method called hybrid genetic algorithm (HGA) is proposed and applied for a fully automatic optimisation of QTFDs. The HGA comprises a genetic algorithm and Nelder-Mead algorithm.

- *HGA offers a 'black box' approach that needs no extra input from the user except the signal itself:* One of the salient features of the proposed HGA is that the optimal kernel parameter(s) for producing a high resolution (t, f) signature can be obtained by using the HGA algorithm based only on the signal itself, without need for any additional user input.

- *The fully automatic optimisation of QTFDs is useful for the non-specialist user to optimally use QTFDs in many disciplines:* The HGA can automatically choose the optimal kernel parameter(s). It can be useful for all types of non-stationary signals from different areas including communications, radar, sonar and acoustics, and machine condition monitoring. However, the performance of different QTFDs is different and users need to decide which QTFD is the most useful for them. This indicates that, although this study automatically optimises the TFDs in terms of balancing concentration, resolution, cross-terms and computation, it still requires specialist knowledge in terms of correct interpretation of the results.

- *Fast and memory-efficient implementation of TFDs is advantageous in 'big data' science:* The fast and optimised implementation of TFDs provides another contribution by which computation time is reduced significantly, e.g. computation time reduces from 104.6 ± 1.95 s to 3.97 ± 0.15 s for a signal length of 512. This process not only optimises TFD in terms of optimising the TFD parameters but also in terms of computational time. This contribution can be advantageous in optimising the use of TFDs in 'big data' science, as in the case of multichannel EEG signal.

- Fractional Fourier transform (FrFT) provides a generalisation of the classical FT and offers improved flexibility when designing high resolution time-frequency signatures: The FrFT is the generalisation of the classical Fourier transform (FT). It can be regarded as a rotation by an arbitrary angle α in the (t, f) plane. This approach provides improved flexibility when designing high resolution time-frequency signatures as the signal chirp rate (i.e. angle) can be adapted using this approach.

- The Locally Optimised Spectrogram (LOS) defines a novel method for obtaining a high-resolution (t, f) representation based on the short-time fractional Fourier transform (STFrFT): The proposed LOS defines a novel method for obtaining a high-resolution (t, f) representation based on the short-time fractional Fourier transform (STFrFT). The key novelty and strength of the LOS is that it automatically determines the locally optimal window parameters and fractional order (angle) for all signal components, leading to a high resolution and cross-term free time-frequency representation. This method is applicable to multicomponent and non-stationary signals without any a priori signal information.

- *The LOS is one of the most effective and robust QTFDs for non-stationary signal characterisation and classification:* The efficiency of the LOS has been tested under various SNR values by comparing the accuracy of IF estimation with other TFDs; the logarithmic

MSE of the proposed method is -4 at $\mathrm{SNR} = -3\mathrm{dB}$ when determining the IF of a multicomponent noisy (AWGN) signal, whereas at the same SNR the WVD and ordinal spectrogram provide -2.9 and -2.3 respectively. These results indicate that local adaptation and optimisation by the LOS can not only enhance the resolution and concentration but also shows the robustness under different noise conditions. The LOS also increases the detection and accuracy rate of different non-stationary signals. For example, the LOS achieves an average AUC of 0.887 in detecting burst and suppression patterns of a multichannel neonatal EEG dataset.

From Chapter 4:

- *Neonatal EEG background patterns can be well-characterised using state-of-the-art optimised and high-resolution QTFDs:* Optimised and high resolution QTFDs, developed in Chapter 4, were used to characterise neonatal EEG background patterns in the (t, f) domain. Different background patterns show different (t, f) characteristics in the (t, f) plane. In this way, different background patterns were classified manually by visual inspection.

- High-resolution QTFDs are not only useful for non-stationary signal representation but also useful to extract information for classification: High-resolution QTFDs are rich in information and take the signal's non-stationarity into account. Therefore, various dynamic features such as instantaneous frequency and instantaneous spectral ratio (ISPR), as well as (t, f) image-related features, were extracted for classification purpose. These dynamic features are expected to increase the classification accuracy rate.

- Simultaneous selection of feature subset and classifier parameter(s) ensures optimal use of a classifier and provides better accuracy: Classification using visual inspection is both time-consuming and subjective; therefore, automatic classification is needed. Feature and classifier parameter selection is one of the mandatory tasks. A feature selection technique is able to simplify the models to make them easier to interpret by researchers or users, to shorten training times and to enhance generalisation by reducing overfitting. In addition, tuning parameter(s) of a classifier plays an important role and needs to be optimised simultaneously with feature selection. A novel algorithm called hybrid feature selection (HFS) was proposed for this purpose.

- The proposed novel hybrid feature selection (HFS) algorithm, comprising mRMR and genetic algorithm, can simultaneously select prominent feature subset and tuning parameter(s) of a classifier: Feature selection and classifier tuning parameter selection are application-

dependent, meaning that the feature(s) and parameter(s) that are dominant in one application may be different in another. The proposed HFS algorithm simultaneously selects prominent feature subset and tuning parameter(s) of a classifier for the specific application. The mRMR method is used as a pre-processing stage in the HFS algorithm, whereas the genetic algorithm is used as wrapper-based feature selection algorithm. This has the advantage of reducing the computation load and the severity of the search operations involved in the genetic algorithm.

- Optimised RBF-SVM can provide better classification performance in neonatal EEG background pattern classification compared to optimal linear discriminant analysis (LDA), naive Bayes, logistic regression, neural network, and Bayesian network: The RBF-SVM optimised by HFS algorithm shows better classification accuracy in neonatal EEG background pattern classification than optimal linear discriminant analysis (LDA), naive Bayes, logistic regression, neural network, and Bayesian network. For example, the multichannel EEG with channel fusion techniques achieved 91.7% accuracy using optimised RBF-SVM classifier, versus 72.2%, 65.7%, 73.7%, 80.7%, and 73.9% in case of optimal LDA, naive Bayes, logistic regression, neural network, and Bayesian network respectively.

- Multichannel neonatal EEG can increase the classification accuracy and different fusion techniques such as channel fusion, feature fusion, decision fusion can be used for combining multichannel information: Multichannel information increases the classification accuracy. In this thesis, three different fusion techniques – channel fusion, feature fusion, and decision fusion – are considered to combine information extracted from newborn multichannel EEG. There is a trade-off between computational load and accuracy. The channel fusion technique is more computationally efficient than other fusion techniques as all the channels are combined before analysis. However, classification accuracy is lower (91.7%) than feature fusion (94.8%) and decision fusion (96.4%). On the other hand, the decision fusion technique is more computationally demanding than all other fusion techniques as it classifies the signal channel-by-channel and then utilises a decision fusion technique such as the Signal Strength Combination (SSC) algorithm. However, it shows the highest classification accuracy (96.4%).

- Class-specific feature subset (method-2) is more accurate than single feature subset (method-1) for neonatal EEG background classification: Method-2 is more accurate than method-1 in terms of classification accuracy. This is because method-1 selects a global features subset for all classes (or patterns), whereas method-2 selects a class-specific feature subset. Interestingly, the computation time for both methods is the same because method-1 utilises

all the EEG patterns (or classes) at the same time while method-2 utilises only two patterns at a time in a one-against-rest iterative fashion.

- A decision support system can be built based on the methodology presented in Chapter 4 which can be helpful in clinical settings. The methodology presented in this chapter (Chapter 4) could be used to implement a computer-based decision support system to monitor EEG patterns.

From Chapter 5:

- *EEG signal-based features can predict the neurodevelopmental outcome in term neonates with HIE:* Both local and global features contribute to accurate outcome prediction. In addition, several (t, f) features have been selected by the HFS algorithm, suggesting that joint time-frequency analysis, i.e. optimised TFD, can improve the prediction of neurodevelopmental outcome. The cross-validation accuracy of this system is 88.4% in predicting good/poor outcome and 89.6% in predicting optimal/suboptimal outcome.

- An early decision support system could be a very effective tool for clinicians both to counsel parents about likely future outcome and to guide early clinical management decisions such as the treatment of hypotension or seizures: A decision support system (DSS) has been created that provides a continuous output of probable neurodevelopmental outcome that, within the first 24 h after birth, predicts the neurodevelopmental outcome at 2 years.

6.3 Future directions

There are few firm end-points in science; rather it is constantly modified, updated and applied to new applications with the aim of hopefully making the world a better place, or simply reducing the unknowns. With this in mind, I propose this short list of the most useful work to implement or extend the approaches described in this project.

- *Applications of the proposed optimised QTFDs to other problems*: The proposed optimised QTFDs could be applied to characterise and classify different non-stationary signal analyses in various settings or industries such as HRV analysis, fetal movement analysis, machine condition monitoring etc.

- *Validation of the proposed approach using a larger database*: Further study should be carried out to further validate the proposed methods on a different and larger clinical dataset.

- *Implementation of a computer-based decision support system for EEG monitoring*: This possible application of the EEG background patterns classification system is briefly described in Chapter 4. This could be implemented in a standalone computer for bed-side monitoring in NICUs.

- *Implementation of an outcome prediction model in the clinical setting*: An early prediction of neurodevelopmental outcome would be beneficial to guide early clinical management of newborns with HIE and at risk of cerebral injury. This could reduce morbidity and mortality rates and improve health outcomes.

- *Improvement of the accuracy of neurodevelopmental outcome prediction:* As mentioned in section 5.6.1, the accuracy of neurodevelopmental outcome prediction could be further improved by fusing the EEG-based features with the features extracted from HRV, MRI, perinatal events, Apgar score and blood biochemistry. In addition, ordinal (4-level) rather than binary outcome prediction could be used. Furthermore, separate outcomes for 'motor' and 'cognitive' as well as statistical comparison could strengthen the analysis.

- *Packaging:* Integrating and packaging the proposed methods, and any future modifications, in a user-friendly graphical user interface (GUI) would be very useful in future research.

- *The use of a 'deep learning' approach:* Recently 'deep learning' has gained popularity due to its potential for superior classification performance. One of the advantages of deep learning is that, unlike other machine learning approaches (e.g. SVM), users do not need to extract features (called 'Feature Engineering'), which is one of the time-consuming elements of machine learning. The deep learning approach automatically extracts features from the data itself and provides decision output. In a sense, this approach is 'optimal', i.e. it is end-to-end from data to decision. However, it requires a large amount of data and many hyper-parameters need to be optimised. This approach is computationally expensive and, unlike SVM, does not yet have a strong theoretical background. In addition, the features created by its deep hidden networks are often clinically non-interpretable and it can be difficult to determine the underlying reasoning process leading to a decision. Therefore, this approach is not clinical-friendly. However, it could be a useful engineering contribution to explore this approach and compare it with the methods developed in this project.

- *The use of an 'online learning' approach:* Online learning might be very useful if the research is conducted in multiple sites where a continuous stream of data is generated and available. In such settings, the learning model could be updated with the incoming data. Such a dynamic

model would be more adapted to classify difficult examples than a fixed model and increase the classification accuracy. Finally, such an optimised system would need to be evaluated in further clinical trials.

Appendices

Appendix 3.A: Computer codes used in this study

All analyses have been done in Matlab environment and the TFSAP toolbox has been used to generate the QTFDs. The TFSAP toolbox can be downloaded from the link: http://time-frequency.net/tf/ .

Appendix 3.B: FrSM calculation (Pseudocode)

Function FrSM=FrSM_calculation (STFrFT, L, LOS)

Calculate total time index (N) of STFrFT i.e. N=size (STFrFT, 1).

Calculate STFrFT and LOS using Eqn. (3.20) and (3.21). *respectively. LOS is the initial distribution i.e.* $FrSM_0(n,k) = LOS$.

Assign $FrSM(n,k) = FrSM_0(n,k)$.

Define L

Iterate $j = 1$ *to L*

Calculate $FrSM(1+j:N-j,:) = FrSM(1+j:N-j,:) + 2Re[STFrFT(1:N-2*j,:).*$
$conj(STFrFT(1+2*j:N,:))]$

end of the loop.

end of the function.

Appendix 3.C: IF estimation method

The main algorithm of the components extraction method are given below [129, 247]:

1. Initialise $k = 1, k \in K$, where K is the number of components.
2. Find the maximum energy location,(t_0, f_0) in the $TFD(t,f)$ plane. The IF of the kth component at t_0 is computed $IF_k(t_0) = f_0$. For predetermined l, assign zero around this maximum point, i.e., $TFD(t_0, f) = 0$ for $f \in [f_0 - l, f_0 + l]$.
3. For sampling frequency f_s, assign $t_r = t_0 + (1/f_s)$, $t_l = t_0 - (1/f_s)$, $f_l = f_0$ and $f_r = f_0$. IF of the signal component confined by the TFD area $t < t_0$ and $t > t_0$ is computed by the following procedures:
    a) Find the frequency f_0' around this maximum point in $TFD(t_r, f)$ for $f \in [f_r - l, f_r + l]$ and f_1' around the maximum point in $TFD(t_l, f)$ for $f \in [f_l - l, f_l + l]$. IF at t_r and t_l are calculated to be $IF_k(t_r) = f_0'$ and $IF_k(t_l) = f_1'$.

b) Assign zero around the maximum point, i.e.,$TFD(t_r, f)$=0 for $f \in [f_r - l, f_r + l]$ and $TFD(t_l, f)$=0 for $f \in [f_l - l, f_l + l]$ and then assign $t_r = t_r + (1/f_s)$, $t_l = t_l - (1/f_s)$.

c) Repeat step (a) and (b) until the TFD boundary is reached.

4. Increment k from step 2 untill all the components have been extracted.

Appendix 3.D: EEG pre-processing and feature extraction

Pre-processing and channels combination

Firstly, the DC component was removed from EEG signals. Then, following the approach presented in [248], multichannel signals were combined into one to reduce the overall runtime and computational complexity using:

$$y[n] = \frac{1}{C}\sqrt{\sum_{i=1}^{C} x_i^2[n]}$$

where $x_i[n]; i = 1, 2, \dots, C$ is the signal acquired from the i^{th} channel and $C = 12$ is the number of channels. Finally, $y[n]$ is band pass filtered within the frequency band of $[0.5 - 16]$ Hz as most of the signal energies of B-S signals are within this frequency bands.

T-F Feature Extraction

TFDs seek to act as a bi-dimensional energy density in the time-frequency domain [138]. Therefore, the main advantage of (t, f) domain-based classification is the flexibility to form features from a 2D representation. Seven features were extracted from TFDs and these features have already been justified and used in several studies [249]. They are discussed below.

Singular Value Decomposition (SVD)-based features are extracted from the singular values of $\rho[n, k]$, i.e. $s_i, i = 1,2, \dots .. N$, which has been shown to be useful in EEG signal classification [250]. The maximum of the singular values was also used as first feature, F_1. The second feature was defined as the Shannon entropy of the singular values given by:

$$F_2 = -\sum_{i=1}^{N} \bar{s_i} \, log \, \bar{s_i}$$

$\bar{s_i}$ is the normalised value of the singular values. The energy concentration measure is used as a feature, F_3.

The IF of the signal can be estimated using the first order moment of its TFD as [249]:

$$f_i[n] = \frac{f_s}{2M}\frac{\sum_{k=1}^{M} k.\rho[n,k]}{\sum_{k=1}^{M}\rho[n,k]}$$

The deviation of IF was used as a feature, i.e.

$$F_4 = max(f_i[n]) - min(f_i[n])$$

(t, f) Renyi entropy is used as the feature, F_5.

Two time-domain signal features (mean and standard deviation) were extended to (t, f) domain and used in this study. The feature F_6 and F_7 can be expressed as:

$$F_6 = m_{(t,f)} = \frac{1}{NM}\sum_{n=1}^{N}\sum_{K=1}^{M} TFD[n,k]$$

$$F_7 = \sigma^2_{(t,f)} = \frac{1}{NM}\sum_{n=1}^{N}\sum_{K=1}^{M}\left(TFD[n,k] - m_{(t,f)}\right)^2$$

Appendix 4. A Time-frequency matching pursuit (TFMP)

The TFMP is the matching pursuit (MP) method equipped with a TF dictionary. MP is a greedy iterative atomic decomposition technique which represents a signal x using an over-complete dictionary D. This can be mathematically expressed as [171] :

$$x = \sum_{i=1}^{m} \alpha_{\gamma_i} d_{\gamma i} + R_m = \hat{x}_m + R_m$$

where $R_m = x - \hat{x}_m$ denotes the residual of the signal after m iterations, $\hat{x}_m$ is signal approximation and α_{γ_i} is the expansion coefficient associated with decomposed atom $d_{\gamma i} \in D$.

Assuming $R_0 = x$ and $\hat{x}_0 = 0$, the MP algorithm computes all inner products of the signal with atoms in dictionary D. The atom $d_{\gamma i}$ for which the largest magnitude of the inner product is selected as the first element with the following criteria:

$$\left|\langle x, d_{\gamma i}\rangle\right| \geq \left|\langle x, d_{\gamma(i-1)}\rangle\right|$$

After the first iteration, the selected atom is subtracted from x to yield residual R_1. At each iteration $(k, k = 1,2, \ldots \ldots . m)$, the MP algorithm identifies the atom that best correlates with the signal residual and then adds the scalar multiple of that atom to the current approximation.

Appendix 4. B Maximum relevance minimum redundancy (mRMR)

The mRMR method uses mutual information for calculating the correlation/dependency between variables (or features) and target classes, and among features. The dependency $I(x, y)$ between any two features x and y is defined as [194, 251]:

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy$$

where $p(x)$, $p(y)$ and $p(x, y)$ represent the probability density functions of x ,y and joint (x, y) respectively. Thus, the *maximum relevance* between feature x and target class $C = (1,2, \dots, Q)$ represented by $D = I(x, c)$ is used to rank the features. *Minimum redundancy* selects the maximally dissimilar features and can be represented by $R = \frac{1}{m}\sum_{j=1}^{m} I(x, x_j)$ between candidate features x and ranked features $x_j (i = 1, \dots, m)$. Both constraints formulate the following optimisation problem [194]:

$$\max \Phi(D, R) = D - R$$

A search method is used to find the optimal feature subset defined by $\Phi(.)$ operator. The mRMR implementation can be downloaded at: http://penglab.janelia.org/proj/mRMR/.

Appendix 4. C The equivalency between diagnostic tests and clinical research

Introduction

In Chapter 4 and 5, results are mainly represented in terms of sensitivity, specificity and PPV (positive predictive value), Fmeasure[8] and accuracy. These are the standard quality measures used in machine learning and classification problems [29]. While machine learning researchers use these measures to assess a test (classification) performance, clinical researchers are interested to know how significant a test is in terms of p-value, power etc. Sensitivity, specificity and PPV are related to diagnostic test (DT), whereas power and p-value are related to research studies i.e. statistical hypothesis testing (SHT). DT and SHT are two scenarios of making conditional decision. There exists an equivalent relationship between them and the purpose of this appendix is to show this relationship [252]. Finally,

[8] Fmeasure is defined as

$$Fmeasure, F_\beta = \frac{(1 + \beta^2)(PPV \times Sensitivity)}{\beta^2 . (PPV + Sensitivity)}$$

where β is the weight parameter that controls weight between PPV and $Sensitivity$. It is usually set to 1 in the $Fmeasure$ calculation to provide same importance to $PPV$and $Sensitivity$ and hence used in the thesis.

this appendix provides the justification of choosing sensitivity, specificity and PPV, Fmeasure and accuracy used in Chapter 4 and 5. Note that, since DT and SHT are two different scenarios, we cannot equalize them directly. We can provide an equivalent relationship between them. For example, there exists an equivalent relationship between specificity and p-value which is descried below:

Definitions

Specificity is the (conditional) probability that a healthy person tests negative: $Prob(Negative \mid Healthy)$ [253]. Specificity is also known as true negative rate (TNR).

P-value is the conditional probability of observing a value of the test statistic at least as extreme as that observed in the study, if in fact the null hypothesis is true [253]. SHT is performed on the null hypothesis[9], and DT is performed on the alternative hypothesis. An example of null hypothesis is given below:

Hypothesis (H): "Support vector Machine (SVM) can classify seizure pattern in newborn EEG"

Null hypothesis (H_0): "SVM cannot classify seizure pattern in newborn EEG"

Example: Let us take an example from this thesis i.e., classification of seizure/non-seizure EEG pattern in newborns. In the 2×2 confusion matrix, seizure is a positive class while non-seizure is a negative class. In this case, specificity is defined as the proportion of non-seizure EEG that correctly classified as non-seizure by a classifier. However, misclassification can occur by a classifier, e.g., a proportion of non-seizure EEG can incorrectly classify as seizure. In terms of DT, this error is called false-positive rate (FPR) and equals to $(1 - specificity)$ or (0.05 in the case of 95% specificity). Mathematically,

$$FPR = (1 - specificity)$$

Equivalently, in terms of SHT, this error is called the probability of type-I error [253]. So, judging the evidence by rejecting the null hypothesis (declaring statistical significance) based on p = 0.05 alone is analogous to accepting a positive diagnostic test result based on its 95% specificity alone [253]. In this way, we can relate p-value with Type-I error or false positive rate $(1 - Specificity)$ or α value. Here, $\alpha = (1 - Specificity)$ would be a maximum allowed false positive rate. This is called "cut-off" value or statistical significant level and the common choice of α is 0.01

[9] A null hypothesis is a hypothesis that says there is no statistical significance between the two variables in the hypothesis. The null hypothesis (H_0) is a hypothesis which the researcher tries to disprove, reject or nullify. On the other hand, an alternative hypothesis simply is the inverse, or opposite, of the null hypothesis.

or 0.05 i.e. 1% or 5% significant level. We can relate the other measures in the same way as shown in the table below:

Table Appendix 4.C: The equivalency between diagnostic tests and statistical hypothesis testing (SHT) in research studies (Adapted from [252])

Diagnostic Tests	Statistical Hypothesis Testing (SHT)
Absence of disease	Truth of null hypothesis
Presence of disease	Truth of research (alternative)
Hypothesis Positive result (outside normal limits)	Positive result (reject null hypothesis)
Negative result (within normal limits)	Negative result (fall to reject null hypothesis)
Sensitivity	Power
False-positive rate (1 -specificity)	p value
Prior probability of disease	Prior probability of research hypothesis
Predictive value of a positive (or negative) test result	Predictive value of a positive or negative study

Discussion

A study with a very low P value is like a test with very high specificity: both give few false-positive results, but may require careful consideration of other possible explanations such as power and sample size etc. [252]. Despite the wide use of p-value in research study, there are certain limitations such as, it ignores the concept of prior probability and power of the study [254]. Likewise, specificity only provides the true negative rate, e.g., proportion of non-seizure EEG that correctly classified as non-seizure. It does not tell about any other aspect of a classifier; e.g., true positive rate, total accuracy etc. Therefore, one should not solely rely on p-value or specificity. Different performance measures such as sensitivity, PPV, accuracy etc. describe a classifier from different perspectives. Moreover, what the clinician wants to know is the probability that the subject (e.g. newborn) has the disease given that the test is positive. This is the PPV [186].

Conclusion

For the above rationale, several quality measures such as sensitivity, precision (=PPV), Fmeasure and accuracy have been used in Chapter 4 and Chapter 5 to describe the classification performance from different perspectives. A detailed definition of the above quality measures has already been provided in Section 4.4.5.

Appendix 5. A: The HFS selected t- and f- features predicting good and poor outcome

Selected t- and f- features for the prediction of good and poor outcome
Kurtosis
Bout percentage in delta band
Bout percentage in theta band
Peak Coherence index
Shannon entropy
Mean bout in alpha band
Mean PLV
Hurst Exponential
Mean IBI
Mean bout in theta band
Bout percentage in alpha band
SEF 90
Mean of lower envelope
Difference in Spectral Flatness between left and right channel
RMS

Appendix 5. B: The HFS selected t- and f- features predicting optimal and suboptimal outcome.

Selected t- and f- features for the prediction of optimal and suboptimal outcome
Kurtosis
Bout percentage in delta
Bout percentage in theta
Bout mean alpha
PLV mean
Peak coherence index
Skewness
Spectral Flatness
SEF 80
Shannon entropy

References

[1] V. E. Johnson, W. Stewart, and D. H. Smith, "Axonal pathology in traumatic brain injury," *Experimental Neurology,* vol. 246 pp. 35-43, 2013.

[2] M. Douglas-Escobar and M. D. Weiss, "Hypoxic-ischemic encephalopathy: A review for the clinician," *JAMA Pediatrics,* vol. 169 (4), pp. 397-403, 2015.

[3] R. Lawrence, A. Mathur, S. Nguyen The Tich, J. Zempel, and T. Inder, "A pilot study of continuous limited-channel aEEG in term infants with encephalopathy," *Journal of Pediatrics,* vol. 154 (6), pp. 835-41.e1, 2009.

[4] J. Bryce, C. Boschi-Pinto, K. Shibuya, and R. E. Black, "WHO estimates of the causes of death in children," *The Lancet,* vol. 365 (9465), pp. 1147-1152, 2005.

[5] J. E. Lawn, S. Cousens, and J. Zupan, "4 million neonatal deaths: when? Where? Why?," *The Lancet,* vol. 365 (9462), pp. 891-900, 2005.

[6] B. H. Walsh, D. I. Broadhurst, R. Mandal, D. S. Wishart, G. B. Boylan, L. C. Kenny, *et al.*, "The metabolomic profile of umbilical cord blood in neonatal hypoxic ischaemic encephalopathy," *PloS One,* vol. 7 (12), p. e50520, 2012.

[7] M. D. Donovan, B. T. Griffin, L. Kharoshankaya, J. F. Cryan, and G. B. Boylan, "Pharmacotherapy for Neonatal Seizures: Current Knowledge and Future Perspectives," *Drugs,* vol. 76 (6), pp. 647-661, 2016.

[8] C. Robertson, N. Finer, and M. Grace, "School performance of survivors of neonatal encephalopathy associated with birth asphyxia at term," *The Journal of Pediatrics,* vol. 114 (5), pp. 753-760, 1989.

[9] S. Shankaran, E. Woldt, T. Koepke, M. P. Bedard, and R. Nandyal, "Acute neonatal morbidity and long-term central nervous system sequelae of perinatal asphyxia in term infants," *Early Human Development,* vol. 25 (2), pp. 135-148, 1991.

[10] L. Filippi, P. Fiorini, M. Daniotti, S. Catarzi, S. Savelli, C. Fonda, *et al.*, "Safety and efficacy of topiramate in neonates with hypoxic ischemic encephalopathy treated with hypothermia (NeoNATI)," *BMC Pediatrics,* vol. 12 (1), p. 144, 2012.

[11] (2016). *Queensland Clinical Guideline: Hypoxic-ischaemic encephalopathy (HIE), https://www.health.qld.gov.au/qcg/documents/g-hie.pdf*.

[12] B. H. Walsh, D. M. Murray, and G. B. Boylan, "The use of conventional EEG for the assessment of hypoxic ischaemic encephalopathy in the newborn: a review," *Clinical Neurophysiology,* vol. 122 (7), pp. 1284-94, 2011.

[13] S. P. Miller, B. Latal, H. Clark, A. Barnwell, D. Glidden, A. J. Barkovich, *et al.*, "Clinical signs predict 30-month neurodevelopmental outcome after neonatal encephalopathy," *American Journal of Obstetrics and Gynecology,* vol. 190 (1), pp. 93-9, 2004.

[14] N. Robertson and J. Wyatt, "The magnetic resonance revolution in brain imaging: impact on neonatal intensive care," *Archives of Disease in Childhood Fetal and Neonatal Edition,* vol. 89 (3), pp. F193-F197, 2004.

[15] E. Shany, E. Goldstein, S. Khvatskin, M. D. Friger, N. Heiman, M. Goldstein, *et al.*, "Predictive value of amplitude-integrated electroencephalography pattern and voltage in asphyxiated term infants," *Pediatric Neurology,* vol. 35 (5), p. 335, 2006.

[16] L. G. van Rooij, M. C. Toet, D. Osredkar, A. C. van Huffelen, F. Groenendaal, and L. S. de Vries, "Recovery of amplitude integrated electroencephalographic background patterns within 24 hours of perinatal asphyxia," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 90 (3), pp. F245-FF251, 2005.

[17] H. J. Ter Horst, C. Sommer, K. A. Bergman, J. M. Fock, T. W. Van Weerden, and A. F. Bos, "Prognostic significance of amplitude-integrated EEG during the first 72 hours after birth in severely asphyxiated neonates," *Pediatric Research,* vol. 55 (6), pp. 1026-1033, 2004.

[18] L. M. Douglass, J. Y. Wu, N. P. Rosman, and C. E. Stafstrom, "Burst suppression electroencephalogram pattern in the newborn: predicting the outcome," *Journal of Child Neurology,* vol. 17 (6), pp. 403-408, 2002.

[19] H. C. Glass and D. M. Ferriero, "Treatment of hypoxic-ischemic encephalopathy in newborns," *Current Treatment Options in Neurology,* vol. 9 (6), pp. 414-23, 2007.

[20] V. Apgar, "A proposal for a new method of evaluation of the newborn infant," *Curr Res Anesth Analg,* vol. 32 (4), pp. 260-7, 1953.

[21] S. Cnattingius, M. Norman, F. Granath, G. Petersson, O. Stephansson, and T. Frisell, "Apgar Score Components at 5 Minutes: Risks and Prediction of Neonatal Mortality," 2017.

[22] N. Foldvary-Schaefer and M. M. Grigg-Damberger, "Identifying Interictal and Ictal Epileptic Activity in Polysomnograms," *Sleep Medicine Clinics,* vol. 7 (1), pp. 39-58, 2012.
[23] G. L. Holmes and C. T. Lombroso, "Prognostic value of background patterns in the neonatal EEG," *Journal of Clinical Neurophysiology,* vol. 10 (3), pp. 323-352, 1993.
[24] K. Palmu, N. Stevenson, S. Wikström, L. Hellström-Westas, S. Vanhatalo, and J. M. Palva, "Optimization of an NLEO-based algorithm for automated detection of spontaneous activity transients in early preterm EEG," *Physiological Measurement,* vol. 31 (11), p. N85, 2010.
[25] K. Palmu, "Automatic detection of spontaneous activity transients in preterm electroencephalography," PhD Thesis, Aalto University, 2013, http://lib.tkk.fi/Lic/2013/urn100758.pdf.
[26] J. C. Rowe, G. L. Holmes, J. Hafford, D. Baboval, S. Robinson, A. Philipps, *et al.*, "Prognostic value of the electroencephalogram in term and preterm infants following neonatal seizures," *Electroencephalography and Clinical Neurophysiology,* vol. 60 (3), pp. 183-196, 1985.
[27] H. Witte, B. Schack, M. Helbig, P. Putsche, C. Schelenz, K. Schmidt, *et al.*, "Quantification of transient quadratic phase couplings within EEG burst patterns in sedated patients during electroencephalic burst-suppression period," *Journal of Physiology-Paris,* vol. 94 (5-6), pp. 427-434, 2000.
[28] M. Hosoya, H. Ushiku, H. Arakawa, and A. Morikawa, "Low-voltage activity in EEG during acute phase of encephalitis predicts unfavorable neurological outcome," *Brain and Development,* vol. 24 (3), pp. 161-165, 2002.
[29] B. Boashash and S. Ouelha, "Automatic signal abnormality detection using time-frequency features and machine learning: a newborn EEG seizure case study," *Knowledge-Based Systems,* vol. 106 pp. 38-50, 2016.
[30] L. C. Weeke, G. B. Boylan, R. M. Pressler, B. Hallberg, M. Blennow, M. C. Toet, *et al.*, "Role of EEG background activity, seizure burden and MRI in predicting neurodevelopmental outcome in full-term infants with hypoxic-ischaemic encephalopathy in the era of therapeutic hypothermia," *European Journal of Paediatric Neurology,* vol. 20 (6), pp. 855-864, 2016.
[31] S. Johnson, T. Moore, and N. Marlow, "Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used?," *Pediatric Research,* vol. 75 (5), pp. 670-674, 2014.
[32] J. J. Volpe, *Neurology of the Newborn*, 5th ed.: Elsevier Health Sciences, 2008.
[33] B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, 2nd ed.: Academic Press, Dec. 2015.
[34] Y. Yang, Z. K. Peng, X. J. Dong, W. M. Zhang, and G. Meng, "General Parameterized Time-Frequency Transform," *IEEE Transactions on Signal Processing,* vol. 62 (11), pp. 2751-2764, 2014.
[35] B. Boashash, G. Azemi, and N. Ali Khan, "Principles of time–frequency feature extraction for change detection in non-stationary signals: Applications to newborn EEG abnormality detection," *Pattern Recognition,* vol. 48 (3), pp. 616-627, 2015.
[36] S. R. Mathieson, N. J. Stevenson, E. Low, W. P. Marnane, J. M. Rennie, A. Temko, *et al.*, "Validation of an automated seizure detection algorithm for term neonates," *Clinical Neurophysiology,* vol. 127 (1), pp. 156-168, 2016.
[37] I. Despotovic, P. J. Cherian, M. Vos, H. Hallez, W. Deburchgraeve, P. Govaert, *et al.*, "Relationship of EEG sources of neonatal seizures to acute perinatal brain lesions seen on MRI: a pilot study," *Human Brain Mapping,* vol. 34 (10), pp. 2402-2417, 2013.
[38] M. Čić, J. Šoda, and M. Bonković, "Automatic classification of infant sleep based on instantaneous frequencies in a single-channel EEG signal," *Computers in Biology and Medicine,* vol. 43 (12), pp. 2110-2117, 2013.
[39] B. Mehta, R. Hunt, K. Walker, and N. Badawi, "Evaluation of Preoperative Amplitude-Integrated Electroencephalography (aEEG) Monitoring for Predicting Long-Term Neurodevelopmental Outcome Among Infants Undergoing Major Surgery in the Neonatal Period," *Journal of Child Neurology,* vol. 31 (11), pp. 1276-81, 2016.
[40] R. A. Shellhaas, "Continuous long-term electroencephalography: The gold standard for neonatal seizure diagnosis," *Seminars in Fetal and Neonatal Medicine,* vol. 20 (3), pp. 149-153, 2015.
[41] J. M. Dunne, D. Wertheim, P. Clarke, O. Kapellou, P. Chisholm, J. P. Boardman, *et al.*, "Automated electroencephalographic discontinuity in cooled newborns predicts cerebral MRI and neurodevelopmental outcome," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* pp. F2015-F309697, 2016.

[42] M. Perlman and P. S. Shah, "Hypoxic-ischemic encephalopathy: challenges in outcome and prediction," *The Journal of Pediatrics,* vol. 158 (2), pp. e51-e54, 2011.
[43] M.-C. Lai and S.-N. Yang, "Perinatal hypoxic-ischemic encephalopathy," *BioMed Research International,* vol. 2011 p. , 2011, http://doi.org/10.1155/2011/609813.
[44] H. van Laerhoven, T. R. de Haan, M. Offringa, B. Post, and J. H. van der Lee, "Prognostic tests in term neonates with hypoxic-ischemic encephalopathy: a systematic review," *Pediatrics,* vol. 131 (1), pp. 88-98, 2013.
[45] S. T. Bjorkman, S. M. Miller, S. E. Rose, C. Burke, and P. B. Colditz, "Seizures are associated with brain injury severity in a neonatal model of hypoxia-ischemia," *Neuroscience,* vol. 166 (1), pp. 157-67, 2010.
[46] M. A. Navakatikyan, P. B. Colditz, C. J. Burke, T. E. Inder, J. Richmond, and C. E. Williams, "Seizure detection algorithm for neonates based on wave-sequence analysis," *Clinical Neurophysiology,* vol. 117 (6), pp. 1190-203, 2006.
[47] S. Hamelin, N. Delnard, F. Cneude, T. Debillon, and L. Vercueil, "Influence of hypothermia on the prognostic value of early EEG in full-term neonates with hypoxic ischemic encephalopathy," *Clinical Neurophysiology,* vol. 41 (1), pp. 19-27, 2011.
[48] D. L. Schriger, D. G. Altman, J. A. Vetter, T. Heafner, and D. Moher, "Forest plots in reports of systematic reviews: a cross-sectional study reviewing current practice," *International Journal of Epidemiology,* vol. 39 (2), pp. 421-429, 2010.
[49] R. T. C. Collaboration, "Review Manager (RevMan). 5.0," *Copenhagen, The Nordic Cochrane Centre: The Cochrane Collaboration,* 2008.
[50] S. Stata, "Release 11," *Statistical Software,* 2009.
[51] K. Nash, S. Bonifacio, H. Glass, J. Sullivan, A. Barkovich, D. Ferriero, *et al.*, "Video-EEG monitoring in newborns with hypoxic-ischemic encephalopathy treated with hypothermia," *Neurology,* vol. 76 (6), pp. 556-562, 2011.
[52] B. Hallberg, K. Grossmann, M. Bartocci, and M. Blennow, "The prognostic value of early aEEG in asphyxiated infants undergoing systemic hypothermia treatment," *Acta Paediatrica,* vol. 99 (4), pp. 531-536, 2010.
[53] A. Csekő, M. Bangó, P. Lakatos, J. Kárdási, L. Pusztai, and M. Szabó, "Accuracy of amplitude-integrated electroencephalography in the prediction of neurodevelopmental outcome in asphyxiated infants receiving hypothermia treatment," *Acta Paediatrica,* vol. 102 (7), pp. 707-11, 2013.
[54] D. Azzopardi, I. Guarino, C. Brayshaw, F. Cowan, D. Price-Williams, A. D. Edwards, *et al.*, "Prediction of neurological outcome after birth asphyxia from early continuous two-channel electroencephalography," *Early Human Development,* vol. 55 (2), p. 113, 1999.
[55] M. Thoresen, L. Hellström-Westas, X. Liu, and L. S. de Vries, "Effect of hypothermia on amplitude-integrated electroencephalogram in infants with asphyxia," *Pediatrics,* vol. 126 (1), pp. e131-e39, 2010.
[56] G. Ancora, E. Maranella, S. Grandi, F. Sbravati, E. Coccolini, S. Savini, *et al.*, "Early predictors of short term neurodevelopmental outcome in asphyxiated cooled infants. A combined brain amplitude integrated electroencephalography and near infrared spectroscopy study," *Brain and Development,* vol. 35 (1), pp. 26-31, 2011.
[57] K. Gucuyener, S. Beken, E. Ergenekon, S. Soysal, İ. Hirfanoglu, O. Turan, *et al.*, "Use of amplitude-integrated electroencephalography (aEEG) and near infrared spectroscopy findings in neonates with asphyxia during selective head cooling," *Brain and Development,* vol. 34 (4), pp. 280-286, 2012.
[58] L. Zhang, Z. Y. Zong, Y. B. Liu, H. Ye, and X. J. Lv, "Pcr versus serology for diagnosing mycoplasma pneumoniae infection: A systematic review & meta-analysis," *Indian Journal of Medical Research,* vol. 134 (9), pp. 270-280, 2011.
[59] J. M. Rennie, G. Chorley, G. B. Boylan, R. Pressler, Y. Nguyen, and R. Hooper, "Non-expert use of the cerebral function monitor for neonatal seizure detection," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 89 (1), pp. F37-F40, 2004.
[60] R. A. Shellhaas, P. R. Gallagher, and R. R. Clancy, "Assessment of neonatal electroencephalography (EEG) background by conventional and two amplitude-integrated EEG classification systems," *The Journal of Pediatrics,* vol. 153 (3), pp. 369-74, 2008.
[61] G. Boylan, L. Burgoyne, C. Moore, B. O'Flaherty, and J. Rennie, "An international survey of EEG use in the neonatal intensive care unit," *Acta Paediatrica,* vol. 99 (8), pp. 1150-1155, 2010.
[62] D. M. Murray, G. B. Boylan, C. A. Ryan, and S. Connolly, "Early EEG Findings in Hypoxic-Ischemic Encephalopathy Predict Outcomes at 2 Years," *Pediatrics,* vol. 124 (3), pp. E459-E467, 2009.

[63] R. Pressler, G. Boylan, M. Morton, C. Binnie, and J. Rennie, "Early serial EEG in hypoxic ischaemic encephalopathy," *Clinical Neurophysiology,* vol. 112 (1), pp. 31-37, 2001.
[64] E. Zeinstra, J. M. Fock, J. H. Begeer, T. W. van Weerden, N. M. Maurits, and M. J. Zweens, "The prognostic value of serial EEG recordings following acute neonatal asphyxia in full-term infants," *European Journal of Paediatric Neurology,* vol. 5 (4), pp. 155-160, 2001.
[65] H. v. Lieshout, J. Jacobs, J. Rotteveel, W. Geven, and M. v. t. Hof, "The prognostic value of the EEG in asphyxiated newborns," *Acta Neurologica Scandinavica,* vol. 91 (3), pp. 203-207, 1995.
[66] K. Watanabe, S. Miyazaki, K. Hara, and S. Hakamada, "Behavioral state cycles, background EEGs and prognosis of newborns with perinatal hypoxia," *Electroencephalography and Clinical Neurophysiology,* vol. 49 (5), pp. 618-625, 1980.
[67] D. Selton and M. Andre, "Prognosis of Hypoxic-Ischemic Encephalopathy in Full-term newborns-Value of Neonatal Electroencephalography," *Neuropediatrics,* vol. 28 (5), pp. 276-280, 1997.
[68] E. Mariani, B. Scelsa, L. Pogliani, P. Introvini, and G. Lista, "Prognostic value of electroencephalograms in asphyxiated newborns treated with hypothermia," *Pediatric Neurology,* vol. 39 (5), pp. 317-324, 2008.
[69] G. Holmes, J. Rowe, J. Hafford, R. Schmidt, M. Testa, and A. Zimmerman, "Prognostic value of the electroencephalogram in neonatal asphyxia," *Electroencephalography and Clinical Neurophysiology,* vol. 53 (1), pp. 60-72, 1982.
[70] N. Finer, C. Robertson, R. Richards, L. Pinnell, and K. Peters, "Hypoxic-ischemic encephalopathy in term neonates: perinatal factors and outcome," *The Journal of pediatrics,* vol. 98 (1), pp. 112-117, 1981.
[71] N. N. Finer, C. M. Robertson, K. L. Peters, and J. H. Coward, "Factors affecting outcome in hypoxic-ischemic encephalopathy in term infants," *Archives of Pediatrics and Adolescent Medicine,* vol. 137 (1), pp. 21-25, 1983.
[72] C. Pezzani, M.-F. Radvanyi-Bouvet, J. Relier, and N. Monod, "Neonatal electroencephalography during the first twenty-four hours of life in full-term newborn infants," *Neuropediatrics,* vol. 17 (1), pp. 11-18, 1986.
[73] S. Shankaran, A. Pappas, S. A. McDonald, A. R. Laptook, R. Bara, R. A. Ehrenkranz, *et al.*, "Predictive value of an early amplitude integrated electroencephalogram and neurologic examination," *Pediatrics,* vol. 128 (1), pp. e112-e120, 2011.
[74] A. R. Horn, G. H. Swingler, L. Myer, L. L. Linley, M. S. Raban, Y. Joolay, *et al.*, "Early clinical signs in neonates with hypoxic ischemic encephalopathy predict an abnormal amplitude-integrated electroencephalogram at age 6 hours," *BMC Pediatrics,* vol. 13 (1), p. 52, 2013.
[75] L. Hellström-Westas, I. Rosén, L. De Vries, and G. Greisen, "Amplitude-integrated EEG classification and interpretation in preterm and term infants," *NeoReviews,* vol. 7 (2), pp. e76-e87, 2006.
[76] K. Aso, M. S. Scher, and M. A. Barmada, "Neonatal electroencephalography and neuropathology," *Journal of Clinical Neurophysiology,* vol. 6 (2), pp. 103-23, 1989.
[77] D. Wertheim, E. Mercuri, J. Faundez, M. Rutherford, D. Acolet, and L. Dubowitz, "Prognostic value of continuous electroencephalographic recording in full term infants with hypoxic ischaemic encephalopathy," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 71 (2), pp. F97-F102, 1994.
[78] N. Laroia, R. Guillet, J. Burchfiel, and M. C. McBride, "EEG Background as Predictor of Electrographic Seizures in High-Risk Neonates," *Epilepsia,* vol. 39 (5), pp. 545-51, 1998.
[79] E. Biagioni, L. Bartalena, A. Boldrini, R. Pieri, and G. Cioni, "Constantly discontinuous EEG patterns in full-term neonates with hypoxic-ischaemic encephalopathy," *Clinical Neurophysiology,* vol. 110 (9), pp. 1510-1515, 1999.
[80] M. El-Ayouty, H. Abdel-Hady, S. El-Mogy, H. Zaghlol, M. El-Beltagy, and H. Aly, "Relationship between electroencephalography and magnetic resonance imaging findings after hypoxic-ischemic encephalopathy at term," *American Journal of Perinatology,* vol. 24 (08), pp. 467-473, 2007.
[81] M. C. Toet, W. van der Meij, L. S. de Vries, C. S. Uiterwaal, and K. C. van Huffelen, "Comparison between simultaneously recorded amplitude integrated electroencephalogram (cerebral function monitor) and standard electroencephalogram in neonates," *Pediatrics,* vol. 109 (5), pp. 772-79, 2002.
[82] M. Toet, L. Hellström-Westas, F. Groenendaal, P. Eken, and L. De Vries, "Amplitude integrated EEG 3 and 6 hours after birth in full term neonates with hypoxic–ischaemic encephalopathy," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 81 (1), pp. F19-F23, 1999.

[83] M. D. Bourez-Swart, L. van Rooij, C. Rizzo, L. S. de Vries, M. C. Toet, T. A. Gebbink, *et al.*, "Detection of subclinical electroencephalographic seizure patterns with multichannel amplitude-integrated EEG in full-term neonates," *Clinical Neurophysiology,* vol. 120 (11), pp. 1916-1922, 2009.
[84] N. al Naqeeb, A. D. Edwards, F. M. Cowan, and D. Azzopardi, "Assessment of neonatal encephalopathy by amplitude-integrated electroencephalography," *Pediatrics,* vol. 103 (6), pp. 1263-71, 1999.
[85] L. F. Shalak, A. R. Laptook, S. C. Velaphi, and J. M. Perlman, "Amplitude-integrated electroencephalography coupled with an early neurologic examination enhances prediction of term infants at risk for persistent encephalopathy," *Pediatrics,* vol. 111 (2), pp. 351-57, 2003.
[86] C. C. Menache, B. F. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology,* vol. 27 (2), pp. 93-101, 2002.
[87] M. M. Grigg-Damberger, S. B. Coker, C. L. Halsey, and C. L. Anderson, "Neonatal burst suppression: its developmental significance," *Pediatric Neurology,* vol. 5 (2), pp. 84-92, 1989.
[88] E. Niedermeyer, D. L. Sherman, R. J. Geocadin, H. C. Hansen, and D. F. Hanley, "The burst-suppression electroencephalogram," *Clinical EEG (electroencephalography),* vol. 30 (3), pp. 99-105, 1999.
[89] K. Klebermass, S. Kuhle, C. Kohlhauser-Vollmuth, A. Pollak, and M. Weninger, "Evaluation of the Cerebral Function Monitor as a tool for neurophysiological surveillance in neonatal intensive care patients," *Child's Nervous System,* vol. 17 (9), pp. 544-50, 2001.
[90] D. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. Robertson, "EEG and long-term outcome of term infants with neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology,* vol. 110 (4), pp. 655-659, 1999.
[91] M. Polat, A. Simsek, N. Tansug, R. G. Sezer, M. Ozkol, P. Baspinar, *et al.*, "Prediction of neurodevelopmental outcome in term neonates with hypoxic-ischemic encephalopathy," *European Journal of Paediatric Neurology,* vol. 17 (3), pp. 288-93, 2013.
[92] T. Takeuchi and K. Watanabe, "The EEG evolution and neurological prognosis of perinatal hypoxia neonates," *Brain and Development,* vol. 11 (2), pp. 115-120, 1989.
[93] H. F. Prechtl, F. Ferrari, and G. Cioni, "Predictive value of general movements in asphyxiated fullterm infants," *Early Human Development,* vol. 35 (2), pp. 91-120, 1993.
[94] P. D. Gluckman, J. S. Wyatt, D. Azzopardi, R. Ballard, A. D. Edwards, D. M. Ferriero, *et al.*, "Selective head cooling with mild systemic hypothermia after neonatal encephalopathy: multicentre randomised trial," *The Lancet,* vol. 365 (9460), pp. 663-670, 2005.
[95] E. Thornberg and B. Ekström Jodal, "Cerebral function monitoring: a method of predicting outcome in term neonates after severe perinatal asphyxia," *Acta Paediatrica,* vol. 83 (6), pp. 596-601, 1994.
[96] R. Griffith, "The abilities of babies," ed: University of London Press, London, 1954.
[97] N. Bayley, *Bayley scales of infant development: Manual*. San Antonio, TX: Psychological Corporation, 1993.
[98] W. K. Frankenburg and J. B. Dodds, "The Denver developmental screening test," *The Journal of Pediatrics,* vol. 71 (2), pp. 181-91, 1967.
[99] P. Snyder, J. M. Eason, D. Philibert, A. Ridgway, and T. McCaughey, "Concurrent validity and reliability of the Alberta Infant Motor Scale in infants at dual risk for motor delays," *Physical & Occupational Therapy in Pediatrics,* vol. 28 (3), pp. 267-82, 2008.
[100] W. H. Organization, "International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29. 35 of the Twenty-ninth World Health Assembly, May 1976," pp. 1-207, 1980.
[101] A. Jose, J. Matthai, and S. Paul, "Correlation of EEG, CT, and MRI brain with neurological outcome at 12 months in term newborns with hypoxic ischemic encephalopathy," *Journal of clinical neonatology,* vol. 2 (3), p. 125, 2013.
[102] C. Amiel-Tison and J. Gosselin, *Neurological development from birth to six years: guide for examination and evaluation*. Baltimore: Johns Hopkins University Press, 2001.
[103] L. Haataja, E. Mercuri, R. Regev, F. Cowan, M. Rutherford, V. Dubowitz, *et al.*, "Optimality score for the neurologic examination of the infant at 12 and 18 months of age," *The Journal of Pediatrics,* vol. 135 (2), pp. 153-61, 1999.
[104] L. M. Dubowitz, *The neurological assessment of the preterm and full-term newborn infant*. London: Cambridge University Press, 1999.
[105] B. Touwen, *Neurological development in infancy*: Butterworth-Heinemann Ltd, 1976.

[106] P. Eken, M. Toet, F. Groenendaal, and L. De Vries, "Predictive value of early neuroimaging, pulsed Doppler and neurophysiology in full term infants with hypoxic-ischaemic encephalopathy," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 73 (2), pp. F75-F80, 1995.
[107] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *Annals of Internal Medicine*, vol. 151 (4), pp. 264-69, 2009.
[108] R. E. Spitzmiller, T. Phillips, J. Meinzen-Derr, and S. B. Hoath, "Amplitude-integrated EEG is useful in predicting neurodevelopmental outcome in full-term infants with hypoxic-ischemic encephalopathy: a meta-analysis," *Journal of Child Neurology*, vol. 22 (9), pp. 1069-78, 2007.
[109] F. Pisani, C. Facini, E. Pavlidis, C. Spagnoli, and G. Boylan, "Epilepsy after neonatal seizures: Literature review," *European Journal of Paediatric Neurology*, vol. 19 (1), pp. 6-14, 2015.
[110] L. Hellström Westas, G. Boylan, and J. Ågren, "Systematic review of neonatal seizure management strategies provides guidance on antiepileptic treatment," *Acta Paediatrica*, vol. 104 (2), pp. 123-9, 2015.
[111] A. D. Edwards, P. Brocklehurst, A. J. Gunn, H. Halliday, E. Juszczak, M. Levene, *et al.*, "Neurological outcomes at 18 months of age after moderate hypothermia for perinatal hypoxic ischaemic encephalopathy: synthesis and meta-analysis of trial data," *British Medical Journal*, vol. 340 p. c363, 2010.
[112] S. E. Jacobs, R. Hunt, W. O. Tarnow Mordi, T. E. Inder, and P. G. Davis, "Cochrane Review: Cooling for newborns with hypoxic ischaemic encephalopathy," *Evidence Based Child Health: A Cochrane Review Journal*, vol. 3 (4), pp. 1049-1115, 2008.
[113] N. S. Abend, R. Mani, T. N. Tschuda, T. Chang, A. A. Topjian, M. Donnelly, *et al.*, "EEG monitoring during therapeutic hypothermia in neonates, children, and adults," *Neurodiagnostic Journal*, vol. 51 (3), pp. 141-164, 2011.
[114] E. Kochs, "Electrophysiological monitoring and mild hypothermia," *Journal of Neurosurgical Anesthesiology*, vol. 7 (3), pp. 222-228, 1995.
[115] R. A. Shellhaas, T. Chang, T. Tsuchida, M. S. Scher, J. J. Riviello, N. S. Abend, *et al.*, "The American Clinical Neurophysiology Society's guideline on continuous electroencephalography monitoring in neonates," *Journal of Clinical Neurophysiology*, vol. 28 (6), pp. 611-617, 2011.
[116] L. Hellström-Westas, L. S. De Vries, and I. Rosén, *An atlas of amplitude-integrated EEGs in the newborn*: Parthenon Publishing, 2003.
[117] C. Wusthoff, R. Shellhaas, and R. Clancy, "Limitations of single-channel EEG on the forehead for neonatal seizure detection," *Journal of Perinatology*, vol. 29 (3), pp. 237-242, 2008.
[118] M. S. Khlif, P. B. Colditz, and B. Boashash, "Effective implementation of time-frequency matched filter with adapted pre and postprocessing for data-dependent detection of newborn seizures," *Medical Engineering and Physics*, vol. 35 (12), pp. 1762-9, 2013.
[119] E. Thornberg and K. Thiringer, "Normal pattern of the cerebral function monitor trace in term and preterm neonates," *Acta Paediatrica*, vol. 79 (1), pp. 20-25, 1990.
[120] J. S. Hahn, H. Monyer, and B. R. Tharp, "Interburst interval measurements in the EEGs of premature infants with normal neurological outcome," *Electroencephalography and Clinical Neurophysiology*, vol. 73 (5), pp. 410-418, 1989.
[121] T. K. Hon and A. Georgakis, "Enhancing the resolution of the spectrogram based on a simple adaptation procedure," *IEEE Transactions on Signal Processing*, vol. 60 (10), pp. 5566-5571, 2012.
[122] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: optimal kernel design," *IEEE Transactions on Signal Processing*, vol. 41 (4), pp. 1589-1602, 1993.
[123] L. Stanković, "Measuring Time-Frequency Distributions Concentration," in *Time-Frequency Signal Analysis and Processing*, B. Boashash, Ed., 2nd ed: Academic Press, Dec. 2015, pp. 401-408.
[124] B. Boashash and V. Sucic, "Resolution measure criteria for the objective assessment of the performance of quadratic time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 51 (5), pp. 1253-1263, 2003.
[125] D. Malnar, V. Sucic, and J. O'Toole, "Automatic quality assessment and optimisation of quadratic time–frequency representations," *Electronics Letters*, vol. 51 (13), pp. 1029-1031, 2015.
[126] K.-S. Tang, K.-F. Man, S. Kwong, and Q. He, "Genetic algorithms and their applications," *IEEE Signal Processing Magazine*, vol. 13 (6), pp. 22-37, 1996.
[127] (October, 2014). *Global optimization toolbox*. Available: http://www.mathworks.com/products/global-optimization/index.html

[128] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7 (4), pp. 308-313, 1965.
[129] B. Barkat and K. Abed-Meraim, "Algorithms for blind components separation and extraction from the time-frequency distribution of their mixture," *EURASIP Journal on Applied Signal Processing*, vol. 2004 pp. 2025-2033, 2004.
[130] N. S. Rao and P. S. Moharir, "A Signal-Dependent Evolution Kernel for Cohen Class Time–Frequency Distributions," *Digital Signal Processing*, vol. 8 (3), pp. 158-165, 1998.
[131] L. B. Almeida, "The fractional Fourier transform and time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 42 (11), pp. 3084-3091, 1994.
[132] A. T. Catherall and D. P. Williams, "High resolution spectrograms using a component optimized short-term fractional Fourier transform," *Signal Processing*, vol. 90 (5), pp. 1591-1596, 2010.
[133] A. Bultheel and H. Martinez Sulbaran, "A shattered survey of the Fractional Fourier Transform," *TW Reports*, 2002, https://nalag.cs.kuleuven.be/papers/ade/frft/manuscript.pdf.
[134] L. Stanković, T. Alieva, and M. J. Bastiaans, "Time–frequency signal analysis based on the windowed fractional Fourier transform," *Signal Processing*, vol. 83 (11), pp. 2459-2468, 2003.
[135] L. Stankovic, M. Daković, and T. Thayaparan, *Time-frequency signal analysis with applications*: Artech House, 2013.
[136] S. Starosielec and D. Hägele, "Discrete-time windows with minimal RMS bandwidth for given RMS temporal width," *Signal Processing*, vol. 102 pp. 240-246, 2014.
[137] A. Akan and E. Önen, "A discrete fractional Gabor expansion for multi-component signals," *AEU - International Journal of Electronics and Communications*, vol. 61 (5), pp. 279-285, 2007.
[138] R. G. Baraniuk, P. Flandrin, A. J. Janssen, and O. J. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Transactions on Information Theory*, vol. 47 (4), pp. 1391-1409, 2001.
[139] V. Sucic, N. Saulig, and B. Boashash, "Estimating the number of components of a multicomponent nonstationary signal using the short-term time-frequency Rényi entropy," *EURASIP Journal on Advances in Signal Processing*, vol. 2011 (1), pp. 1-11, 2011.
[140] I. Djurović, L. Stanković, and M. Simeunović, "Robust time–frequency representation based on the signal normalization and concentration measures," *Signal Processing*, vol. 104 (0), pp. 424-431, 2014.
[141] N. Hurley and S. Rickard, "Comparing Measures of Sparsity," *IEEE Transactions on Information Theory*, vol. 55 (10), pp. 4723-4741, 2009.
[142] P. Flandrin and P. Borgnat, "Time-Frequency Energy Distributions Meet Compressed Sensing," *IEEE Transactions on Signal Processing*, vol. 58 (6), pp. 2974-2982, 2010.
[143] L. Stanković, I. Djurović, S. Stanković, M. Simeunović, S. Djukanović, and M. Daković, "Instantaneous frequency in time–frequency analysis: Enhanced concepts and performance of estimation algorithms," *Digital Signal Processing*, vol. 35 (0), pp. 1-13, 2014.
[144] L. Stankovic, S. Stankovic, and M. Dakovic, "From the STFT to the Wigner Distribution [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 31 (3), pp. 163-174, 2014.
[145] A. Gholami, "Sparse Time-Frequency Decomposition and Some Applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51 (6), pp. 3598-3604, 2013.
[146] G. Van Hoey, W. Philips, and I. Lemahieu, "Time-frequency analysis of EEG signals," in *Proceedings of the ProRISC Workshop on Circuits, Systems and Signal Processing*, 1997.
[147] E. Sejdić, I. Djurović, and J. Jiang, "Time--frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19 (1), pp. 153-183, 2009.
[148] R. Ahmed, A. Temko, W. Marnane, G. Lightbody, and G. Boylan, "Grading hypoxic–ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine," *Clinical Neurophysiology*, vol. 127 (1), pp. 297-309, 2016.
[149] J. E. Lawn, K. Kerber, C. Enweronu-Laryea, and S. Cousens, "3.6 million neonatal deaths—what is progressing and what is not?," in *Seminars in Perinatology*, 2010, pp. 371-386.
[150] R. D. Folkerth, "Neuropathologic substrate of cerebral palsy," *Journal of Child Neurology*, vol. 20 (12), pp. 940-949, 2005.
[151] H. B. Sarnat and M. S. Sarnat, "Neonatal encephalopathy following fetal distress," *Obstetrical and Gynecological Survey*, vol. 32 (5), p. 295, 1977.
[152] V. Matić, P. J. Cherian, K. Jansen, N. Koolen, G. Naulaers, R. M. Swarte, *et al.*, "Improving Reliability of Monitoring Background EEG Dynamics in Asphyxiated Infants," *IEEE Transactions on Biomedical Engineering*, vol. 63 (5), pp. 973-983, 2016.

[153] N. Koolen, K. Jansen, J. Vervisch, V. Matic, M. De Vos, G. Naulaers, *et al.*, "Line length as a robust method to detect high-activity events: Automated burst detection in premature EEG recordings," *Clinical Neurophysiology*, vol. 125 (10), pp. 1985-1994, 2014.
[154] J. A. Roberts, K. K. Iyer, S. Finnigan, S. Vanhatalo, and M. Breakspear, "Scale-Free Bursting in Human Cortex following Hypoxia at Birth," *The Journal of Neuroscience*, vol. 34 (19), pp. 6557-6572, 2014.
[155] N. Stevenson, I. Korotchikova, A. Temko, G. Lightbody, W. Marnane, and G. Boylan, "An automated system for grading EEG abnormality in term neonates with hypoxic-ischaemic encephalopathy," *Annals of Biomedical Engineering*, vol. 41 (4), pp. 775-785, 2013.
[156] F. Amzica, "Basic physiology of burst-suppression," *Epilepsia*, vol. 50 (s12), pp. 38-39, 2009.
[157] S. Zanelli, H. Goodkin, S. Kowalski, and J. Kapur, "Impact of transient acute hypoxia on the developing mouse EEG," *Neurobiology of Disease*, vol. 68 pp. 37-46, 2014.
[158] D. L. Sherman, A. M. Brambrink, R. N. Ichord, V. K. Dasika, R. C. Koehler, R. J. Traystman, *et al.*, "Quantitative EEG during early recovery from hypoxic-ischemic injury in immature piglets: burst occurrence and duration," *Clinical EEG and Neuroscience*, vol. 30 (4), pp. 175-183, 1999.
[159] R. S. Fisher, "Animal models of the epilepsies," *Brain Research Reviews*, vol. 14 (3), pp. 245-278, 1989.
[160] S. Chu, S. Narayanan, and C. J. Kuo, "Environmental sound recognition with time–frequency audio features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17 (6), pp. 1142-1158, 2009.
[161] P. Karthick and S. Ramakrishnan, "Surface electromyography based muscle fatigue progression analysis using modified B distribution time–frequency features," *Biomedical Signal Processing and Control*, vol. 26 pp. 42-51, 2016.
[162] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Classification of burst and suppression in the neonatal electroencephalogram," *Journal of Neural Engineering*, vol. 5 (4), p. 402, 2008.
[163] N. J. Stevenson, M. Mesbah, G. B. Boylan, P. B. Colditz, and B. Boashash, "A Nonlinear Model of Newborn EEG with Nonstationary Inputs," *Annals of Biomedical Engineering*, vol. 38 (9), pp. 3010-3021, 2010.
[164] M. Borowska, "Entropy-Based Algorithms in the Analysis of Biomedical Signals," *Studies in Logic, Grammar and Rhetoric*, vol. 43 (1), pp. 21-32, 2015.
[165] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. W. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review," *Knowledge-Based Systems*, vol. 88 pp. 85-96, 2015.
[166] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278 (6), pp. H2039-H2049, 2000.
[167] M. T. Peiris, P. R. Davidson, P. J. Bones, and R. D. Jones, "Detection of lapses in responsiveness from the EEG," *Journal of Neural Engineering*, vol. 8 (1), p. 016003, 2011.
[168] K. Mohanchandra, S. Saha, and K. S. Murthy, "Evidence of Chaos in EEG Signals: An Application to BCI," in *Advances in Chaos Theory and Intelligent Control*, ed: Springer, 2016, pp. 609-625.
[169] D. Song, P. Jegatheesan, S. Weiss, B. Govindaswami, J. Wang, J. Lee, *et al.*, "Modulation of EEG spectral edge frequency during patterned pneumatic oral stimulation in preterm infants," *Pediatric Research*, vol. 75 (1-1), pp. 85-92, 2014.
[170] H. J. Niemarkt, W. Jennekens, J. W. Pasman, T. Katgert, C. van Pul, A. W. D. Gavilanes, *et al.*, "Maturational Changes in Automated EEG Spectral Power Analysis in Preterm Infants," *Pediatric Research*, vol. 70 (5), pp. 529-534, 2011.
[171] L. Rankine, M. Mesbah, and B. Boashash, "A matching pursuit-based signal complexity measure for the analysis of newborn EEG," *Medical and Biological Engineering and Computing*, vol. 45 (3), pp. 251-260, 2007.
[172] B. Jokanovic and M. Amin, "Sparse and cross-term free time-frequency distribution based on Hermite functions," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 3696-3700.
[173] A. Uppuluri. (June, 2016). *GLCM toolbox*. Available: https://au.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features
[174] X. Ou, W. Pan, and P. Xiao, "In vivo skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM)," *International Journal of Pharmaceutics*, vol. 460 (1), pp. 28-32, 2014.

[175] A. Costa. (June,2016). *SFTA Toolbox*. Available: https://au.mathworks.com/matlabcentral/fileexchange/37933-alceufc-sfta
[176] S. Bhattacharyya, A. Biswas, J. Mukherjee, A. K. Majumdar, B. Majumdar, S. Mukherjee, *et al.*, "Detection of artifacts from high energy bursts in neonatal EEG," *Computers in Biology and Medicine*, vol. 43 (11), pp. 1804-1814, 2013.
[177] A. Şengür, Y. Guo, and Y. Akbulut, "Time–frequency texture descriptors of EEG signals for efficient detection of epileptic seizure," *Brain Informatics*, vol. 3 (2), pp. 101-108, 2016.
[178] A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina, "An Efficient Algorithm for Fractal Analysis of Textures," in *SIBGRAPI Conference on Graphics, Patterns and Images*, 2012, pp. 39-46.
[179] A. K. Seghouane and J. L. Ong, "Efficient feature selection for polyp detection," in *IEEE International Conference on Image Processing*, 2010, pp. 2285-2288.
[180] H. Chih-Wei and L. Chih-Jen, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13 (2), pp. 415-425, 2002.
[181] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press, 2000.
[182] H.-T. Lin, C.-J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Machine learning*, vol. 68 (3), pp. 267-276, 2007.
[183] M. S. Khlif, M. Mesbah, B. Boashash, and P. Colditz, "Multichannel-Based Newborn EEG Seizure Detection using Time-Frequency Matched Filter," in *IEEE Engineering in Medicine and Biology Society (EMBC) Conference*, 2007, pp. 1265-1268.
[184] C. O'Reilly and T. Nielsen, "Revisiting the ROC curve for diagnostic applications with an unbalanced class distribution," in *8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)* 2013, pp. 413-420.
[185] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27 (8), pp. 861-874, 2006.
[186] P. D. P. Pharoah, "Re: Understanding P values," *BMJ*, vol. 349 p. g4550, 2014.
[187] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.
[188] I. Amarreh, M. E. Meyerand, C. Stafstrom, B. P. Hermann, and R. M. Birn, "Individual classification of children with epilepsy using support vector machine with multiple indices of diffusion tensor imaging," *NeuroImage: Clinical*, vol. 4 pp. 757-764, 2014.
[189] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial Intelligence in Medicine*, vol. 34 (2), pp. 113-127, 2005.
[190] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3 pp. 1157-1182, 2003.
[191] J. L. Ong and A.-K. Seghouane, "Feature selection using mutual information in CT colonography," *Pattern Recognition Letters*, vol. 32 (2), pp. 337-341, 2011.
[192] J. LaRocco, C. R. Innes, P. J. Bones, S. Weddell, and R. D. Jones, "Optimal EEG feature selection from average distance between events and non-events," in *IEEE Medicine and Biology Society (EMBC) Conference* 2014, pp. 2641-4.
[193] Z. M. Hira and D. F. Gillies, "A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data," *Advances in Bioinformatics*, vol. 2015 p. 13, 2015.
[194] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27 (8), pp. 1226-1238, 2005.
[195] M. Milošević, A. Van de Vel, K. Cuppens, B. Bonroy, B. Ceulemans, L. Lagae, *et al.*, "Feature selection methods for accelerometry-based seizure detection in children," *Medical and Biological Engineering and Computing*, pp. 1-15, 2016.
[196] Y.-W. Chen and C.-J. Lin, "Combining SVMs with various feature selection strategies," in *Feature extraction*, ed: Springer, 2006, pp. 315-324.
[197] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36 (2, Part 2), pp. 3240-3247, 2009.
[198] F. Pedro, L. Xi, R. Mustafa, H. Reinder, M. A. Ronald, and R. Jérôme, "Sleep stage classification with ECG and respiratory effort," *Physiological Measurement*, vol. 36 (10), p. 2027, 2015.
[199] B. B. Pineda-Bautista, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "General framework for class-specific feature selection," *Expert Systems with Applications*, vol. 38 (8), pp. 10018-10024, 2011.

[200] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering,* vol. 21 (9), pp. 1263-1284, 2009.
[201] H. He, Y. Bai, E. A. Garcia, and L. Shutao, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *IEEE Conference on Neural Networks*, 2008, pp. 1322-1328.
[202] H. He and Y. Cao, "SSC: A classifier combination method based on signal strength," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 23 (7), pp. 1100-1117, 2012.
[203] B. Boashash, H. Barki, and S. Ouelha, "Performance evaluation of time-frequency image feature sets for improved classification and analysis of non-stationary signals: Application to newborn EEG seizure detection," *Knowledge-Based Systems,* vol. 132 pp. 188-203, 2017.
[204] V. Matic, P. J. Cherian, N. Koolen, G. Naulaers, R. M. Swarte, P. Govaert, *et al.*, "Holistic approach for automated background EEG assessment in asphyxiated full-term infants," *Journal of Neural Engineering,* vol. 11 (6), p. 066007, 2014.
[205] A. Temko, O. Doyle, D. Murray, G. Lightbody, G. Boylan, and W. Marnane, "Multimodal predictor of neurodevelopmental outcome in newborns with hypoxic-ischaemic encephalopathy," *Computers in Biology and Medicine,* vol. 63 pp. 169-177, 2015.
[206] A. R. Laptook, S. Shankaran, N. Ambalavanan, W. A. Carlo, S. A. McDonald, R. D. Higgins, *et al.*, "Outcome of Term Infants Using Apgar Scores at 10 Minutes Following Hypoxic-Ischemic Encephalopathy," *Pediatrics,* vol. 124 (6), pp. 1619-1626, 2009.
[207] P. K. Misra, N. Srivastava, G. K. Malik, R. K. Kapoor, K. L. Srivastava, and S. Rastogi, "Outcome in relation to Apgar score in term neonates," *Indian Pediatrics,* vol. 31 (10), pp. 1215-8, 1994.
[208] "The Apgar Score," *Pediatrics*, vol. 117 (4), pp. 1444-1447, 2006.
[209] V. Ramaswamy, J. Horton, B. Vandermeer, N. Buscemi, S. Miller, and J. Yager, "Systematic review of biomarkers of brain injury in term neonatal encephalopathy," *Pediatric Neurology,* vol. 40 (3), pp. 215-26, 2009.
[210] T. Alderliesten, L. S. de Vries, L. Staats, I. C. van Haastert, L. Weeke, M. J. N. L. Benders, *et al.*, "MRI and spectroscopy in (near) term neonates with perinatal asphyxia and therapeutic hypothermia," *Archives of Disease in Childhood - Fetal and Neonatal Edition,* 2016.
[211] L. Hellström-Westas, I. Rosen, and N. W. Svenningsen, "Predictive value of early continuous amplitude integrated EEG recordings on outcome after severe birth asphyxia in full term infants," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 72 (1), pp. F34-F38, 1995.
[212] D. A. Pizzagalli, "Electroencephalography and high-density electrophysiological source localization," *Handbook of psychophysiology,* vol. 3 pp. 56-84, 2007.
[213] M. S. Scher, D. A. Steppe, R. J. Sclabassi, and D. L. Banks, "Regional differences in spectral EEG measures between healthy term and preterm infants," *Pediatric Neurology,* vol. 17 (3), pp. 218-223, 1997.
[214] Q. Noirhomme, R. Lehembre, Z. del Rosario Lugo, D. Lesenfants, A. Luxen, S. Laureys, *et al.*, "Automated analysis of background EEG and reactivity during therapeutic hypothermia in comatose patients after cardiac arrest," *Clinical EEG and Neuroscience,* p. 1550059413509616, 2014.
[215] K. K. Iyer, J. A. Roberts, L. Hellstrom-Westas, S. Wikstrom, I. Hansen Pupp, D. Ley, *et al.*, "Cortical burst dynamics predict clinical outcome early in extremely preterm infants," *Brain,* vol. 138 (Pt 8), pp. 2206-18, 2015.
[216] A. Dereymaeker, N. Koolen, K. Jansen, J. Vervisch, E. Ortibus, M. De Vos, *et al.*, "The Suppression Curve as a quantitative approach for measuring brain maturation in preterm infants," *Clinical Neurophysiology,* vol. 127 (8), pp. 2760-65, 2016.
[217] A. F. Bos, "Bayley-II or Bayley-III: what do the scores tell us?," *Developmental Medicine and Child Neurology,* vol. 55 (11), pp. 978-979, 2013.
[218] S. Jary, A. Whitelaw, L. Walloe, and M. Thoresen, "Comparison of Bayley-2 and Bayley-3 scores at 18 months in term infants following neonatal encephalopathy and therapeutic hypothermia," *Developmental Medicine and Child Neurology,* vol. 55 (11), pp. 1053-9, 2013.
[219] I. Daly, F. Pichiorri, J. Faller, V. Kaiser, A. Kreilinger, R. Scherer, *et al.*, "What does clean EEG look like?," in *IEEE Engineering in Medicine and Biology Society (EMBS) Conference*, 2012, pp. 3963-6.
[220] M. Zima, P. Tichavský, K. Paul, and V. Krajča, "Robust removal of short-duration artifacts in long neonatal EEG recordings using wavelet-enhanced ICA and adaptive combining of tentative reconstructions," *Physiological Measurement,* vol. 33 (8), p. N39, 2012.
[221] U. Jose Antonio and G.-Z. Begoña, "EEG artifact removal—state-of-the-art and guidelines," *Journal of Neural Engineering,* vol. 12 (3), p. 031001, 2015.

[222] C. A. Teixeira, B. Direito, H. Feldwisch-Drentrup, M. Valderrama, R. P. Costa, C. Alvarado-Rojas, *et al.*, "EPILAB: A software package for studies on the prediction of epileptic seizures," *Journal of Neuroscience Methods,* vol. 200 (2), pp. 257-271, 2011.
[223] B. Yao, J. Z. Liu, R. W. Brown, V. Sahgal, and G. H. Yue, "Nonlinear features of surface EEG showing systematic brain signal adaptations with muscle force and fatigue," *Brain Research,* vol. 1272 pp. 89-98, 2009.
[224] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, *et al.*, "On the predictability of epileptic seizures," *Clinical Neurophysiology,* vol. 116 (3), pp. 569-587.
[225] A. Suppiej, I. Festa, L. Bartolini, A. Cappellari, E. Cainelli, M. Ermani, *et al.*, "Power spectral analysis of two-channel EEG in very premature infants undergoing heat loss prevention," *Neurophysiologie Clinique/Clinical Neurophysiology,* vol. 44 (3), pp. 239-244, 2014.
[226] S. L. Olsen, M. DeJonge, A. Kline, E. Liptsen, D. Song, B. Anderson, *et al.*, "Optimizing Therapeutic Hypothermia for Neonatal Encephalopathy," *Pediatrics,* vol. 131 (2), pp. e591-e603, 2013.
[227] K. Murphy, N. J. Stevenson, R. M. Goulding, R. O. Lloyd, I. Korotchikova, V. Livingstone, *et al.*, "Automated analysis of multi-channel EEG in preterm infants," *Clinical Neurophysiology,* vol. 126 (9), pp. 1692-1702, 2015.
[228] I. A. Williams, A. R. Tarullo, P. G. Grieve, A. Wilpers, E. F. Vignola, M. M. Myers, *et al.*, "Fetal Cerebrovascular Resistance and Neonatal EEG Predict 18-month Neurodevelopmental Outcome in Infants with Congenital Heart Disease," *Ultrasound in Obstetrics and Gynecology,* vol. 40 (3), pp. 304-309, 2012.
[229] M. Videman, A. Tokariev, S. Stjerna, R. Roivainen, E. Gaily, and S. Vanhatalo, "Effects of prenatal antiepileptic drug exposure on newborn brain activity," *Epilepsia,* vol. 57 (2), pp. 252-262, 2016.
[230] C. Vidaurre, T. H. Sander, and A. Schlögl, "BioSig: The Free and Open Source Software Library for Biomedical Signal Processing," *Computational Intelligence and Neuroscience,* vol. 2011 p. 12, 2011.
[231] Y. Wang, E. M. Sokhadze, A. S. El-Baz, X. Li, L. Sears, M. F. Casanova, *et al.*, "Relative Power of Specific EEG Bands and Their Ratios during Neurofeedback Training in Children with Autism Spectrum Disorder," *Frontiers in Human Neuroscience,* vol. 9 p. 723, 2015.
[232] S. Finnigan, A. Wong, and S. Read, "Defining abnormal slow EEG activity in acute ischaemic stroke: Delta/alpha ratio as an optimal QEEG index," *Clinical Neurophysiology,* vol. 127 (2), pp. 1452-1459, 2016.
[233] M. J. A. M. van Putten, J. M. Peters, S. M. Mulder, J. A. M. de Haas, C. M.A. Bruijninckx, and D. L. J. Tavy, "A brain symmetry index (BSI) for online EEG monitoring in carotid endarterectomy," *Clinical Neurophysiology,* vol. 115 (5), pp. 1189-1194, 2004.
[234] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping,* vol. 8 (4), pp. 194-208, 1999.
[235] L. Li, W. Chen, X. Shao, and Z. Wang, "Analysis of Amplitude-Integrated EEG in the Newborn Based on Approximate Entropy," *IEEE Transactions on Biomedical Engineering,* vol. 57 (10), pp. 2459-2466, 2010.
[236] G. Lee, S. Fattinger, A.-L. Mouthon, Q. Noirhomme, and R. Huber, "Electroencephalogram approximate entropy influenced by both age and sleep," *Frontiers in Neuroinformatics,* vol. 7 (33), 2013.
[237] D. Zhang, H. Ding, Y. Liu, C. Zhou, H. Ding, and D. Ye, "Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram," *Physiological Measurement,* vol. 30 (5), pp. 491-504, 2009.
[238] M. A. Awal, S. S. Mostafa, and M. Ahmad, "Performance analysis of Savitzky-Golay smoothing filter using ECG signal," *International Journal of Computer and Information Technology,* vol. 1 (02), 2011.
[239] M. Odabaee, W. J. Freeman, P. B. Colditz, C. Ramon, and S. Vanhatalo, "Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes," *NeuroImage* . vol. 68 pp. 229-35, 2013.
[240] B. Vasiljevic, S. Maglajlic-Djukic, and M. Gojnic, "The prognostic value of amplitude-integrated electroencephalography in neonates with hypoxic-ischemic encephalopathy," *Vojnosanitetski Pregled,* vol. 69 (6), pp. 492-9, 2012.
[241] S. Dong, B. Boashash, G. Azemi, B. E. Lingwood, and P. B. Colditz, "Automated detection of perinatal hypoxia using time–frequency-based heart rate variability features," *Medical and Biological Engineering and Computing,* vol. 52 (2), pp. 183-191, 2014.

[242] R. M. Goulding, N. J. Stevenson, D. M. Murray, V. Livingstone, P. M. Filan, and G. B. Boylan, "Heart rate variability in hypoxic ischemic encephalopathy: correlation with EEG grade and 2-y neurodevelopmental outcome," *Pediatric Research*, vol. 77 (5), pp. 681-7, 2015.
[243] J. L. Y. Cheong, D. K. Thompson, A. J. Spittle, C. R. Potter, J. M. Walsh, A. C. Burnett, *et al.*, "Brain Volumes at Term-Equivalent Age Are Associated with 2-Year Neurodevelopment in Moderate and Late Preterm Children," *The Journal of Pediatrics*, vol. 174 pp. 91-97.e1, 2016.
[244] A. M. Pagnozzi, N. Dowson, J. Doecke, S. Fiori, A. P. Bradley, R. N. Boyd, *et al.*, "Automated, quantitative measures of grey and white matter lesion burden correlates with motor and cognitive function in children with unilateral cerebral palsy," *NeuroImage : Clinical*, vol. 11 pp. 751-759, 2016.
[245] N. Wiberg, K. Kallen, A. Herbst, and P. Olofsson, "Relation between umbilical cord blood pH, base deficit, lactate, 5-minute Apgar score and development of hypoxic ischemic encephalopathy," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 89 (10), pp. 1263-9, 2010.
[246] S. S. D. P. Ayyagari, R. D. Jones, and S. J. Weddell, "Optimized echo state networks with leaky integrator neurons for EEG-based microsleep detection," in *IEEE Engineering in Medicine and Biology Society (EMBC) Conference*, 2015, pp. 3775-3778.
[247] N. A. Khan and B. Boashash, "Instantaneous Frequency Estimation of Multicomponent Nonstationary Signals Using Multiview Time-Frequency Distributions Based on the Adaptive Fractional Spectrogram," *IEEE Signal Processing Letters*, vol. 20 (2), pp. 157-160, 2013.
[248] Y. Wang and R. Agarwal, "Automatic detection of burst suppression," in *IEEE Engineering in Medicine and Biology Society (EMBS) Conference* 2007, pp. 553-556.
[249] B. Boashash, G. Azemi, and J. M. O'Toole, "Time-Frequency Processing of Nonstationary Signals: Advanced TFD Design to Aid Diagnosis with Highlights from Medical Applications," *IEEE Signal Processing Magazine*, vol. 30 (6), pp. 108-119, 2013.
[250] H. Hassanpour, M. Mesbah, and B. Boashash, "Time-frequency feature extraction of newborn EEG seizure using SVD-based techniques," *EURASIP Journal on Applied Signal Processing*, vol. 2004 pp. 2544-2554, 2004.
[251] Z. Cai, D. Xu, Q. Zhang, J. Zhang, S.-M. Ngai, and J. Shao, "Classification of lung cancer using ensemble-based feature selection and machine learning methods," *Molecular BioSystems*, vol. 11 (3), pp. 791-800, 2015.
[252] W. S. Browner and T. B. Newman, "Are all significant p values created equal?: The analogy between diagnostic tests and clinical research," *JAMA*, vol. 257 (18), pp. 2459-2463, 1987.
[253] G. L. Grunkemeier, Y. Wu, and A. P. Furnary, "What is the Value of a p Value?," *The Annals of Thoracic Surgery*, vol. 87 (5), pp. 1337-1343, 2009.
[254] N. G. Adams and G. O'Reilly, "A likelihood-based approach to P-value interpretation provided a novel, plausible, and clinically useful research study metric," *Journal of Clinical Epidemiology*, 2017.