



**HAL**  
open science

# An e-health system for personalized automatic sleep stages classification

Chen Chen

► **To cite this version:**

Chen Chen. An e-health system for personalized automatic sleep stages classification. Emerging Technologies [cs.ET]. Université Pierre et Marie Curie - Paris VI, 2016. English. NNT : 2016PA066664 . tel-01587766

**HAL Id: tel-01587766**

**<https://theses.hal.science/tel-01587766>**

Submitted on 14 Sep 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité

**Informatique**

École doctorale Informatique, Télécommunications et Électronique (Paris)

Présentée par

**Chen CHEN**

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

Sujet de la thèse :

**An E-health System for Personalized Automatic Sleep Stages  
Classification**

Soutenue le 12 décembre 2016

Devant le jury composé de :

M. Patrick GARDA	Directeur de thèse
M. Kurosh MADANI	Rapporteur
M. Etienne SICARD	Rapporteur
M. Jérôme BOUDY	Examineur
M. Christophe MARSALA	Examineur
Mme. Marie-Christine JAULENT	Examineur
M. Andrea PINNA	Examineur

# Acknowledgments

I would like to sincerely thank my thesis director, Prof. Patrick Garda, for providing me with the opportunity to pursue my PhD in SYstèmes ELelectroniques (SYEL), Laboratoire d'Informatique de Paris 6 (LIP6), Université Pierre et Marie Curie (UPMC).

I would also express my great thanks to my supervisors: Dr. Andrea Pinna, Dr. Xun Zhang and Prof. Amara Amara. Especially to Dr. Pinna Andrea for guidance, encouragement and advice he has provided throughout my PhD. And Prof. Amara Amara, for his support and cooperation with Institut Supérieur d'Electronique de Paris (ISEP).

I would also like to thank Prof. Chih-ming Ho, Dr. Ding Xianting and Prof. Kenneth Yun for their valuable time and suggestion on my work; I am thankful to Dr. Carole Philippe and Dr. Amina Kotti, for the collaboration with the hospital and for helping in the field of medicine.

Also, I am very appreciative to all faculty members, colleagues and friends in LIP6 and ISEP, especially for M. Adrien Ugon and M. Navneet Gupta.

I extend my sincere gratitude for the financial support China Scholarship Council.

Last but not the least, I would like to thank my family: my parents and my boyfriend for supporting me spiritually throughout my thesis and my life in general.

# Abstract

In this thesis, a personalized automatic sleep staging system is proposed by combining symbolic fusion and feedback system control technique. Symbolic fusion is inspired by the decision-making process of clinical sleep staging. It starts from the extraction of digital parameters from raw polysomnography signals and it goes up to a high-level symbolic interpretation through a features extraction process. At last, the decision is generated using rules inspired by international guidelines in sleep medicine. Meanwhile, the symbols and the features computations depend on a set of thresholds, whose determination is a key issue. In this thesis, two different FSC algorithms, Differential Evolution and Cross Entropy, were studied to compute these thresholds automatically.

Individual variability was often ignored in existing automatic sleep staging systems. However, an individual variability was observed in many aspects of sleep research (such as polysomnography recordings, sleep patterns, sleep architecture, sleep duration, sleep events, etc.). In order to improve the effectiveness of the sleep stages classifiers, a personalized automatic sleep staging system that can be adapted the different persons and take individual variability into consideration was explored and evaluated.

The perspectives of this work are based on evaluating the complexity and the performances of these algorithms in terms of latencies and hardware resource requirements, in order to target an personalized automated embedded sleep staging system.

# Résumé

Dans cette thèse, un système personnalisé de stadification automatique du sommeil est proposé, combinant fusion symbolique et système de contrôle rétroactif. La fusion symbolique est inspirée par le processus décisionnel mis en oeuvre par les cliniciens experts du sommeil lors la reconnaissance visuelle des stades de sommeil. Il commence par l'extraction de paramètres numériques à partir des signaux polysomnographiques bruts. L'interprétation symbolique de haut niveau se fait par l'intermédiaire de l'extraction de caractéristiques à partir des paramètres numériques. Enfin, la décision est générée en utilisant des règles inspirées par les recommandations internationales en médecine du sommeil. Les symboles et les valeurs des caractéristiques dépendent d'un ensemble de seuils, dont la détermination est une question clé. Dans cette thèse, deux algorithmes de recherche différents, Differential Evolution et Cross Entropy ont été étudiés pour calculer la valeur de ces seuils automatiquement.

La variabilité individuelle a souvent été ignorée dans les systèmes automatiques de stadification du sommeil existants. Cependant, elle a été démontrée dans plusieurs travaux de recherche vis à vis de nombreux aspects du sommeil (comme les enregistrements polysomnographiques, les habitudes de sommeil, l'architecture du sommeil, la durée du sommeil, les événements liés au sommeil, etc.). Afin d'améliorer l'efficacité des classificateurs des stades de sommeil, un système automatisé de sommeil automatique adapté aux différentes personnes et tenant compte de la variabilité individuelle a été exploré et évalué.

# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	Introduction . . . . .	13
1.2	Thesis Outline . . . . .	14
<b>2</b>	<b>Problem Statement</b>	<b>15</b>
2.1	Problem Statement . . . . .	15
2.2	Context . . . . .	15
<b>3</b>	<b>State of the Art</b>	<b>17</b>
3.1	Introduction on Sleep Analysis . . . . .	17
3.1.1	Polysomnography (PSG) . . . . .	17
3.1.2	Clinical Sleep Staging Manual . . . . .	18
3.1.3	Hypnogram . . . . .	19
3.2	Automatic Sleep Staging Systems . . . . .	20
3.2.1	Machine Learning Methods . . . . .	20
3.2.1.1	Decision Tree . . . . .	20
3.2.1.2	Artificial Neural Network . . . . .	22
3.2.1.3	Support Vector Machine . . . . .	23
3.2.2	Rule-based Methods . . . . .	24
3.2.3	Hybrid Methods . . . . .	25
3.2.4	Symbolic Fusion Method . . . . .	25
3.3	Thresholds Setting-Up Problem and Methods . . . . .	26
3.3.1	Thresholds Setting-Up Problem . . . . .	26
3.3.2	Thresholds Setting-Up Method: Fuzzy Logic . . . . .	27
3.3.3	Thresholds Setting-Up Method: Feedback System Control . . . . .	27
3.3.3.1	Typical Stochastic Search Algorithms for Feedback System Control . . . . .	28
3.4	Conclusion . . . . .	30
<b>4</b>	<b>Symbolic Fusion-based Sleep Staging System</b>	<b>32</b>
4.1	Symbolic Fusion Conception . . . . .	32
4.1.1	JDL Architecture . . . . .	32
4.1.2	Dasarathy Architecture . . . . .	33
4.2	Existing SF-SSS . . . . .	35
4.2.1	Data Fusion . . . . .	35
4.2.2	Feature Fusion . . . . .	36
4.2.3	Decision Fusion . . . . .	37
4.3	Existing SF-SSS Performances and Limitations . . . . .	38
4.4	Modifications of Existing SF-SSS . . . . .	40
4.4.1	Sleep Staging Design Flow . . . . .	40
4.4.2	Pre-processing and Segmentation . . . . .	41

4.4.3	Data Fusion: K-Complex . . . . .	42
4.4.3.1	Teager Energy Operator . . . . .	43
4.4.3.2	K-Complex Digital Parameter Extraction . . . . .	43
4.4.4	Feature Fusion: K-Complex . . . . .	45
4.4.5	Decision Fusion: K-Complex . . . . .	45
4.4.6	Decision Fusion: Stage N1 . . . . .	45
4.4.7	Smoothing . . . . .	46
4.5	Conclusion . . . . .	48
<b>5</b>	<b>Thresholds and Symbolic Fusion-based Sleep Staging System</b>	<b>49</b>
5.1	Introduction on Thresholds . . . . .	49
5.1.1	Explanation of Thresholds . . . . .	49
5.1.1.1	Description in AASM . . . . .	49
5.1.1.2	From AASM to SF-SSS model using Thresholds . . . . .	50
5.1.2	Manual Thresholds Setting-Up Method . . . . .	51
5.2	Issues on Thresholds . . . . .	52
5.3	Thresholds in SF-SSS . . . . .	53
5.3.1	Thresholds & Sleep Stages . . . . .	54
5.3.2	Thresholds & Patients . . . . .	56
5.4	Thresholds Configuration . . . . .	57
5.5	Conclusion . . . . .	59
<b>6</b>	<b>Towards to a Personalized Sleep Staging System</b>	<b>60</b>
6.1	Automatic Thresholds Setting-Up Method for SF-SSS . . . . .	60
6.1.0.1	Feedback System Control technique . . . . .	61
6.1.0.2	Automatic Thresholds Setting-Up Method . . . . .	62
6.1.0.3	Differential Evolution . . . . .	64
6.1.0.4	Cross Entropy . . . . .	68
6.2	Personalized Sleep Staging System . . . . .	70
6.2.1	Toward to a Hybrid Expert System for Sleep Staging . . . . .	71
6.2.1.1	Personalized Thresholds and Generalized Thresholds . . . . .	71
6.2.2	Personalized Sleep Staging System . . . . .	71
6.2.3	Differential Evolution-based Personalized Sleep Staging System . . . . .	72
6.2.3.1	Epoch Selection of DE-PSSS . . . . .	73
6.2.4	Cross Entropy-based Personalized Sleep Staging System . . . . .	73
6.2.4.1	Epoch Selection of CE-PSSS . . . . .	74
6.3	Conclusion . . . . .	74
<b>7</b>	<b>Evaluation of Sleep Staging System</b>	<b>75</b>
7.1	Database . . . . .	75
7.1.1	Database 1 . . . . .	75
7.1.1.1	PSG Recordings . . . . .	75
7.1.1.2	Subjects Description . . . . .	75
7.1.2	Database 2 . . . . .	76
7.1.2.1	PSG Recording Device . . . . .	77
7.1.2.2	PSG Recordings . . . . .	77
7.1.2.3	Expert Analysis . . . . .	77
7.1.2.4	Subjects Description . . . . .	77
7.2	Evaluation of SF-SSS with Modifications . . . . .	78
7.2.0.1	F-Measure of Existing SF-SSS VS SF-SSS with Modifications . . . . .	79

7.2.0.2	Agreement Rate of Existing SF-SSS <i>VS</i> SF-SSS with Modifications . . . . .	81
7.2.0.3	Confusion Matrix of Existing SF-SSS <i>VS</i> SF-SSS with Modifications . . . . .	83
7.3	Evaluation of Automatic Thresholds Setting-Up Method . . . . .	84
7.3.1	Evaluation of Automatic Thresholds Setting-Up Method . . . . .	84
7.3.1.1	Parameters Selection of Differential Evolution . . . . .	84
7.3.1.2	Parameters Selection of Cross Entropy . . . . .	88
7.3.2	Evaluation of Personalized Sleep Staging System . . . . .	89
7.3.2.1	Training Set Selection of DE-PSSS . . . . .	90
7.3.2.2	Training Set Selection of CE-PSSS . . . . .	91
7.3.3	Comparison between DE-PSSS and CE-PSSS . . . . .	91
7.3.3.1	Algorithm Comparison . . . . .	91
7.3.3.2	Computational Time . . . . .	92
7.3.3.3	F-Measure Comparison . . . . .	92
7.4	F-Measure Comparison between SF-SSS and DE-PSSS/CE-PSSS . . . . .	93
7.4.1	F-Measure of SF-SSS using MTSU . . . . .	93
7.4.2	Comparison among SF-SSS, CE-PSSS and DE-PSSS . . . . .	93
7.5	Agreement Rate Comparison between PSSSs and Existing Works . . . . .	95
7.6	Agreement Rate of PSSSs . . . . .	95
7.7	Agreement Rate Comparison between PSSSs and Inter-raters . . . . .	97
7.8	Agreement Rate Comparison between PSSSs and Other Works . . . . .	98
7.9	Conclusion . . . . .	99
<b>8</b>	<b>Conclusions and Perspectives</b> . . . . .	<b>100</b>
8.1	Conclusions . . . . .	100
8.2	Perspectives . . . . .	103
8.3	Appendix 1 . . . . .	105
8.4	Appendix 2 . . . . .	106
8.5	Appendix 3 . . . . .	106
8.6	Appendix 4 . . . . .	107
8.7	Appendix 5 . . . . .	107
8.8	Appendix 6 . . . . .	108
8.9	Appendix 7 . . . . .	108
8.10	Appendix 8 . . . . .	109
8.11	Appendix 9 . . . . .	109
8.12	Appendix 10 . . . . .	110
8.13	Appendix 11 . . . . .	110
8.14	Appendix 12 . . . . .	111
8.15	Appendix 13 . . . . .	111
8.16	Appendix 14 . . . . .	112



# Acronyms

- AASM** American Academy of Sleep Medicine. 13, 75, 76
- ANN** Artificial Neural Network. 20
- AR** Agreement Rate. 81
- ATSU** Automatic Thresholds Setting-Up. 60
- CE** Cross Entropy. 84, 91
- CE-PSSS** Cross Entropy-based Personalized Sleep Staging System. 70, 75
- DE** Differential Evolution. 84, 91
- DE-PSSS** Differential Evolution-based Personalized Sleep Staging System. 70, 75
- FSC** Feedback System Control. 60
- G-FSC** Generic Feedback System Control. 61
- JDL** Joint Directors of Laboratories. 32
- MLP** MultiLayer Perceptron. 22
- MTSU** Manual Thresholds Setting-Up. 60
- PSG** Polysomnography. 13
- PSSS** Personalized Sleep Staging System. 60, 75, 84
- R&K** Rechtschaffen & Kales. 14, 75, 76
- SAS** Sleep Apnea Syndrome. 75
- SF-SSS** Symbolic Fusion-based Sleep Staging System. 32
- SSA** Stochastic Search Algorithm. 62
- SVM** Support Vector Machine. 20
- TEO** Teager Energy Operator. 43

# List of Figures

1.1	Prevalence of Insomnia in Europe . . . . .	13
3.1	PSG Device . . . . .	17
3.2	PSG Recordings . . . . .	17
3.3	PSG Signals . . . . .	18
3.4	An Example of Hypnogram . . . . .	20
3.5	An Example of Hypnogram of Sleep Apnea Patient . . . . .	20
3.6	An example of Decision Tree . . . . .	21
3.7	Structure of an Artificial Neural Network in [62] . . . . .	23
3.8	Symbolic Fusion-based Sleep staging System Model . . . . .	26
4.1	Fusion I/O Characterizations of Dasarathy Architecture . . . . .	34
4.2	Sleep Staging Design Flow . . . . .	35
4.3	First Level of Feature Fusion - DAI-FEO Fusion . . . . .	37
4.4	Second Level of Feature Fusion - FEI-FEO Fusion . . . . .	37
4.5	Classification of Stage N2 . . . . .	38
4.6	Movement Artifact . . . . .	39
4.7	Power Line Artifact . . . . .	39
4.8	Sleep Staging Design Flow . . . . .	41
4.9	Elimination of Movement Artifact . . . . .	42
4.10	Elimination of Power-Line Artifact . . . . .	42
4.11	K-Complex . . . . .	43
4.12	Block Diagram of K-Complex Parameter Extraction . . . . .	43
4.13	EEGKComplex Extraction(EEG K-Complex) . . . . .	44
4.14	EEGKComplex Extraction(Movements) . . . . .	44
4.15	Classification of Stage N1 . . . . .	46
4.16	Sleep Transitions Detection and Correction . . . . .	46
4.17	Sleep Transitions Described by AASM . . . . .	47
4.18	Sleep Transitions: Impossible (Red Arrow) and Irregular (Blue Arrow) . . . . .	47
4.19	Irregular Transitions Reporting . . . . .	48
5.1	Digital Parameter: EMGActivity . . . . .	50
5.2	Thresholds for EMGActivity . . . . .	51
5.3	Dependencies among Thresholds and other conditions . . . . .	53
5.4	Thresholds for EMGActivity (Stage W) of Patient 3774 . . . . .	54
5.5	Thresholds for EMGActivity (Stage N1) of Patient 3774 . . . . .	55
5.6	Thresholds for EMGActivity (Stage R) of Patient 3774 . . . . .	55
5.7	Thresholds for EMGActivity (Stage W) of Patient 55341 . . . . .	56
5.8	Thresholds for EMGActivity (Stage N1) of Patient 55341 . . . . .	57
5.9	Thresholds for EMGActivity (Stage R) of Patient 55341 . . . . .	57

6.1	Thresholds Setting-Up for Symbolic Fusion based Sleep Staging System . . .	60
6.2	Generic Feedback System Control . . . . .	61
6.3	Automatic Thresholds Setting-Up for Symbolic Fusion based Sleep Staging System . . . . .	63
6.4	Flow Chart of Differential Evolution used in ATSU . . . . .	67
6.5	Flow Chart of Cross Entropy used in ATSU . . . . .	70
6.6	Work Flow of Hybrid Sleep Staging System . . . . .	71
6.7	Personalized Sleep Staging System . . . . .	72
6.8	Differential Evolution-based Personalized Sleep Staging System . . . . .	73
6.9	Cross Entropy-based Personalized Sleep Staging System . . . . .	74
7.1	Evaluation of SF-SSS . . . . .	79
7.2	F-Measure of Existing SF-SSS VS SF-SSS with Modifications for stage N2	81
7.3	Total Agreement Rate of Existing SF-SSS VS SF-SSS with Modifications .	82
7.4	Agreement Rate of Existing SF-SSS VS SF-SSS with Modifications for Subject 639 . . . . .	82
7.5	DE Control Parameters Selection: DE/rand/1/bin, NP=5D . . . . .	86
7.6	DE Control Parameters Selection: DE/rand/1/bin, NP=10D . . . . .	86
7.7	DE Control Parameters Selection: DE/rand/1/bin, NP=500 . . . . .	87
7.8	DE Control Parameters Selection: DE/rand/1/bin . . . . .	87
7.9	CE Control Parameters Selection (Iteration=200) . . . . .	89
7.10	DE Control Parameters Selection (Iteration=50) . . . . .	89
7.11	Comparison between SF-SSS and PSSS . . . . .	94
7.12	Hypnogram of Subject 9 using DE-PSSS . . . . .	95
7.13	Heatmap of Confusion Matrix of Database 2 using DE-PSSS . . . . .	96
7.14	Heatmap of Confusion Matrix of Database 2 using CE-PSSS . . . . .	97

# List of Tables

3.1	AASM rules for sleep staging . . . . .	19
3.2	Comparison among Decision Tree methods . . . . .	22
3.3	Comparison among Artificial Neural Network methods . . . . .	23
3.4	Comparison among Support Vector Machine method . . . . .	24
3.5	Comparison among Rule-based methods . . . . .	25
3.6	Comparison among Hybrid methods . . . . .	25
3.7	Comparison among Typical Stochastic Search Algorithms . . . . .	30
4.1	Correlation Between Digital and First-Level Feature Parameters via Thresholds Setting-Up . . . . .	36
4.2	AASM Recommendations for Filter Settings . . . . .	41
4.3	Impossible Sleep Transitions Correction . . . . .	48
5.1	Thresholds used in SF-SSS model . . . . .	54
5.2	Thresholds of EMGActivity including Sleep Stages Dependencies . . . . .	56
5.3	Thresholds Configuration for Each Sleep Stage . . . . .	58
6.1	Main Control Parameters of Differential Evolution . . . . .	68
7.1	Subject Information of Database 1 . . . . .	76
7.2	Information of Database 1 . . . . .	76
7.3	Subject Information of Database 2 . . . . .	78
7.4	Information of Database 2 . . . . .	78
7.5	Confusion Matrix of Stage W . . . . .	80
7.6	F-Measure on Database 1 . . . . .	80
7.7	Confusion Matrix of Database 1 using Existing SF-SSS . . . . .	83
7.8	Confusion Matrix of Database 1 using Modified SF-SSS . . . . .	83
7.9	Main Control Parameters of Differential Evolution . . . . .	85
7.10	Parameters used for DE Control Parameters Selection Simulation . . . . .	85
7.11	Parameters used for CE Control Parameters Selection Simulation . . . . .	88
7.12	Training Set & Evaluation Set: 5% & 95 % for Subject ID=1 . . . . .	90
7.13	F-Measure Comparison among training set & evaluation set of DE-PSSS . . . . .	90
7.14	F-Measure Comparison among training set & evaluation set of CE-PSSS . . . . .	91
7.15	Comparisons between DE and CE . . . . .	92
7.16	Computational Time (s) of DE for one iteration . . . . .	92
7.17	Computational Time (s) of CE for one iteration . . . . .	92
7.18	F-Measure . . . . .	93
7.19	Average F-Measures of SF-SSS (with modifications) using MTSU on Database 2 . . . . .	93
7.20	F-Measure of SF-SSS and PSSS . . . . .	94
7.21	Confusion Matrix of Database 2 using DE-PSSS . . . . .	96

7.22	Confusion Matrix of Database 2 using CE-PSSS . . . . .	97
7.23	Agreement Rate Comparison between PSSSs and Inter-raters . . . . .	98
7.24	Agreement Rate Comparison between PSSSs and Other Works . . . . .	98
8.1	Recall Comparison between SF-SSS (with modifications) and existing SF-SSS on Database 1 . . . . .	105
8.2	Precision Comparison between SF-SSS (with modifications) and existing SF-SSS on Database 1 . . . . .	106
8.3	Agreement Rate of existing SF-SSS on Database 1 . . . . .	106
8.4	Agreement Rates of SF-SSS (with modifications) on Database 1 . . . . .	107
8.5	F-Measures of Evaluation Set using 5% Training Set (DE-PSSS) on Database 2 . . . . .	107
8.6	F-Measures of Evaluation Set using 10% Training Set (DE-PSSS) on Database 2 . . . . .	108
8.7	F-Measures of Evaluation Set using 15% Training Set (DE-PSSS) on Database 2 . . . . .	108
8.8	F-Measures of Evaluation Set using 20% Training Set (DE-PSSS) on Database 2 . . . . .	109
8.9	F-Measures of Evaluation Set using 5% Training Set (CE-PSSS) on Database 2 . . . . .	109
8.10	F-Measures of Evaluation Set using 10% Training Set (CE-PSSS) on Database 2 . . . . .	110
8.11	F-Measures of Evaluation Set using 15% Training Set (CE-PSSS) on Database 2 . . . . .	110
8.12	F-Measures of Evaluation Set using 20% Training Set (CE-PSSS) on Database 2 . . . . .	111
8.13	F-Measures of SF-SSS (with modifications) using MTSU on Database 2 . .	111
8.14	Agreement Rate and Cohen's Kappa of PSSSs . . . . .	112

# Chapter 1

## Introduction

### 1.1 Introduction

Sleep is an indispensable part of daily life and contributes to self-repairing and self-recovering. However, sleep disorders involving signs and symptoms like excessive daytime sleepiness, irregular breathing or increased movement during sleep, difficulty in sleeping, and abnormal sleep behaviors are affecting more and more people. It deteriorates the quality of life and becomes a significant cause of morbidity and mortality [28]. Figure 1.1 presents the prevalence of one typical sleep disorders: insomnia in Europe. Over 26% people in France were reported suffering with insomnia.

In clinical sleep analysis, an overnight Polysomnography (PSG) test is required for physicians to give an accurate diagnosis and appropriate treatment. It is usually done by a technician by placing sensors on the patient's body for recording multiple bio-signals. Based on recorded PSG signals, a detailed analysis and interpretation will be given by a physician (Europe) or a sleep expert (USA) with recommendations for the diagnosis and treatment. Sleep staging, as a fundamental step of PSG interpretation, needs 3 to 4 hours for physicians to interpret an overnight PSG recordings. American Academy of Sleep Medicine (AASM) manual, as the gold standard for the sleep study, describes the definitions of sleep events and sleep stages. Physicians can realize clinical sleep staging by combining different sleep events they observed from PSG signals into each stage according to AASM manual.

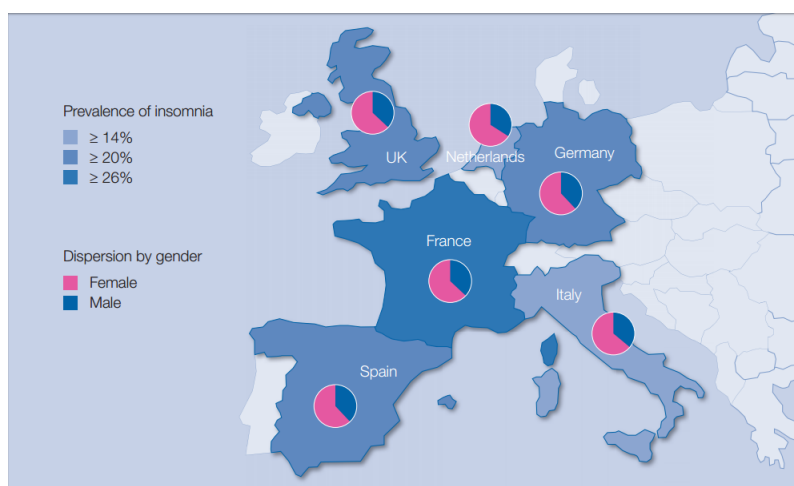


Figure 1.1 – Prevalence of Insomnia in Europe

Clinical sleep staging is a time-consuming task. Meanwhile, inter-rater reliability concerns also exist due to subjective interpretation and decision by different physicians. In

[10], 80.6% and 82.0% inter-rater reliability was reported by using Rechtschaffen & Kales (R&K) (old gold standard for sleep study) and AASM (new gold standard for sleep study), respectively.

In order to reduce the burden of physicians, automatic sleep staging systems have attracted extensive attention. Numerous attempts have been undertaken to automate interpretation of PSG recordings. Most of existing automatic sleep staging systems are using numerical classification methods without taking medical knowledge or physicians experience into consideration. Numerical methods are used to learn the patterns between features and corresponding stages classes. The pattern recognition is usually established by interaction with a set of training data. Patterns used in the classification of sleep stages are mainly dependent on the training data. Insignificant patterns may be selected independent of medical knowledge and without validation from physicians. However, for physicians, they cannot be easily convinced by these kinds of systems without involving any medical knowledge; thus, these kinds of systems are not really used by physicians in clinical practice.

From physician's perspective, the major issues for most of existing automatic sleep staging systems can be described as: 1) a system without taking any medical knowledge into consideration cannot win fully trust from physicians; 2) most of systems are not yet accepted and validated by physicians. Only by addressing these issues, a system can really be used in clinical practice for help physicians in diagnosis and treatment of sleep disorders.

## 1.2 Thesis Outline

The next chapter explains the major issues of existing sleep staging systems. In chapter 3, state of the art on existing sleep staging systems and thresholds setting-up methods are presented. In chapter 4, an existing Symbolic Fusion-based Sleep Staging System (SF-SSS) is presented followed by an improved release in order to enrich the SF-SSS model. In chapter 5, Manual Thresholds Setting-Up (MTSU) method that used in existing SF-SSS is introduced, followed by detailed analysis of these thresholds. In chapter 6, an Automatic Thresholds Setting-Up (ATSU) method based on Feedback System Control (FSC) is proposed. Two different search algorithms for FSC are evaluated: (1) Differential Evolution; (2) Cross Entropy. Inspired by ATSU method, Personalized Sleep Staging System (PSSS) is investigated by combining symbolic fusion and FSC in an original hybrid expert system. Evaluations of this system are presented in chapter 7. At last, the conclusion and perspectives are given in chapter 8.

## Chapter 2

# Problem Statement

### 2.1 Problem Statement

In order to mimic clinical sleep staging process by translating AASM and medical knowledge into computer logic, a Symbolic Fusion-based Sleep Staging System (SF-SSS) was proposed by Ugon, Isabelle, et al. [76]. SF-SSS starts from the extraction of digital parameters from raw polysomnography (PSG) signals and goes up to high-level symbolic interpretation of feature parameters while using thresholds. At last, decision is generated using rules inspired by international guidelines in sleep medicine and applied to feature parameters. SF-SSS is a knowledge-based approach, from low-level digital parameters extraction to high-level symbolic interpretation of feature parameters and decision-making rules in decision-level, which is directly inspired from the gold standard AASM manual and medical knowledge. It can be understood, accepted and validated by physicians according to their knowledge and experience.

SF-SSS model is the first attempt by using symbolic fusion to realize sleep staging. However, it is only a proof from conception to validation of symbolic fusion for sleep staging application. There still exist several points which need to be improved in order to enrich the knowledge-based used in the sleep staging system: (1) Pre-processing of PSG signals was not included; (2) Not all the sleep patterns described by AASM manual were implemented; (3) Classification of stage N1 was not performed; (4) After obtaining Hypnogram based on the output of classifiers, smoothing of Hypnogram to detect and remove irregular and incorrect sleep transitions was not included; (5) Manual Thresholds Setting-Up (MTSU) method was performed.

Beside these points motioned above for SF-SSS, there exist several issues which need further research. These issues exist not only in SF-SSS but also in most of sleep staging systems. They are described in below.

1. How to realize automatic sleep staging by taking individual variability into consideration, from the conception to implementation, validation and final practical use?
2. How to assess the effectiveness of personalized sleep staging system?
3. Is it necessary to propose a personalized sleep staging system which can take individual variability into consideration? Or a generalized sleep staging system is sufficient?

### 2.2 Context

In this thesis, SF-SSS model is adopted to develop a sleep staging system which can really be accepted and used for physicians in clinical practice. We start from the comple-



tion of SF-SSS model by proposing modifications in order to improve the performance of existing SF-SSS.

Then, thresholds used to transform low-level digital parameters into the high-level symbolic interpretation of feature parameters are studied. In SF-SSS model, methods used to extract the digital parameters, symbolic interpretation of feature parameters and decision rules in making classification of sleep stage are generic. While, thresholds used to transform digital parameters into a symbolic interpretation of feature parameters are different. The issues we mentioned above turned out to be thresholds setting-up problems. To address these issues, we researched from the following parts:

1. **Semantic interpretations of sleep events by using thresholds:** details of thresholds dependencies are analyzed; the number of thresholds used to realize classification of each sleep stage are studied.

Thresholds dependencies among sleep stages have been evaluated to understand whether same thresholds can be used in different sleep stages or different thresholds are required among different sleep stages.

Thresholds dependencies among patients have been evaluated to fully understand whether generalized thresholds are sufficient, or personalized thresholds are required to take individual variability into consideration.

Number of thresholds needed to realize classification of each sleep stage have been researched to understand how many thresholds are required for classifying each stage.

2. **An automatic way to set-up thresholds is researched.**

An automatic thresholds setting-up method based on Feedback System Control (FSC) is proposed. Inspired by the automatic method we proposed, a personalized sleep staging conception is presented by combining symbolic fusion and FSC. Symbolic fusion is dedicated to mimic the decision-making process of clinical sleep staging. Feedback System Control (FSC) is designed to provide personalized thresholds for Symbolic Fusion in transforming digital parameters into a symbolic interpretation of feature parameters while taking individual variability into consideration. Differential Evolution-based Personalized Sleep Staging System (DE-PSSS) and Cross Entropy-based Personalized Sleep Staging System (CE-PSSS) are presented by using two different search algorithms Differential Evolution and Cross Entropy respectively.

3. **Evaluations on a new clinical database** for proposed personalized sleep staging systems are presented.

The proposed systems are evaluated on a clinical database of PSG recording from 16 subjects (9 males and 7 females) ranging from 22 to 82 years old. These PSG recordings are fully scored by a physician following the international guidelines in sleep medicine.

## Chapter 3

# State of the Art

In this chapter, basic information of sleep analysis is introduced. Meanwhile, the state of the art of existing automatic sleep staging systems is presented. Finally, thresholds setting-up problems in sleep staging systems are described followed by a review of existing thresholds setting-up methods.

### 3.1 Introduction on Sleep Analysis

This section mainly introduces basic information of sleep analysis involving, polysomnography (PSG) signals, manuals of clinical sleep staging for better understanding the field of sleep analysis.

#### 3.1.1 Polysomnography (PSG)

Polysomnography is the gold standard test by recording the biophysiological changes that occur during sleep to diagnose sleep disorders. PSG test generally is conducted by a technician using PSG device (Figure 3.1). Several sensors (Figure 3.2) will be put on a patient for recording multi-signals like: EEG<sup>1</sup>, EOG<sup>2</sup>, EMG<sup>3</sup>, SpO<sub>2</sub><sup>4</sup>, ECG<sup>5</sup>, nasal and oral airflow, limb movement, and body position as shown in Figure 3.3.



Figure 3.1 – PSG Device

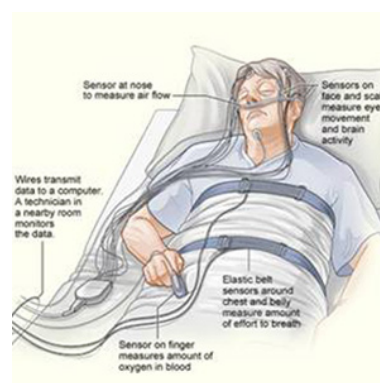


Figure 3.2 – PSG Recordings

---

<sup>1</sup>Electroencephalogram

<sup>2</sup>Electrooculogram

<sup>3</sup>Electromyogram

<sup>4</sup>Oxygen Saturation

<sup>5</sup>Electrocardiogram

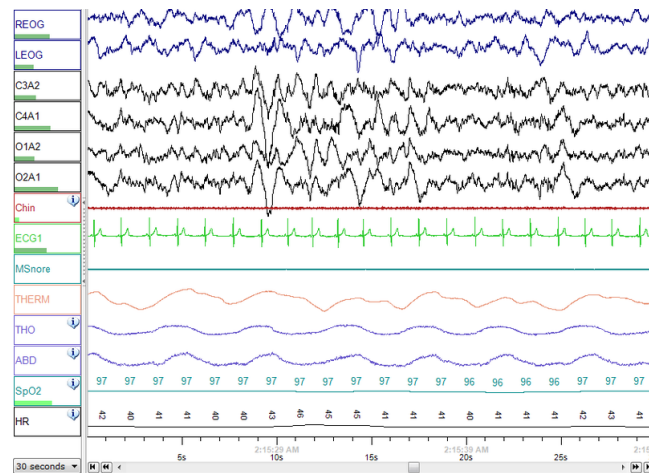


Figure 3.3 – PSG Signals

In sleep staging, EEG, EOG, and EMG are three classical physiological signals which are used for the fundamental sleep analysis and they are described in below:

- **EEG** is a recording of the electrical activity of the brain from the scalp.
- **EOG** is a recording of the voltages generated by eyes movements.
- **EMG** is a recording of chin muscle tone activity.

### 3.1.2 Clinical Sleep Staging Manual

In 1968, Rechtschaffen and Kales (R&K) sleep scoring manual was published as the first standardized criteria for sleep staging [55]. On the basis of R&K manual, there are 6 sleep stages, namely Stage Wake, S1, S2, S3, S4 and REM. Even R&K manual was considered as the only widely standard for describing the human sleep process for approximately 40 years, it has also been criticized for leaving plenty of room for subjective interpretation, which leads to a great variability in the visual evaluation of sleep stages [26].

In 2007, American Academy of Sleep Medicine (AASM) modified the R&K standard guidelines for sleep classification and released a new edition of sleep staging criterion. AASM manual [67] is the result of a review of the literature, analysis and consensus which addresses 7 topics: reporting parameters for PSG, technical and digital specifications, visual scoring, arousal, cardiac and respiratory events, movements and pediatric scoring. According to AASM manual, there are five sleep stages: stage W, Non-Rapid Eye Movement (stage N1, N2 and N3) and Rapid Eye Movement (stage R). Different stages are used to describe different state from awake to deep sleep.

- **Stage W** represents the wake state, ranging from full alertness through early stages of drowsiness.
- **Stage N1** is considered as a transition between wake and sleep.
- **Stage N2** muscle activity decreases and conscious awareness of the outside world begins to fade completely.
- **Stage N3** often referred as deep sleep or slow-wave sleep and during this period the sleeper is even less responsive to the outside environment.
- **Stage R** is the stage where dreams occur and eyes move rapidly. While the chin muscle tone activity stays low.

In AASM manual, it describes the definitions of sleep events (e.g. Sleep Spindles, K-Complex, etc.) and rules involving sleep events in order to score sleep stages. Details of main rules described in AASM for guidance the classification of different sleep stages are shown in Table 3.1. Physicians can realize clinical sleep staging by combining different sleep events they observed from PSG signals into each stage according to AASM manual.

Sleep Stages	AASM
<b>W</b>	<p><b>A.</b> Score epochs as stage W when more than 50% of the epoch has alpha rhythm over the occipital region.</p> <p><b>B.</b> Score epochs without visually discernible alpha rhythm as stage W if ANY of the following are present:</p> <ol style="list-style-type: none"> <li>Eye blinks at a frequency of 0.5-2 Hz</li> <li>Reading eye movements</li> <li>Irregular, conjugate rapid eye movements associated with normal or high chin muscle tone</li> </ol>
<b>N1</b>	<p><b>A.</b> In subjects who generate alpha rhythm, score stage N1 if the alpha rhythm is attenuated and replaced by low-amplitude, mixed-frequency activity for more than 50% of the epoch.</p> <p><b>B.</b> In subjects who do not generate alpha rhythm, score stage N1 commencing with the earliest of ANY of the following phenomena:</p> <ol style="list-style-type: none"> <li>EEG activity in range of 4-7 Hz with slowing of background frequencies by <math>\geq 1</math> Hz from those of stage W</li> <li>Vertex sharp waves</li> <li>Slow eye movements</li> </ol>
<b>N2</b>	<p>If EITHER OR BOTH of the following occur during the first half of that epoch or the last half of the previous epoch:</p> <ol style="list-style-type: none"> <li>One or more K complexes unassociated with arousals</li> <li>One or more trains of sleep spindles</li> </ol>
<b>N3</b>	Score stage N3 when $\geq 20\%$ of an epoch consists of slow wave activity, irrespective of age.
<b>R</b>	Score stage R sleep in epochs with ALL of the following phenomena: <ol style="list-style-type: none"> <li>Low-amplitude, mixed-frequency EEG</li> <li>Low chin EMG tone</li> <li>Rapid eye movements</li> </ol>

Table 3.1 – AASM rules for sleep staging

### 3.1.3 Hypnogram

Hypnogram is used to represent the stages of sleep as a function of time. It provides a visual depiction of the symbolic behavior of sleep. In clinical sleep staging, physicians can realize sleep scoring by interpreting PSG recordings according to AASM manual. Meanwhile, physicians also take the succession of stages through time into consideration in order to integrate all scored stages for generating a Hypnogram. A Hypnogram normally consists of 5 to 6 sleep cycles. A complete sleep cycle takes an average of 90 to 110 minutes. The first sleep cycle has relatively short REM sleep and long periods of deep sleep but later in the night, REM periods lengthen and deep sleep time decreases. Figure 3.4 illustrates hypnogram of a health person. However, in reality, hypnograms deviate from person to person. Especially, for the person with sleep disorders, hypnogram is more disrupted. Figure 3.5 shows a Hypnogram of a patient who suffers from severe sleep apnea.

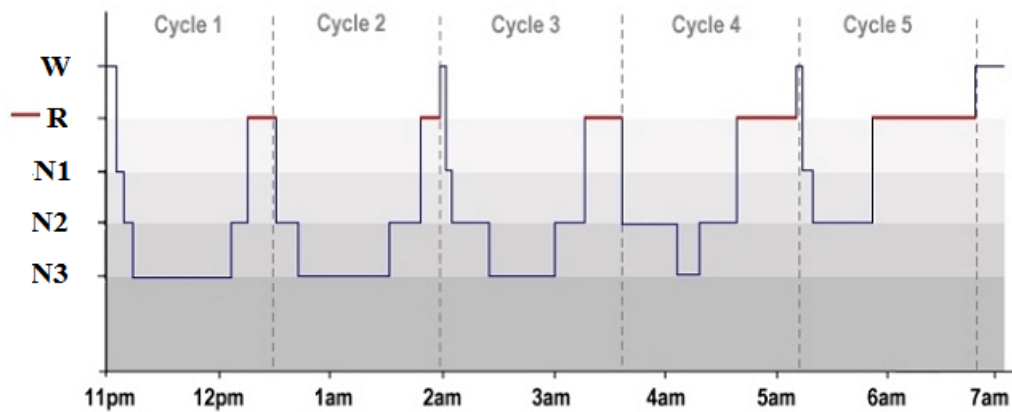


Figure 3.4 – An Example of Hypnogram

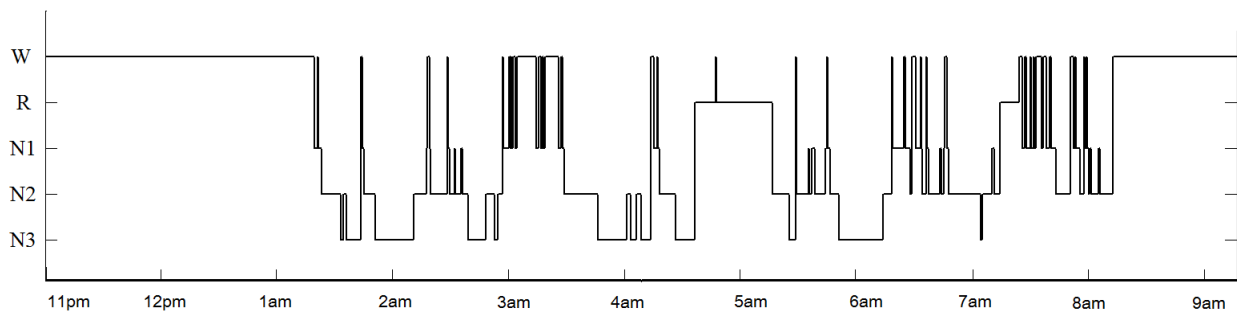


Figure 3.5 – An Example of Hypnogram of Sleep Apnea Patient

## 3.2 Automatic Sleep Staging Systems

Due to the time-consuming and inter-rater reliability of clinical sleep staging, automatic sleep staging systems have attracted extensive attention. Methods used in current sleep staging systems can be classified in machine learning methods, rule-based methods, hybrid method and symbolic fusion method. The details are discussed in the following section.

### 3.2.1 Machine Learning Methods

Machine learning is the most widespread method in sleep staging. Typical approaches in machine learning such as Decision Tree, Artificial Neural Network (ANN) and Support Vector Machine (SVM) have been widely used in classifying sleep stages. Digital parameters are extracted from PSG signals using time-domain, frequency-domain or non-linear analysis; then different approach is used to learn the patterns between the digital parameters and the corresponding classes. The pattern recognition is usually established by interaction with a set of training data. For the training set, it needs to be scored by physicians in advance. Brief description of these methods is presented in below.

#### 3.2.1.1 Decision Tree

The history of tree methods can be traced back to 1963 [51], since then, Decision tree became very popular. It applies the tree structure which classifies instances from the root to leaf. Each node in the tree specifies an attribute of the instance; each branch from

the node corresponds to one of the possible values for this attribute and leaves represent class labels. An instance is classified by starting at the root node of the tree, checking the specified attribute by this node, then moving down the branch which corresponds to the value of the attribute. This process is repeated until it reaches the leaf. Figure 3.6 shows an example of a simple decision tree.

In 2002, an automatic sleep staging system using decision tree was proposed by Masaaki [25]. Six feature parameters are extracted from five biological signals: 2 EEG channels, 2 EOG channels and one EMG channel. In this approach, it takes a tree structure in which the explanatory attributes (feature parameters) correspond to the nodes, their values correspond to the branches, and the classes (sleep stages) are assigned to the leaves. Five decision trees are built based on the training set which involving all the database. Then, classification is made based on the majority decision from the results of these five trees.

In 2012, an automated sleep staging system was proposed based on a single EEG channel using random forest [20]. In 2013, based on the heart rate variability, an automatic system to identify Stage W, REM and NREM was proposed using random forest [84].

Besides these Decision Tree methods, there exist several works also using a tree structure to make the decision for classifying different sleep stages [46, 44]. Figure 3.6 presents a tree build in [46] while using 13 nodes. Comparisons among sleep staging systems based on Decision Tree methods are shown in Table 3.2.

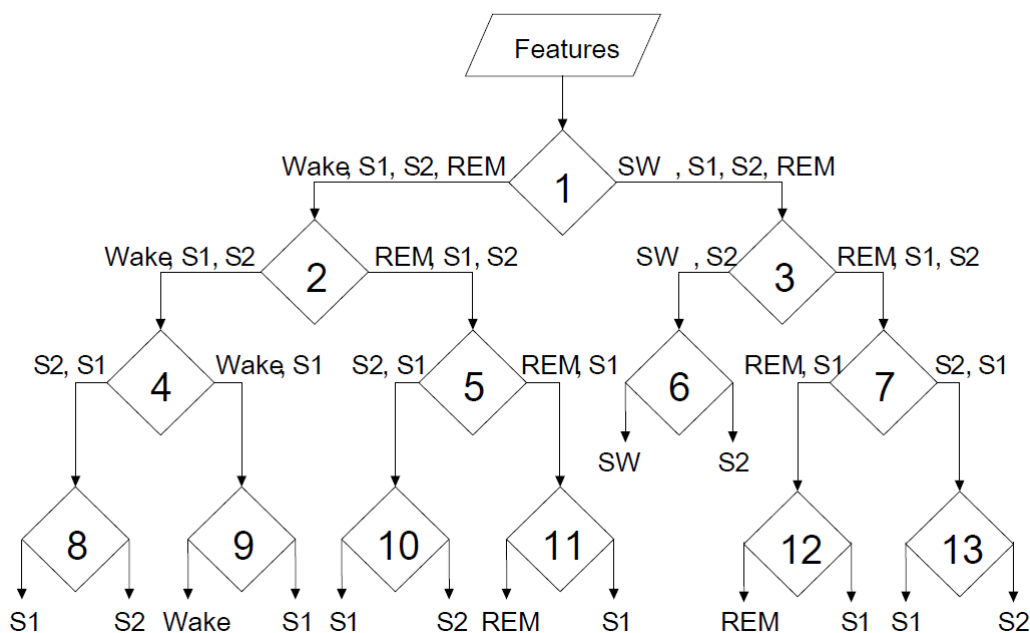


Figure 3.6 – An example of Decision Tree

As one of the most popular machine learning methods, decision tree is very intuitive and easy to explain with a tree structure. It also can predict a classification with fast speed. However, to build a tree, it requires a large training set. Meanwhile, without proper pruning or limiting tree growth, it tends to over-fit the training data which may make poor classifications.

Authors	Signals	Database	Classification	Results
Masaaki et al. [25]	EEG, EOG EMG	1 male training: 80% & testing: 20%	SWS(S3,S4),SREM, S1,S2,S3,S4,MT <sup>6</sup>	Agreement Rate: 81.5%
Luay et al. [20]	EEG	16 subjects training: 66% & testing: 34%	W,N1,N2,N3,R	Accuracy: 83%
Meng et al. [84]	ECG	45 subjects training: 98% & testing: 2%	W,NREM,REM	Accuracy: 72.6%
Sheng-Fu et al. [46]	EEG,EOG EMG	20 subjects training: NM <sup>7</sup> & testing: 80%	W,S1,S2, SWS(S3,S4),R	Agreement Rate: 87%
Khai et al. [44]	EEG,EOG EMG	5 subjects training: NM <sup>7</sup> & testing: NM <sup>7</sup>	W,N1,N2, N3,R,MT <sup>6</sup>	Agreement Rate: 78.8%

Table 3.2 – Comparison among Decision Tree methods

### 3.2.1.2 Artificial Neural Network

ANN, as an effective approach for pattern recognition and classification, inspired by the architecture and function of the human brain. As the human brain can be described as a biological neural network (an interconnected web of neurons transmitting elaborate patterns of electrical signals. Dendrites receive input signals and, based on those inputs, fire an output signal via an axon), ANN can be described as a network of neurons that receives inputs, processes those inputs, and generates an output [64]. The network are typically organized in three layers: input, hidden and output layer. The input layer serves to introduce the information into the network. The hidden layers represent a connection between the input and output layers. The final output of the whole network forms in the output layer. All neurons in adjacent layers are interconnected and each connection is defined as a weight and is represented with a rational number.

In 1993, MultiLayer Perceptron (MLP) as a typical technique of Artificial Neural Network was proposed to realize automatic sleep staging in [62]. The architecture of this work is shown in Figure 3.7. Seventeen feature parameters were extracted from 4 biological signals: 2 EEG channels, one EOG horizontal channel and one EMG channel. Firstly, a three-layer perceptron network architecture with a 17-unit input layer, a 10-unit hidden layer and a 6-unit output layer was build based on a training set. Secondly, ambiguity epochs and artifacts epochs were rejected before the classification. Finally, automatic sleep staging was performed.

---

<sup>6</sup>Movement

<sup>7</sup> Not Mentioned

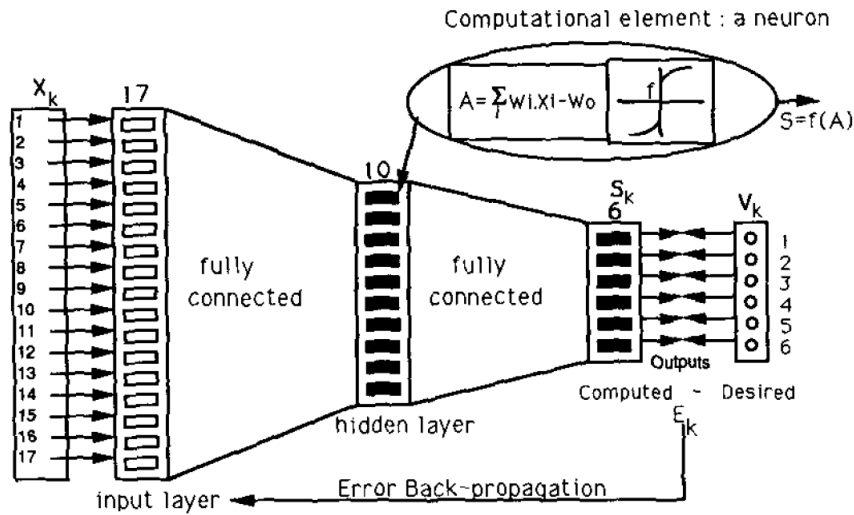


Figure 3.7 – Structure of an Artificial Neural Network in [62]

Since 1993, several research works have also explored the MLP technique in automatic sleep staging [37, 71] which associates different input patterns to a set of output classes during the training set. In 2005, an automatic sleep staging system architecture with 5-unit input layer, a 6-unit hidden layer and a 6-unit output layer based on one channel EEG was proposed [37]. In 2010, a multi-layer neural network with two hidden layers that simultaneously employs EEG, EMG and EOG was proposed [71]. Comparisons among sleep staging systems based on ANN methods are shown in Table 3.3.

ANN has the learning ability to form its own structure automatically based on training instances, which is robustness to noise and environment. However, when additional learning is performed based on some new training instances, the weights in the network are often changes. A long learning time is required if additional learning is performed by integrating all training instances [25].

Authors	Signals	Database	Classification	Results
Nicolas et al. [62]	EEG,EOG EMG	12-night recordings training: 51% & testing: 49%	W, MT <sup>6</sup> , S1, S2, S3, S4, REM	Agreement Rate: 80. 6%
Nizar et al. [37]	EEG	1 male training: NM <sup>7</sup> & testing: NM <sup>7</sup>	W,N1,N2, N3,REM	Agreement Rate: 76%
Tagluk et al. [71]	EEG, EOG, EMG	21 subjects training: 98% & testing: 2%	S1,S2, S3,R	Accuracy: 74.7%

Table 3.3 – Comparison among Artificial Neural Network methods

### 3.2.1.3 Support Vector Machine

Support Vector Machine was introduced by Vladimir N. Vapnik which constructs a linear classifier by finding the hyperplane that maximizes the margin between two classes in 1963 [78]. In 1999, nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes were proposed in non-linear classification problems [79]. As SVM is originally designed for binary classification, in order to solve multi-class scenario, different frameworks are proposed involving one-against-one and one-against-all. By constructing and combining several binary classifiers, these frameworks are able to perform multi-class classification by reducing a multi-class problem to binary ones. The one-against-all



framework constructs  $k$  separate binary classifiers for  $k$ -class classification. The  $i$ -th binary classifier is trained using the data from the  $i$ -th class as positive examples and the remaining  $(k - 1)$  classes as negative examples. During the test, the class label is determined by the binary classifier that gives maximum output value. Another framework: one-against-one, it evaluates all possible pairwise classifiers and thus induces  $k(k - 1)/2$  individual binary classifiers. Applying each classifier to a test example would give one vote to the winning class. A test example is labeled to the class with the most votes.

In 2005, six binary classifiers were constructed in classifying sleep stages into four classes (Wake, Light Sleep, Slow Wave Sleep and REM) by applying one-against-one framework [23]. A common Radial Basis Function (RBF) kernel was used to construct each binary classifier based on twenty-three feature parameters which were extracted from EEG. In classification part, each binary classifier is considered to be a voting, the class label of highest number of votes was assigned as the final class label. Based on one-against-one framework, a sleep staging system with ten binary classifiers combining EEG, EOG and EMG was proposed in 2010 [50]. Sleep stages were classified into five different stages (Wake, S1, S2, S3&S4 and REM).

In 2014, a sleep staging system based on one-against-all framework were proposed [1]. Five classes binary classifiers were constructed on basis of linear kernel function in classifying Stage Wake, S1, S2, S3 and S4. Each classifier was constructed to separate one class from the rest four classes. In the classification, the final class label was determined by the binary classifier that gives maximum output value. Comparisons among sleep staging systems based on SVM methods are shown in Table 3.4.

For SVM-based sleep staging system, it has high algorithmic complexity and requires extensive memory. For one-against-one framework-based sleep staging system, more binary classifiers need to be constructed. While, for one-against-all framework-based sleep staging system, it imbalances the training set and does not take the competence of classifiers into consideration in the classification part.

Authors	Signals	Database	Classification	Results
Steinn et al. [23]	EEG,EOG EMG	4-night recordings training: 75% & testing: 25%	W, LS(S1,S2), SWS(S3,S4),R	Agreement Rate: 76%
Antonio et al. [50]	EEG,EOG EMG	9-night recordings training: 33% & testing: 33%	W,S1,S2 SWS(S3,S4),R	Agreement Rate: 70%
Khalid et al. [1]	EEG	13 subjects training: 80% & testing: 20%	W,S1,S2, S3,S4	Agreement Rate: 92%

Table 3.4 – Comparison among Support Vector Machine method

### 3.2.2 Rule-based Methods

In 2011, a rule-based inference method was proposed by Sheng-Fu Liang [45]. Nine digital parameters are extracted according to R&K manual. Total 14 fuzzy rules based on these digital parameters are applied in a fuzzy inference system to realize sleep staging.

Sleep staging based on inference system using a set of rules is also proposed in [18]. Eight digital parameters are extracted according to AASM manual. Based on these digital parameters, total 111 fuzzy rules are used in a fuzzy inference system to perform sleep staging.

In [45, 18], digital parameters are extracted based on the medical guidance. However, instead of directly translating R&K or AASM manual into machine rules which can be recognized by the computer, set rules are manually defined according to the physician's

experience and the distributions of digital parameters respecting to various sleep stages. Directly translating R&K or AASM manual into machine rules is very challenge because of the ambiguous and implicit of information content from the medical guidance.

Authors	Signals	Database	Classification	Results
Sheng-Fu et al. [45]	EEG,EOG EMG	16 subjects training: 20% & testing: 80%	W,S1,S2, SWS(S3,S4),R	Agreement Rate: 87%
Diego et al. [18]	EEG,EOG EMG	33 subjects training: NM <sup>7</sup> & testing: 100%	W,N1,N2, N3,R	Agreement Rate: 84%

Table 3.5 – Comparison among Rule-based methods

### 3.2.3 Hybrid Methods

Consider advantages/disadvantage of machine learning and rule-based methods, hybrid methods have been proposed in the last few years. Hybrid method either combines two different machine learning methods or combine machine learning method with rule-based method.

In 2011, Haoyu Ma et al. proposed a hybrid classification method using an artificial neural network and decision tree to realize automatic sleep staging [48]. However, inherent limitation in this method is that it cannot separate W and S1 stages.

In 2015, Tarek Lajnef et al. presented a hybrid classification method using decision tree and support vector machine [40]. The overall accuracy report for this method is high except for stage S1. This method is not suitable to distinguish stage S1 with stage W with high accuracy.

Beside [48, 40], there is also another hybrid method combine rules with machine learning method [19]. However, the improvement of accuracy of the hybrid method is not clearly instead of the complexity of algorithm increases. Comparisons among sleep staging systems based on hybrid methods are shown in Table 3.6.

Authors	Signals	Database	Classification	Results
Haoyu et al. [48]	EEG	NM <sup>7</sup> training: NM <sup>7</sup> & testing: NM <sup>7</sup>	W,(W,S1),S2, SWS(S3,S4), R	Agreement Rate: 88.8%
Tarek et al. [40]	EEG,EOG EMG	15 subjects training: 67% & testing: 33%	W,S1,S2, SWS(S3,S4),R	Agreement Rate: 74%
Farag et al. [19]	EEG	13 subjects training: 55% & testing: 45%	W,S1,S2, S3,R	Accuracy: 85.18%

Table 3.6 – Comparison among Hybrid methods

### 3.2.4 Symbolic Fusion Method

Symbolic fusion is an efficient decision-making technique involving interdisciplinary among signal processing, artificial intelligent, statics and so on. It has been proven to be efficient to fuse information from different sources, possibly heterogeneous. Considering own limitations and uncertain perceptions from every single source, symbolic fusion provides enhanced and complementary perceptions combining different sources which increase overall accuracy.

Instead of only depending on numerical classification methods like machine learning methods, Symbolic Fusion based method was proposed to realize sleep staging [76] in 2013 which adopts three level architecture proposed by Dasarathy as shown in Figure 3.8. It uses AASM manual as a guidance to imitate clinical sleep staging process. With the cooperation between engineers and clinicians, a set of digital parameters was extracted using signal processing methods; then these digital parameters were transferred into symbolic features using manual interpretation of boundaries definition; lastly, a composite decision for sleep staging based on symbolic features and pre-defined rules in accordance with AASM was performed. In comparison to other works, this is the only method based on knowledge instead of based on data.

It has been evaluated on a database of 16 subjects (4 males and 12 females) ranging from 26 to 67 years old. Agreement Rate can reach 76%, 54%, 60% and 71% for stage W, N2, N3 and R respectively.

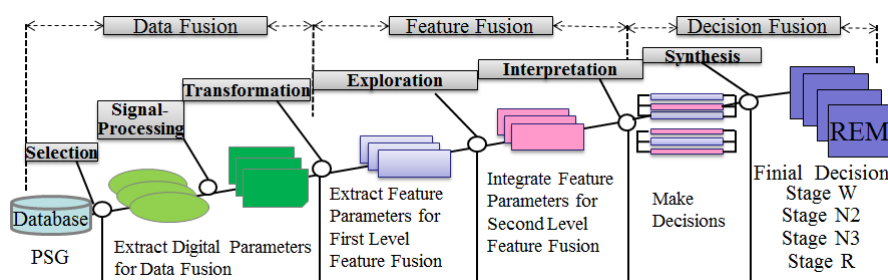


Figure 3.8 – Symbolic Fusion-based Sleep staging System Model

### 3.3 Thresholds Setting-Up Problem and Methods

#### 3.3.1 Thresholds Setting-Up Problem

Thresholds have been widely used in existing sleep staging systems. Three typical applications of thresholds are described below.

- **Artifacts Detection**

To detect and reject artifacts, thresholds have been used in [72]. A pre-defined value was defined as the threshold to differentiate the boundary of artifacts and useful signals. Recorded PSG data which exceeds the threshold is considered to be artifact, otherwise, it considered as useful data.

- **Body Movement Detection**

In [47], thresholds were used to detect the body movement. A fixed value was defined as threshold to differentiate body movement and non-body movement.

- **Differentiate Different Features**

In rule-based and Symbolic Fusion sleep staging systems, thresholds are widely applied in transforming digital parameters into linguistic or symbolic features in order to model inference process under the guidance of medical knowledge [45, 46, 76].

The boundaries of linguistic or symbolic features are very flexible. In clinical practice, physician may adjust the boundaries for each linguistic or symbolic feature according to his or her experience and patient information. In sleep staging systems, thresholds used to determine the boundaries in differentiating different features also should have high flexibility which can take individual variability into consideration.

Despite thresholds have been widely applied, there exists very limited research on how to set up these thresholds due to the following reasons: 1) To build a mathematical model or a threshold function to differentiate boundary requires a set of data with sufficient quantity and adequate quality; 2) There is lack of uniformity between subjects and thresholds variability exists.

Several methods to set up thresholds used in existing research are described below:

- **Set up a fixed value for all the subjects.**

It is a simply way to set up thresholds, as shown in [47]. However, it is not suitable to deal with high inter-subject variability PSG signals in sleep staging analysis. Even with the normalization process, a fixed value is not sufficient.

- **Manually assigned different values to the thresholds for different subjects.**

It takes individual variability into consideration, as shown in [45, 46, 76], which is more precise than using a fixed value for all subjects. While, it requires manual efforts and it is a time-consuming process.

- **List several possible thresholds values and choose the optimal values for the thresholds.**

Several possible thresholds values are enumerated, and the values which can reach highest agreement rate among enumerated values are selected as the optimal values for the thresholds [80]. The accuracy of this method highly depends on values it enumerates. With the number and range of thresholds increase, this method cannot be an effective method.

Besides the methods listed above, Fuzzy Logic also be proposed in setting-up threshold as shown in next section. Meanwhile, a possible Thresholds Setting-Up Method using Feedback System Control technique is presented followed by Fuzzy Logic.

### 3.3.2 Thresholds Setting-Up Method: Fuzzy Logic

To solve Thresholds Setting-Up problem, fuzzy logic was proposed in recognition of electroencephalogram pattern [29, 31]. Instead of directly setting values of thresholds for the boundaries to differentiate different features, soft boundaries were applied in fuzzy logic. A generic fuzzy logic system is composed of three parts: Fuzzification, Fuzzy Inference Engine and Defuzzification. Crisp inputs are fuzzified into linguistic values to be associated with the input linguistic variables. After fuzzification, the inference engine refers to the fuzzy rule base containing fuzzy IF-THEN rules to derive the linguistic values for the intermediate and output linguistic variables [69]. Once the output linguistic values are available, the defuzzifier produces the final crisp values from the output linguistic values.

However, membership functions need to define in Fuzzification and thresholds still requires to be set to transfer linguistic values into the final crisp values in Defuzzification.

### 3.3.3 Thresholds Setting-Up Method: Feedback System Control

In 2008, a generic platform technology Feedback System Control was proposed by Pro.Chih-Ming Ho [82]. It can be used to find optimal input parameters combination for guiding the complex system toward to the desired state, which involving four parts: Input Stimuli, Complex System, Objective Functions and Stochastic Search Algorithm.

With its robustness and rapid ability in searching and discovering combinatorial parameters, this technique has been applied in biological cells domain, it has been demonstrated to be very effective in optimizing the combinatorial drugs for eradicating cancers [65], inhibiting viral infections [14, 82] and maintaining human embryonic stem cells [74].

### 3.3.3.1 Typical Stochastic Search Algorithms for Feedback System Control

There are several typical Stochastic Search Algorithms: Gur Game, Evolutionary Algorithm, Simulated Annealing, Tabu Search, Ant Colony Optimization and Cross Entropy. In this section, different types of SSA are introduced and the comparisons among these SSAs are also presented.

- **Gur Game**

Gur Game is a simple, but very popular approach [73, 75]. It is useful for the global optimization problems, where the objective function may be non-convex, non-differentiable, and possibly discontinuous over a continuous or discrete domain.

Gur Game is based on biased random walks of finite-state automate. The automate describe a set of parameters with assigned values and a set of rules is included for determining how the different values of the parameter switch from one value to the other. Each value of the parameter is referred as a state of the automaton. The overall goal of the automated design is to have the parameters to self-organize (choose the optimal values) in an attempt to maximize the overall performances.

- **Evolutionary Algorithm**

Evolutionary Algorithm is a generic population-based metaheuristic optimization algorithm proposed in 1996 [4]. It is inspired by biological evolution which proposed by Charles Darwin [11].

In Darwin's Theory of Evolution, the adaptive change of species obeys the principle of natural selection, which means individuals that are more fit have better potential for survival, as in the well-known phrase "survival of the fittest". In a nutshell, as random genetic mutations occur within an organism's genetic code, the beneficial mutations are preserved because they aid survival. These beneficial mutations are passed on to the next generation. Over time, beneficial mutations accumulate and the result is an entirely different organism. In Evolutionary Algorithm, individuals among a population which has higher fitness function values in term of solutions quality can be survived. Survived individuals are taking place after the mutation, recombination, and selection in simulating biological evolution. Typical Evolutionary Algorithm like Differential Evolution and Genetic Algorithm have been widely applied in System Optimization [56], Scheduling [53], and Feature Selection[85].

- **Simulated Annealing**

Simulated Annealing (SA) is a popular stochastic algorithm proposed by Scott Kirkpatrick [39] in 1983. It has been widely applied in Network Design [15], Sequencing [32].

SA is inspired from the physical process of annealing in metal. Annealing involves heating and cooling a material to alter its physical properties due to the changes in its internal structure. Heat a material over melting point and then cool it, the material properties of the substance depend on the rate of cooling. To obtain a perfect or nearly perfect crystal, it must be annealed by first melting and then cooling very slowly. If the substance is cooled very quickly, a crystal with many defects will

be formed. Simulated annealing interprets slow cooling as a slow decrease in the probability of accepting worse solutions as it explores the solution space. It models the physical process of heating a material and then slowly lowering the temperature to decrease defects, thus minimizing the system energy.

- **Tabu Search**

Tabu Search, proposed by Fred W. Glover in 1986 [21] and formalized in 1989 [22]. It is a metaheuristic search method employing local search methods which can be used for solving combinatorial optimization problems. It has been widely applied in VLSI design [60], financial analysis [17] and scheduling [16].

Tabu search uses a local search procedure to iteratively move from one potential solution to its best neighbor, even if this results in a deterioration of the performance measure value. It uses memory structures to record information about solution properties that change in moving from one solution to another. If a potential solution has been previously visited within a certain short-term period or if it has violated a rule, it is marked as "tabu" (forbidden) so that the algorithm does not consider that possibility repeatedly.

- **Cross Entropy**

Cross Entropy was initially proposed to estimate probabilities of rare events for complex stochastic networks by Rubinstein [59] in 1997. It has been extended to solve combinatorial optimization problems in 1999 [58], which turned out to be an effective method.

Cross Entropy involves an iterative procedure where each iteration can be broken into two phases: 1) Generate a random data sample (trajectories, vectors, etc) according to a specific mechanism. 2) Update the parameters of the random mechanism based on the data to produce the "better" sample in the next iteration. By applying these two phases in each iterative procedure, it has been widely applied in Buffer Allocation [3], Medical Image Segmentation [83], Network Reliability Estimation [30] and so on.

- **Ant Colony Optimization**

Ant Colony Optimization (ACO) proposed by Marco Dorigo in 1992 which initially aimed to search for an optimal path in a graph, based on the behavior of ants seeking a path between their colony and a source of food [9].

Within ants, the media to communicate among individuals information regarding paths and used to decide where to go called pheromone trail. Initially, ants wander randomly, and upon finding food return to their colony while laying down pheromone trails. If other ants find such a path, they are likely to follow the trail instead of keep traveling randomly. If they eventually find food, the pheromone trail will be reinforced. However, the pheromone trail also will evaporate over time which results in reducing its attractive strength. The more time it takes for an ant to travel down the path and back again, the more pheromones will evaporate. In comparison, a short path can get marched over more frequently, and the pheromone density is higher than longer paths.

Based on the original idea, ACO has also been diversified to be applied in Water Distribution Systems [49], Routing in Telecommunication Network [13] and so on.

SSAs like Tabu Search, Gur Game, Simulated Annealing, Evolutionary Algorithm and Cross Entropy, it has its own pros and cons. While in order to apply SSA for sleep staging system applications, the following constraints should be considered: 1) The search

algorithm can search for the optimal thresholds combination efficiently and should be robust. For SSAs, initial values of thresholds will be generated to start searching of the optimal thresholds. Without knowledge of initial values, the search algorithm used in sleep staging system applications should be insensitive to initial values. 2) The search algorithm should be flexible. The search algorithm used in sleep staging system applications should be flexible and adaptive to the increment or decrement of thresholds. 3) In consideration of the further work like an implementation of sleep staging system applications or integrated it with other portable PSG systems for sleep evaluation or home care application, constraints of hardware implementation should also be considered. Table 3.7 shows the comparison among several typical SSAs.

Properties	Gur Game	DE	SA	Tabu	ACO	CE
<b>Robustness</b> (Insensitive to initial values)	×	✓	✓	×	✓	✓
<b>Parallel Searching</b>	×	✓	×	×	✓	✓
<b>Non-Memory-Based Structure</b>	✓	✓	✓	×	✓	✓
<b>Easy to Implement</b>	✓	✓	✓	×	×	✓
<b>Flexible</b>	✓	✓	✓	✓	×	✓

Table 3.7 – Comparison among Typical Stochastic Search Algorithms

Among these SSAs, Differential Evolution and Cross Entropy are more suitable for sleep applications. Differential Evolution (DE) is a popular and efficient method of evolutionary algorithm, it owns many advantages like: 1) DE can mimic natural biological evolution and provide a fast and stable convergence; 2) It is less sensitive to initial population; 3) It is a parallel search method; and 4) It can improve fitness function value iteratively. Cross Entropy (CE) also owns several advantages: 1) CE is a parallel search method; 2) It is based on rigorous mathematical and statistical principles; 3) It provides a sample adaptive procedure.

### 3.4 Conclusion

Compared with other automatic sleep staging methods, symbolic fusion can mimic decision-making process of clinical sleep staging and it has the following features:

1. it can integrate data from different sources, like EEG, EMG and EOG which can provide enhanced and complementary decision in comparison to single source based methods;
2. it can deduce a composite decision because it is based on the cooperation between engineers and clinical experts which can involve medical knowledge and physician's experience;
3. it has less complexity and high flexibility which can represent information by using low-level to high-level architecture.

However, to improve existing symbolic fusion-based sleep staging system, the following points should be taken into consideration:

1. Pre-processing of PSG signals;
2. Involve more sleep patterns described by American Academy of Sleep Medicine (AASM) manual;

3. Classification of stage N1;
4. After obtaining Hypnogram based on the output of classifiers, Smoothing of Hypnogram in order to detect and remove incorrect sleep transitions.

Besides of these points, there exist an inherent flaw: manual thresholds setting up method was applied in transforming digital parameters into the symbolic interpretation of feature parameters.

In this thesis, a personalized automatic sleep staging system is proposed based on a hybrid expert system that combines Symbolic Fusion and Feedback System Control (FSC) technique. Symbolic fusion is used to mimic sleep staging process by integrating knowledge of experts in scoring of PSG signals according to AASM manual; meanwhile, Feedback System Control technique is adopted to realize automatic thresholds setting up with the consideration of individual variability. In order to directly apply existing symbolic fusion method, some modifications are presented in order to improve the accuracy and overcome the limitations. Personalized thresholds can be provided automatically using FSC technique into order to realize personalized automatic sleep staging.



## Chapter 4

# Symbolic Fusion-based Sleep Staging System

In this chapter, Symbolic Fusion paradigm is introduced. Then, an existing Symbolic Fusion-based Sleep Staging System (SF-SSS) model (proposed by Ugon, Isabelle, et al.) is presented. This model was reached good classification results on a small database. However, this model is not yet completed. In this chapter, I present this model and my contributions in order to enrich this model.

### 4.1 Symbolic Fusion Conception

Symbolic fusion is an efficient decision-making technique involving interdisciplinary among signal processing, artificial intelligent, inference methods, statics and so on. It has been widely applied in image processing [7, 41], medical analysis [77], TV program [43] which proven to be efficient to fuse information from different sources. Considering uncertain perceptions from every single source, symbolic fusion can provide enhanced and complementary perceptions combining different sources which increase overall accuracy.

There exist several different architectures for symbolic fusion, two typical architectures are introduced in the following paragraph.

#### 4.1.1 JDL Architecture

In 1991, the Joint Directors of Laboratories (JDL) provided a definition for symbolic fusion as "*A multi-level process dealing with the association, correlation, combination of data and information from single and multiple sources to achieve the refined position, identify estimates and complete and timely assessments of situations, threats and their significance*" [34]. Meanwhile, JDL also presented a five-level architecture for symbolic fusion.

- **Level 0: Source Preprocessing**

Source Preprocessing is the lowest level of the process, it provides estimation and prediction of signal/object observable states on the basis of signal level data association and characterization. It includes signal detection and feature extraction which can reduce the amount of data and maintain useful information for the high-level processes.

- **Level 1: Object Assessment**

Object Assessment provides estimation and prediction of entity states on the basis of observation-to-track association, continuous state estimation and discrete state

estimation. It includes spatiotemporal alignment, association, correlation, clustering or grouping techniques, state estimation, identity fusion, and the combining of features that were extracted from signals.

- **Level 2: Situation Assessment**

Situation Assessment estimates and predicts of relations among entities, to include force structure and cross force relations, communications and perceptual influences, physical context, etc.

- **Level 3: Impact Assessment**

Impact Assessment performs estimation and prediction of effects on situations of planned or estimated/predicted actions by the participants; to include interactions between action plans of multiple players.

- **Level 4: Process Refinement**

Process Refinement involves adaptive data acquisition and processing to support mission objectives.

JDL architecture can be used to support human decision-making by refining and reducing the quantity of information from the data instead of knowledge.

#### 4.1.2 Dasarathy Architecture

In 1997, a three-level architecture was proposed by Dasarathy [12] involving data fusion, feature fusion and decision fusion.

- **Data Fusion**

Data fusion performs signal detection and digital parameters extraction which can maximize useful information and minimize noise and artifacts. It is conducted immediately after the data are gathered from the sensors; then signal processing algorithms are employed to extract features or characteristics that describe an entity in the environment.

- **Feature Fusion**

Feature fusion performs feature set uniformization and normalization, feature reduction and concatenation, and feature matching. It is used to improve, refine or obtain a set of features based on the digital parameters which were extracted in data fusion.

- **Decision Fusion**

In decision fusion, a composite decision is generated based on inference methods. Inference methods which are tolerant of imprecision, uncertainty, partial truth and approximation are usually applied.

Different from JDL architecture, Dasarathy architecture can be used to mimic human decision-making by modeling directly from the knowledge. In comparison with Dasarathy Architecture, JDL Architecture focused on providing communications among fusion researchers and implementation engineers, rather than a prescription for implementing fusion system or exhaustive enumeration of fusion functions and techniques [24]. While Dasarathy Architecture defined a very useful categorization of fusion functions in terms of the types of information that are processed as shown in Figure 4.1. These fusion functions are:

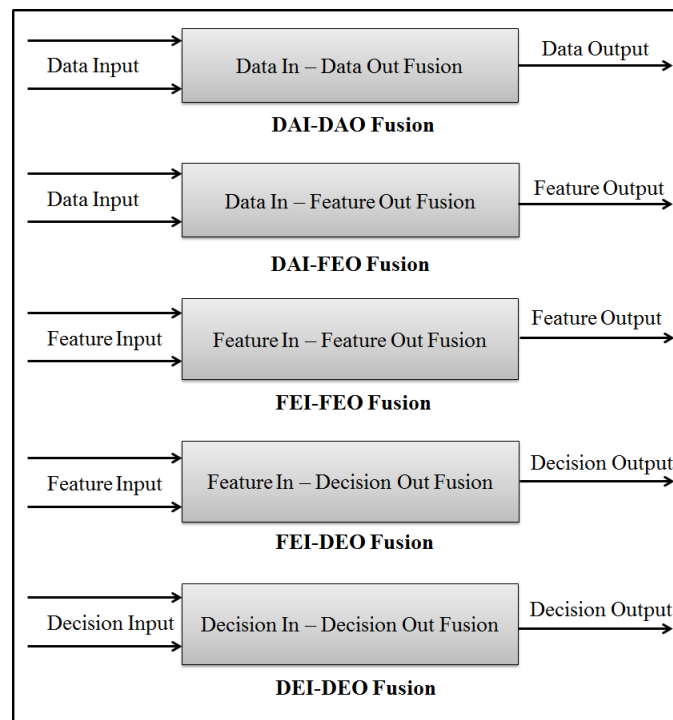


Figure 4.1 – Fusion I/O Characterizations of Dasarathy Architecture

1. **Data In-Data Out (DAI-DAO) Fusion:** This is the basic and elementary form of fusion which begins from data inputs and resulting in a form of data output. It has been commonly referred to as *data fusion* and conducted immediately after the data are gathered from the sensors. Signal and image processing algorithms are commonly used in this level.
2. **Data In-Feature Out (DAI-FEO) Fusion:** Data from different sensors or different bands of one sensor are combined to derive a feature of the object in the environment or a descriptor of the phenomenon under observation. It has been looked upon either *data fusion* or *feature fusion*.
3. **Feature In-Feature Out (FEI-FEO) Fusion:** In this level, both input and output of the fusion process are features. It addressed a set of features with to improve, refine or obtain new features. This process is also known as *feature fusion*.
4. **Feature In-Decision Out (FEI-DEO) Fusion:** It obtains a set of features as inputs and provides a set of decisions as outputs. It has been referred to as either *feature fusion* or *decision fusion*.
5. **Decision In-Decision Out (DEI-DEO) Fusion:** In this level, it fuses input decisions to obtain a new decision. It has been commonly referred as *decision fusion*.

Compare JDL Architecture and Dasarathy Architecture, Dasarathy Architecture is more suitable for the formalization of medical knowledge and experience for sleep staging application. With the guidance of AASM, it can be adopted to mimic the clinical sleep staging process, from low-level digital parameters extraction to high-level feature interpretation and final decision-making.

## 4.2 Existing SF-SSS

In 2013, Adrien, Isabelle, et al. [76] proposed Symbolic Fusion-based Sleep Staging System (SF-SSS) using Dasarathy Architecture, as shown in Figure 4.2. It starts from the extraction of digital parameters from raw PSG signals and goes up-to high-level symbolic interpretation of feature parameters. Finally, rules are used to make the decision. Digital parameters, symbolic interpretation and rules in SF-SSS are inspired by international guidelines in sleep medicine.

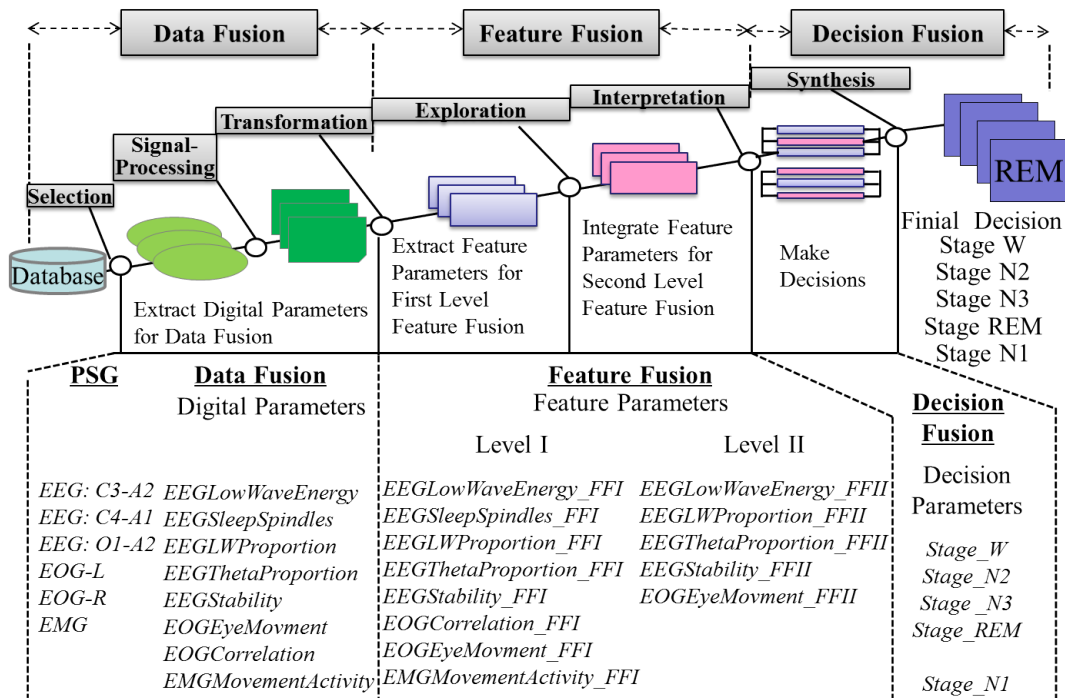


Figure 4.2 – Sleep Staging Design Flow

### 4.2.1 Data Fusion

In data fusion, it adopts DAI-DAO fusion, eight digital parameters were extracted using time-domain and frequency-domain signal processing methods. These parameters are used to represent the sleep events which are described in AASM manual. Parameters *EEGLowWaveEnergy*, *EEGSleepSpindles*, *EEGLWProportion*, *EEGThetaProportion* and *EEGStability* were extracted from three different EEG channels: C3-A2, C4-A1, O1-A2; parameters *EOGEyeMovement* and *EOGCorrelation* were extracted from 2 EOG channels: EOG-L, EOG-R; *EMGMovementActivity* parameter was extracted from EMG. Below is the brief description of each parameter. These parameters are inspired from AASM manual.

- **EEGLowWaveEnergy**

*EEGLowWaveEnergy* indicates the energy of the slow wave of frequency between 0.5 Hz to 2 Hz in EEG signal.

- **EEGSleepSpindles**

Sleep Spindle is a train of distinct waves with frequency 11-16 Hz and duration more than 0.5 seconds. It is a significant indicator of stage N2. *EEGSleepSpinles* represents the number of occurrence of sleep spindles.

- **EEGLWProportion**

*EEGLWProportion* signifies the power ratio between slow wave (frequency smaller than 2 Hz) and the total power of each 30-s epoch.

- **EEGThetaProportion**

*EEGThetaProportion* signifies the power ratio between Theta frequency band (frequency between 4 Hz and 7 Hz) and the total power of each 30-s epoch.

- **EEGStability**

*EEGStability* signifies the power ratio between the fast wave (frequency more than 18 Hz) and the total power. The fast wave is one characteristic of Stage W.

- **EOGEyeMovement**

*EOGEyeMovement* represents the number of times eye moves during sleep. This parameter is a significant indicator to distinguish between REM and NREM stage.

- **EOGCorrelation**

*EOGCorrelation* is the correlation between the left and right eye movements. It indicates whether the movement of the two eyes is independent or not.

- **EMGActivity**

*EMGActivity* represents the mean absolute value of the EMG signal. This parameter is used to indicate the activity level of EMG, which can be used as an indicator of the muscle tone movement in sleep staging.

#### 4.2.2 Feature Fusion

In feature fusion, 2-level fusion was applied. In the first level, digital parameters are transformed into the symbolic interpretation of feature parameters. Eight digital parameters were transferred into 21 different features via 13 thresholds as shown in Table 4.1. In the second level, new feature parameters were integrated either combining 3 EEG channels or 2 EOG channels. The symbolic interpretation of feature parameters is used to represent the semantic description of sleep events in AASM manual.

Digital Parameters	Features	Number of Thresholds
EEGLowWaveEnergy	High - Middle - Low	2
EEGSleepSpindles	Confidently Have - Not Confident	1
EEGLWProportion	High - Low	1
EEGThetaProportion	High - Low	1
EEGStability	Stable - Not Confident - Unstable	2
EOGEyeMovement	High - Middle - Low - Lowest	3
EOGCorrelation	Conjugate - Disconjugate	1
EMGActivity	High - Normal - Low	2

Table 4.1 – Correlation Between Digital and First-Level Feature Parameters via Thresholds Setting-Up

In first level of feature fusion, DAI-FEO fusion is performed. Figure 4.3 shows an example of first level of feature fusion. In second level of feature fusion, FEI-FEO fusion is adopted, as an example shown in Figure 4.4. At least two channels feature parameters are High, the output is High.

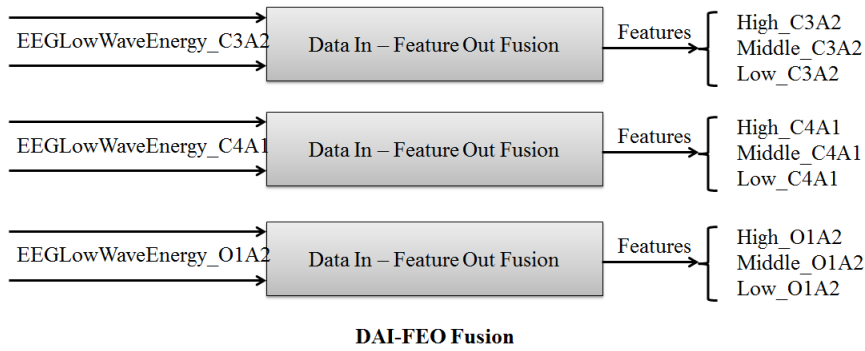


Figure 4.3 – First Level of Feature Fusion - DAI-FEO Fusion

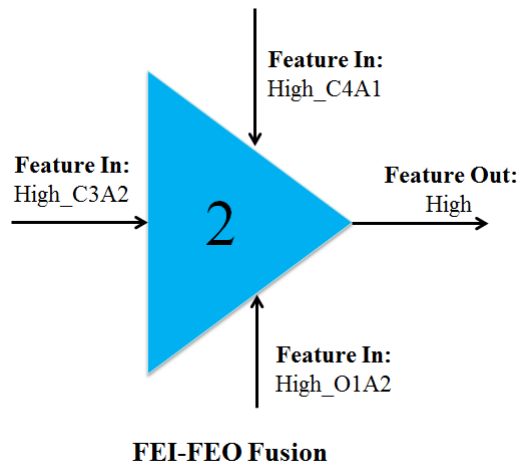


Figure 4.4 – Second Level of Feature Fusion - FEI-FEO Fusion

### 4.2.3 Decision Fusion

In decision fusion, inference method is used to fulfill sleep staging task on the basis of feature parameters. In order to generate a composite decision, a set of rules were defined for classifying of stage W, N2, N3 and R under the cooperation between clinical experts and engineers according to the guidance from AASM manual. Rules used to make the decision of each stage are different. Four different classifiers which combined by different rules are proposed for classifying stage W, N2, N3 and R respectively.

Figure 4.5 illustrates the classification of stage N2 when one of the following rules is satisfied:

1) *EOGCorrelation* is *Disconjugate* and *EEGSleepSpindles* is *Confidently Have* and *EEGStability* is *Stable*.

2) *EEGSleepSpindles* is *Confidently Have* and *EOGEyeMovement* is *Lowest* or *Low* and *EEGLowWaveEneergy* is *Low* and *EEGThetaProportion* is *High*.

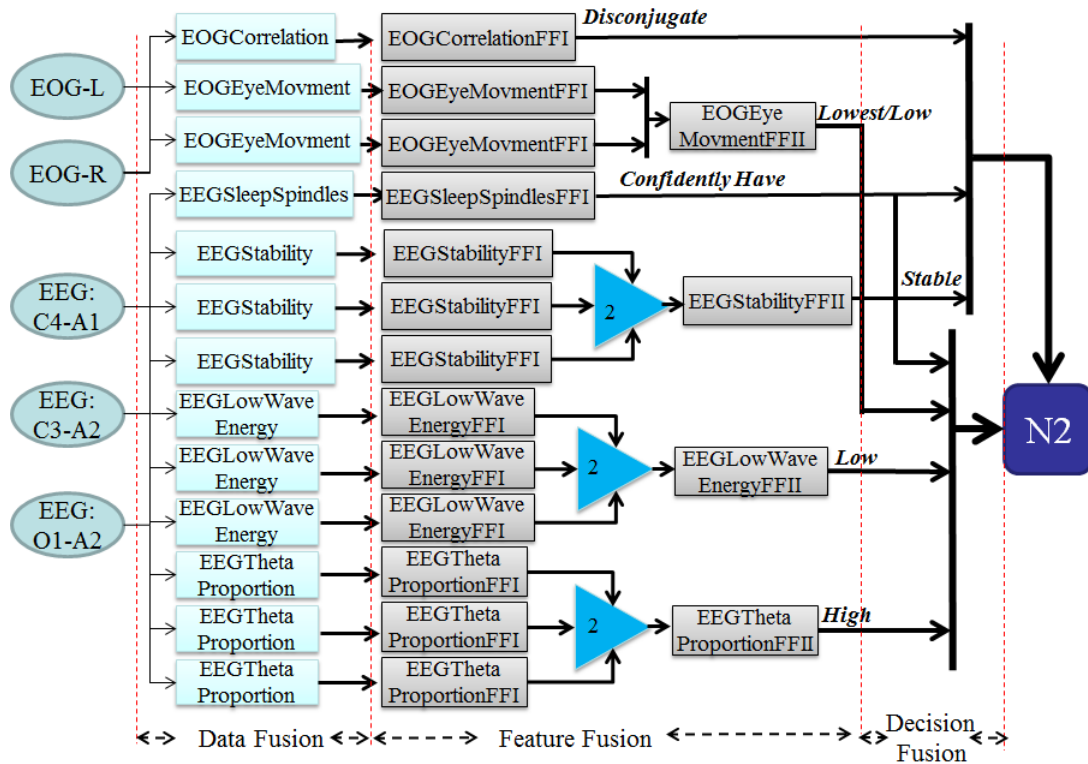


Figure 4.5 – Classification of Stage N2

### 4.3 Existing SF-SSS Performances and Limitations

SF-SSS starts from extraction of digital parameters from raw polysomnography (PSG) signals and goes up-to high-level symbolic interpretation of feature parameters while using thresholds. At last, decision is generated using rules inspired by international guidelines in sleep medicine and applied to feature parameters. It owns many advantages like:

- It integrates data from different sources, like EEG, EMG and EOG which can provide enhanced and complementary decision in comparison to signal source based methods;
- It can deduce a composite decision in accordance with medical knowledge under the cooperation between engineers and clinical experts; thus, 4 classifiers were designed to classify 4 out of 5 sleep stages (stage W, N2, N3 and R);
- It can be implemented in embedded systems. SF-SSS has been evaluated on a database of 16 subjects (4 males and 12 females) ranging from 26 to 67 years old. Agreement Rate can reach 76%, 54%, 60% and 71% for stage W, N2, N3 and R respectively.

However, this model is not yet complete and it needs to be enriched, below are some points need to be completed:

#### 1. Pre-processing of PSG signals

Although PSG like EEG, EOG and EMG are designed to record brain activity, eye movement activity and submental muscle activity, it may also record unwanted patterns which are not caused by underlying physiological event of interest. These unwanted patterns are called artifacts. In terms of artifacts, it can be divided into physiologic artifacts and extra-physiologic artifacts. Physiologic artifacts

are generated from body like head or body movement. Extra-physiologic artifacts are generated from outside the body like environment or instrument. Two typical artifacts are shown in Figure 4.6 and Figure 4.7.

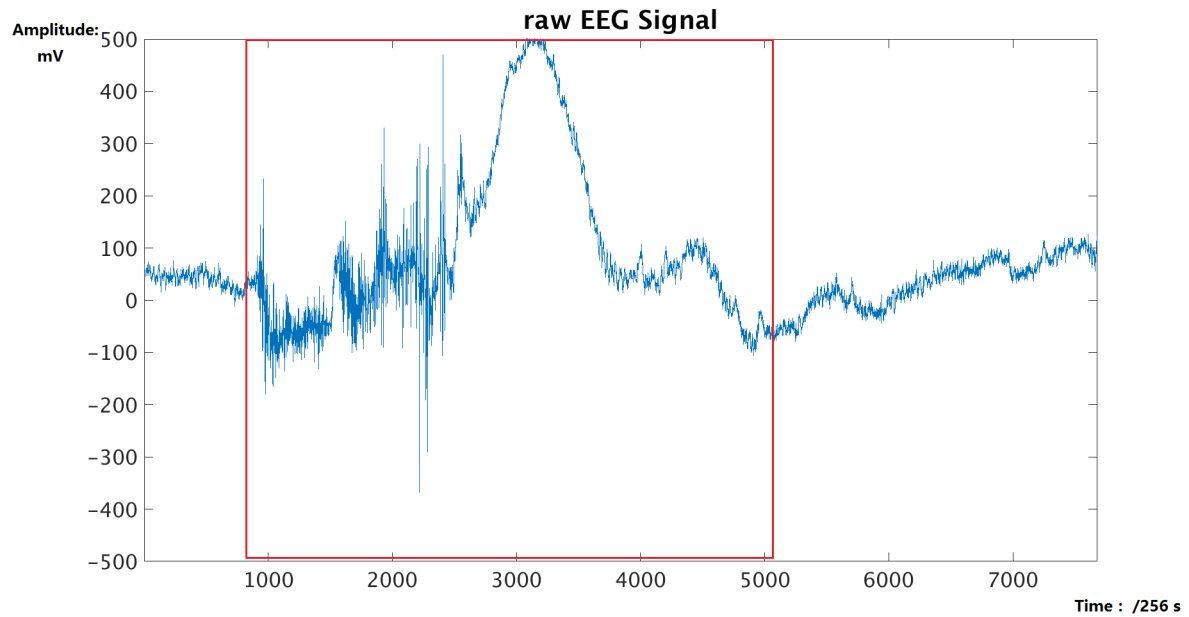


Figure 4.6 – Movement Artifact

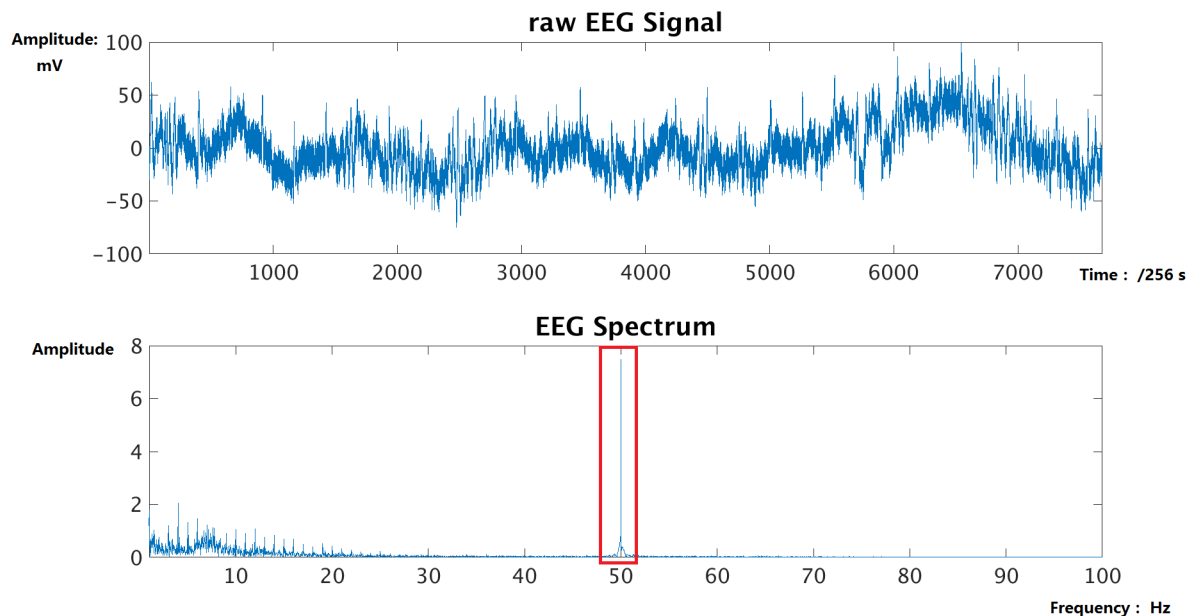


Figure 4.7 – Power Line Artifact

According to AASM manual, band pass filters with a cut of frequency of 0.3 - 35 Hz, 0.3 - 35 Hz and 10 - 100 Hz are suggested to perform pre-process for EEG, EOG and EMG respectively. Pre-processing can be used to eliminate some noise and artifacts.

## 2. Smoothing of sleep stages

Temporal contextual information was not considered, however, sleep staging is a time-dependent classification problem. A sleep stage could be influenced by the



previous sleep stage and it could influence the next sleep stage. Sleep transitions were also not considered. However, there exists some impossible transitions and irregular transitions.

### 3. Some missing sleep events and rules described by AASM

Most of the sleep events described by AASM manual were included in [76], however, there is still some sleep events which are not included. E.g. K-Complex, a well-delineated negative sharp wave immediately followed by a positive component standing out from the background EEG signal, with total duration  $\geq 0.5$  sec [67], which is also a significant indicator to classify Stage N2.

### 4. Classification of Stage N1

In [76], classification of Stage W, N2, N3 and R were performed, while classification of Stage N1 was not performed. Stage N1 is considered as a transition between wake and sleep. It occurs upon falling asleep and during brief arousal periods within sleep and usually accounts for 2 - 5% of total sleep time.

The detection of N1 is always the most problematic aspect of the sleep stages [80] in both clinical sleep staging and automatic sleep staging system. Only 63.0 % inter-scoring reliability for stage N1 is reported among different scorers in [57]. Moreover, finding a significant feature that could separate N1 from W, N2, N3 and R, is rather difficult for automatic sleep staging system, because N1 is a transition phase in the changes of wakefulness and other sleep stages.

## 4.4 Modifications of Existing SF-SSS

In this section, modifications for improving existing SF-SSS are proposed.

### 4.4.1 Sleep Staging Design Flow

A completion of sleep staging design flow is proposed, as shown in Figure 4.8. It consists of three main parts: Pre-processing and Segmentation, Symbolic Fusion and Smoothing. Pre-processing and Segmentation is designed to eliminate noise and artifacts and segment PSG recording into 30-s epochs in accordance with AASM manual. Symbolic Fusion is used to realize sleep staging. Smoothing is proposed to consider the temporal effects of sleep staging process, and to detect and correct falsely sleep transitions. Details of each part are described below.

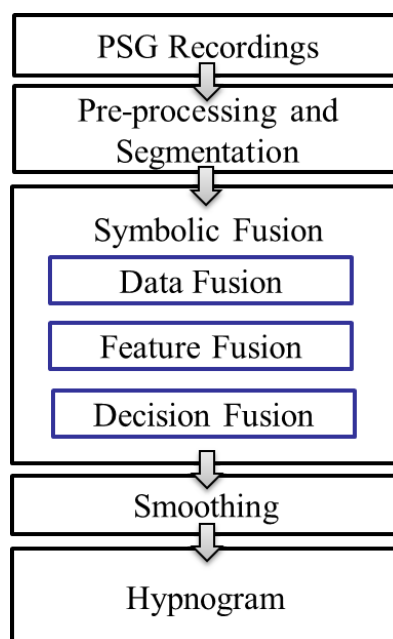


Figure 4.8 – Sleep Staging Design Flow

#### 4.4.2 Pre-processing and Segmentation

Pre-processing is designed to eliminate noise and artifacts. Details of filter setting recommended by AASM are shown in Table 4.2. For EEG and EOG, a band pass filter between 0.3 Hz and 35 Hz is suggested by AASM; For EMG, a band pass filter between 10 Hz and 100 Hz is recommended.

AASM Recommendations	Low Frequency	High Frequency
<b>EEG</b>	0.3 Hz	35 Hz
<b>EOG</b>	0.3 Hz	35 Hz
<b>EMG</b>	10 Hz	100 Hz

Table 4.2 – AASM Recommendations for Filter Settings

As recommend by AASM, a Butterworth bandpass filter between 0.3 Hz and 35 Hz is designed for EEG and EOG by using Filter Design and Analysis Tool in Matlab. Meanwhile, a Butterworth band-pass filter between 10 Hz and 100 Hz and a band-stop filter with a cut-off frequency of 50 Hz are designed for EMG.

By applying filters proposed in Pro-processing, it can effectively eliminate some noise and artifacts like movement artifacts and power-line artifacts. Figure 4.9 and Figure 4.10 presents the elimination of movement artifacts and power-line artifacts, respectively.

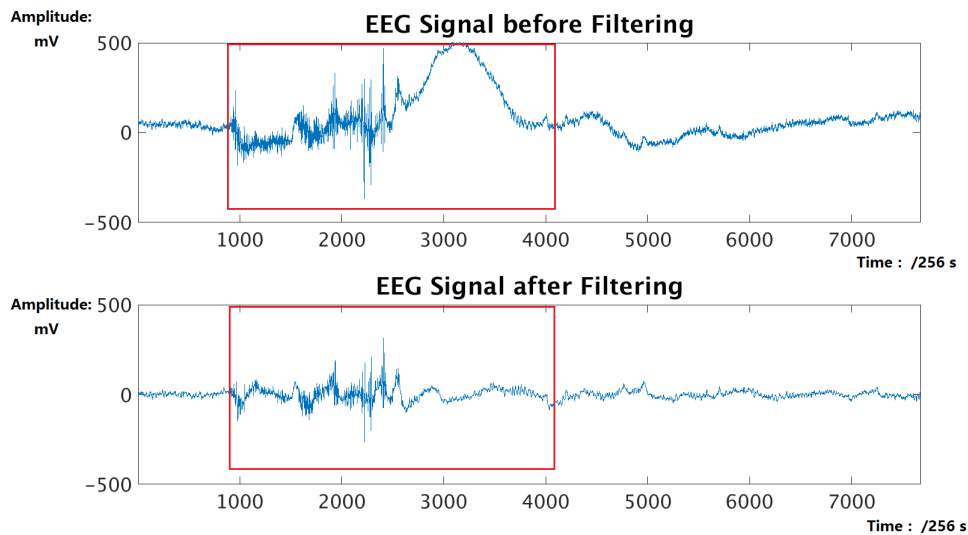


Figure 4.9 – Elimination of Movement Artifact

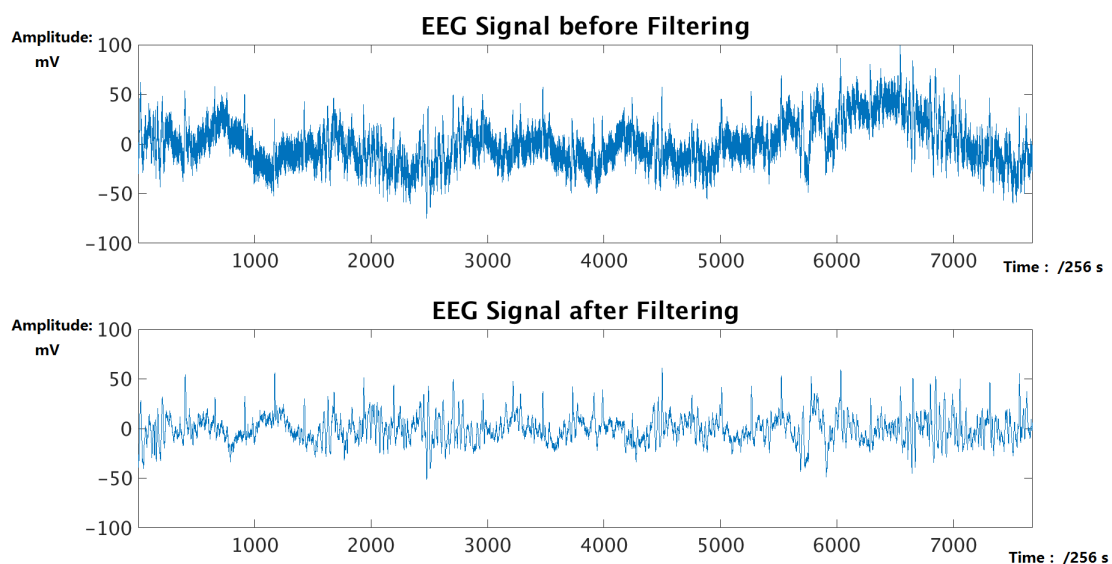


Figure 4.10 – Elimination of Power-Line Artifact

#### 4.4.3 Data Fusion: K-Complex

K-Complex is a significant indicator of stage N2 which was not involved in existing SF-based sleep staging system. To improve the classification performance of stage N2, another significant parameter K-Complex is included and fused to existing SF-SSS model in modifications. In this section, K-Complex is introduced firstly followed by the method we proposed to use for extracting the digital parameter of K-Complex.

In AASM manual, K-Complex is described as *"A well-delineated, negative, sharp wave immediately followed by a positive component standing out from the background EEG, with total duration  $\geq 0.5$  sec, usually maximal in amplitude when recorded using frontal derivations"* as shown in Figure 4.11. To extract K-Complex, Teager Energy Operator is introduced and adopted.

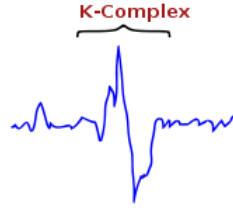


Figure 4.11 – K-Complex

#### 4.4.3.1 Teager Energy Operator

Teager Energy Operator (TEO) is a non-linear quadratic operator proposed by Kaiser [35] as a measure to assess the instantaneous energy of the signal that incorporates both amplitude and frequency of the signal. It owns the advantages for the detection of instantaneous changes in the signal such as discontinuities, increases or decreases of amplitude and frequency. TEO has been widely applied in speech analysis [33, 5] and bio-signal analysis [52, 2, 68].

The definition of continuous Teager Energy Operator is shown in Equation 4.1 where  $x(t)$  is the signal,  $\dot{x}(t)$  and  $\ddot{x}(t)$  are the first and second derivatives of the signal respectively.

$$\Psi_c[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \quad (4.1)$$

In the discrete domain, Teager Energy Operator is expressed as Equation 4.2, where  $x(n-1)$ ,  $x(n)$ ,  $x(n+1)$  are the  $(n-1)$ th,  $n$ th and  $(n+1)$ th sample value of the signal.

$$\Psi_d[x(n)] = x^2(n) - x(n-1)x(n+1) \quad (4.2)$$

TEO can be used to assess the instantaneous energy of the continuous and discrete signal that incorporates both amplitude and frequency of the signal.

#### 4.4.3.2 K-Complex Digital Parameter Extraction

Due to the specific characteristics of K-Complex such as sharp rising and falling edges and long duration, we adopt TEO to extract KComplex parameter as shown in Figure 4.12. A Butterworth low pass filter with the cut off frequency of 5 Hz is designed to eliminate high frequency waveforms. Then, TEO is calculate as *KComplex Parameter* to obtain rapid changes of amplitude and suppressed background activity.

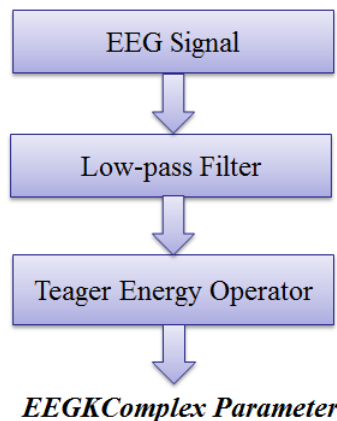


Figure 4.12 – Block Diagram of K-Complex Parameter Extraction

Figure 4.13 shows the TEO output of a 30-s EEG epoch which contains a K-Complex wave. From the figure, we can see, where the K-Complex wave occurs there is a high TEO value. TEO can be effectively present instantaneous changes in the signal like K-Complex in stage N2. Meanwhile, it also can present instantaneous changes like movement in stage W as shown in Figure 4.14.

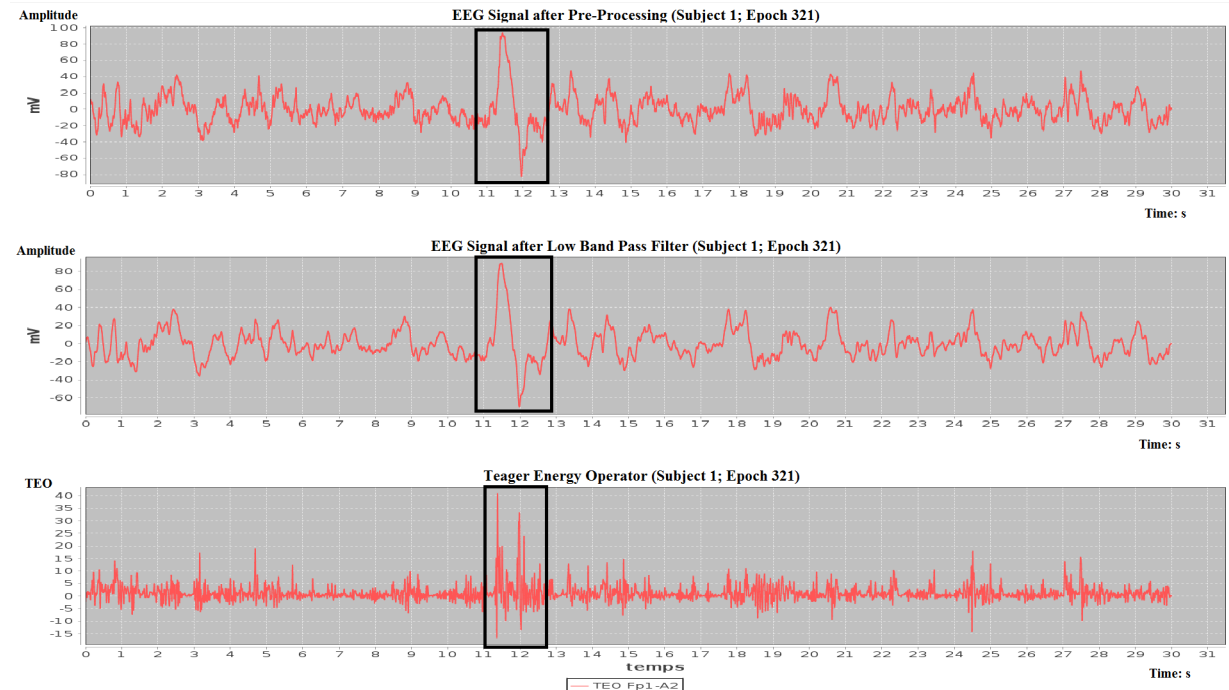


Figure 4.13 – EEGKComplex Extraction(EEG K-Complex)

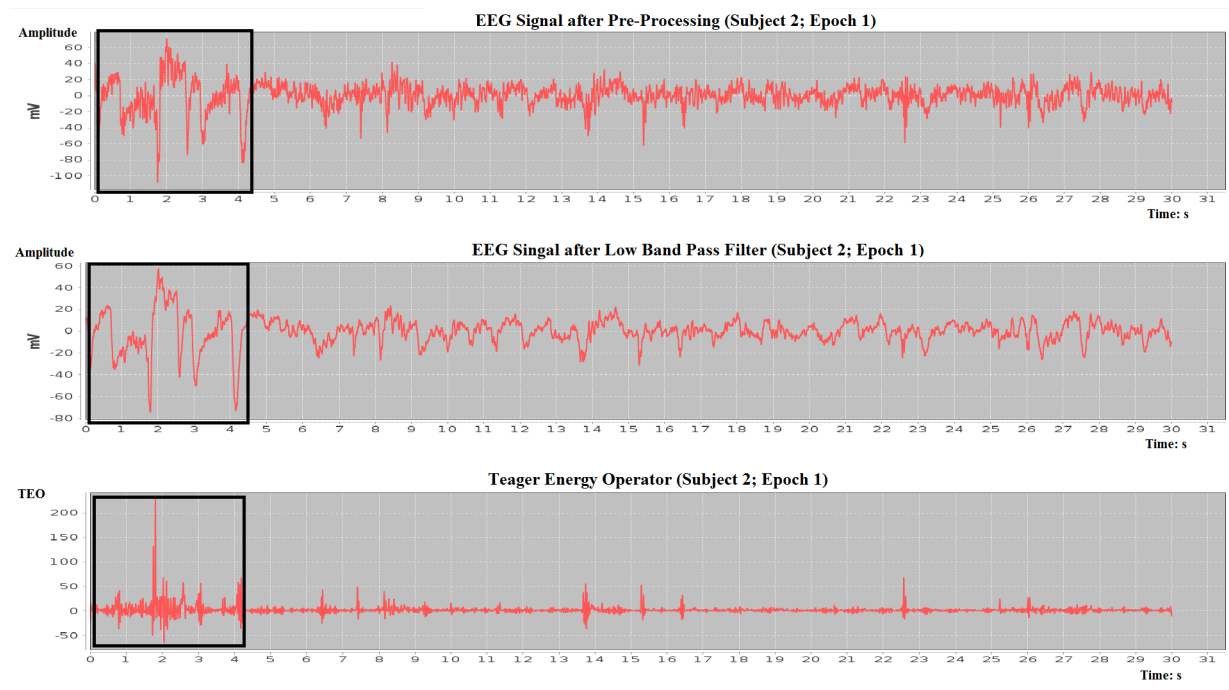


Figure 4.14 – EEGKComplex Extraction(Movements)

In comparison of TEO for K-Complex in stage N2 and TEO for movement in stage W, TEO of movement is much higher than TEO of K-Complex (almost 5 times higher). For

each epoch, the maximum value of TEO is represented as the value for digital parameter *EEGKComplex*.

#### 4.4.4 Feature Fusion: K-Complex

In order to realize semantic interpretation of K-Complex: digital parameter *EEGKComplex* based on TEO is extracted in data fusion; transformation of *EEGKComplex* into symbolic features are performed in feature fusion.

In feature fusion, two thresholds (*EEGKCA* and *EEGKCB*) are used in transforming *EEGKComplex* into three symbolic features: *High*, *Middle* and *Low*. Epochs with values of digital parameter *EEGKComplex* that higher than *EEGKCA* are transformed as *High*; values between *EEGKCA* and *EEGKCB* are transformed as *Middle*; values lower than *EEGKCB* are transformed as *Low*.

For the symbolic interpretation of *EEGKComplex*, three features are used to distinguish the occurrence of K-Complex and the occurrence of instantaneous movement. TEO value for an instantaneous movement is relatively higher than TEO value for a K-Complex. Meanwhile, TEO value for K-Complex is relatively higher than TEO value for a no occurrence of K-Complex and movement. Only symbolic feature equals to *Middle* is used to describe as the occurrence of K-Complex. Symbolic feature *High* is used to reflect the occurrence of movements.

#### 4.4.5 Decision Fusion: K-Complex

In decision fusion, two rules are added based on feature parameter of *EEGKComplex*.

For stage W, *EEGKComplex* is *High*.

For stage N2, *EEGKComplex* is *Middle*.

For stage N1, *EEGKComplex* is *Low*.

#### 4.4.6 Decision Fusion: Stage N1

Guidance from AASM manual for physicians to classify stage were described in Chapter 2. Vertex sharp waves and slow eye movements are typical characteristics in distinguishing stage N1. To fully mimic clinical sleep staging process, detection of vertex sharp waves and slow eye movements are required. However, as far as we known, there exists very limit and effective method in detection of these characteristics.

Instead of directly extract typical characteristics of stage N1 and fully use existing digital parameters, a rule to classify stage N1 is proposed as shown in Figure 4.15. It combines supplementary descriptions from AASM in classifying stage N1 and existing parameters.

E.g. In AASM, *Rule F.Note 3*, During stage N1, the chin EMG amplitude is variable, but often lower than in stage W. In translating this rule into exiting SF-SSS model, *EMGActivity* is used and the rule for classifying stage N1 is composed by one condition: *EMGActivity* is *Low*.

Once *EMGActivity* is *Low*, *EEGSleepSpindles* is *Not Confident*, *EEGStability* is *Not Confident*, *EEGLowWaveEnergy* is *Low* and *EEGKComplex* is *Low* then this epoch can be considered as stage N1.

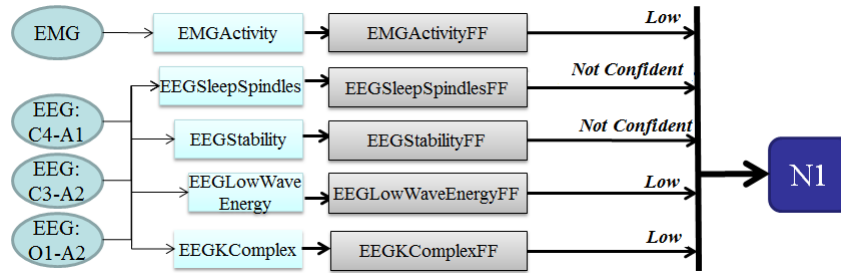


Figure 4.15 – Classification of Stage N1

#### 4.4.7 Smoothing

The smoothing function is proposed to consider the temporal effects of sleep staging process, and to detect and correct false sleep transitions. In smoothing, temporal contextual information and sleep transitions are considered.

- **Temporal Contextual Information Smoothing**

According to AASM manual, current sleep stage can be influenced by the previous sleep stage and can influence the next sleep stage. In smoothing, we implement generally accepted smoothing rules: the "3-minute rule" [38]. If a sequence of six epochs has only one epoch (*isolated sleep stage*) scored differently from the others (*major sleep stage*), this one could be replaced according to the following criteria: if the major sleep stage is R/W, then the *isolated sleep stage* is changed into the *major sleep stage* R/W.

- **Sleep Transitions Detection and Correction**

In smoothing, we propose impossible transitions and irregular transitions detection and impossible transitions correction and irregular transitions reporting. Impossible transitions and irregular transitions detection are used to detect the impossible transitions and irregular transitions. Impossible transitions correction is used to correct impossible sleep transitions and irregular transitions reporting is used to report the location of irregular transitions.

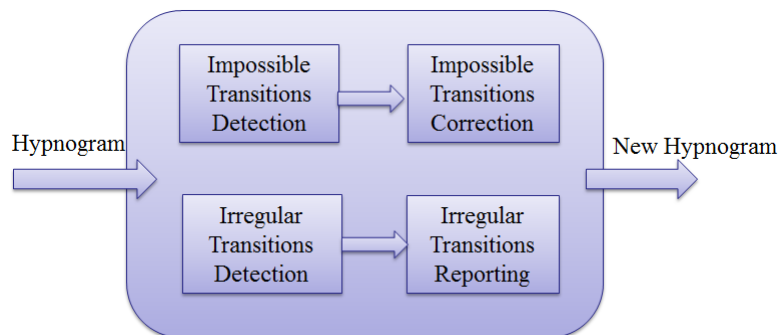


Figure 4.16 – Sleep Transitions Detection and Correction

For five different stages, there exist 25 possible sleep transitions. If there is no new sleep event that can be observed, physicians will keep score as the same stage. Meanwhile, according to AASM manual, 6 combinations of sleep transitions have been described in observing new sleep event: N2 to N1, N2 to N2, N2 to R, R to N1, R to N2, R to R as shown in Figure 4.17. Where:

- *a* represents an arousal or a major body movement followed by slow eye movement occurs.
- *b* is a major body movement followed by slow eye movement occurs.
- *c* represents absence of non-arousal associated K complexes and absence of sleep spindles.
- *d* is an increase in chin EMG or an arousal occurs followed by slow eye movement or a major body movement followed by slow eye movement.
- *e* represents K complex or sleep spindle occurs.
- *f* is an arousal without slow eye movement or A major body movement without slow eye movement.

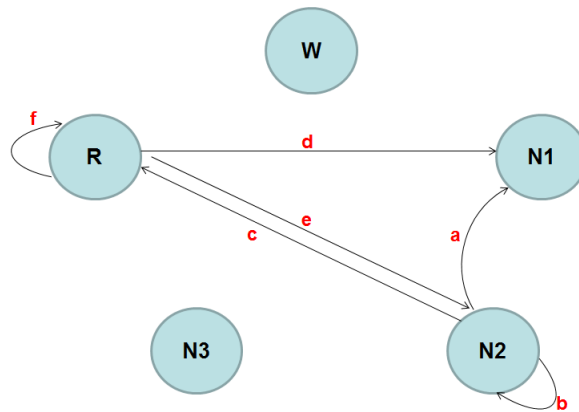


Figure 4.17 – Sleep Transitions Described by AASM

For another 14 possible sleep transitions which have not been described by AASM, we verified with physicians about the possibility of the transitions. Impossible transitions and irregular transitions are defined as shown in Figure 4.18.

**Impossible Transitions** are used to describe the transitions that will never happen. Transition from Stage W to N3, N1 to N3 and R to N3 are impossible transition.

**Irregular Transitions** are used to describe the transitions that rarely happen. Irregular transitions include transition from Stage N3 to N1.

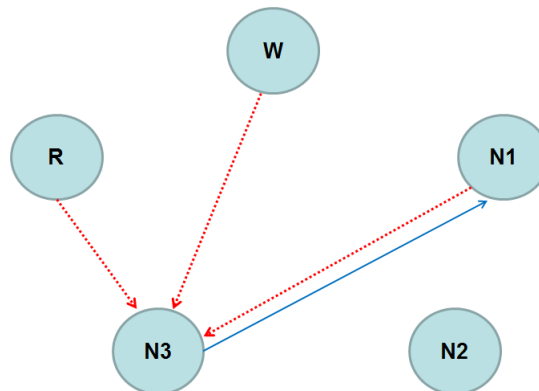


Figure 4.18 – Sleep Transitions: Impossible (Red Arrow) and Irregular (Blue Arrow)

To deal with impossible transitions, impossible transitions correction is performed. For impossible transition Stage W to N3, if there are three consecutive epochs like W,



N3, W, it will be replaced with the sequence W, W, W; if there are three consecutive epochs like W, N3, N3, it will be replaced with the sequence N3, N3, N3; otherwise N3 epoch after W will be replaced with N2. For impossible transition Stage N1 to N3, if there are three consecutive epochs like N3, N1, N3, it will be replaced with the sequence N3, N3, N3; if there are three consecutive epochs like N1, N1, N3, it will be replaced with the sequence N1, N1, N1; otherwise N1 epoch before N3 will be replaced with N2. For impossible transition Stage R to N3, if there are three consecutive epochs like N3, R, N3, it will be replaced with the sequence N3, N3, N3; if there are three consecutive epochs like R, R, N3, it will be replaced with the sequence R, R, R; otherwise N3 will be replaced with N2.

Details of correction of impossible sleep transitions are described in Table 4.3.

Sleep Transitions	Impossible Transition	Correction
<b>W to N3</b>	W, N3, W	W, W, W
<b>W to N3</b>	W, N3, N3	N3, N3, N3
<b>W to N3</b>	W, N3, N1/N2/R	W, N2, N1/N2/R
<b>N1 to N3</b>	N1, N3, N3	N3, N3, N3
<b>N1 to N3</b>	N1, N3, N1	N1, N1, N1
<b>N1 to N3</b>	N1, N3, W/N2/R	N2, N3, W/N2/R,
<b>R to N3</b>	R, N3, N3	N3, N3, N3
<b>R to N3</b>	R, N3, R	R, R, R
<b>R to N3</b>	R, N3, W/N1/N2	R, N2, W/N1/N2

Table 4.3 – Impossible Sleep Transitions Correction

To deal with irregular transitions, irregular transitions reporting is performed. It gives a warning notice with irregular transitions location at the end of program as shown in Figure 4.19.

```

Command Window
>> Smoothing_SF_SSS
Warning: Irregular Transitions of N3 to N1.
Patient: 627; Location: Epoch 515.

```

Figure 4.19 – Irregular Transitions Reporting

## 4.5 Conclusion

In this chapter, Symbolic Fusion-based Sleep Staging model was presented. It can mimic the clinical sleep staging process by translating AASM manual into computer logic. Limitations of the existing SF-SSS were analyzed and modifications were proposed in this chapter. In order to fully understand how SF-SSS use thresholds to transform digital parameters into symbolic interpretation and propose an automatic way to set up these thresholds, thresholds are presented in next chapter.

## Chapter 5

# Thresholds and Symbolic Fusion-based Sleep Staging System

Thresholds are used by Symbolic Fusion-based Sleep Staging System (SF-SSS) model to transform low-level digital parameters into the high-level symbolic interpretation of feature parameters. In this chapter, these thresholds are investigated. Firstly, the Manual Thresholds Setting-Up (MTSU) method, applied in SF-SSS for setting-up these thresholds is presented. Then, issues on these thresholds are discussed. Lastly, thresholds dependencies among sleep stages and patients for SF-SSS model are analyzed in order to understand how many thresholds are required.

### 5.1 Introduction on Thresholds

Thresholds are widely applied in most of the automatic sleep staging system for transforming digital parameters into linguistic or symbolic features for modeling inference process under the guidance of medical knowledge [45, 46, 76]. In this section, thresholds that used in SF-SSS are introduced.

#### 5.1.1 Explanation of Thresholds

SF-SSS adopts three-level architecture of symbolic fusion to mimic the decision-making process of clinical sleep staging in accordance with AASM as we described in previous chapter.

In this section, we describe how to define and use thresholds to transform digital parameters into the symbolic interpretation of feature parameters. According to AASM rules, we analyzed how it is possible to generate different symbols from one sleep event through thresholds.

##### 5.1.1.1 Description in AASM

Take chin EMG as an example, in AASM there exist several rules regarding chin EMG which described as below.

- *Rule E3.c* Score epoch as stage W when Irregular, conjugate rapid eye movements associated with **normal or high chin muscle tone**.
- *Rule F.N3* During stage N1, **the chin EMG amplitude** is variable, but often **lower than in stage W**.
- *Rule I.2b* Score stage R sleep in epochs with the following phenomena: **low chin EMG tone** for the majority of the epoch.

In AASM, chin EMG/chin muscle tone has been mentioned in three rules for guiding physicians to score stage W, N1 and R respectively.

### 5.1.1.2 From AASM to SF-SSS model using Thresholds

In order to formalize these rules into SF-SSS model, the description of **chin muscle tone**, **the chin EMG amplitude** and **chin EMG tone** in AASM are assumed as the same semantic/linguistic representation of *chin EMG activity* in SF-SSS. Data Fusion, Feature Fusion and Decision Fusion on *chin EMG activity* are described as follows.

- Data Fusion of *chin EMG activity*

In SF-SSS, digital parameter *EMGActivity* is extracted by using the mean absolute value of the chin EMG signal as shown in the Equation 5.1; where  $x(t)$  is chin EMG signal.

$$EMGActivity = mean(abs(x(t))) \quad (5.1)$$

This parameter can be used to indicate the activity level of chin EMG, which can be used as an indicator of the chin muscle activity during sleep.

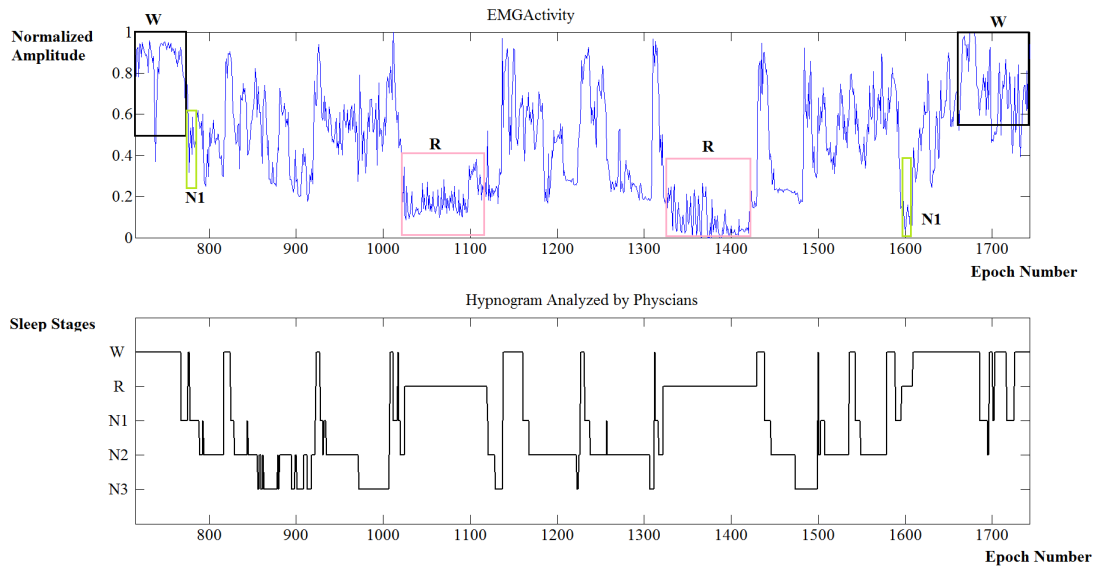


Figure 5.1 – Digital Parameter: EMGActivity

Figure 5.1 shows the digital parameter *EMGActivity* of one subject. For stage W, digital parameter *EMGActivity* is relative higher as shown in black box; for stage N1 and R, *EMGActivity* is relative lower as shown in green and pink box respectively.

- Feature Fusion of *chin EMG activity*

In feature fusion, thresholds are used to transform digital parameters into symbolic interpretation of feature parameters. To correspond symbolic interpretation of feature parameter with AASM manual, three symbolic interpretation of feature parameters are used: *High*, *Normal* and *Low*.

In order to transform *EMGActivity* into symbolic interpretation of *High*, *Normal* and *Low*, two thresholds *EMGTh1* and *EMGTh2* are used as shown in Figure 5.2. Values of digital parameter *EMGActivity* higher than *EMGTh1* are interpreted as

*High*, values between  $EMGTh1$  and  $EMGTh2$  are interpreted as *Normal*, values lower than  $EMGTh2$  are interpreted as *Low*.

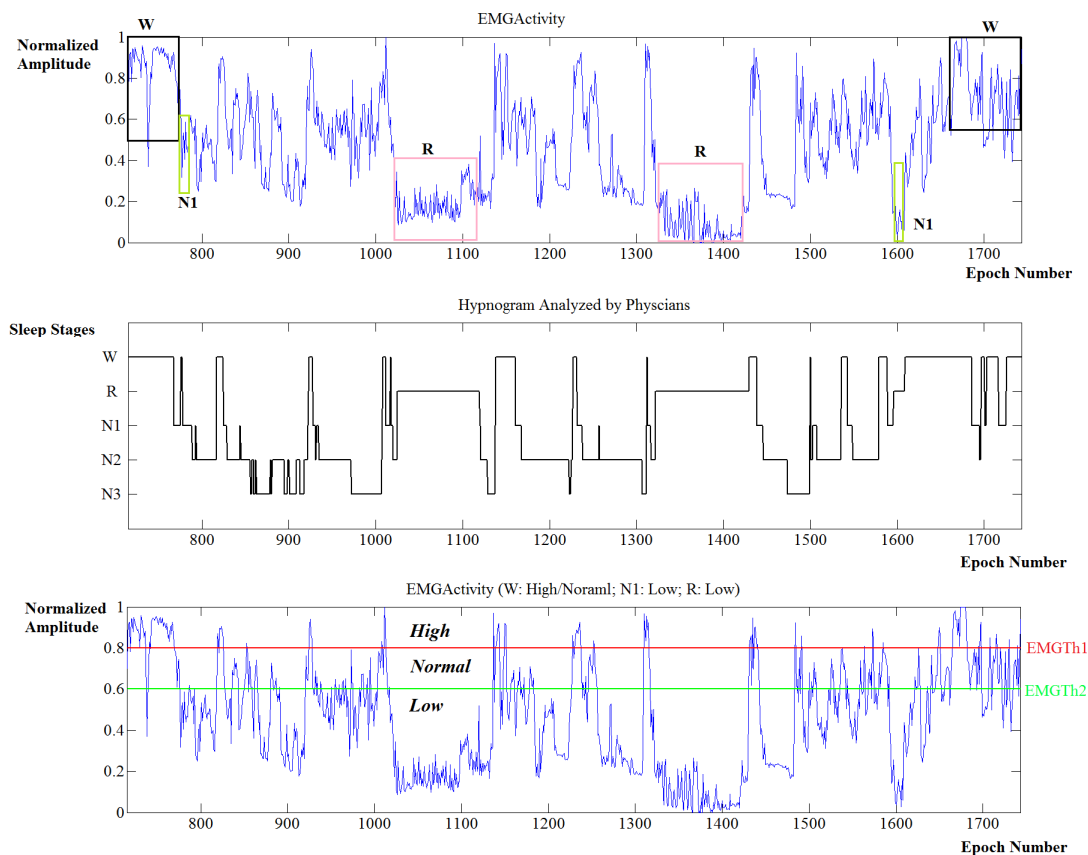


Figure 5.2 – Thresholds for EMGActivity

- Decision Fusion of *chin EMG activity*

In decision fusion, rules inspired from AASM to make decisions are shown as below.

For stage W:  $EMGActivity$  is *Normal* or *High*.

For stage N1:  $EMGActivity$  is *Low*.

For stage R:  $EMGActivity$  is *Low*.

In SF-SSS model, digital parameters (e.g.  $EMGActivity$ ) in the low-level, symbolic interpretation (e.g. *High*, *Normal*, *Low*) and decision rules in the high-level are inspired from AASM. However, in medical guidance, there has no definition or description on transform process of low-level digital parameters to high-level symbolic interpretation. In other words, from medical guidance, there is no definition or description for guiding setting-up these thresholds. In the first release of SF-SSS model, a Manual Thresholds Setting-Up (MTSU) method was used, this method is described in the next section.

### 5.1.2 Manual Thresholds Setting-Up Method

Manual Thresholds Setting-Up (MTSU) method was applied in existing SF-SSS [76]. Values of thresholds are visually determined by the author. MTSU method mainly involves the following steps:

1. Correspond each digital parameter to AASM manual (e.g correspond digital parameter *EMGActivity* to *chin EMG activity* in AASM manual).
2. Correspond each digital parameter to the Hypnogram that analyzed by physicians (e.g correspond the digital parameter *EMGActivity* with Hypnogram).
3. Estimate values of thresholds for each digital parameter (e.g by comparing the digital parameter *EMGActivity* with Hypnogram to estimate the values of *EMGTh1* and *EMGTh2*).
4. Adjust values of thresholds until appropriate thresholds are found (e.g adjust values of *EMGTh1* and *EMGTh2* until appropriate thresholds which can reach high classification performance are found).

Manual Thresholds Setting-Up (MTSU) was used in order to have a first proof of concept for the SF-SSS model. For this reason, the method implies several limitations:

1. It is a time-consuming process and requires manual effort.

Total 15 thresholds need to be manually set for each subject for transforming nine digital parameter into 24 symbolic features. Using MTSU method to set up these thresholds is a time-consuming process. Moreover, the thresholds values remain subjective.

2. The whole Hypnogram analyzed by experts is required.

In order to set up thresholds for transforming digital parameters into different symbolic features, a whole Hypnogram which analyzed by physicians is required. Without this Hypnogram, thresholds cannot be easily set by MTSU method. Instead of reduce the burden of physician, MTSU needs the analysis from physicians as the necessary prerequisite.

3. It can only provide estimated values of thresholds.

MTSU can only provide estimated values of thresholds instead of precise values. Most important, to realize the automated process of sleep staging, an automatic way to set-up these thresholds is required.

To release the burden of manual effort in setting-up thresholds and to automate the sleep staging process, an automatic method to set-up thresholds is required. Before studying which algorithm could be used to implement an automatic thresholds setting-up system, an analysis of thresholds is needed. This analysis covers the discussion on thresholds issues to figure out the number of thresholds needed to be configured for SF-SSS model.

## 5.2 Issues on Thresholds

In SF-SSS, thresholds are used to transform low-level digital parameters into the high-level symbolic interpretation of feature parameters. An automatic thresholds setting-up method is required. There exist several issues need further research:

**1) Whether there exists any dependency between thresholds and other conditions like: sleep stages, patients or any other factor?**

Figure 5.3 represents several possible conditions may exist dependency with thresholds. In this section, we mainly focus on the dependency between Thresholds and Sleep Stages, Thresholds and Patients.

**Thresholds & Sleep Stages:** For classifying each sleep stage in accordance with AASM, how many thresholds are required? Whether same thresholds can be reused in different sleep stages or different thresholds are required?

**Thresholds & Patients:** For different patients, whether generalized thresholds are sufficient or specific thresholds for different patients are required need to be researched to understand whether there is a dependency between Thresholds and Patients or not.

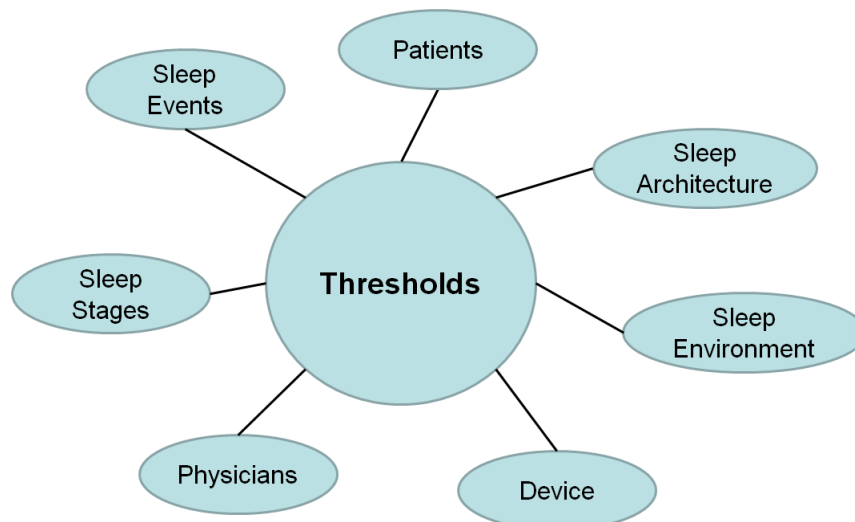


Figure 5.3 – Dependencies among Thresholds and other conditions

## 2) How many thresholds are required for SF-SSS model?

For SF-SSS, number of thresholds for classifying each stage needs to be studied. Increase the number of thresholds may increase the precision of symbolic interpretation of feature parameters but it also increase the complexity of SF-SSS. Finding appropriate number of thresholds can balance the precision and the complexity of SF-SSS.

## 5.3 Thresholds in SF-SSS

In this section, thresholds dependencies among sleep stages and patients are discussed. In analysis of thresholds dependencies among sleep stages can help us to understand whether same thresholds can be used in different sleep stages or different thresholds are required for different sleep stages. Meanwhile, in analysis of thresholds dependencies among patients can help us to understand whether generalized thresholds are sufficient or specific thresholds for different patients are required.

In SF-SSS model, nine digital parameters are transformed into 24 features via 15 thresholds as shown in Table 5.1.

Digital Parameters	Features	Number of Thresholds
EEGLowWaveEnergy	High - Middle - Low	2
EEGSleepSpindles	Confidently Have - Not Confident	1
EEGLWProportion	High - Low	1
EEGThetaProportion	High - Low	1
EEGStability	Stable - Not Confident - Unstable	2
EEGKComplex	High - Middle - Low	2
EOGEyeMovement	High - Middle - Low - Lowest	3
EOGCorrelation	Conjugate - Disconjugate	1
EMGActivity	High - Normal - Low	2

Table 5.1 – Thresholds used in SF-SSS model

Take *EMGActivity* as an example, 2 thresholds *EMGTh1* and *EMGTh2* are used in distinguish three different symbolic features: *High*, *Normal* and *Low* as shown in Figure 5.2. Symbolic features of *EMGActivity High*, *Normal* and *Low* are used in classifying stage W, N1 and R. The number of thresholds only depends on the digital parameters and not on sleep stages.

### 5.3.1 Thresholds & Sleep Stages

In existing SF-SSS model, values of *EMGTh1* and *EMGTh2* are the same for each stage. In existing SF-SSS model, 0.8 and 0.55 are used as the appropriate values for *EMGTh1* and *EMGTh2* for classifying stage W, N1 and R of patient 3774 (Figure 5.4). However, thresholds dependencies among stages are ignored in MTSU method.

In Figure 5.4, 0.8 and 0.55 can be considered as the appropriate vales for Thresholds *EMGTh1* and *EMGTh2* respectively in classifying stage W. However, only one thresholds *EMGTh2* is required from the technique perspective. Because no matter how we adjust *EMGTh1*, it has no impact on the classification result for stage W, N1 or R.

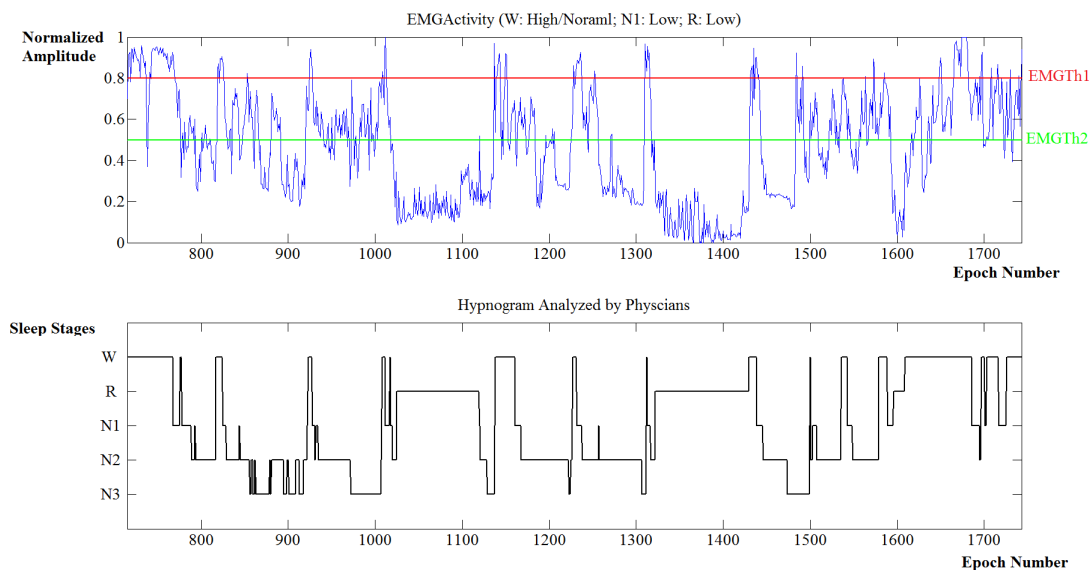


Figure 5.4 – Thresholds for EMGActivity (Stage W) of Patient 3774

For classifying stage N1, increment of *EMGTh2* value can be more precise as shown in Figure 5.5. 0.8 and 0.6 can be considered as the appropriate vales for Thresholds

$EMGTh1$  and  $EMGTh2$  respectively in classifying stage N1. By increasing the value of  $EMGTh2$ , more epochs can be correctly classified into stage N1.

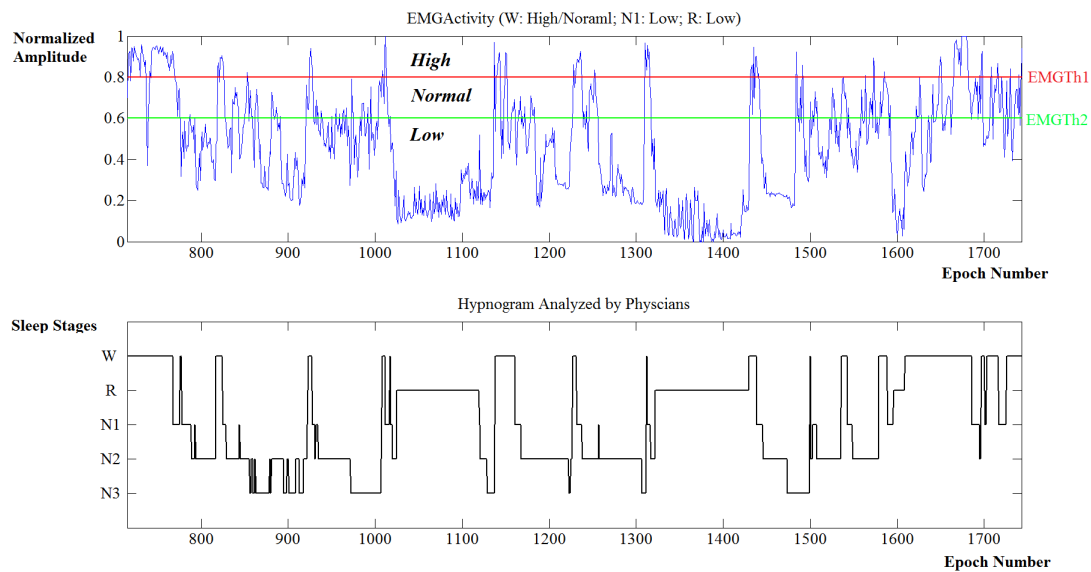


Figure 5.5 – Thresholds for EMGActivity (Stage N1) of Patient 3774

As for stage R, decrease the value of  $EMGTh2$  value can be more precise as shown in Figure 5.6. 0.8 and 0.38 can be considered as the appropriate vales for Thresholds  $EMGTh1$  and  $EMGTh2$  respectively in classifying stage R. By decreasing the value of  $EMGTh2$ , less epochs will be misclassified into stage R.

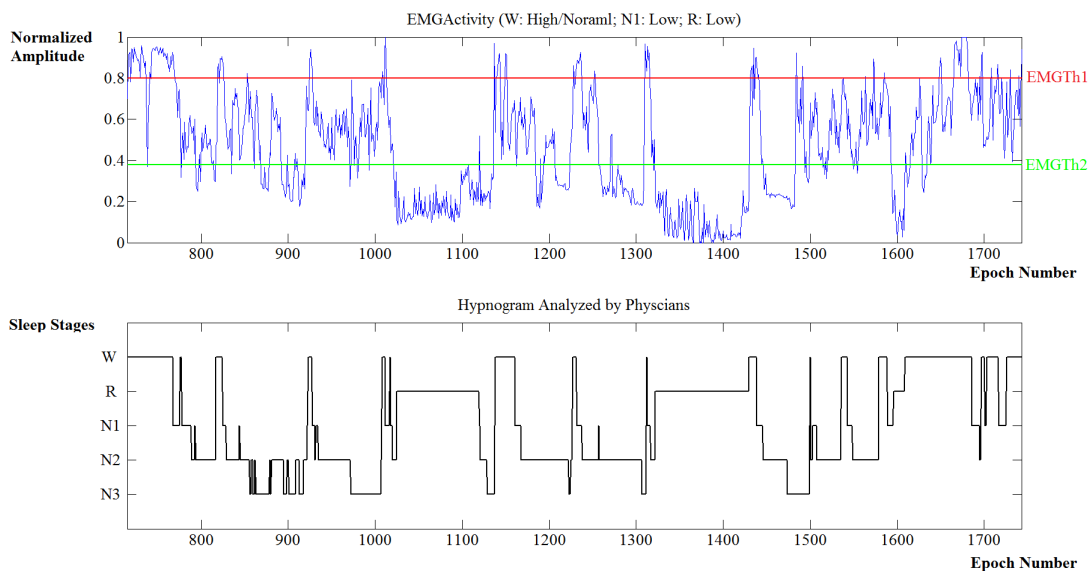


Figure 5.6 – Thresholds for EMGActivity (Stage R) of Patient 3774

In SF-SSS, MTSU method ignores thresholds dependencies among sleep stages. However, Figure 5.4, 5.5 and 5.6 show thresholds variability among sleep stages. For classifying stage W, N1 and R, different thresholds are required. In order to improve the classification performance for each stage and remove useless ( $EMGTh1$ ) thresholds, different thresholds are required for different stages while taking thresholds dependencies among sleep stages



into consideration. Instead of only use two thresholds for *EMGActivity*, three thresholds are required as shown in Table 5.2.

stages	Symbolic Features of <i>EMGActivity</i>	Thresholds
W	High or Normal	<i>EMGThN1</i>
N1	Low	<i>EMGThN2</i>
R	Low	<i>EMGThN3</i>

Table 5.2 – Thresholds of *EMGActivity* including Sleep Stages Dependencies

### 5.3.2 Thresholds & Patients

In this section, thresholds dependencies among patients are analyzed to understand whether generalized thresholds are sufficient or personalized thresholds are required.

Take patient 55341 as an example, *EMGTh1*=0.7 and *EMGTh2*=0.1 can be considered as appropriate thresholds values in classifying stage W as shown in Figure 5.7.

However, values of *EMGTh1* and *EMGTh2* between patient 3774 and patient 55341 are quite different because of the individual variability of chin EMG signals. For patient 3774, *EMGTh1*=0.8 and *EMGTh2*=0.55 are considered as appropriate thresholds values in classifying stage W.

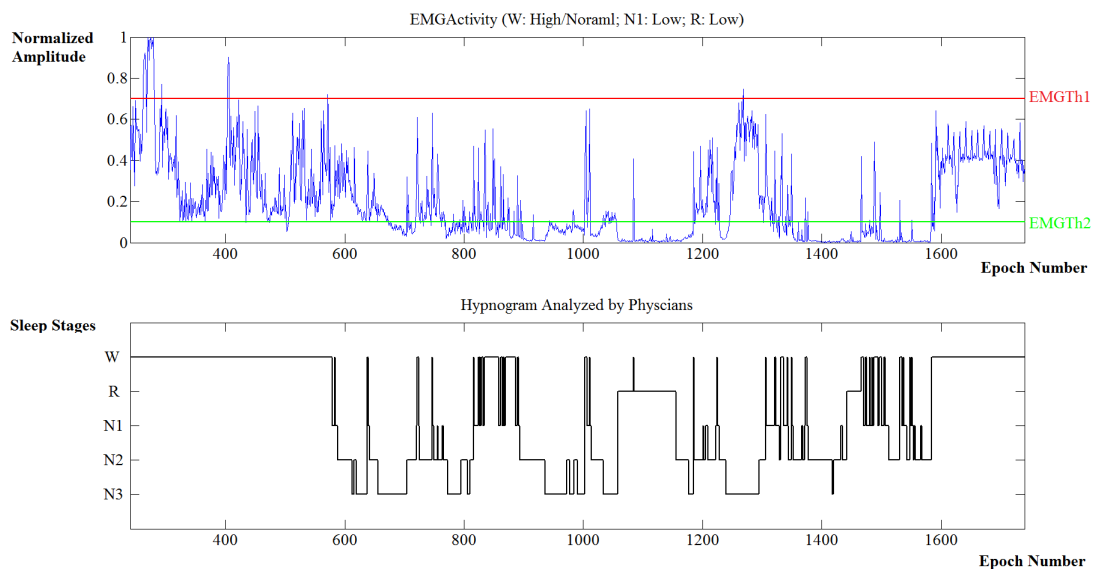


Figure 5.7 – Thresholds for *EMGActivity* (Stage W) of Patient 55341

In Figure 5.8, *EMGTh1*=0.7 and *EMGTh2*=0.2 can be considered as appropriate thresholds values in classifying stage N1 for patient 55341. However, for patient 3774, *EMGTh1*=0.8 and *EMGTh2*=0.6 are considered as appropriate thresholds values in classifying stage N1.

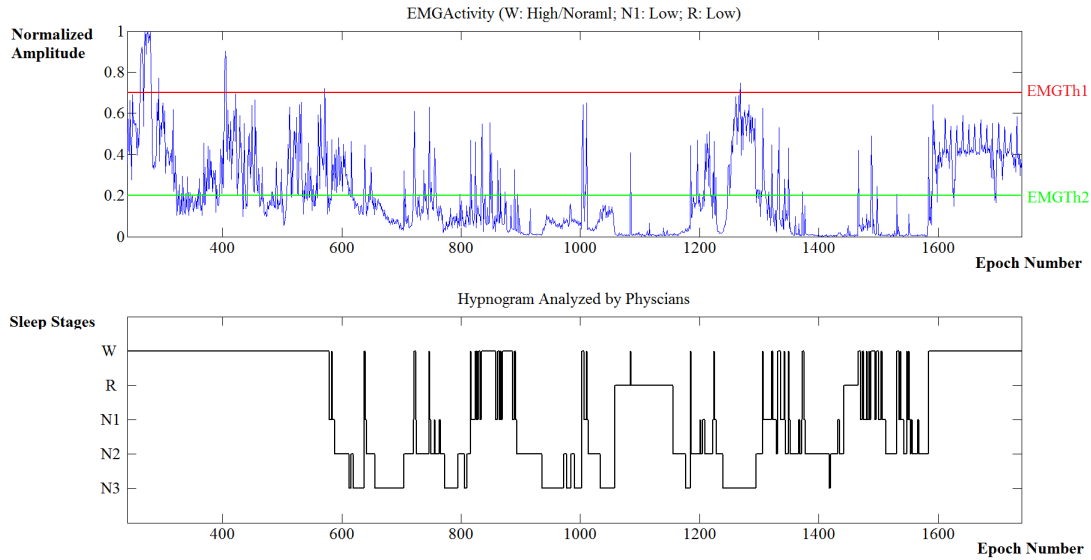


Figure 5.8 – Thresholds for EMGActivity (Stage N1) of Patient 55341

In Figure 5.9,  $EMGTh1=0.7$  and  $EMGTh2=0.05$  can be considered as appropriate thresholds values in classifying stage R for patient 55341. However, for patient 3774,  $EMGTh1=0.8$  and  $EMGTh2=0.38$  are considered as appropriate thresholds values in classifying stage R.

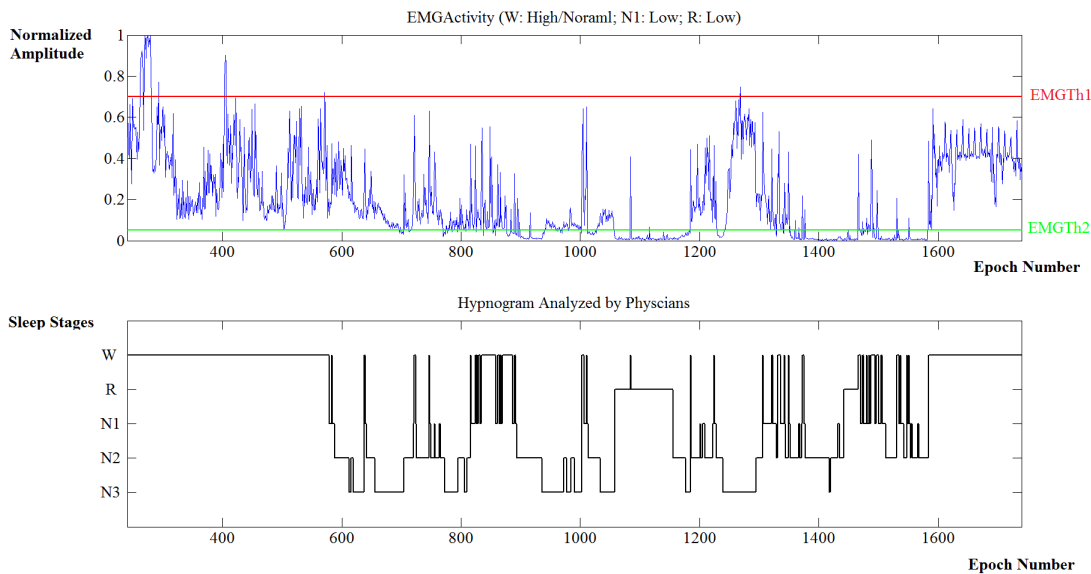


Figure 5.9 – Thresholds for EMGActivity (Stage R) of Patient 55341

For each patient the number of thresholds is same while the values of thresholds may be different.

## 5.4 Thresholds Configuration

In MTSU, total 15 thresholds are used to transform nine digital parameters into 24 symbolic features as shown in Table 5.1 without taking thresholds dependencies among stages into consideration.

For example, *EMGActivity* are used in classification stage W, N1 and R. Same thresholds *EMGTh1* and *EMGTh2* are applied in transforming *EMGActivity* into *High*, *Normal* and *Low* without taking thresholds dependencies among sleep stages into consideration while using MTSU method. However, thresholds variability was observed among sleep stages. In order to improve the classification performance of each sleep stage, different thresholds of *EMGActivity* are required for stage W, N1 and R. In conclusion, in MTSU method, two thresholds are required for *EMGActivity*. However, to take thresholds dependencies among stages into consideration, three thresholds are required for *EMGActivity*.

Take all the parameters and thresholds dependencies among stages into consideration, the numbers of thresholds need to be set-up for each stage are listed in Table 5.3. E.g. to classify stage W, five digital parameters are extracted and 5 thresholds are required in transforming these digital parameters into different symbolic features. For stage N1, N2, N3 and R, the numbers of thresholds need to be set are 6, 8, 7 and 8 respectively. Total 34 thresholds are required.

	Digital Parameters	Symbols	Number of Thresholds
Stage W	EEGThetaProportion EEGStability EEGKComplex EOGEyeMovement EMGActivity	Low Unstable High High(Middle) High(Normal)	5
Stage N1	EEGStability EEGLowWaveEnergy EEGKComplex EEGSleepSpindles EMGActivity	Not Confident Low Low Not Confident Low	6
Stage N2	EEGThetaProportion EEGStability EEGLowWaveEnergy EEGKComplex EEGSleepSpindles EOGEyeMovement EOGCorrelation	High Stable Low Middle Confidently Have Low(Lowest) Disconjugate	8
Stage N3	EEGLWProportion EEGStability EEGLowWaveEnergy EEGSleepSpindles EOGEyeMovement EOGCorrelation	High Stable High Not Confident Low(Middle) Disconjugate	7
Stage R	EEGThetaProportion EEGLWProportion EEGStability EEGLowWaveEnergy EEGSleepSpindles EOGCorrelation EMGActivity	Low Low Not Confident Low Not Confident Conjugate Low	8

Table 5.3 – Thresholds Configuration for Each Sleep Stage

## 5.5 Conclusion

In this chapter, a brief introduction on thresholds is presented firstly which involves how to use thresholds for formalizing AASM manual into symbolic fusion model. Then Manual Thresholds Setting-Up (MTSU) method used in existing SF-SSS model is introduced with the analysis of its limitations. Then issues on thresholds are described and followed the discussion of thresholds dependencies among sleep stages and patients of SF-SSS. Lastly, thresholds that need to be set up for classifying each stage are listed with the consideration of thresholds dependencies among stages.

## Chapter 6

# Towards to a Personalized Sleep Staging System

In this chapter, Automatic Thresholds Setting-Up (ATSU) method is introduced to provide optimal thresholds for Symbolic Fusion-based Sleep Staging System (SF-SSS) in accordance with the conclusion of the previous chapter. It is inspired by Feedback System Control (FSC) technique and dedicates in searching optimal thresholds combination for SF-SSS. We propose two different search algorithms: Differential Evolution Cross Entropy of FSC for ATSU. Based on proposed ATSU method, Personalized Sleep Staging System (PSSS) conception is proposed by taking individual variability into consideration.

### 6.1 Automatic Thresholds Setting-Up Method for SF-SSS

In existing SF-SSS model, Manual Thresholds Setting-Up (MTSU) was adopted between Data Fusion and Feature Fusion as shown in Figure 6.1. Details of how to set up these thresholds manually are described in chapter 4. To release the burden of manually setting of thresholds and take individual variability into consideration, an Automatic Thresholds Setting Up (ATSU) for SF-SSS is proposed.

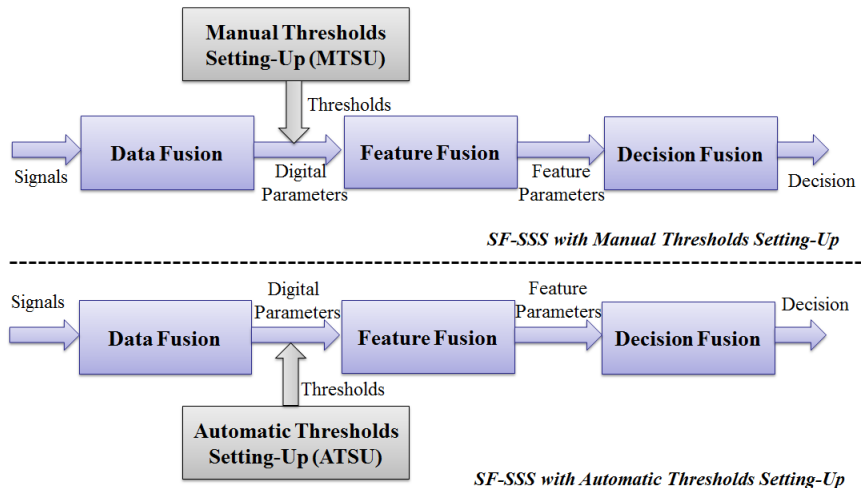


Figure 6.1 – Thresholds Setting-Up for Symbolic Fusion based Sleep Staging System

ATSU is proposed to provide thresholds automatically in transforming digital parameters into the symbolic interpretation of feature parameters for SF-SSS. It is inspired from Feedback System Control (FSC) technique. To fully understand how ATSU works, a brief introduction of FSC is introduced followed by the details of ATSU.

### 6.1.0.1 Feedback System Control technique

Feedback System Control (FSC) is initially proposed to navigate through the large parametric space of different drugs to identify optimal low-dose drug combinations in manipulating the cellular network toward a therapeutic goal [82] in 2008. In each living cell, the interactions among the biomolecules, proteins and nucleic acids intrinsically serve as the foundation of the extensive networks of signal and regulatory pathways. To explore and understand the cell functions, the bottom-up reductionist approach is very challenge due to the sheer magnitude of pathway processes and pathway crosstalk. FSC manipulates the cellular network as a whole, rather than analyzing the processes through individual signaling pathways to circumvent the need for detailed information of biological signaling and regulatory networks. It proved to be an efficient combinatorial drug screening method in finding optimal drug combinations which can improve treatment efficacy and enable the reduction of side effects and drug resistance.

In 2014, FSC is extended to be a Generic Feedback System Control (G-FSC) technique [27], which can be used to find input variable combination for guiding the complex system toward to a desired state. It can be applied not only in biological cells domain but also many other domains like Internet and financial activities. Without the requirement of detailed information from complex system or how the complex system responds to input stimuli, G-FSC provides a solution to in searching for an appropriate input stimuli which can reach optimal objective value for the complex system.

As shown in Figure 6.2, it consists of 4 parts: Input Stimuli, Complex System, Objective Assessment and Stochastic Search Algorithm. Input stimuli are arbitrarily selected to apply to the complex system; if the specific system output is not met, G-FSC uses an engineering search algorithm that selects the next group of inputs to iteratively feedback to the complex system. Details of each part are described below.

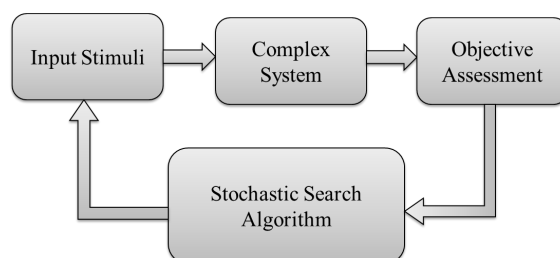


Figure 6.2 – Generic Feedback System Control

- **Input Stimuli**

Input Stimuli provides the inputs for the complex system. It can be composed of a parameter or a combination of several parameters. For each parameter, it has its own space (range). The space of each parameter can be continuous or discrete. For example,  $N$  parameters with  $M$  possible values from each space would result in  $M^N$  potential input combinations and the input stimuli are from these potential input combinations.

- **Complex System**

A complex system is composed of a large number of interacting building blocks or elements which self-organize, generating emerging properties that are usually not directly linked to those of the individual building elements. Biological cells, the Internet and financial activities are all examples of complex systems.

- **Objective Assessment**

Objective Assessment is used to evaluate the input stimuli impact on the complex system. It can provide systematic quantitative characterization of complex system response to different input stimuli.

- **Stochastic Search Algorithm**

Without knowing the exact mechanism of the complex system and how the complex system responds to manipulation of the inputs, set up a model to stimulate complex system response induced by the inputs is very challenge. However, Stochastic Search Algorithm (SSA) can provide a fast and stable convergence in finding an optimal input for regulating complex systems without the requirement of detailed information about the complex system or how the system responds to input stimuli. Furthermore, it is also robust to random noise and nonlinear changes in the system and the environment, which are commonly observed in complex systems like a biological system.

To solve large-scale combinatorial optimization problems of highly complex systems, FSC provides a rapid and stable search and discovery ability in finding an optimal combination among potential combinations space without the requirement of detailed information about the complex system.

#### 6.1.0.2 Automatic Thresholds Setting-Up Method

Thresholds are widely applied in decision support systems in transforming digital parameters into linguistic or symbolic features to make the final decision. Sleep staging systems, as one of the typical clinical applications of decision support systems, widely adopts thresholds in transforming digital parameters into linguistic or symbolic features to model inference process under the guidance of medical knowledge [45, 46, 76]. While, in clinical practice, boundaries of linguistic or symbolic features are very flexible. Physicians may adjust the boundaries for each linguistic or symbolic feature according to their experience and patient information.

However, as far as we know, there are no fully satisfying automatic setting up thresholds methods in existing sleep staging systems. Manually predefined values of thresholds have been widely due to the following reasons: 1) To build a mathematical model or a threshold function in setting up thresholds is very challenge which requires a set of data with sufficient quantity and adequate quality; 2) There is a lack of uniformity between subjects; 3) Interaction among thresholds is not clear. Instead of building a mathematical model or a threshold function, we propose a new solution to thresholds setting-up problems. Assume, there exists a thresholds combination which can reach the highest objective value of sleep staging systems; to set up these thresholds can be described as to find the optimal thresholds combination among the possible thresholds combinations space which can reach the highest objective value of sleep staging system. Then thresholds setting-up problems can be described as a combinatorial optimization problem in finding optimal thresholds combination among possible thresholds combinations space regarding the objective value of sleep staging systems.

Based on FSC technique, ATSU we proposed mainly consists of the following parts: *Thresholds Combinations (TC)*, *Symbolic Fusion-based Sleep Staging System (SF-SSS)*, *Assessment of SF-SSS (A-SF-SSS)* and *Stochastic Search Algorithm (SSA)*. *TC* provides a possible thresholds combinations space. *SF-SSS* is used to performed sleep staging. *A-SF-SSS* is used to assess the impact of thresholds combination on *SF-SSS*. Initially, *SSA* provides randomly thresholds combinations from *TC* and pass it to *SF-SSS*. Then *A-SF-SSS* is performed to evaluate the impact of thresholds combination on *SF-SSS* with an

objective value. With this objective value, *SSA* will generate new thresholds combinations for the next loop. This process is repeated until optimal thresholds combination is found which can provide good assessment in *A-SF-SSS*. Details of each part are described in below.

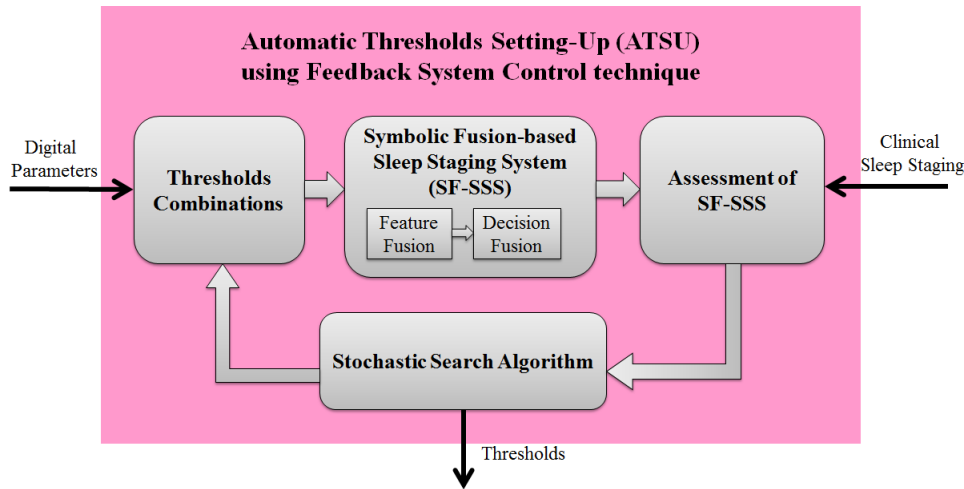


Figure 6.3 – Automatic Thresholds Setting-Up for Symbolic Fusion based Sleep Staging System

- **Thresholds Combinations**

Thresholds Combinations (TC) provides possible thresholds combinations space for SF-SSS. Suppose if these are 15 thresholds needed to set up, each threshold has 10 possible discrete values, it would result in  $10^{15}$  potential thresholds combinations. TC provides this potential combination space and every thresholds combination select by SSA and passes to SF-SSS should within this space.

- **Symbolic Fusion-based Sleep Staging System**

Symbolic Fusion-based Sleep Staging System (SF-SSS) is performed to realize sleep staging as shown in the previous chapter which consists of data fusion, feature fusion and decision fusion. In ATSU method, data fusion only needs to perform once at the beginning to provide digital parameters of selected signals. This is because *Thresholds* have no impact on Data Fusion. SF-SSS in ATSU is mainly used to perform Feature Fusion and Decision Fusion. Feature Fusion and Decision Fusion are repeated in setting up thresholds until optimal thresholds are found.

- **Assessment of SF-SSS**

Assessment of SF-SSS (A-SF-SSS) is used to evaluate the thresholds combination impact on SF-SSS. In analysis of a classification, F-Measure is widely applied by considering both precision and recall. In A-SF-SSS, F-Measure is used to evaluate the impact of different thresholds on SF-SSS. The traditional F-Measure also called balanced F-score ( $F_1$ ) is shown in Equation 7.1 which can balance the precision and recall.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6.1)$$

In F-Measures, Precision and Recall are used which as shown in Equation 7.2 and Equation 7.3, respectively. Precision (also called positive predictive value) is the ratio of all positive predictions among predicted events. Recall (also known as sensitivity or true positive rate) is the ratio of all positive predictions among all true



events.

$$Precision = \frac{TP}{TP + FP} \quad (6.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (6.3)$$

TP, FN, FP and TN are used to evaluate how good the observations (predictions) reflect the actual events for a classification. *True Positives* and *True Negatives* are the observations which were correctly predicted. *False Negatives* are observations are negative where the actual events are positive. *False Positive* are observations are positive where the actual events are negative.

### • Stochastic Search Algorithm

In solving large-scale combinatorial optimization problems as finding an optimal combination from a finite set of combinations, exhaustive search is not feasible. Instead of applying exhaustive search, SSA is used to search for the optimal thresholds combination for SF-SSS in ATSU.

Among typical SSAs, Differential Evolution and Cross Entropy are more suitable for sleep staging application based on thresholds. Both of them are robustness and support parallel searching. Moreover, both of them are flexible which allow to deal with low or high dimension of parameters. For the consideration of hardware implementation, they are not based on memory structure and easy to implement. The SSA is not only used to generate thresholds combination but also controls the stop conditions. The stop criteria for these search algorithms is described in next section.

For Differential Evolution and Cross Entropy, they have own advantages as follows. Differential Evolution is a popular and efficient method of the evolutionary algorithm which has been successfully applied to solve multi-parameter problems in diverse domains like mechanical engineering design or chemical engineering [70]. It owns many advantages like: 1) DE can mimic natural biological evolution and provide a fast and stable convergence; 2) It is less sensitive to initial population; 3) It is a parallel search method; and 4) It can improve fitness function value iteratively. Cross Entropy also owns several advantages: 1) CE is a parallel search method; 2) It is based on rigorous mathematical and statistical principles; 3) It provides a simple adaptive procedure.

Details of Differential Evolution and Cross Entropy are presented in next section with the procedure and main control parameters analysis.

#### 6.1.0.3 Differential Evolution

Differential Evolution (DE) was proposed by Storn and Price [70] in 1997, is one of the most effective and efficient stochastic optimization technique. Since 1997, it obtained great develop and was successfully applied in diverse domains like electronic engineering [63], chemical engineering [81].

To deal with optimization problems, DE starts with a set of initial population (as parents) which are usually drawn randomly from the uniform distribution with the variable space. Then DE operators (mutation and crossover) are applied to each individual in the population to produce another population (as offspring). Both populations then evaluated using fitness function. The individual which can reach better fitness function survives for further reproduction, evaluation and selection until termination criteria meet.

In DE, different strategies can be expressed by  $DE/x/y/z$ .  $DE$  stands for Differential Evolution,  $x$  represents a string denoting the solution to be perturbed,  $y$  is the number of different solution considered for the perturbation of  $x$ , and  $z$  is the type of crossover. Details of DE can be described as follows:

- **Initialization**

As a population based search algorithm, DE starts with an initial population vector shown in (6.4), where the index  $i$  denotes the  $i^{th}$  individual of the population;  $G$  denotes the generation to which the population belongs and  $NP$  denotes population size.

$$X_{i,G}, i = 1, 2, \dots, NP \quad (6.4)$$

The initial population vector is chosen randomly and assumed to cover the entire parameter space using Equation (6.5), where  $x_{ij}^L, x_{ij}^U$  denote the lower and upper limits of the variable in the population respectively,  $rand_{ij}(0, 1)$  represents a uniformly distributed random value within  $[0,1]$ .

$$x_{ij} = rand_{ij}(0, 1) * (x_{ij}^U - x_{ij}^L) + x_{ij}^L \quad (6.5)$$

- **Mutation**

A mutant vector  $V_{i,G+1}$  is generated according to different strategies which are listed below. The indexes  $r_1, r_2, r_3, r_4, r_5$  are mutually exclusive integers randomly chosen from the range  $[1, NP]$  and all are different from base index  $i$ . The  $X_{best,G}$  is the individual having the best fitness function values in the population at generation  $G$ . Mutation Scale Factor  $F$  is a real and constant factor belongs to  $[0,2]$  which controls the amplification of the differential variation.

$$DE/rand/1 : V_{i,G+1} = X_{r_1,G} + F * (X_{r_2,G} - X_{r_3,G}) \quad (6.6a)$$

$$DE/rand/2 : V_{i,G+1} = X_{r_1,G} + F * (X_{r_2,G} - X_{r_3,G}) + F * (X_{r_4,G} - X_{r_5,G}) \quad (6.6b)$$

$$DE/best/1 : V_{i,G+1} = X_{best,G} + F * (X_{r_1,G} - X_{r_2,G}) \quad (6.6c)$$

$$DE/best/2 : V_{i,G+1} = X_{best,G} + F * (X_{r_1,G} - X_{r_2,G}) + F * (X_{r_3,G} - X_{r_4,G}) \quad (6.6d)$$

$$DE/randtobest/1 : V_{i,G+1} = X_{r_1,G} + F * (X_{best,G} - X_{r_2,G}) + F * (X_{r_3,G} - X_{r_4,G}) \quad (6.6e)$$

- **Crossover**

In order to increase the diversity of the DE population, crossover is introduced in Equation (6.7) for generating a crossover vector  $U_{i,G+1}$ . There are two types crossover schemes: *exponential(exp)* and *binomial(bin)* in DE.

$$U_{i,G+1} = (u_{1i,G+1}, u_{2i,G+1}, \dots, u_{Di,G+1}) \quad (6.7)$$

The *binomial* scheme takes parameters from the mutation vector  $v_{ji,G+1}$  every time that the generated random number is equal or less than the Crossover Rate as given by  $randb(j) \leq CR$ , else all parameters come from  $x_{ji,G}$ . Details of *bin* scheme is shown in Equation (6.8). Figure 6.8 presents how *binomial* scheme crossover works.

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} : if(randb(j) \leq CR) \text{ or } j = rnbr(i) \\ x_{ji,G} : if(randb(j) > CR) \text{ and } j \neq rnbr(i) \end{cases} \quad (6.8)$$

In Equation (6.8),  $randb(j)$  is the  $j$ th evaluation of a uniform random number generator with outcome  $\in [0, 1]$ .

The *exponential* scheme, an integer  $rnbr(i)$  among  $[1, D]$  is randomly generated. This integer acts as a starting point to take  $v_{ji,G+1}$  until random number is greater than CR, otherwise all parameters come from  $x_{ji,G}$ . Details of *exp* scheme is shown in Equation (6.9).

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} : \text{from } j = rnbr(i) \text{ while } randb(j) \leq CR \\ x_{ji,G} : \text{otherwise} \end{cases} \quad (6.9)$$

- **Selection**

To decide whether or not it should become a member of Generation  $G + 1$ , greedy criterion is used by comparing  $u_{i,G+1}$  to  $x_{ji,G}$  as shown in Equation (6.10). Where  $f$  is the function to evaluate the population.

$$x_{i,G+1} = \begin{cases} v_{i,G+1} : \text{if } f(v_{i,G+1}) \geq f(x_{i,G}) \\ x_{i,G} : \text{otherwise} \end{cases} \quad (6.10)$$

### Procedure of Different Evolution

The procedure of Differential Evolution in ATSU is described as follows and the flow chart is shown in Figure 6.4.

- **Step 1. Initialization** Determine several control parameters: population size, mutation factor, crossover factor. Generate initial population (initial thresholds combinations). Details of these control parameters (population size, mutation factor, crossover factor) of DE are explained in next section.
- **Step 2. Assessment** F-Measure is widely used in binary classification, it considers both precision and recall. In this step, F-Measure is used to assess thresholds impact on SF-SSS and it generates from the A-SF-SSS part of ATSU.
- **Step 3. Check** Check whether the terminate condition is satisfied or not. If one of the terminate conditions like, F-Measure reaches to the desired value (FM=0.98) or iteration reaches to the pre-defined maximum iteration number (G=200), is satisfied, then the procedure stops.
- **Step 4. Mutation and Crossover** Do mutation operation and crossover operation to generate provisional population (provisional thresholds combinations).
- **Step 5. Selection** Evaluate F-Measure of the provisional population. Compare F-Measure of the initial population with the F-Measure of the provisional population to generate new population (new thresholds combinations).
- **Step 6.** Repeat from step 2.

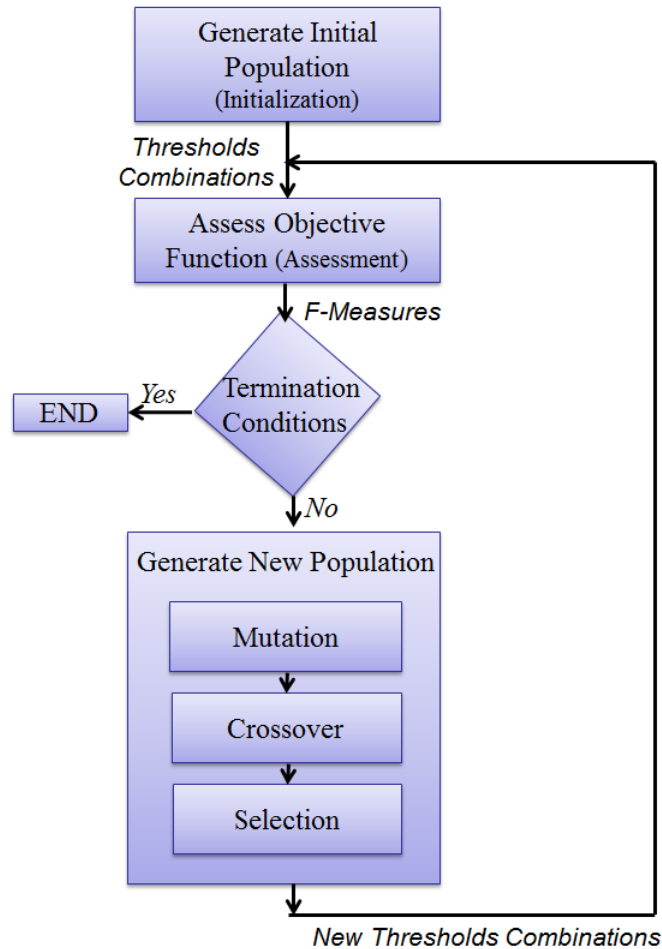


Figure 6.4 – Flow Chart of Differential Evolution used in ATSU

**Main Control Parameters of Different Evolution** There are three main control parameters of DE-PSSS: Population Size (NP), Mutation Scale Factor (F) and Crossover Rate (CR). NP may play a crucial role in the efficiency and effectiveness: large population size potentially increases the population diversity and computational complexity. F controls the amplification of the differential variation. CR controls the number of components inherited from the mutant vector which can be interpreted as a mutation probability. Details of each control parameter is described as follows:

1. **Population Size (NP)**: NP may play a crucial role in the efficiency and effectiveness of DE. Large population size potentially increases the population diversity. However, when computational budget is limited, increasing the population size will decrease the number of iterations (generations) but also increase the number of operations in each iteration.
2. **Mutation Scale Factor (F)**: F controls the amplification of the differential variation. Too small F values increase the risk of premature convergence (i.e. converge to an undesirable point), while too large F values decrease the convergence speed that degrades DE efficiency and may result in early termination [36].
3. **Crossover Rate (CR)**: CR controls the number of components inherited from the mutant vector it can be interpreted as a mutation probability. Small CR values can boost convergence speed when a few decision variables are interacting with each other. In turn, large CR values are more effective when lots of decision variables are interacting.

The selection of appropriate parameters can affect the efficiency of ATSU directly. Due to the variability of the underlying mathematical properties of different problems, a fixed set of control parameters that suits well for one problem, or a class of problems does not guarantee that it will work well for another class, or range of problems [61]. That is, the selection of control parameters is problem dependent. To ensure the performance of ATSU, selection of control parameters is extremely important.

To select the appropriate control parameters, trial-and-error approach, is widely used. Several sets of control parameters are tested, then appropriate control parameters can be selected based on the average performance of the problem.

Main Control Parameters of DE	Range
Population Size(NP)	[5D, 10D] <sup>1</sup>
Mutation Scale Factor(F)	[0, 2]
Crossover Rate(CR)	[0, 1]

Table 6.1 – Main Control Parameters of Differential Evolution

Three different population sizes (NP=5D, 10D and 500; where D is the numbers of thresholds which differs for different stage), three different values of mutation factor (F=0.5, 1 and 1.5), three different values of crossover factor (CR=0.1, 0.5 and 0.9) are analyzed to select optimal control parameters for DE-PSSS with analysis results shown in Chapter 6.

#### 6.1.0.4 Cross Entropy

Cross Entropy was initially proposed to estimate probabilities of rare events for complex stochastic networks by Rubinstein [59] in 1997. It has been extended to solve combinatorial optimization problems in 1999 [58], which turned out to be an effective method.

CE is briefly introduced as follows. In solving the combinatorial optimization problem, a maximization problem can be described as shown in Equation 6.11.

$$\gamma^* = \max_{x \in \chi} [S(x)] \quad (6.11)$$

$\gamma^*$  represents the maximal value of  $S$  on the domain space  $\chi$ . To proceed with CE,  $f(\cdot; v)$  is defined as a family of Probability Density Functions (PDFs) on  $\chi$ , with respect to some base measure  $v$ . Then  $\gamma^*$  can be estimated by  $\ell(\gamma)$  defined in Equation 6.12.

$$\ell(\gamma) = \mathbb{P}_u(S(X) \geq \gamma) = \mathbb{E}_u I_{\{S(X) \geq \gamma\}} \quad (6.12)$$

where  $X$  is a random vector generated by PDFs with parameter  $v$  in  $f(x, v)$ .  $\mathbb{P}_u$  is the probability of the state  $\{S(X) \geq \gamma\}$ ,  $\mathbb{E}_u$  is the corresponding expectation operator and  $I(\cdot)$  is the indicator function, i.e.,  $I_{\{S(X) \geq \gamma\}} = 1$  only if  $S(X) \geq \gamma$ , otherwise, it equals to zero.

Based on the important sampling: take  $N$  random samples  $X = (X_1, X_2, \dots, X_N)$  from an important sampling density  $g$  on  $\chi$ , the unbiased estimator  $\hat{\ell}(\gamma)$  of  $\ell(\gamma)$  can be defined as shown in Equation 6.13.

$$\hat{\ell}(\gamma) = \frac{1}{N} \sum_{i=1}^N I_{\{S(X_i) \geq \gamma\}} \frac{f(X_i; v)}{g(X_i)} = \ell(\gamma) \quad (6.13)$$

From Equation 6.13, only one sample suffices to estimate  $\ell(\gamma)$  since it is true for all  $i$ . While it is difficult to directly compute  $g$  because this  $g$  depends on the unknown

parameter  $\ell$ . Moreover, it is convenient to choose a  $g$  in the family of densities  $f(\cdot; v)$ . The CE method solves this efficiently by finding the minimal Kullback-Leibler distance as shown in Equation 6.14 which defines the distance between  $g$  and  $f(\cdot; v)$ .

$$\mathcal{D}(g, f) = \int g(x) \ln g(x) dx - \int g(x) \ln f(x; v) dx \quad (6.14)$$

In solving combinatorial optimization problems, CE creates a sequence of  $f(\cdot; v_1), f(\cdot; v_2), \dots$  of PDFs that are driven in the direction of the theoretically optimal density  $f(\cdot; v^*)$ .  $f(\cdot; v^*)$  corresponds to the degenerate density of the optimal solution. In each iteration, it generates a set of samples and the elite samples (in terms of solution quality) would be selected to update the parameters of the PDF  $f(x; v)$  parameterized by  $v$ . Since the elite samples are selected in each iteration,  $\gamma$  would be improved and can converge quickly to the optimal solution  $\gamma^*$ .

**Procedure of Cross Entropy** The procedure of Cross Entropy in ATSU is described as follows and the flow chart is shown in Figure 6.5.

- **Step 1. Initialization:** Define a specified mechanism to generate PDFs. Defined the sample size  $N$ , elite sampling rate  $\rho$ .

To search for the optimal thresholds combinations for the SF-SSS, normal distribution mechanism is performed to generate PDFs for each threshold which can be presented as  $\mathcal{N}(\mu, \sigma^2)$ .  $\mu$  is the mean of the distribution and  $\sigma$  is the standard deviation. According to normal distribution density function  $\mathcal{N}(\mu, \sigma^2)$ , a set of samples (thresholds combinations) are generated.

- **Step 2. Assessment:** F-Measure is used to evaluate the impact of different thresholds on the sleep staging system. F-Measures used in assessment are provided by A-SF-SSS part of ATSU.
- **Step 3. Check terminate conditions:** If one of the terminate conditions like, F-Measure (FM=0.98) reaches to a desired value or iteration reaches to the pre-defined value (G=200) or the standard deviation  $\sigma$  is close to zero, is satisfied, then the procedure stops.
- **Step 4. Selection:** Rank the values of F-Measure and select elite samples in term of F-Measure. In this step, a number of  $\rho N$  samples with higher F-Measure are selected as elite samples.
- **Step 5. Updating:** Update the PDFs parameters  $\mu$  and  $\sigma$ . According to the elite samples, new  $\hat{\mu}$  and  $\hat{\sigma}$  are calculated.
- **Step 6. Repeat:** Repeat from Step 2 to Step 6 until one of the terminate conditions is satisfied.

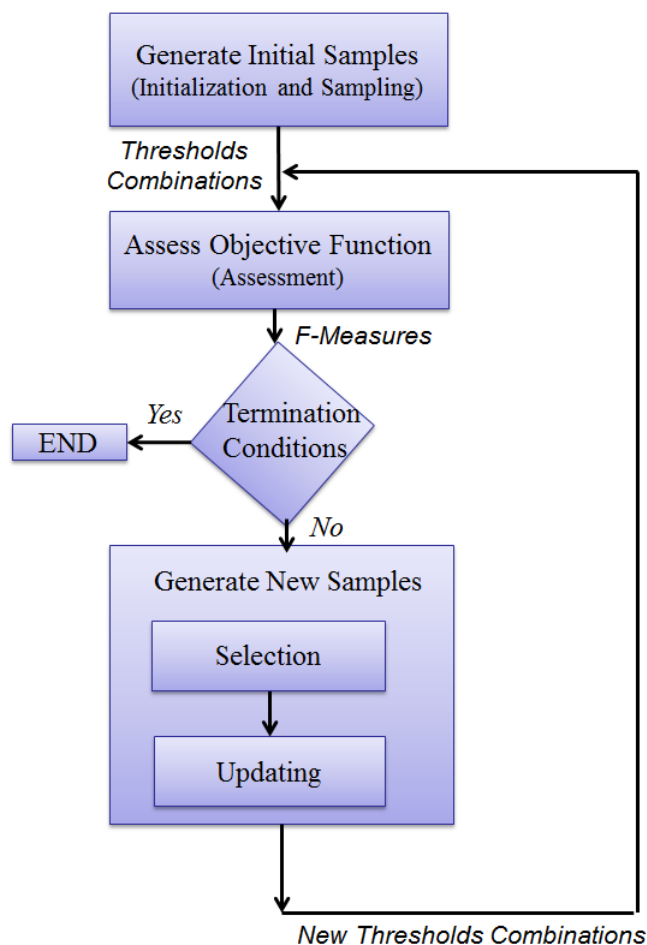


Figure 6.5 – Flow Chart of Cross Entropy used in ATSU

**Main Control Parameters of Cross Entropy** There are two main control parameters of CE: Sample Size ( $N$ ) and Elite Sampling Rate ( $\rho$ ).  $N$  controls the diversity of the samples.  $\rho$  controls the number of elite samples that will be selected to generate a new PDF. Details of these two main control parameters are described as below.

1. **Sample Size ( $N$ ):** Large sample size potentially increases the sample diversity. However, large sample size also increases computational time.
2. **Elite Sampling Rate ( $\rho$ ):**  $\rho$  controls the number of elite samples which belongs to  $(0,1)$ . It should ensure the elite samples to be larger enough for obtaining a reliable parameter update for PDFs. In practice,  $\rho$  is suggested to select from  $[0.1, 0.5]$  in [8, 6].

In order to select appropriate control parameters of CE-PSSS, different sample sizes ( $N=100, 500, 1000$  and  $10000$ ) and different values of elite sampling rate ( $\rho=0.1, 0.3, 0.5$  and  $0.7$ ) are analyzed with results shown in Chapter 7.

## 6.2 Personalized Sleep Staging System

In this section, Personalized Sleep System (PSSS) conception is presented followed by the details of two different PSSS we proposed: Differential Evolution-based Personalized Sleep Staging System (DE-PSSS) and Cross Entropy-based Personalized Sleep Staging System (CE-PSSS).

### 6.2.1 Toward to a Hybrid Expert System for Sleep Staging

As shown in Figure 6.6, a new work flow for Hybrid Sleep Staging System is presented. In this work flow, several epochs will be selected and analyzed by a physician. Based on the selected epochs and partial Hypnogram that analyzed by the physician, thresholds are set up based on ATSU we proposed in the previous section. Then, these thresholds are used in scoring the full epochs by using SF-SSS model and full Hypnogram can be generated at the end. Proposed hybrid sleep staging system combines symbolic fusion and FSC. Symbolic fusion is dedicated to the mimic decision-making process of clinical sleep staging. It starts from the extraction of digital parameters from raw polysomnography (PSG) signals and goes up to high-level symbolic interpretation of feature parameters. At last, decision is generated using rules inspired by international guidelines in sleep medicine and applied to feature parameters. Meanwhile, Feedback System Control (FSC) is designed to provide thresholds automatically for Symbolic Fusion in transforming digital parameters into symbolic interpretation of feature parameters

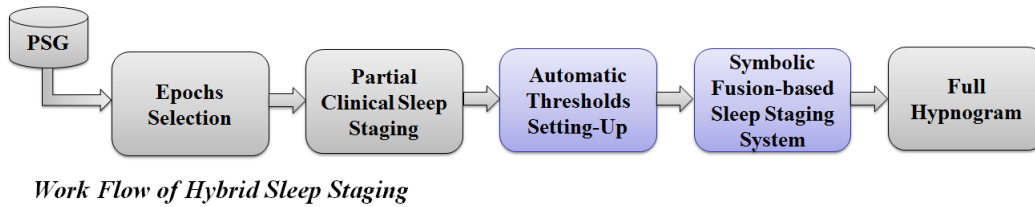


Figure 6.6 – Work Flow of Hybrid Sleep Staging System

Based on proposed work flow, two different types of thresholds can be set up: Personalized Thresholds and Generalized Thresholds. These two different types are explained in the next section.

#### 6.2.1.1 Personalized Thresholds and Generalized Thresholds

By applying ATSU method to different training set (which described as selected epoch in Epochs Selection of Figure 6.6), two different kinds of thresholds can be set up: personalized thresholds and generalized thresholds.

- **Personalized Thresholds**

Personalized Thresholds are thresholds that are specific to each person. Personalized thresholds would be more precise for each person by taking individual variability into consideration. However, it would also increase the complexity in setting up thresholds because different thresholds are needed to be set up for the different person. To set up personalized thresholds, epochs selected from each person are required as the training set to provide personalized thresholds which are specific to each person.

- **Generalized Thresholds**

Generalized Thresholds are thresholds that are generic to all the person. Generalized thresholds may result in less precise of the classification. While it would also be less complexity in setting up thresholds.

### 6.2.2 Personalized Sleep Staging System

In this thesis, a Personalized Sleep Staging System which is able to take individual variability into consideration is proposed. The Details of the PSSS is shown in Figure



6.7. It mainly consists of two parts: Personalized Thresholds Setting-Up part and Personalized Automated Sleep Staging part. For Personalized Thresholds Setting-Up part, it dedicates to provide Personalized Thresholds. *Personal Epoch Selection* is performed for selecting epochs of each person. Then *Personalized Thresholds Setting-Up* based on ATSU method is applied to set up personalized thresholds for a different person. After obtaining these personalized thresholds, the scoring of whole PSG signal can be performed in an automated way.

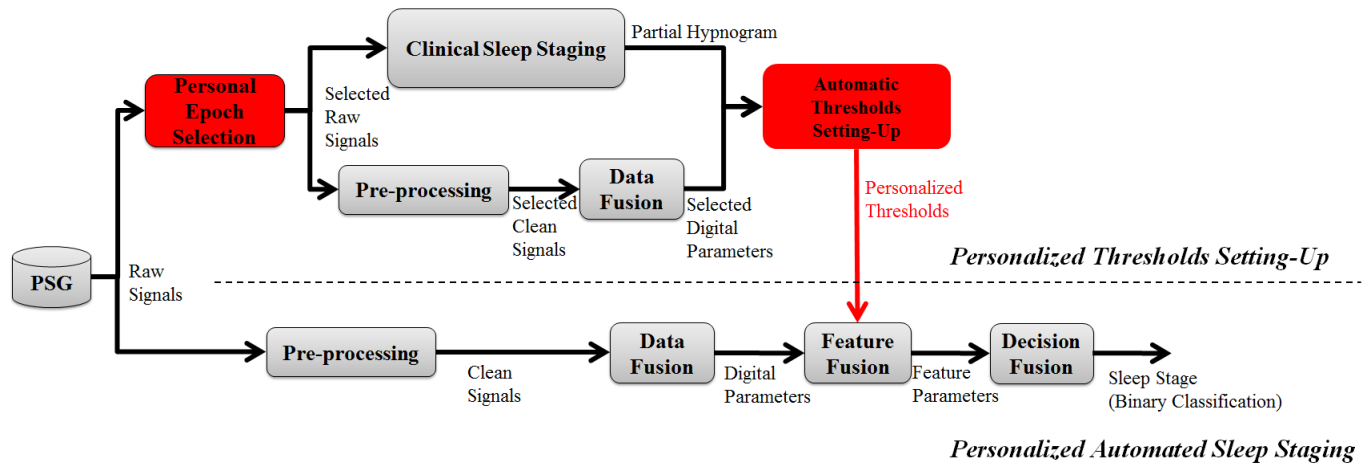


Figure 6.7 – Personalized Sleep Staging System

Based on the conception of PSSS, two different PSSS are proposed: Differential Evolution-based Personalized Sleep Staging System (DE-PSSS) and Cross Entropy-based Personalized Sleep Staging System (CE-PSSS) by adopting Differential Evolution and Cross Entropy respectively in ATSU of Personalized Thresholds Setting-Up part of PSSS. Details of DE-PSSS and CE-PSSS are described in next section.

### 6.2.3 Differential Evolution-based Personalized Sleep Staging System

In this section, DE-PSSS is introduced. It adopts the conception of PSSS we proposed in Figure 6.7, details of ATSU is shown in Figure 6.8.

In DE-PSSS, five ATSU modules are used to provide optimal thresholds combination for five classifiers. ATSU for stage W is dedicated to search for optimal thresholds combination which can reach high F-Measure in classifying stage W and non-W. While ATSU for stage N1, N2, N3 and R are specific to find optimal thresholds combination for stage N1, N2, N3 and R respectively. At last, personalized thresholds composed of five thresholds combinations are provided by PTSU and passed to evaluation set.

In chapter 5, it presents how many thresholds need to be set in each ATSU. E.g. to classifying stage W, five digital parameters are extracted and 5 thresholds are required in transforming these digital parameters into different features. For stage N1, N2, N3 and R, the numbers of thresholds need to be set are 6, 8, 7 and 8 respectively.

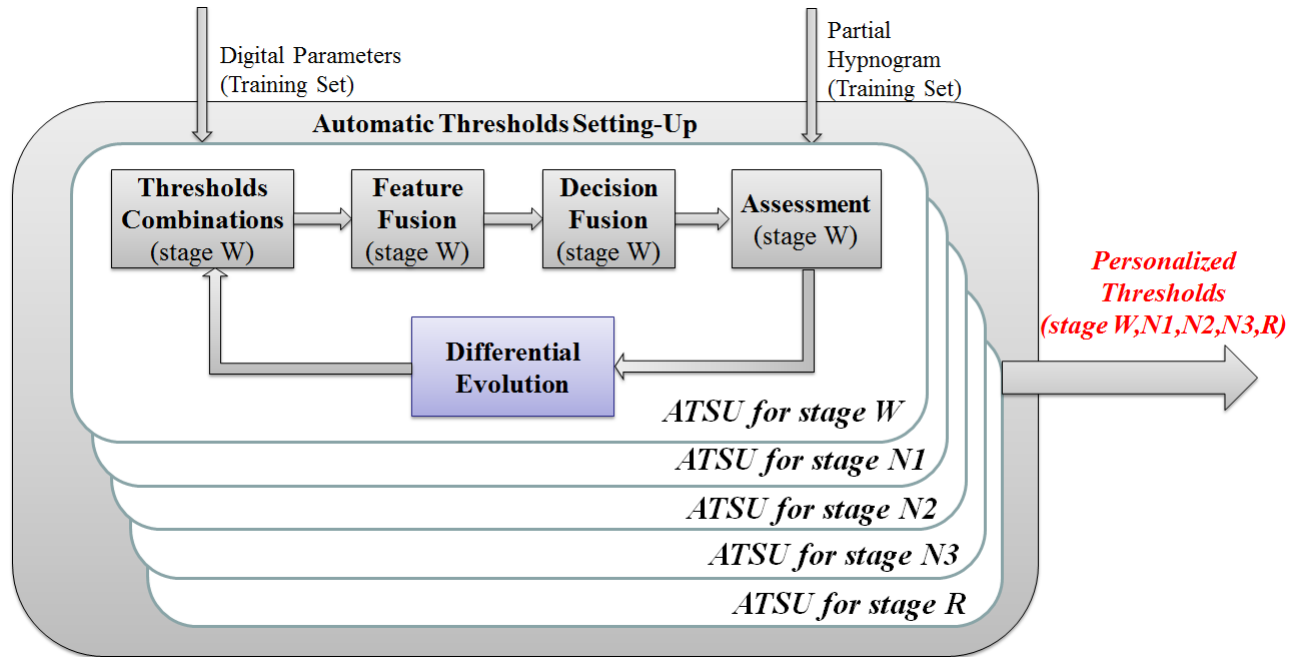


Figure 6.8 – Differential Evolution-based Personalized Sleep Staging System

### 6.2.3.1 Epoch Selection of DE-PSSS

In DE-PSSS, personalized thresholds are set based on the selected epochs and then passed to all epochs to realize the automatic classification.

In the evaluation of DE-PSSS, the selected personal epochs are considered as the training set. Remaining epochs are considered as evaluation set.

A small size of training set consumes less time for physician to analyze, while it may result in finding sub-optimal thresholds combination which may not suitable for the large size of evaluation set. Large size of training set may provide better thresholds combination for evaluation set, while it consumes much more time for expert to analyze.

Select the optimal size of the training set for DE-PSSS which can not only consumes less time for a physician to analyze but also can achieve high classification results for evaluation set is required. Different size of the training set is analyzed in chapter 7 to select the optimal size of the training set.

### 6.2.4 Cross Entropy-based Personalized Sleep Staging System

In this section, CE-PSSS is presented. Instead of using Differential Evolution as the search algorithm in finding optimal thresholds combination for each classifier, CE-PSSS applies Cross Entropy in ATSU as shown in Figure 6.9.

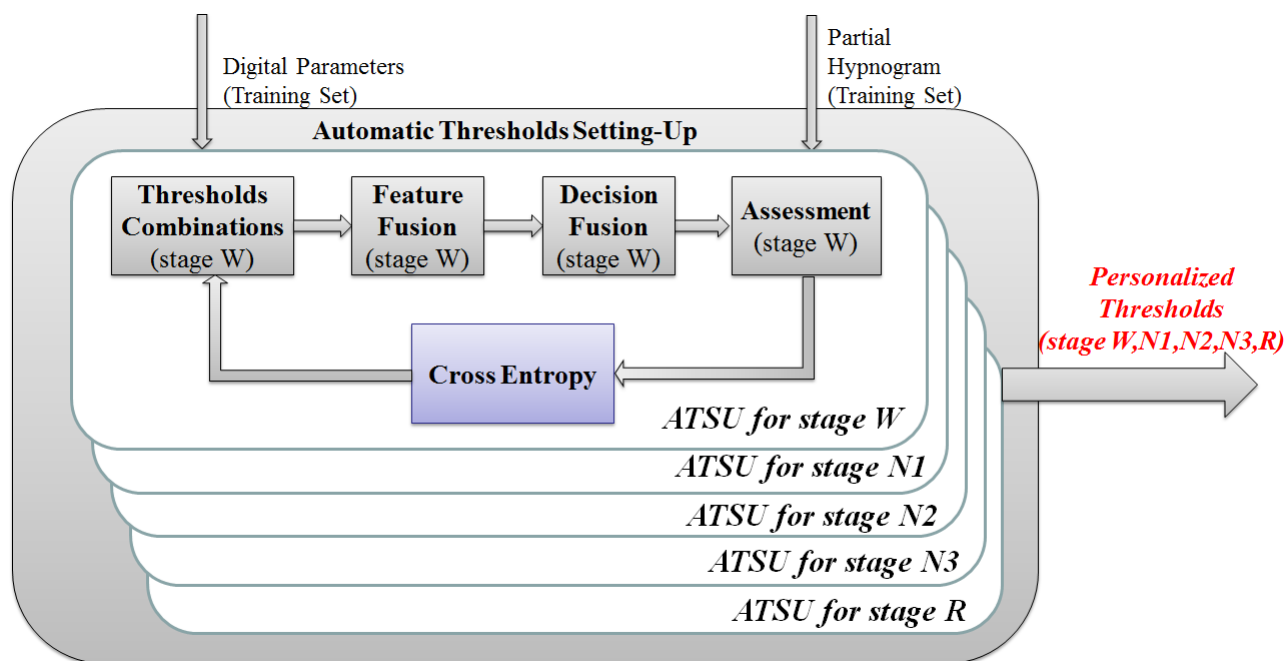


Figure 6.9 – Cross Entropy-based Personalized Sleep Staging System

#### 6.2.4.1 Epoch Selection of CE-PSSS

In CE-PSSS, same as DE-PSSS, personalized thresholds are set based on the selected set and then passed to all epochs to realize automated sleep staging.

In the evaluation of CE-PSSS, the selected personal epochs are considered as the training set. Remaining epochs are considered as evaluation set.

The small size of training set consumes less time for physicians to analyze, while it may result in finding sub-optimal thresholds combination which may not suitable for the large size of evaluation set. Large size of training set may provide better thresholds combination for evaluation set, while it consumes much more time for expert to analyze.

Select the optimal size of the training set for DE-PSSS which can not only consumes less time for physicians to analyze but also can achieve high classification results for evaluation set is required. For selecting optimal training set for CE-PSSS, different size of training set is analyzed in chapter 7.

### 6.3 Conclusion

In this chapter, personalized sleep staging system conception is proposed by combining Feedback System Control (FSC) technique with Symbolic Fusion. Meanwhile, two PSSSs: Differential Evolution-based Personalized Sleep Staging System (DE-PSSS) and Cross Entropy-based Personalized Sleep Staging System (CE-PSSS) are also presented by using Differential Evolution and Cross Entropy respectively. With the feasibility of FSC technique and the flexibility of symbolic intelligence, proposed system can be a reliable computer-assisted tool for assisting clinical sleep analysis and can be integrated with any PSG medical recording device.

# Chapter 7

## Evaluation of Sleep Staging System

In this chapter, evaluation of modified Symbolic Fusion-based Sleep Staging System (SF-SSS) is analyzed at the beginning, followed by comparison with existing SF-SSS. Then, evaluation of DE-PSSS and CE-PSSS is analyzed involving the selection of optimal control parameters and training set. At last, comparison among two different search algorithm methods based PSSS and SF-SSS using manual setting up thresholds are evaluated.

### 7.1 Database

To evaluate SF-SSS model with proposed modifications, the same database (Database 1) which has been used in existing SF-SSS analysis is adopted in order to make the comparisons. However, this database was recorded before 2006 and it was visually scored by physicians according to the old gold standard of sleep study (R & K manual). However, the evaluation of ATSU method and PSSSs are based on a new database (Database 2) which was recorded in 2016.

#### 7.1.1 Database 1

This database was recorded before 2006 which mainly records PSG signals like, three EEG channels, two EOG channels and one chin EMG channel of 16 subjects. PSG recordings are acquired using Embla System and visually scored by ten physicians using Somnologica software according to the old gold standard of sleep study (R & K manual). Subjects in Database 1 are those people that suspected to suffer from Sleep Apnea Syndrome (SAS) by physicians. In existing SF-SSS analysis, stage awake, S1, S2 and REM scored by R&K are considered as stage W, N1, N2 and R of AASM respectively; stage S3 and S4 scored by R&K are considered as stage N3 of AASM. Details of this database are introduced as follows.

##### 7.1.1.1 PSG Recordings

Three EEG channels (C3-A2, C4-A1, O1-A2), two EOG channels (EOG-L, EOG-R) and one chin EMG channel were recorded. The sampling frequency for EEG and EOG is 100 Hz. For EMG, the sampling frequency is 200 Hz.

##### 7.1.1.2 Subjects Description

Overnight PSG signals were recorded from 16 subjects (4 males and 12 females) ranging from 26 to 67 years old ( $\mu = 54.8$ ,  $\sigma = 12.5$ ) in Hôpital-Tenon (AP-HP) before 2006. AHI (average number of apneas and hypopneas per hour of sleep) ranges from 3.8 to 70.5 ( $\mu = 25.5$ ,  $\sigma = 22.7$ ). Details of patient information was shown in Table 7.1. In *ID* column,

the identification number of each subject is listed. In *G* column, it represents *Gender* of subjects: *M* is *Male* and *F* is *Female*. *TE* is the *Total Epoch* number of the subject during sleep. *EASE* is *Expert Analysis Start Epoch* which represents the start epoch analyzed by experts for each subject. *EAE* is *Expert Analysis End Epoch* which represents the last epoch analyzed by expert for each subject. *EATE* is *Expert Analysis Total Epoch* which represents the total epoch number that analyzed by expert for each subject. In *W*, *N1*, *N2*, *N3*, and *R* column, it lists the epoch number of stage *W*, *N1*, *N2*, *N3* and *R* for each subject.

AHI <sup>1</sup>	ID <sup>2</sup>	G <sup>3</sup>	Age	TE <sup>4</sup>	EASE <sup>5</sup>	EAE <sup>6</sup>	EATE <sup>7</sup>	W	N1	N2	N3	R
3.8	627	F	41	1800	724	1798	1075	326	54	383	214	98
4.9	3774	F	26	1830	1	1830	1830	983	165	447	138	97
5.8	4062	M	62	1860	691	1841	1151	314	153	353	115	216
7.6	50897	F	57	1773	479	1713	1235	456	46	364	225	144
8.2	3875	F	29	1830	1	1830	1830	1093	8	445	155	129
9.8	3870	F	65	1800	481	1624	1144	88	192	331	278	255
16.2	50062	M	64	1800	585	1721	1137	144	93	274	473	153
16.3	4064	M	57	1860	628	1731	1104	154	198	306	259	187
16.3	1818	F	67	1800	583	1800	1218	112	49	610	321	126
19.6	45813	F	54	1680	542	1549	1008	253	146	476	94	39
26.3	46618	F	58	1800	751	1731	981	115	272	220	247	127
33.8	3928	F	64	1797	569	1474	906	216	136	329	162	63
42.8	56508	F	60	1560	302	1560	1259	126	39	549	194	351
59	55341	F	51	1740	431	1593	1163	270	133	397	242	121
67.5	48814	F	56	1560	461	1560	1100	73	78	621	178	150
70.5	639	M	66	1740	394	1720	1327	24	229	421	321	332

Table 7.1 – Subject Information of Database 1

Only epochs that analyzed by experted are considered to evaluate the modified SF-SSS. For example, only epochs from 724 to 1798 for subject 627 are used to analyze the modified SF-SSS. Total number of each stage that used to evaluate the modified SF-SSS for all 16 subjects is listed in Table 7.2.

	W	N1	N2	N3	R	Total
Expert Analysis (Epoch Number)	4747	1991	6526	3616	2588	19468

Table 7.2 – Information of Database 1

### 7.1.2 Database 2

To our knowledge, all the database used in existing sleep analysis are visually scored by using R&K. There is no public database that is scored by using AASM. Thus, a new

<sup>1</sup>average number of apneas and hypopneas per hour of sleep

<sup>2</sup>Identification

<sup>3</sup>Gender

<sup>4</sup>Total Epoch

<sup>5</sup>Expert Analysis Start Epoch

<sup>6</sup>Expert Analysis End Epoch

<sup>7</sup>Expert Analysis Total Epoch

database was recorded in La Pitié-Salpêtrière hôpital (AP-HP). This new database was visually scored by physicians using AASM. Instead of only providing the final classification results of each stage, it also provides occurrence and location of sleep events that observed by physicians. Details of this new database is introduced as follows.

### 7.1.2.1 PSG Recording Device

PSG recordings were performed using the Compumedics Grael HD-PSG device which produced by a medical device company called Compumedics Limited from Australia. It recorded the following signals: Fp1-A2, C3-A2, O1-A2 electroencephalogram; bilateral electrooculograms, submental electromyogram, electrocardiogram, oximetry, ribcage and abdominal movement, body position, sound intensity, and bilateral tibial electromyogram.

### 7.1.2.2 PSG Recordings

All PSG recordings were stored in European Data Format (EDF) format. EDF format allows storing multichannel signal with the different sampling frequency. It includes a header and data records. The header contains general information (e.g. patient identification, start time, end time, etc.) and technical specifications of each signal (e.g. sampling frequency, transducer type, etc.) and the data records contain samples. The sampling frequency for EEG, EOG and EMG is 256 Hz.

### 7.1.2.3 Expert Analysis

Based on PSG recordings, expert analysis was performed. PSG recordings were segmented into 30s epoch and manually scored by experts into five different stages : W, N1, N2, and N3 and R according to AASM manual. All the expert analysis results were stored in Extensible Markup Language (XML) file. XML is a markup language which was designed to store and transport data, it defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

### 7.1.2.4 Subjects Description

Overnight PSG signals were recorded from 16 subjects (9 males and 7 females) ranging from 22 to 82 years old ( $\mu = 45.6$ ,  $\sigma = 18.1$ ) in La Pitié-Salpêtrière hôpital (AP-HP) which located in Paris. AHI (average number of apneas and hypopneas per hour of sleep) ranges from 0 to 40.2 ( $\mu = 22.0$ ,  $\sigma = 16.1$ ). Details of patient information are shown in Table 7.3. In  $G$  column,  $M$  represents *Male* and  $F$  represents *Female*. In *Total Epoch* column, it lists the number of total epochs which was analyzed by experts for each subject. In  $W$ ,  $N1$ ,  $N2$ ,  $N3$ , and  $R$  column, it lists the number of epochs of Stage W, N1, N2,N3 and R for each subject. Total number of each stage for all 16 subjects is listed in Table 7.4.

<b>AHI<sup>1</sup></b>	<b>ID<sup>2</sup></b>	<b>G<sup>3</sup></b>	<b>Age</b>	<b>Total Epoch</b>	<b>W</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>R</b>
0	10	F	30	1083	105	80	448	250	200
0	12	F	44	1079	112	50	477	186	254
0	13	M	33	1237	173	31	530	262	241
0.5	11	F	23	1018	100	38	420	250	210
0.7	8	F	28	1004	180	27	488	137	172
1.1	4	M	22	1214	182	82	520	205	225
1.6	3	F	61	1071	102	18	539	209	203
2.4	9	F	40	1247	212	54	417	196	368
3.3	18	M	61	1184	116	154	527	175	212
3.9	29	M	29	845	48	24	407	167	199
4.8	1	M	52	1001	180	45	301	263	212
5.7	7	M	64	1331	151	104	627	237	212
22.7	63	M	36	1195	141	111	544	153	246
28.2	41	F	82	1317	811	80	331	74	21
35.3	5	M	65	1105	126	158	315	329	177
40.2	49	M	59	1203	453	266	437	16	31

Table 7.3 – Subject Information of Database 2

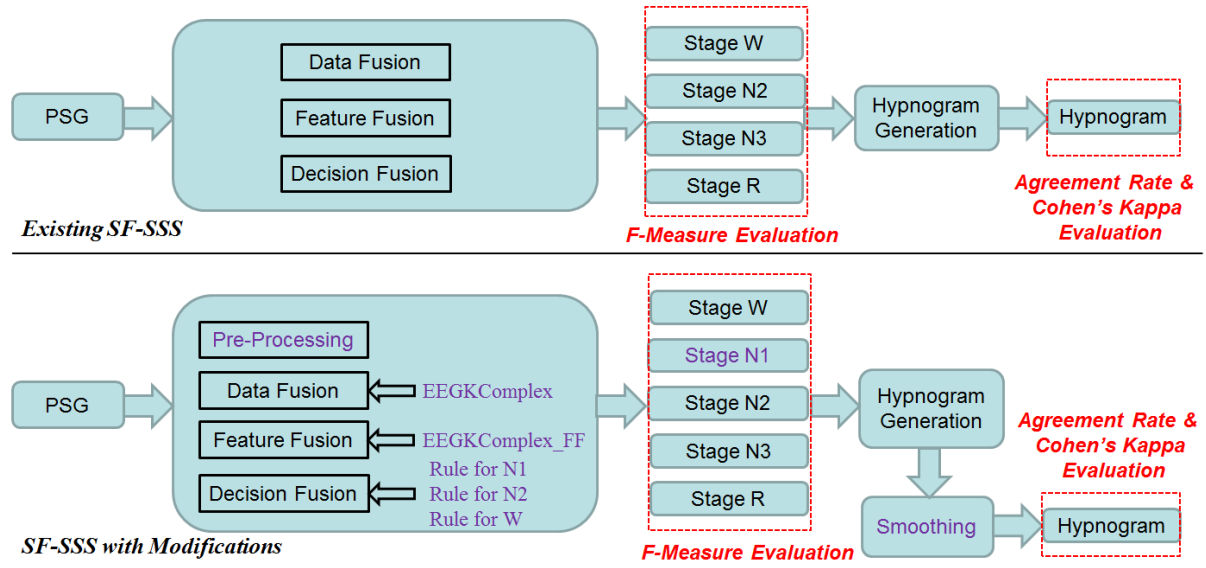
	<b>W</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>R</b>	<b>Total</b>
Expert Analysis (Epoch Number)	3192	1322	7328	3109	3183	18134

Table 7.4 – Information of Database 2

## 7.2 Evaluation of SF-SSS with Modifications

To evaluate SF-SSS with proposed modifications, F-Measure, Agreement Rate and Cohen's Kappa are used. As shown in Figure 7.1, classification result of each stage is obtained after Symbolic Fusion. F-Measure is used evaluate the classification performance of each stage. Based on the classification result of each stage, Hypnogram can be generated. Agreement Rate and Cohen's Kappa are calculated according to Hypnogram for evaluating the classification result of whole sleep stage.

Figure 7.1 – Evaluation of SF-SSS



### 7.2.0.1 F-Measure of Existing SF-SSS VS SF-SSS with Modifications

F-Measure is widely used in binary classification, it considers both precision and recall. The traditional F-Measure also called balanced F-score ( $F_1$ ) is used evaluate for each stage of modified SF-SSS as shown Equation 7.1. It balances the precision and recall.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7.1)$$

In F-Measure, Precision and Recall are used which as shown in Equation 7.2 and Equation 7.3, respectively. Precision (also called positive predictive value) is the ratio of all positive predictions among the predictions. Recall (also known as sensitivity or true positive rate) is the ratio of positive predictions among true events.

$$Precision = \frac{TP}{TP + FP} \quad (7.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (7.3)$$

TP, FN, FP and FN are used to evaluate how good the observations (predictions) reflect the actual events for a classification. True Positives and True Negatives are the observations which were correctly predicted. False Negatives are observations are negative where the actual events are positive. False Positive are observations are positive where the actual events are negative.

To evaluate SF-SSS, F-Measure is calculated for each stage: W, N1, N2, N3 and R. Table 7.5 list the confusion matrix of the classification result of stage W as an example. In column 3, **W** represent the classification result which classified by modified SF-SSS as stage W. In column 4, **NW** represent the classification result which classified by SF-SSS as not stage W.



		SF-SSS	
		W	NW
Expert Analysis	W	True Positive	False Negative
	NW	False Positive	True Negative

Table 7.5 – Confusion Matrix of Stage W

F-Measure of each stage using existing SF-SSS and SF-SSS with modifications are listed in Table 7.6. The last column is the mean value of F-Measure for 16 subjects of each stage. By adding modifications we proposed, F-Measure for stage W, N1 and N3 are improved. For stage W, mean F-Measure is improved from 0.6768 to 0.6802; for stage N1, mean F-Measure is improved from 0 to 0.1424 and for stage N2, mean F-measure is improved from 0.3787 to 0.4762.

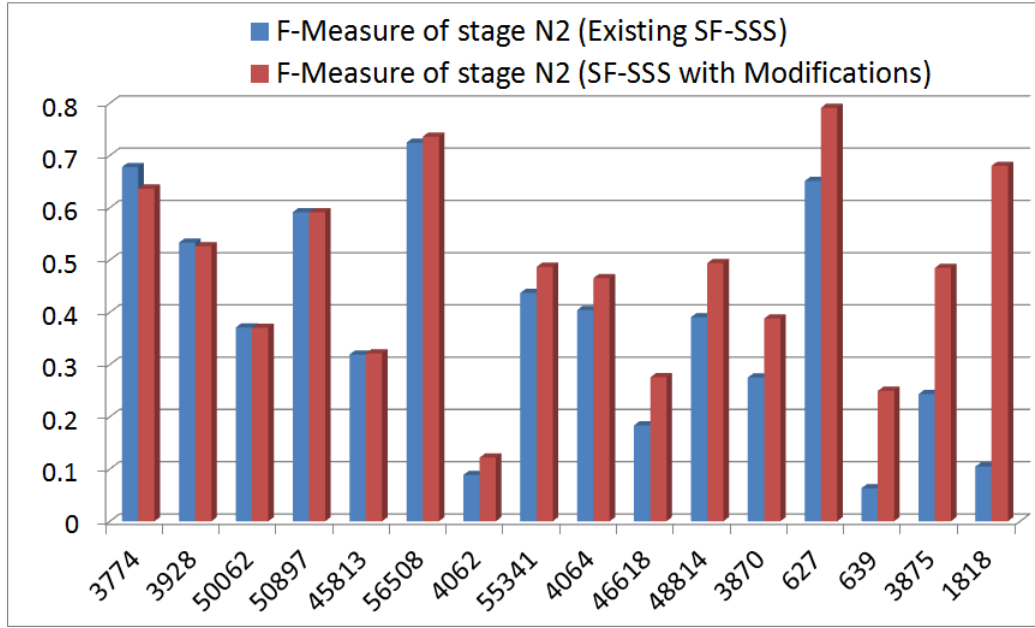
ID	F-Measure of Existing SF-SSS					F-Measure of Modified SF-SSS				
	W	N1	N2	N3	R	W	N1	N2	N3	R
627	0.8089	0	0.6512	0.8327	0.2694	0.8146	0.1467	0.7913	0.8327	0.2694
639	0.5806	0	0.0635	0.8425	0.5336	0.5806	0.3575	0.2496	0.8425	0.5336
1818	0.7813	0	0.1047	0.6027	0.2772	0.7732	0.0000	0.6803	0.6027	0.2772
3774	0.8819	0	0.6780	0.7298	0.3771	0.8831	0.3200	0.6366	0.7298	0.3771
3870	0.7322	0	0.2750	0.7873	0.5049	0.7322	0.2982	0.3878	0.7873	0.5049
3875	0.9447	0	0.2432	0.7653	0.2891	0.9447	0.0352	0.4846	0.7653	0.2891
3928	0.5503	0	0.5331	0.7967	0.3017	0.5503	0.1897	0.5263	0.7967	0.3017
4062	0.6117	0	0.0887	0.7424	0.6667	0.6117	0.0846	0.1218	0.7424	0.6667
4064	0.7051	0	0.4043	0.7514	0.2783	0.7051	0.0591	0.4652	0.7514	0.2783
45813	0.4701	0	0.3192	0.3958	0.0000	0.4888	0.0000	0.3212	0.3958	0.0000
46618	0.4686	0	0.1835	0.8077	0.6031	0.4686	0.1425	0.2755	0.8077	0.6031
48814	0.2626	0	0.3904	0.8375	0.4171	0.2913	0.1156	0.4939	0.8375	0.4171
50062	0.7184	0	0.3710	0.7918	0.6897	0.7184	0.1271	0.3702	0.7918	0.6897
50897	0.8434	0	0.5914	0.7004	0.5485	0.8434	0.1176	0.5914	0.7004	0.5485
55341	0.8235	0	0.4370	0.7279	0.5637	0.8235	0.2080	0.4868	0.7279	0.5637
56508	0.6462	0	0.7246	0.7143	0.6849	0.6531	0.0769	0.7360	0.7143	0.6849
<b>Mean</b>	<b>0.6768</b>	<b>0</b>	<b>0.3787</b>	<b>0.7391</b>	<b>0.4378</b>	<b>0.6802</b>	<b>0.1424</b>	<b>0.4762</b>	<b>0.7391</b>	<b>0.4378</b>

Table 7.6 – F-Measure on Database 1

F-Measure for stage N2 of each subject is shown in Figure 7.2. 12 out of 16 subjects, F-Measures are improved by using modified SF-SSS. Three of the subjects like subject 3774 and 3928, F-Measures are slightly decreased by using modified SF-SSS. However, by combining Recall Appendix (8.3) and Precision Appendix (8.3), Recall for subject 3774 and 3928 are improved from 0.8949 to 0.9150, 0.5137 to 0.5167 respectively while Precision are decreased. It is because of the precision of detection K-Complex which may also depend on the thresholds used to transforming K-Complex digital parameter into the symbolic interpretation of feature parameter.

After adding modifications of existing SF-SSS, F-Measure is improved. However, for stage N1, N2 and R, F-Measure is still low. Here, we don't evaluate the accuracy of Symbolic Fusion itself, because it also depends on the thresholds that were set up in transforming digital parameters into the symbolic interpretation of feature parameters; moreover, more sleep events can be involved to improve the performance and completion of SF-SSS.

Figure 7.2 – F-Measure of Existing SF-SSS VS SF-SSS with Modifications for stage N2



### 7.2.0.2 Agreement Rate of Existing SF-SSS VS SF-SSS with Modifications

After obtaining the classification results of each stage, a Hypnogram Generation module was applied in existing SF-SSS which simply generates the Hypnogram according to the writing order of stage N1, R, N3, N2 and W. In details, the classification result of stage N1 will write firstly, followed by R, N3, N2, and W will write lastly. The previously written results can be overwritten by the later results it writes. The method applied in existing SF-SSS to generate Hypnogram is simple, while it needs to be improved in further work. Once Hypnogram generates, Agreement Rate (AR) is used to evaluate the overall performance. AR is the proportion of same results classified by modified SF-SSS and physicians in comparative to results classified by physicians. AR for each stage is shown as follows:

$$AR\{W, N1, N2, N3, R\} = \frac{TP\{W, N1, N2, N3, R\}}{TP\{W, N1, N2, N3, R\} + FN\{W, N1, N2, N3, R\}}; \quad (7.4)$$

In Equation 7.4,  $TP\{W, N1, N2, N3, R\}$  is True Positive of stage W,N1,N2,N3,R.  $FN\{W, N1, N2, N3, R\}$  is False Positive of stage W,N1,N2,N3,R.

AR for all stages is defined in Equation 7.5, which equals to the proportion of the sum of same results classified by SF-SSS and physicians for all stages in comparative to results classified by physicians.

$$AR\{Total\} = \frac{TP}{TP + FN}; \quad (7.5)$$

AR for each stage and Total AR using SF-SSS with proposed modifications are listed in Appendix (8.6) in comparison to the results obtained using existing SF-SSS as shown in Appendix (8.5). ARs for stage N1 are still very low, it mainly because of the method we adopted in generating Hypnogram which needs to be improved in our further work.

Total AR of each subject in order to compare modified SF-SSS and existing SF-SSS is shown in Figure 7.3. ARs for 15 out of 16 subjects are increased using modified SF-SSS. Only AR for subject 639 is slightly decreased. However, F-Measure of subject 639 for

stage N1 and N2 are improved from 0 to 0.3575, and 0.0635 to 0.2496 respectively using SF-SSS with modifications as shown in Table Appendix (8.3). For other stages like W, N3 and R, F-Measure keeps the same. AR for subject 639 decreases because the Smoothing part adding for SF-SSS model. In Smoothing, a smoothing rule for three consecutive epochs of R, N2, R are replaced with the R, R, R is performed. Figure 7.4 presents the Hypnogram for subject 639 before and after smoothing. After smoothing, AR for stage N2 decreases slightly due to some N2 epochs are labeled as stage R. The main reason why the Smoothing part in modified SF-SSS lower the AR for subject 637 is due to the low precision and recall which obtained in classification for stage N2.

Figure 7.3 – Total Agreement Rate of Existing SF-SSS VS SF-SSS with Modifications

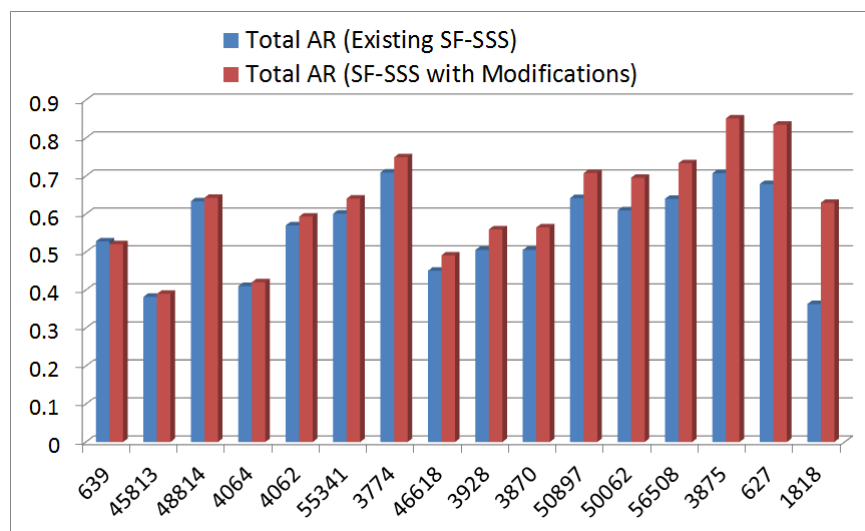
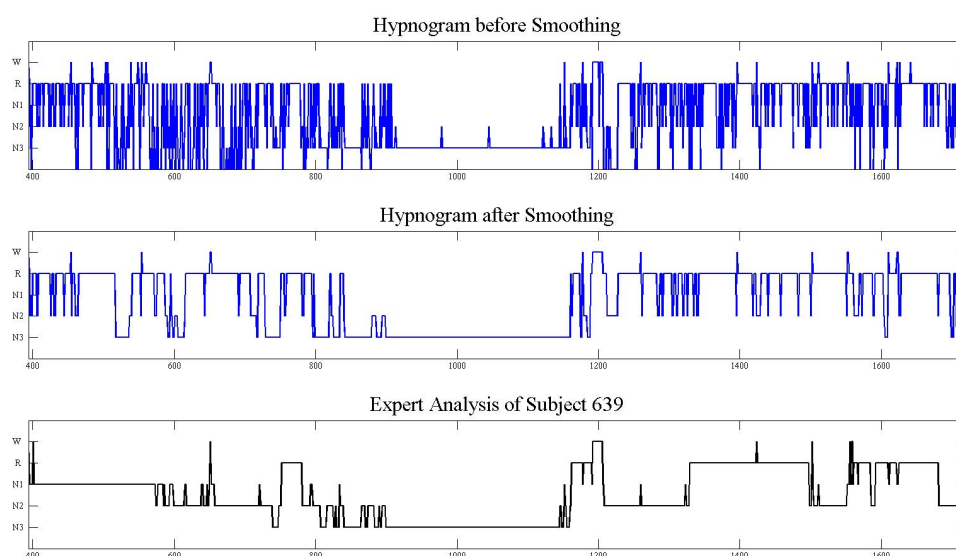


Figure 7.4 – Agreement Rate of Existing SF-SSS VS SF-SSS with Modifications for Subject 639



### 7.2.0.3 Confusion Matrix of Existing SF-SSS VS SF-SSS with Modifications

Besides using AR to evaluate the agreement percentage between SF-SSS and experts analysis, confusion matrix is also adopted to provide more details of SF-SSS classification results. The confusion matrix is a square matrix showing the relation between experts analysis and the results obtained using SF-SSS. The values in the diagonal elements represent the number of correctly identified stages and the off-diagonal values are the number of misclassified. An element of row  $i$  and column  $j$  indicates the number of epochs sleep stage  $i$  is misclassified into stage  $j$ .

Confusion matrix for the Database 1 using modified SF-SSS is listed in Table 7.8, while the confusion matrix which obtained using existing SF-SSS is shown in Table 7.7. In the last column, AR for each stage and AR for the whole database are presented. Meanwhile, in the last row, Cohen's Kappa value is calculated.

Cohen's Kappa coefficient  $\kappa$  is also estimated because it provides a more robust estimate of the sleep staging performance as compared to the simple agreement percentage. As suggested by Landis and Koch [42], Kappa values of 0.21-0.4 indicate fair agreement, 0.41-0.6 moderate agreement, 0.61-0.8 substantial agreement, and 0.81-0.99 almost perfect agreement.

The equation for  $\kappa$  is:

$$\kappa = \frac{p_o - p_e}{1 - p_e}; \quad (7.6)$$

where  $p_o$  is the relative observed agreement among SF-SSS and experts analysis, and  $p_e$  is the hypothetical probability of chance agreement. If results from SF-SSS are in complete agreement with expert analysis then  $\kappa = 1$ . If there is no agreement among SF-SSS classification results and experts analysis other than what would be expected by chance, then  $\kappa \leq 0$ .

		Existing SF-SSS						AR
		W	N1	N2	N3	R	Total	
Expert Analysis	W	3614	0	779	44	310	4747	0.7613
	N1	530	0	815	98	548	1991	0.0000
	N2	377	0	3499	734	1916	6526	0.5362
	N3	18	0	1210	2177	211	3616	0.6020
	R	67	0	562	124	1835	2588	0.7090
	Total	4606	0	6865	3177	4820	19468	0.5715
Kappa		0.4368						

Table 7.7 – Confusion Matrix of Database 1 using Existing SF-SSS

		Modified SF-SSS						AR
		W	N1	N2	N3	R	Total	
Expert Analysis	W	4054	13	339	36	305	4747	0.8540
	N1	771	39	519	103	559	1991	0.0196
	N2	555	12	4246	660	1053	6526	0.6506
	N3	22	0	1321	2205	68	3616	0.6098
	R	61	14	488	67	1958	2588	0.7566
	Total	5463	78	6913	3071	3943	19468	0.6422
Kappa		0.5266						

Table 7.8 – Confusion Matrix of Database 1 using Modified SF-SSS

In comparison between confusion matrix using modified SF-SSS and existing SF-SSS, the number of correctly identified stages in diagonal elements is increased using modified SF-SSS. AR for stage W, N1, N2, N3 and R is increased from 0.7613 to 0.8540, 0 to 0.0196, 0.5362 to 0.6506, 0.6020 to 0.6098 and 0.7090 to 0.7566 respectively. Total AR is increased from 0.5715 to 0.6422. Cohen's Kappa coefficient is also increased.

To evaluate the performance of modified SF-SSS, F-Measure, Agreement Rate and Cohen's Kappa are used. F-Measure is used to assess the performance of each stage; Agreement Rate and Cohen's Kappa are used to evaluate the performance of overall classification. With the modifications proposed in modified SF-SSS, it reaches higher values of F-Measure, Agreement Rate and Cohen's Kappa in comparison to existing SF-SSS.

### 7.3 Evaluation of Automatic Thresholds Setting-Up Method

To evaluate the Automatic Thresholds Setting-Up method, a new database (Database 2) which was recorded in 2015 is used. We start with the evaluation of Differential Evolution (DE) and Cross Entropy (CE) algorithms to select the appropriate control parameter for the sleep staging application. Then, PSSS is evaluated from the selection of training set to the performance of evaluation set. At last, comparison between DE-PSSS and CE-PSSS are made.

#### 7.3.1 Evaluation of Automatic Thresholds Setting-Up Method

Evaluation of ATSU mainly involving the selection of control parameters of DE and CE. In this section, the control parameters selection of DE and CE are analyzed.

##### 7.3.1.1 Parameters Selection of Differential Evolution

There are three main control parameters of Differential Evolution: Population Size (NP), Mutation Scale Factor (F) and Crossover Rate (CR) as shown in Table 7.9. NP plays a crucial role in the efficiency and effectiveness of DE. Large population size potentially increases the population diversity and computational complexity. F controls the amplification of the differential variation. CR controls the number of components inherited from the mutant vector which can be interpreted as a mutation probability.

- **Population Size (NP)**

Population size is probably the most problem-dependent control parameter. It plays a crucial role in the efficiency and effectiveness of DE. Large population size potentially increases the population diversity. In [70], NP is suggested from  $5 \cdot D$  up to  $10 \cdot D$ , where D is the dimension. In ATSU method, D is the number of thresholds.

- **Mutation Scale Factor (F)**

F controls the amplification of the differential variation. Too small F values increase the risk of premature convergence (i.e. converge to an undesirable point), while too large F values decrease the convergence speed that degrades DE efficiency and may result in early termination.

- **Crossover Rate (CR)**

CR controls the number of components inherited from the mutant vector or current vector. It influences the probability for a component to be selected from the mutant vector. CR also heavily depends on the problem, where small CR values are suggested for separable problems and high CR values are suggested for non-separable problems [54].

The selection of appropriate parameters can affect the efficiency of ATSU directly. Due to the variability of the underlying mathematical properties of different problems, a fixed set of control parameters that suits well for one problem, or a class of problems does not guarantee that it will work well for another class, or range of problems [61]. That is, the selection of control parameters is problem dependent. To ensure the performance of ATSU, selection of control parameters is extremely important.

To select the appropriate control parameters, trial-and-error approach, is widely used. Several sets of control parameters are tested, then appropriate control parameters can be selected based on the average performance of the problem.

Main Control Parameters of DE	Range
Population Size(NP)	[5D, 10D] <sup>8</sup>
Mutation Scale Factor(F)	[0, 2]
Crossover Rate(CR)	[0, 1]

Table 7.9 – Main Control Parameters of Differential Evolution

To evaluate the impact of different control parameters and strategies of DE, population size 5\*D, 10\*D and 500; Mutation Scale Factor of 0.5, 1 and 1.5; Crossover Rate of 0.1, 0.5 and 0.9 are used, as shown in Table 7.10. Initially the maximum iteration of DE-PSSS was set to 200, however, the simulation results show after 100 iterations, F-Measure tend to be stable. Thus, the maximum iteration was changed from 200 to 100.

	Parameters and Strategy for Simulation
Population Size (NP)	[5D, 10D, 500]
Mutation Scale Factor (F)	[0.5, 1, 1.5]
Crossover Rate (CR)	[0.1, 0.5, 0.9]

Table 7.10 – Parameters used for DE Control Parameters Selection Simulation

Figure 7.5, Figure 7.6 and Figure 7.7 shows F-Measure dependence on generation number using Strategy *DE/rand/1/bin* for Subject 1 in classification of stage W with population size 5\*D, 10\*D and 500 respectively. Due to the stochastic ability of Differential Evolution, F-Measure in Figure 7.5, Figure 7.6 and Figure 7.7 are mean values of 20 independent runs.

Figure 7.5 – DE Control Parameters Selection: DE/rand/1/bin, NP=5D

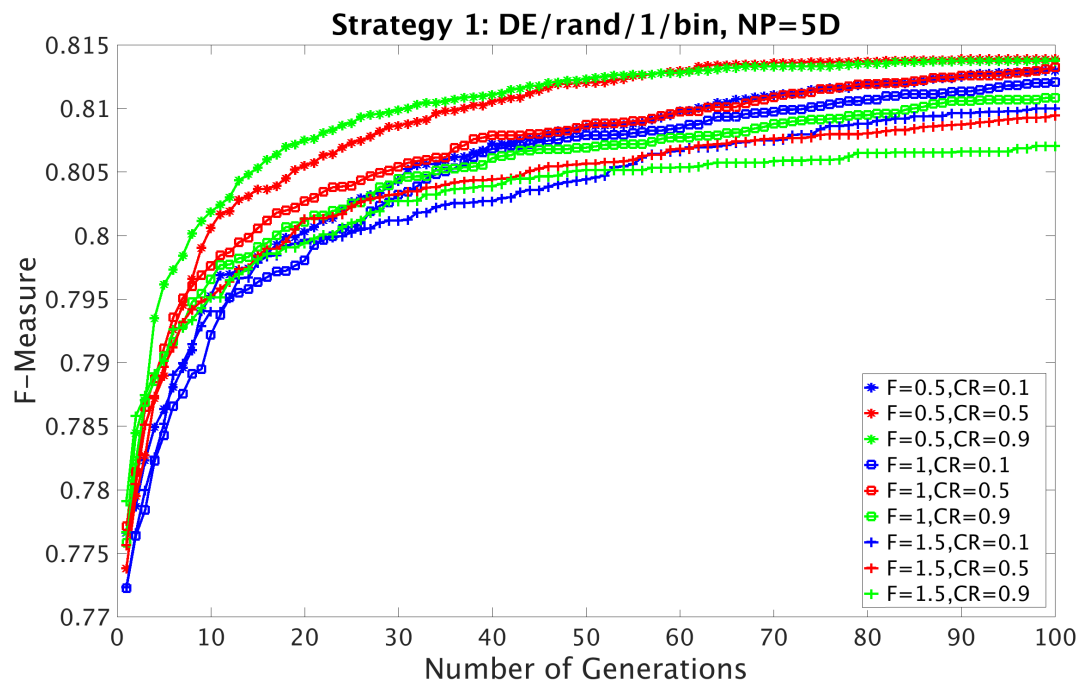


Figure 7.6 – DE Control Parameters Selection: DE/rand/1/bin, NP=10D

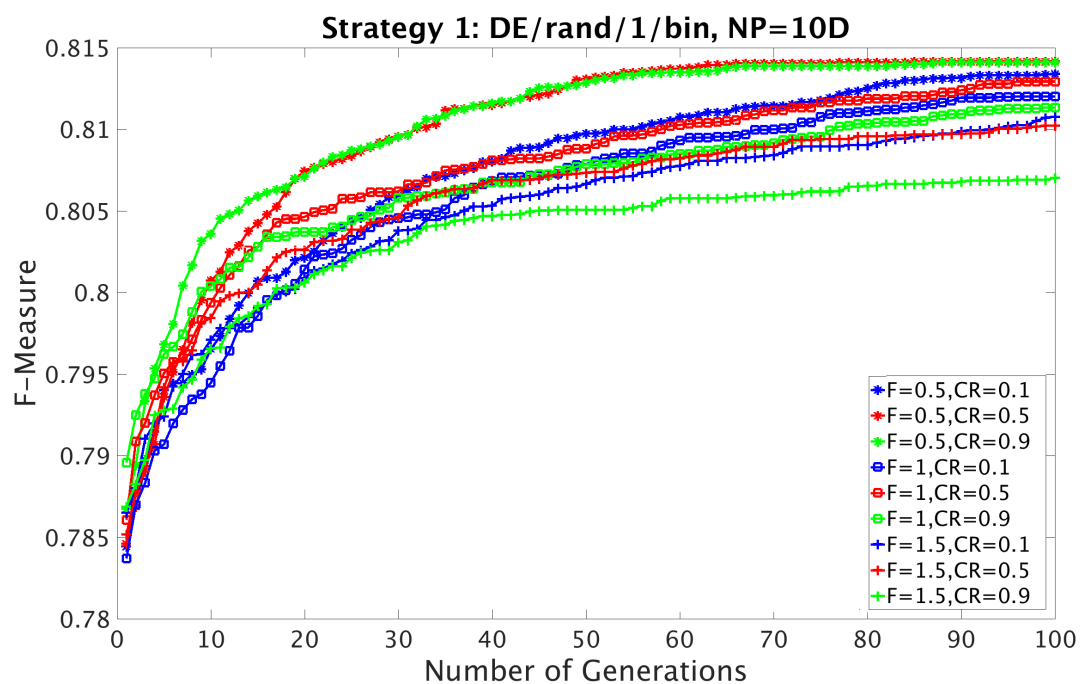
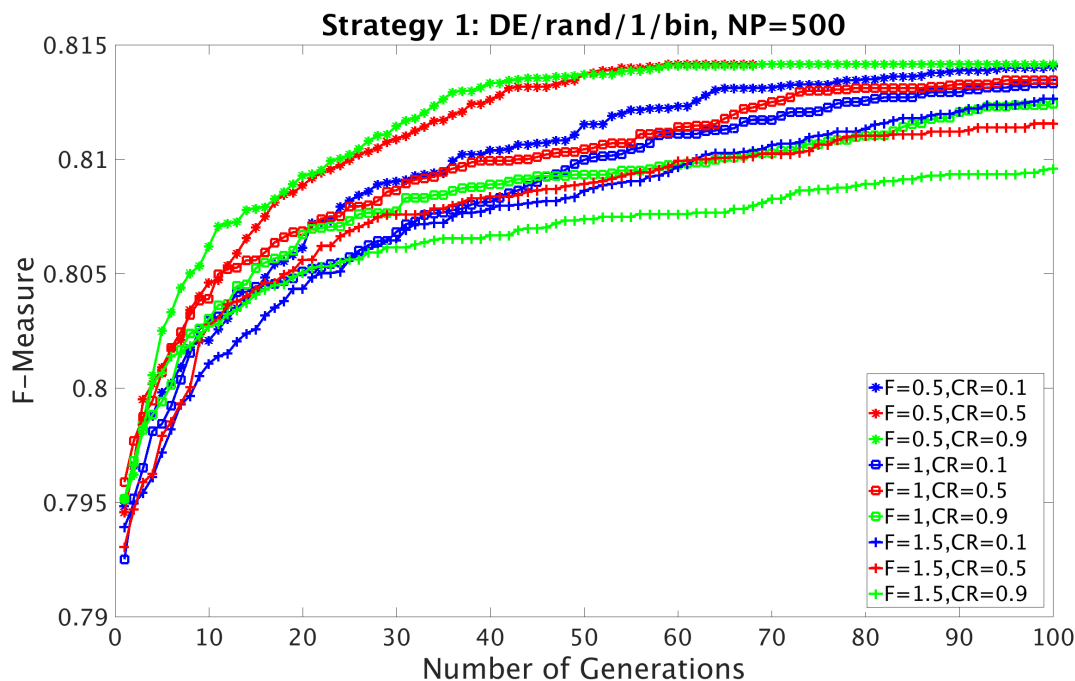
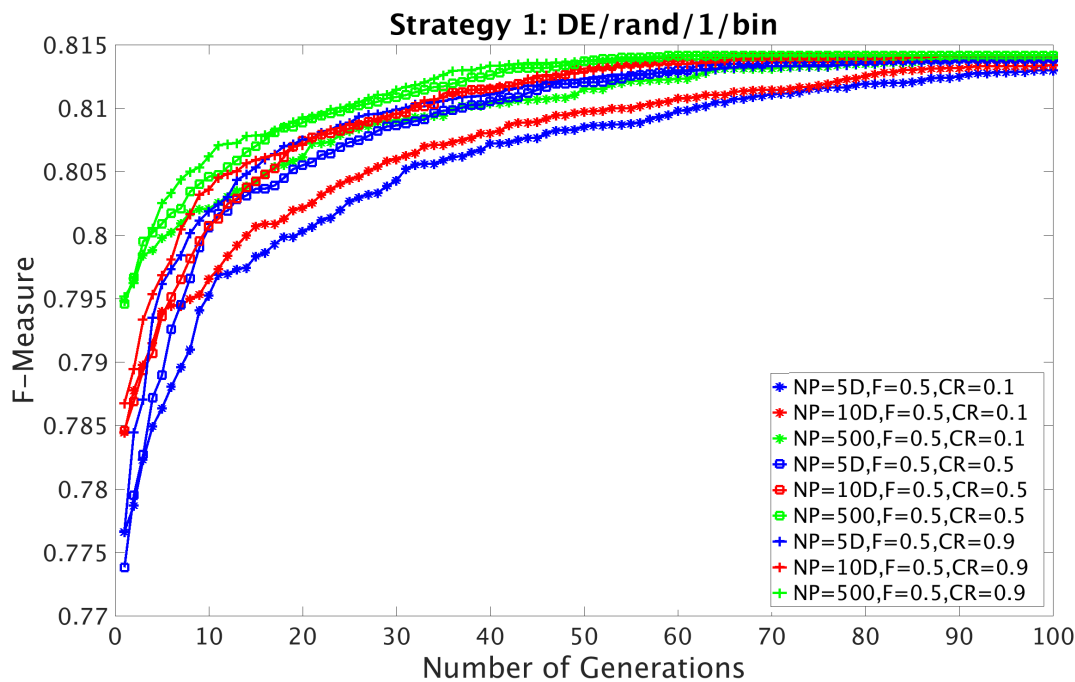


Figure 7.7 – DE Control Parameters Selection: DE/rand/1/bin, NP=500



With increasing of  $F$ , the convergence speed decreases. Among  $F=0.5$ ,  $F=1$  and  $F=1.5$ , 0.5 provides the best convergence speed as shown in Figure 7.5, Figure 7.6 and Figure 7.7. For the same  $NP$  and  $F=0.5$ ,  $CR=0.9$  is relative better than  $CR=0.1/0.5$ .

Figure 7.8 – DE Control Parameters Selection: DE/rand/1/bin



In order to evaluate the impact of different population size, Figure 7.8 shows comparison among different  $NP$  and  $CR$  when  $F=0.5$ . Large population size potentially increases the population diversity and may provide fast convergence speed. However, it also increases the computational time.  $NP=500$ ,  $F=0.5$ ,  $CR=0.5/0.9$  have fast convergence



speed than others. For  $NP=500$ ,  $F=0.5$ ,  $CR=0.1$  almost have same convergence speed as  $NP=5D/10D$ ,  $F=0.5$ ,  $CR=0.9$ .

According to the simulation results, in order to balance F-Measure and computational complexity,  $NP=5D$ ,  $F=0.5$ ,  $CR=0.9$  can be suggested as optimal parameters for Strategy 1 which have less computational complexity compared to  $NP=500$  and have slightly better convergence speed than  $NP=5D$ ,  $F=0.5$ ,  $CR=0.1/0.5$ .

### 7.3.1.2 Parameters Selection of Cross Entropy

There are two main control parameters of Cross Entropy: Sample Size (N) and Elite Sampling Rate ( $\rho$ ).

1. **Sample Size(N)**: Large sample size potentially increases the population diversity. However, large sample size also increases computational time.
2. **Elite Sampling Rate( $\rho$ )**:  $\rho$  controls the number of elite samples which belongs to (0,1). It should ensure the elite samples to be lager enough for obtaining a reliable parameter update for PDFs.

To evaluate impact of different control parameters of CE, Sample Size 100, 500, 1000 and 10000; Elite Sampling Rate 0.1, 0.3, 0.5 and 0.7 are used for different strategies as shown in Table 7.11.

	<b>Parameters and Strategy for Simulation</b>
Sample Size (N)	[100, 500, 1000, 10000]
Elite Sampling Rate( $\rho$ )	[0.1, 0.3, 0.5, 0.7]

Table 7.11 – Parameters used for CE Control Parameters Selection Simulation

Figure 7.9 shows F-Measure dependence on iteration number for different sample sizes of stage W classification for subject 1. Due to the stochastic ability of Cross Entropy, F-Measure in Figure 7.9 are mean values of 20 independent runs. Initially, the maximum iteration number was set to 200. However, the simulation shows that after 50 iterations, F-Measures stay in the stable. So the maximum iteration number changed from 200 to 50.

Figure 7.10 shows F-Measure dependence on the first 50 iterations. Large sample size potentially increases the sample diversity and may provide fast convergence speed. However, it also increases the computational time. For the same sample size, the increase of elite sampling rate slows down the convergence speed. According to the simulation results, sample size of 500 with elite sampling rate of 0.1 provides optimum between F-Measure and computational complexity.

Figure 7.9 – CE Control Parameters Selection (Iteration=200)

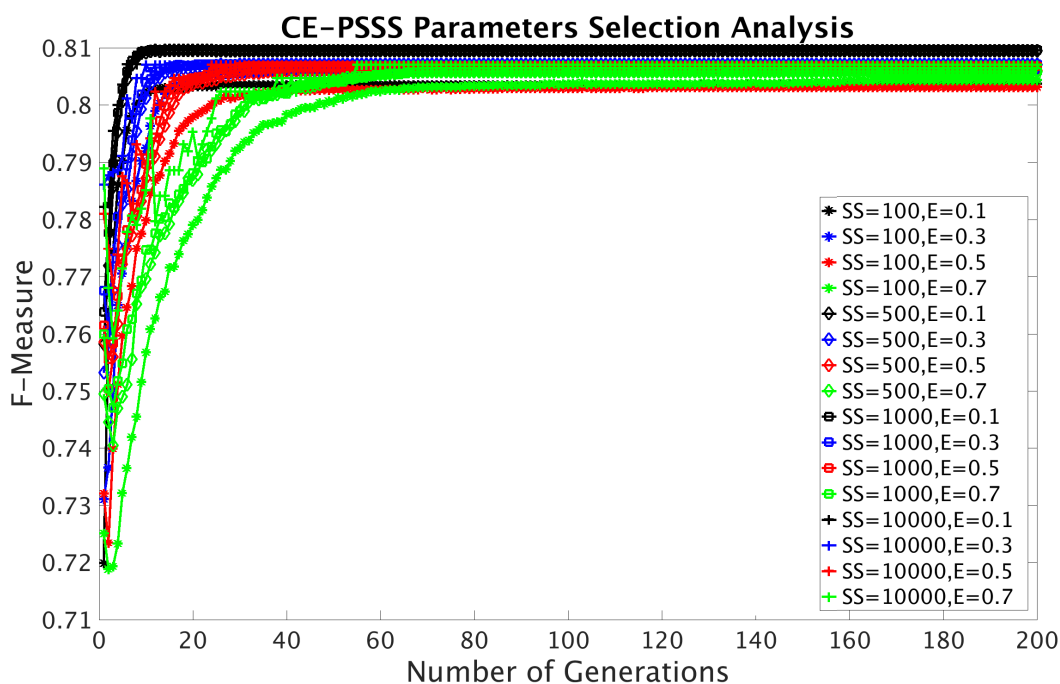
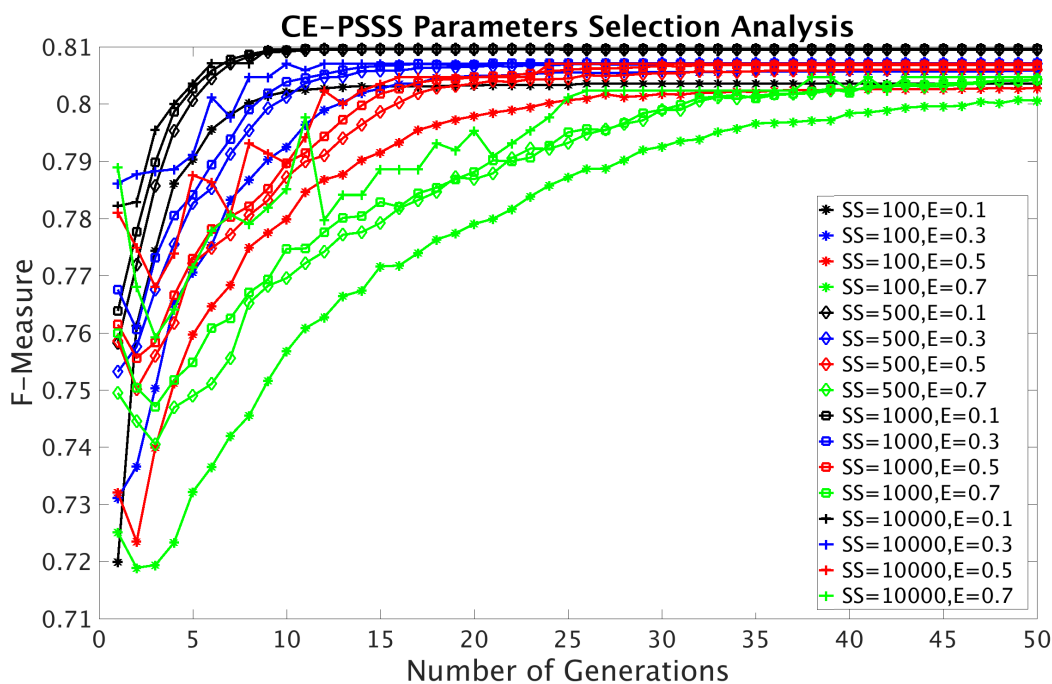


Figure 7.10 – DE Control Parameters Selection (Iteration=50)



### 7.3.2 Evaluation of Personalized Sleep Staging System

In this section, DE-PSSS and CE-PSSS are evaluated involving the selection of training set and the performance of evaluation set.

To fully understand how many epochs are required to be scored by physicians in order to set up personalized thresholds, different training set size are selected.

### 7.3.2.1 Training Set Selection of DE-PSSS

Four different set of values have been used as the size of training set & evaluation set for DE-PSSS: 5% & 95%, 10% & 90%, 15% & 85% and 20% & 80%. DE-PSSS selects the optimal thresholds according to different training sets and passes these thresholds to the corresponding evaluation sets. The selection of 5%, 10%, 15% and 20% of the training set are randomly selected 5%, 10%, 15% and 20% of each stage (W, N1, N2, N3 and R) for each subject.

<b>Subject ID: 1</b>	<b>W</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>R</b>	<b>Total</b>
<b>Total Number</b>	180	45	301	263	212	1001
<b>Training Set</b>	9	3	16	14	11	53
<b>Evaluation Set</b>	171	42	285	249	201	948

Table 7.12 – Training Set & Evaluation Set: 5% & 95 % for Subject ID=1

Table 7.12 gives an example of training set and evaluation set for subject 1 using 5% as training set and remaining 95% as evaluation set. Total number of 53 epochs are randomly selected from the over-night training set and the remaining 948 epochs are used as evaluation set.

Thresholds which can provide highest F-Measure using NP=5D, F=0.5, CR=0.9 control parameters for training set are adopted as the optimal thresholds for evaluation set. F-Measure for each stage of evaluation set is calculated as shown in Appendix (8.7). In order to verify the robustness of DE-PSSS, F-Measures in Appendix (8.7) are the mean value of 20 independently runs. Meanwhile, standard deviation of these 20 runs is also provided Appendix (8.7).

The mean value of F-Measure for 16 subjects is presented in the last column of Appendix (8.7). For stage W, N2 and N3, F-Measure can reach over 0.71. For stage R, the mean value is relative lower than stage W, N2 and N3. This is because for subject 41 and 49, F-Measure is very low which only reaches 0.3138 and 0.1874 respectively. For subject 41, only 2 epochs of stage R are selected among 68 epochs (total epoch number of stage R) as the training set. Thresholds can easily reach high F-Measure for only 2 epochs, while it may not obtain high F-Measure for evaluation set. For subject 49, only 2 epochs of stage R are selected as training set among 62 epochs (total epoch number of stage R).

F-Measure of evaluation set using 10% randomly selected epochs of each stage as training set is shown in Appendix (8.8). F-Measure can reach over 0.73 for stage W, N2 and N3. For stage R, F-Measure can reach 0.6818. F-Measures of evaluation set using 15% and 20% of training are listed in Appendix (8.9) and Appendix (8.10).

In order to select the optimal training set for DE-PSSS, comparison among four different training set & evaluation set is shown in Table 7.13.

<b>DE-PSSS(Training Set &amp; Evaluation Set)</b>	<b>W</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>R</b>
<b>5% &amp; 95%</b>	0.7137	0.2209	0.7229	0.7363	0.6565
<b>10% &amp; 90%</b>	0.7335	0.2299	0.7354	0.7487	0.6818
<b>15% &amp; 85%</b>	0.7411	0.2410	0.7380	0.7568	0.6932
<b>20% &amp; 80%</b>	0.7452	0.2453	0.7419	0.7623	0.6955

Table 7.13 – F-Measure Comparison among training set & evaluation set of DE-PSSS

With the increment of the training set, F-Measures for evaluation set are increased.

For the consideration of consuming less time for physicians in scoring training set, 5% can be considered as the optimal value for the training set. However, for the consideration of reaching higher F-Measure for evaluation set, 15% can be considered as the optimal value for the training set. In comparison to 20%, 5% more epochs should be scored, while F-Measure for each stage only increases less than 0.01.

### 7.3.2.2 Training Set Selection of CE-PSSS

To evaluate impact of training set selection on CE-PSSS, four different set of values have been used to balance the size of training set & evaluation set: 5% & 95%, 10% & 90%, 15% & 85% and 20% & 80%. DE-PSSS selects the optimal thresholds according to different training sets and passes these thresholds to the corresponding evaluation sets.

Details of F-Measure of evaluation set using 5%, 10%, 15% and 20% are shown in Appendix (8.11), Appendix (8.12), Appendix (8.13) and Appendix (8.14) respectively. F-Measure in Appendix (8.11), Appendix (8.12), Appendix (8.13) and Appendix (8.14) for each subject is the mean value of 20 independently runs.

Comparison among F-Measures of evaluation set using different training set is shown in Table 7.14.

CE-PSSS(Training Set & Evaluation Set)	W	N1	N2	N3	R
5% & 95%	0.6652	0.1566	0.7194	0.7385	0.6326
10% & 90%	0.6884	0.1683	0.7309	0.7464	0.6632
15% & 85%	0.6989	0.1804	0.7367	0.7521	0.6756
20% & 80%	0.7051	0.1835	0.7402	0.7573	0.6824

Table 7.14 – F-Measure Comparison among training set & evaluation set of CE-PSSS

As training set increases, F-Measures for evaluation set also increases slowly. For the consideration of consuming less time for physicians in scoring training set, 5% can be considered as the optimal value for training set. For the consideration of reaching higher F-Measure for evaluation set, 15% can be considered as the optimal value for training set. In comparison to 20%, 5% more epochs should be scored, while F-Measure for each stage only increases approximately 0.01 for stage W and R.

### 7.3.3 Comparison between DE-PSSS and CE-PSSS

In this thesis, we proposed two personalized sleep staging systems based on ATSU. Two different stochastic search algorithms: Different Evolution and Cross Entropy are used in ATSU. To compare DE-PSSS and CE-PSSS, we start from analyzing the algorithms.

#### 7.3.3.1 Algorithm Comparison

Table 7.15 lists the comparisons between DE and CE. For DE, three control parameters: Population Size (NP), Mutation Scale Factor (F) and Crossover Rate (CR) are required to be set; for CE, two control parameters: Sample Size (N) and Elite Sampling Rate  $\rho$  are required to be set.

In consideration of the maximum iteration number of DE and CE, 200 is set as initially maximum iteration number. However, for DE, it tends to be stable after 100 iterations and for CE, it tends to be stable after 50 iterations. Thus the maximum iteration number for DE and CE are set to 100 and 50 respectively.

For CE with control parameters (N=500,  $\rho=0.1$ ), within 10 iterations F-Measure can reach approximate to 0.81 as shown in Figure 7.10 and then F-Measure stays as same.

While for DE with control parameter(NP=5D, F=0.5, CR=0.9),it require approximate 30 iterations to reach 0.81 as shown in Figure 7.8 but F-Measure still increases after 30 iterations.

	<b>DE</b>	<b>CE</b>
<b>Control Parameters</b>	NP, F, CR	N, $\rho$
<b>Iterations</b>	100	50
<b>Convergence</b>	slow but stable	fast but may trap into local optimal

Table 7.15 – Comparisons between DE and CE

### 7.3.3.2 Computational Time

In consideration of targeting DE-PSSS and CE-PSSS into embedded systems, the performances of Differential Evolution and Cross Entropy algorithms in terms of latencies are evaluated. Table 7.16 and Table 7.17 gives detailed information on computational time (in seconds) of DE-PSSS and CE-PSSS for one iteration respectively.

For different control parameters of DE-PSSS, only Population Size (NP) affects the computational time. For CE-PSSS, only Sample Size (N) affects the computational time.

<b>DE-PSSS</b>	<b>NP=5D</b>	<b>NP=10D</b>	<b>NP=500</b>
folk <sup>9</sup>	<b>0.20</b>	0.35	1.65
trot <sup>10</sup>	<b>0.27</b>	0.51	2.31

Table 7.16 – Computational Time (s) of DE for one iteration

<b>CE-PSSS</b>	<b>N=100</b>	<b>N=500</b>	<b>N=1000</b>	<b>N=10000</b>
folk <sup>9</sup>	0.38	<b>1.67</b>	3.34	31.41
trot <sup>10</sup>	0.53	<b>2.31</b>	5.01	41.78

Table 7.17 – Computational Time (s) of CE for one iteration

Compare the computational time of DE and CE, the time CE (SS=500) consumes is more 8 times than DE (NP=5D) for one iteration.

### 7.3.3.3 F-Measure Comparison

F-Measure of evaluation set using DE-PSSS and CE-PSSS is also compared by adopting the optimal thresholds which are searched from different size of training set as shown in Table 7.18.

<sup>9</sup> folk sever: CPU 2xXeon E5-2640; 12 coeurs; 128Go RAM

<sup>10</sup> trot sever: CPU 2xXeon E5-2637; 8 coeurs; 128Go RAM

		<b>W</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>R</b>
<b>DE-PSSS</b>	5% & 95%	0.7137	0.2209	0.7229	0.7363	0.6565
	10% & 90%	0.7335	0.2299	0.7354	0.7487	0.6818
	15% & 85%	0.7411	0.2410	0.7380	0.7568	0.6932
	20% & 80%	0.7452	0.2453	0.7419	0.7623	0.6955
<b>CE-PSSS</b>	5% & 95%	0.6652	0.1566	0.7194	0.7385	0.6326
	10% & 90%	0.6884	0.1683	0.7309	0.7464	0.6632
	15% & 85%	0.6989	0.1804	0.7367	0.7521	0.6756
	20% & 80%	0.7051	0.1835	0.7402	0.7573	0.6824

Table 7.18 – F-Measure

With the increment of the training set, F-Measures for evaluation set are also increased. For the consideration of consuming less time for physicians in scoring training set, 5% can be considered as the optimal value for the training set. While if we want to ensure the high performance of evaluation set, 15% can be considered as the optimal value for the training set. In comparison to 20%, 5% more epochs should be scored, while F-Measure for each stage only increases less than 0.01.

Compare DE-PSSS and CE-PSSS, DE-PSSS is much better which can reach higher F-Measure for all stage using the same size of the training set and consume less time.

Proposed DE-PSSS and CE-PSSS are evaluated on the whole night PSG recordings of subjects. While, if a short-time PSG recordings are required to be analyzed (e.g. if we only need to analyze 1 hour PSG recordings), an increment of training set should be taken into consideration to ensure there is enough epochs number for the training set instead of only selecting 5%.

## 7.4 F-Measure Comparison between SF-SSS and DE-PSSS/CE-PSSS

To compare SF-SSS and personalized sleep staging system, F-Measure is used to assess the classification results of each stage.

### 7.4.1 F-Measure of SF-SSS using MTSU

In SF-SSS model, Manual Thresholds Set Up (MTSU) is used to set up thresholds. Table 7.19 provides F-Measure of each stage using MSUT of SF-SSS model for Database 2. More details of F-Measure for each patient is shown in Appendix (8.15).

<b>W</b>	<b>N1</b>	<b>N2</b>	<b>N3</b>	<b>R</b>
<b>0.5397</b>	<b>0.0494</b>	<b>0.3859</b>	<b>0.5768</b>	<b>0.0624</b>

Table 7.19 – Average F-Measures of SF-SSS (with modifications) using MTSU on Database 2

F-Measures for stage W and N3 can reach over 0.5, while F-Measures for stage N1 and R is below 0.1.

### 7.4.2 Comparison among SF-SSS, CE-PSSS and DE-PSSS

Table 7.20 and Figure 7.11 provide F-Measure using modified SF-SSS, DE-PSSS and CE-PSSS. In SF-SSS model, thresholds are setting by MTSU method. While in DE-

PSSS and CE-PSSS, thresholds are setting by ATSU using Differential Evolution and Cross Entropy respectively.

		W	N1	N2	N3	R
<b>SF-SSS using MTSU</b>		0.5397	0.0494	0.3859	0.5768	0.0624
<b>DE-PSSS</b>	<b>5% &amp; 95%</b>	<b>0.7137</b>	<b>0.2209</b>	<b>0.7229</b>	<b>0.7363</b>	<b>0.6565</b>
	10% & 90%	0.7335	0.2299	0.7354	0.7487	0.6818
	<b>15% &amp; 85%</b>	<b>0.7411</b>	<b>0.2410</b>	<b>0.7380</b>	<b>0.7568</b>	<b>0.6932</b>
	20% & 80%	0.7452	0.2453	0.7419	0.7623	0.6955
<b>CE-PSSS</b>	<b>5% &amp; 95%</b>	<b>0.6652</b>	<b>0.1566</b>	<b>0.7194</b>	<b>0.7385</b>	<b>0.6326</b>
	10% & 90%	0.6884	0.1683	0.7309	0.7464	0.6632
	<b>15% &amp; 85%</b>	<b>0.6989</b>	<b>0.1804</b>	<b>0.7367</b>	<b>0.7521</b>	<b>0.6756</b>
	20% & 80%	0.7051	0.1835	0.7402	0.7573	0.6824

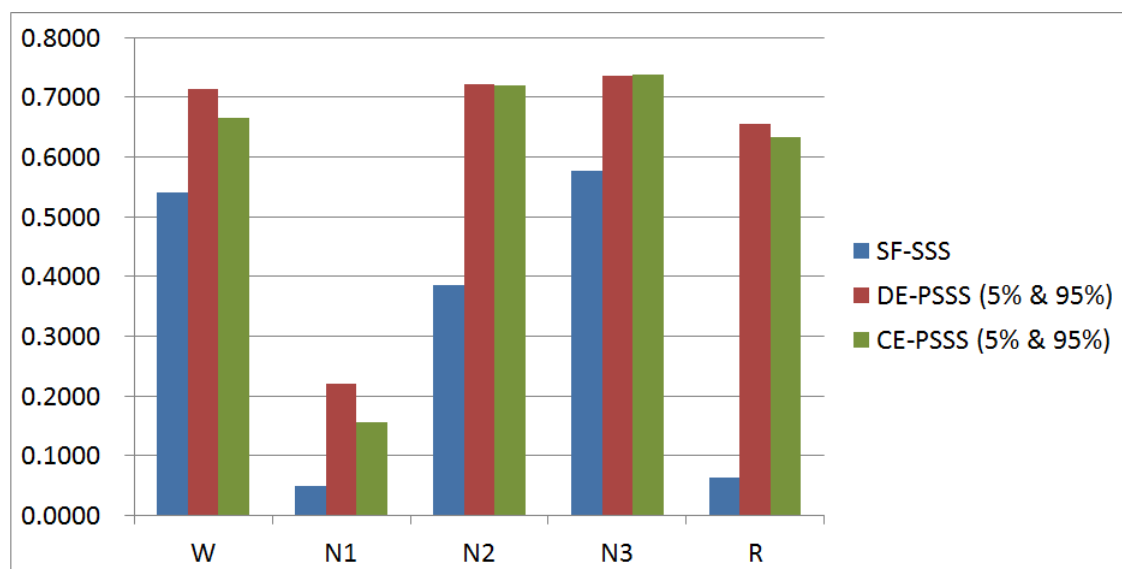
Table 7.20 – F-Measure of SF-SSS and PSSS

Compare SF-SSS using MTSU to PSSS using ATSU, F-Measure increased sharply for each stage by using PSSS.

In comparison to SF-SSS, F-Measures increase 0.17, 0.17, 0.34, 0.16 and 0.59 for stage W, N1, N2, N3 and R respectively, by using 5% as the training set for DE-PSSS. By using 5% as the training set for CE-PSSS, F-Measures increased 0.13, 0.11, 0.33, 0.16 and 0.57 for stage W, N1, N2, N3 and R respectively.

Figure 7.11 presents the comparison of F-Measure using SF-SSS, DE-PSSS and CE-PSSS. 5% is used as the training set for DE-PSSS and CE-PSSS. The results present that PSSS can reach higher performance than SF-SSS for every stage. For stage N2 and N3, DE-PSSS and CE-PSSS can approximately reach the same F-Measure. While for stage W, N1, and R, DE-PSSS can reach higher F-Measure.

Figure 7.11 – Comparison between SF-SSS and PSSS



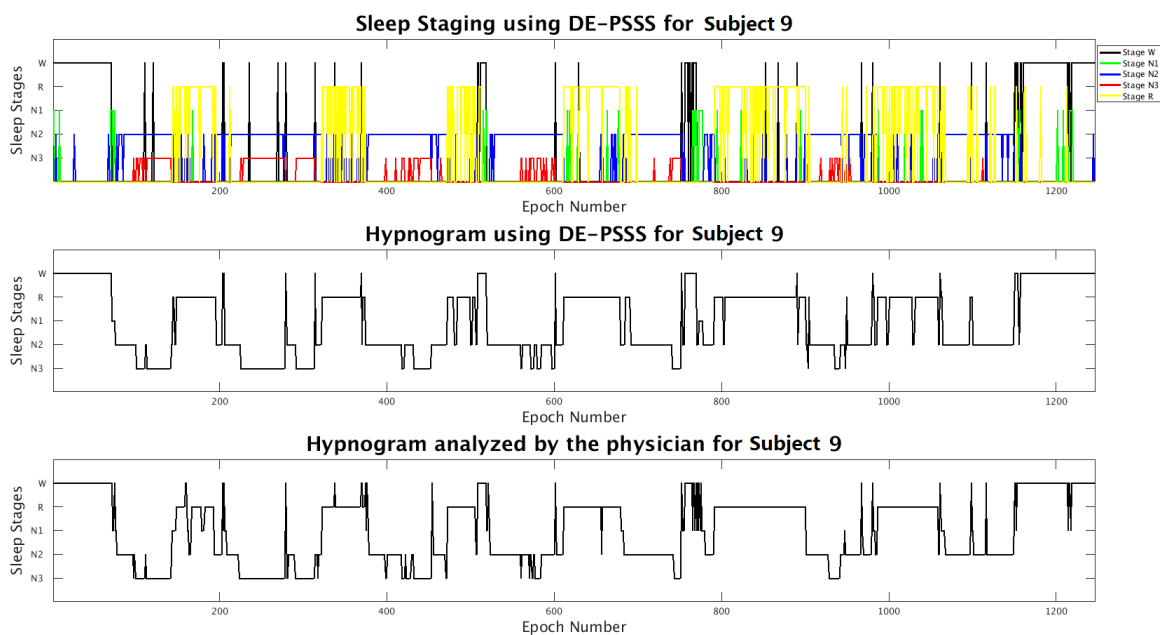
## 7.5 Agreement Rate Comparison between PSSSs and Existing Works

After obtaining the classification results of each stage by using PSSS, Hypnogram is generated according to the writing order of stage N1, R, N3, N2 and W as explained in Section 7.2.0.2. Then Smoothing is applied to detect and correct false sleep transitions for obtaining the final Hypnogram. To evaluate the overall performance of PSSSs, Agreement Rate is adopted.

## 7.6 Agreement Rate of PSSSs

Figure 7.12 shows comparison between the PSSS and the visual analysis by the physician using one patient as an example. Sleep staging results of Subject 9 using DE-PSSS are presented: the above sub-figure is the binary classification of each stage; the middle sub-figure is the Hypnogram after combining and smoothing all the five binary classifications; the bottom sub-figure is the Hypnogram analyzed by the physician. In order to evaluate the overall performance of PSSSs, Agreement Rate is adopted by comparing the Hypnogram generated using PSSS to the Hypnogram analyzed by the physician.

Figure 7.12 – Hypnogram of Subject 9 using DE-PSSS



Appendix (8.16) shows the Agreement Rate and Cohen's Kappa Coefficient for each subject using DE-PSSS and CE-PSSS. For each subject, 5% of the whole epochs are randomly selected as the training set to provide personalized thresholds for scoring all the epochs.

Confusion matrix for the Database 2 using DE-PSSS is presented in Table 7.21 and a heatmap of the same information is shown in Figure 7.13. The ARs for all the stages, except stage N1, are higher than 0.8. The overall AR for five stages also can reach 0.8. Cohen's Kappa shows a substantial agreement (0.7).

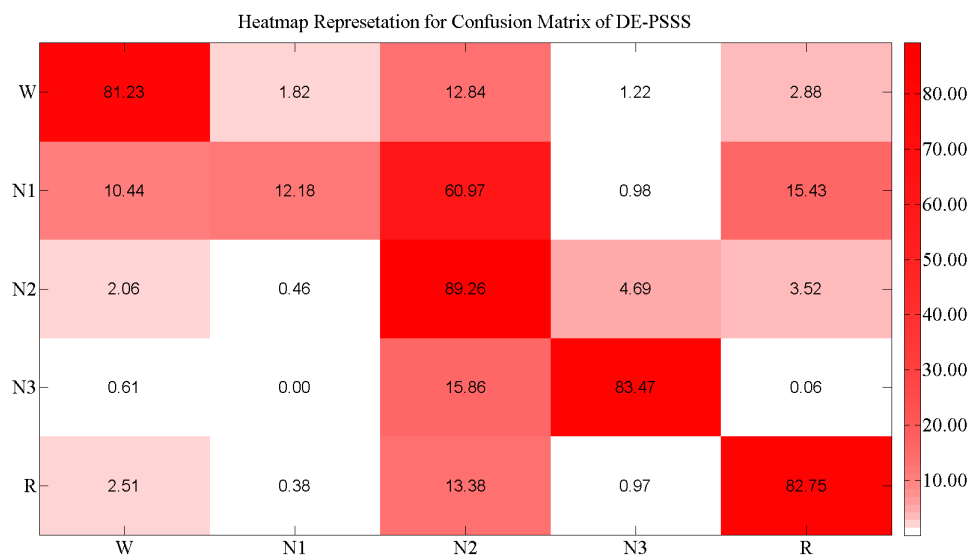


		DE-PSSS						AR
		W	N1	N2	N3	R	Total	
Analysis of the Physician	W	<b>2593</b>	58	410	39	92	3192	0.8123
	N1	138	<b>161</b>	806	13	204	1322	0.1218
	N2	151	34	<b>6541</b>	344	258	7328	0.8926
	N3	19	0	493	<b>2595</b>	2	3109	0.8347
	R	80	12	426	31	<b>2634</b>	3183	0.8275
	Total	2981	265	8676	3022	3190	18134	<b>0.8009</b>
Kappa		0.7224						

Table 7.21 – Confusion Matrix of Database 2 using DE-PSSS

Figure 7.13 shows a simple and clear way to present the confusion matrix of Database 2 using DE-PSSS. The gradient indicates percentage agreement, with red indicating a 100% agreement rate and white at 0%. The gradient legend to the right of the figure shows the percentages that correspond to each gray-scale level. X-axis represents classification results of DE-PSSS and Y-axis shows the analyzed results of the physician.

Figure 7.13 – Heatmap of Confusion Matrix of Database 2 using DE-PSSS



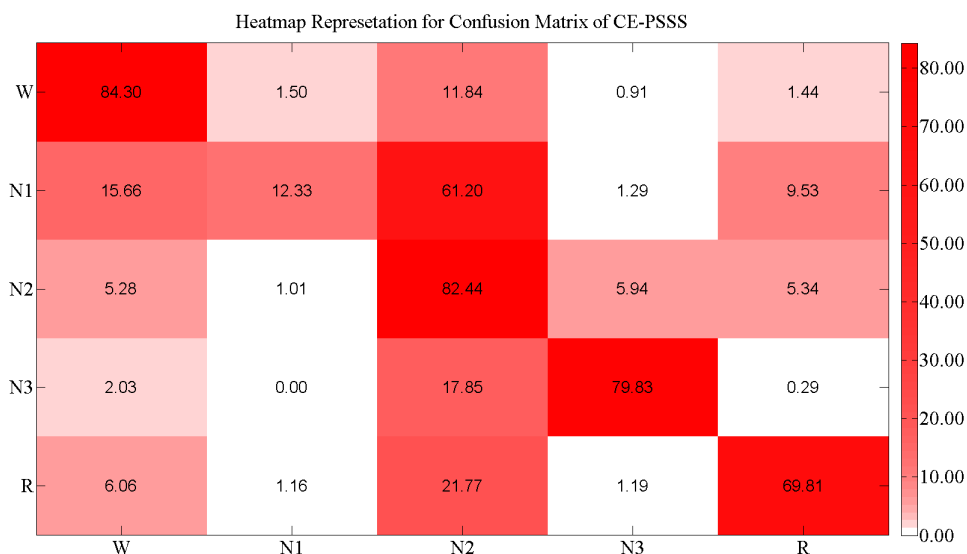
Confusion matrix for the Database 2 using CE-PSSS is also presented, as shown in Table 7.22. Overall AR for Database 2 reaches 0.75. In addition, ARs for stage W and N2 can reach over 0.8. Cohen's Kappa Coefficient for Database 2 is 0.65, which also shows a substantial agreement between CE-PSSS and the analysis from the physician.

		CE-PSSS						AR
		W	N1	N2	N3	R	Total	
Analysis of the Physician	W	2691	48	378	29	46	3192	0.8430
	N1	207	163	809	17	126	1322	0.1233
	N2	387	74	6041	435	391	7328	0.8244
	N3	63	0	555	2482	9	3109	0.7983
	R	193	37	693	38	2222	3183	0.6981
	Total	3541	322	8476	3001	2794	18134	0.7499
Kappa		0.6527						

Table 7.22 – Confusion Matrix of Database 2 using CE-PSSS

Heatmap of confusion matrix of Database 2 using CE-PSSS is shown in Figure 7.14. The gradient indicates percentage agreement, with red indicating a 100% agreement rate and white at 0%. The gradient legend to the right of the figure shows the percentages that correspond to each gray-scale level. X-axis represents classification results of DE-PSSS and Y-axis shows the analyzed results of the physician.

Figure 7.14 – Heatmap of Confusion Matrix of Database 2 using CE-PSSS



### 7.7 Agreement Rate Comparison between PSSSs and Inter-raters

In clinical sleep analysis, inter-raters variability is a major issues for manually sleep staging. Inter-raters agreement for different stages is reported in [66]. Agreement for stage W is among 68%-89%; for stage N1 is among 23%-74%; for stage N2 is among 79%-90%; for stage N3 is around 69%. for stage R is among 78%-94%. Comparisons of PSSSs and inter-rater agreement are presented in Table 7.23.

Sleep Stages	DE-PSSS	CE-PSSS	Inter-raters
<b>W</b>	81.23 %	84.30%	68-89%
<b>N1</b>	12.18%	12.33%	23-74%
<b>N2</b>	89.26%	82.44%	79-90%
<b>N3</b>	83.47%	79.83%	69%
<b>R</b>	82.75 %	69.81%	78-94%

Table 7.23 – Agreement Rate Comparison between PSSSs and Inter-raters

By using DE-PSSS, ARs for stage W, N2 and R are within the inter-raters agreement reported in [66]. For stage N3, AR is higher than inter-raters agreement. However, for stage N1, AR is lower than inter-raters agreement.

For CE-PSSS, ARs for stage W and N2 are within the in-raters agreement. For stage N3, AR is higher than inter-raters agreement. While, for stage N1 and R, AR is relative lower than inter-raters agreement.

## 7.8 Agreement Rate Comparison between PSSSs and Other Works

Methods	Authors	Database	Classification	AR
Decision Tree	Masaaki et al. [25]	1 male training: 80% & testing: 20%	SWS(S3,S4),SREM, S1,S2,S3,S4,MT <sup>6</sup>	82%
Decision Tree	Khai et al. [44]	5 subjects training: NM <sup>7</sup> & testing: NM <sup>7</sup>	W,N1,N2, N3,R,MT <sup>6</sup>	79%
ANN	Nicolas et al. [62]	12-night recordings training: 51% & testing: 49%	W, MT <sup>6</sup> , S1, S2, S3, S4, R	81%
ANN	Nizar et al. [37]	1 male training: NM <sup>7</sup> & testing: NM <sup>7</sup>	W,N1,N2, N3,R	76%
SVM	Steinn et al. [23]	4-night recordings training: 75% & testing: 25%	W, LS(S1,S2), SWS(S3,S4),R	76%
SVM	Antonio et al. [50]	9-night recordings training: 33% & testing: 33%	W,S1,S2 SWS(S3,S4),R	70%
Rule-based Method	Sheng-Fu et al. [45]	16 subjects training: 20% & testing: 80%	W,S1,S2, SWS(S3,S4),R	87%
Rule-based Method	Diego et al. [18]	33 subjects training: NM <sup>7</sup> & testing: 100%	W,N1,N2, N3,R	84%
Hybrid Method	Tarek et al. [40]	15 subjects training: 67% & testing: 33%	W,S1,S2, SWS(S3,S4),R	74%
<b>DE-PSSS</b>	<b>Proposed Method</b>	<b>16 subjects</b> <b>training: 5% &amp; testing: 100%</b>	<b>W,N1,N2,</b> <b>N3,R</b>	<b>80%</b>
<b>CE-PSSS</b>	<b>Proposed Method</b>	<b>16 subjects</b> <b>training: 5% &amp; testing: 100%</b>	<b>W,N1,N2,</b> <b>N3,R</b>	<b>75%</b>

Table 7.24 – Agreement Rate Comparison between PSSSs and Other Works

Table 7.24 lists comparisons among different methods for sleep staging. For DE-PSSS we proposed, it can reach higher AR than most of the existing works while using less

training set. In comparison to rule-based method, [45] and [18] reaches higher AR.

In [45], only young healthy subjects are included. It also used a relative larger training set (20% of the database). Most importantly, in order to reduce the feature variability, for each feature, the mean of maximal 10% data was calculated as the maximum value of the feature, values larger than this maximum value are set to this maximum value. Meanwhile, the mean of minimal 10% data was calculated as the minimum value of the feature, values smaller than this minimum value are set to this minimum value. This process may prevent extremely high or low values for the features, while it may also lose some significant information of the features.

In [18], AR can reach approximate 84 % without mentioning the size of training set they used. It implemented 111 fuzzy rules which are derived from available medical knowledge and author's experience. However, in our work, 13 rules inspired from the AASM manual are implemented. With the increment of rules, it may potentially improve the performance and it also increased the complexity. In further work, more rules will be involved to complete our model.

## 7.9 Conclusion

In this Chapter, evaluation of SF-SSS with modifications is analyzed. F-Measure, Agreement Rate and Cohen's Kappa are used to assess the performance of SF-SSS with modifications. F-Measure, Agreement Rate and Cohen's Kappa are improved by adding modifications for SF-SSS. Meanwhile, evaluations on ATSU by analyzing the selection of control parameters for Differential Evolution and Cross Entropy are also presented. Followed by evaluations on two PSSSs (DE-PSSS and CE-PSSS). Compare SF-SSS with PSSS, PSSS can reach higher performance by using ATSU to set-up thresholds automatically.

## Chapter 8

# Conclusions and Perspectives

### 8.1 Conclusions

In this thesis, a personalized automatic sleep staging system was proposed by combining symbolic intelligence and feedback system control technique. Symbolic fusion is dedicated to mimic decision-making process of clinical sleep staging. It starts from the extraction of digital parameters from raw polysomnography signals and goes up to high-level the symbolic interpretation of feature parameters. At last, decision is generated using rules inspired by international guidelines in sleep medicine. Meanwhile, Feedback System Control technique is designed to provide optimal thresholds for symbolic fusion in transforming digital parameters into symbols while taking individual variability into consideration.

In this thesis, it composed of eight main chapters. The overall description of the first seven chapters is shown below.

In Chapter 1, a brief introduction on sleep staging is presented and the major issues of existing automatic sleep staging systems for the real clinical practical use are discussed. From the physician's perceptive, the major issues can be described as

- A system without taking any medical knowledge into consideration cannot win fully trust from physicians;
- Most of systems are not yet accepted and validated by physicians.

Only by addressing these issues, a system can really be used in clinical practice for help physicians in diagnosis and treatment of sleep disorders.

In Chapter 2, Issues on existing automatic sleep stagings systems were discussed. In consideration of real clinical practical use for physicians, a Symbolic Fusion-based Sleep Staging System (SF-SSS) was proposed by Ugon, Isabelle, et al. It is dedicated to mimic clinical sleep staging process by translating AASM and medical knowledge into computer logic. SF-SSS model attempts to realize sleep staging by using symbolic fusion. However, it is only a proof from conception to validation of symbolic fusion for sleep staging application. This model still can be improved. Thus, several issues need further research for this SF-SSS model:

- How to realize automatic sleep staging by taking individual variability into consideration, from the conception to implementation, validation and final practical use?
- How to assess the effectiveness of personalized sleep staging system?
- Is it necessary to propose a personalized sleep staging system which can take individual variability into consideration? Or a generalized sleep staging system is sufficient?

In Chapter 3, information and terminology for understanding the field of sleep analysis are presented at the beginning. It mainly introduces polysomnography (PSG) signals, different sleep stages, manuals of clinical sleep staging. Meanwhile, existing sleep staging methods are analyzed involving machine learning methods, hybrid methods and symbolic fusion method. In the comparison of existing methods, symbolic fusion method owns several advantages like, it is efficient to fuse information from different sources by considering own limitations and uncertain perceptions from every single source and it can mimic clinical sleep staging process under the guidance of AASM manual. However, either symbolic fusion method or most of the existing methods, there exist two problems:

- To our knowledge, there is no personalized sleep staging system which can take individual variability into consideration.
- Among existing sleep staging systems, thresholds are widely applied in transforming digital parameters into linguistic or symbolic features to model inference process under the guidance of medical knowledge.

As far as we know, there is no fully satisfying automatic method to set up these thresholds instead of manual setting up.

In order to build a threshold function or model to set up thresholds directly, search an appropriate threshold combination among possible combinations space method is inspired by FSC which is proposed as a personalized medicine platform in searching optimal drug combinations for a different patient. Different stochastic search algorithms which can be applied within FSC are analyzed in this chapter. Differential Evolution and Cross Entropy are more suitable to sleep staging in comparison of different properties among all researched search algorithms.

In Chapter 4, an existing Symbolic Fusion-based Sleep Staging System (SF-SSS) is introduced at the beginning. However, there exist several limitations in existing SF-SSS like, pre-processing of PSG signals was not included, smoothing of sleep stages was not included, not all the parameters described by AASM manual were implemented and classification of stage N1 was not performed. To overcome limitations of existing SF-SSS, modifications are proposed in this chapter:

- Band-pass filters recommended by AASM and band-stop filter designed to reject power line impact are applied to raw PSG signals for eliminating noise and artifacts.
- In order to consider temporal effects and sleep transitions of sleep staging, smoothing is proposed. In smoothing, not only temporal contextual information but also sleep transitions detection and correction are performed.
- A significant indicator of stage N2, K-Complex is extracted and fuse to existing SF-SSS in order to improve the classification of stage N2.
- Rules to classify stage N1 by using existing digital parameters is proposed.

In Chapter 5, thresholds used in SF-SSS to transform digital parameters into a symbolic interpretation of feature parameters are introduced. Manual Thresholds Setting-Up (MTSU) method used in existing SF-SSS are described by the analysis of its limitations. Then, major issues on thresholds are discussed and thresholds dependencies (thresholds & sleep stage, thresholds & patients) are discussed for SF-SSS model. Details of thresholds that need to be set-up for SF-SSS model are presented by taking thresholds dependencies among sleep stages into consideration.

In Chapter 6, an Automatic Thresholds Setting-Up (ATSU) method based on Feedback System Control technique is proposed to overcome the limitations of MTSU. Instead of

building a threshold function or model, ATSU provides an idea of finding optimal thresholds combination within possible thresholds combination space. By combining symbolic fusion and feedback system control, a hybrid expert system conception for sleep staging is proposed. Based on this conception, personalized sleep staging system is presented. In personalized sleep staging system, several epochs will be selected and analyzed by physicians. With partial Hypnogram that analyzed by the physicians of selected epochs, personalized thresholds are set up using ATSU. Then, these thresholds are used in scoring all epochs and full Hypnogram can be generated at the end. In this chapter, Differential Evolution-based Personalized Sleep Staging System (DE-PSSS) and Cross Entropy-based Personalized Sleep Staging System (CE-PSSS) are presented by using two different search algorithms Differential Evolution and Cross Entropy respectively.

In Chapter 7, SF-SSS with modifications, ATSU, DE-PSSS and CE-PSSS are evaluated. In comparison existing SF-SSS and SF-SSS with modifications, overall agreement rate is improved from 57.15% to 64.22%, especially for stage N2 which is improved from 53.62% to 65.06%. SF-SSS with proposed modifications can reach higher accuracy than existing SF-SSS. In the analysis of ATSU, impacts of control parameters for Differential Evolution and Cross Entropy are evaluated firstly,  $NP=5D$  ( $D$  is the number of thresholds),  $F=0.5$ ,  $CR=0.9$  and  $N=500$ , Elite Sampling Rate= $0.1$  are the optimal controls parameters for DE and CE respectively. To evaluate the DE-PSSS and CE-PSSS, the selection of training set are studied. From the physician's perspective, 5 % can be considered as the optimal training set for DE-PSSS and CE-PSSS which consumes less time for physicians to score and still can ensure the performance of the classification. However, in consideration of reaching high performance, 15 % can be considered as the optimal training set for DE-PSSS and CE-PSSS. In the comparison of modified SF-SSS which using MTSU method with DE-PSSS and CE-PSSS, DE-PSSS and CE-PSSS can reach higher F-Measure and can take individual variability into consideration. Compare DE-PSSS and CE-PSSS, DE-PSSS is much better which can reach higher F-Measure for all stage using the same size of the training set and consume less time.

Overall, the main contributions of this thesis are described as below:

- Improved the existing SF-SSS.

Existing SF-SSS model is the first attempt to realize sleep staging by using symbolic fusion. In this thesis, with the analysis of existing SF-SSS model, modifications are proposed in order to enrich the existing SF-SSS. With proposed modifications, SF-SSS can reach higher performance in sleep staging.

- Proposed an automatic thresholds setting-up method.

In SF-SSS model, thresholds are used to transform low-level digital parameter into a high-level symbolic interpretation of feature parameters. Details of thresholds dependencies are analyzed and thresholds configuration are studied in this thesis. Then an automatic thresholds setting-up method is proposed to overcome the limitations of manual thresholds setting-up method.

- Proposed a hybrid expert system conception for sleep staging application.

A hybrid expert system conception is proposed by combining symbolic fusion and feedback system control technique. Only several epochs are needed to be selected and analyzed by a physician. Then thresholds can be set up based on the selected epochs and partial Hypnogram that analyzed by the physician using feedback system control technique. Finally, these thresholds are used in scoring the full epochs by symbolic fusion and full Hypnogram can be generated at the end.

- Proposed a personalized automatic sleep staging system.

In consideration of individual variability, personalized sleep staging system based on the hybrid expert system conception is proposed. By selecting several epochs from different person and set up personalized thresholds, personalized automatic sleep staging can be realized by combining symbolic fusion and the personalized thresholds. The performance of the proposed personalized sleep staging system has also been evaluated: in comparison with SF-SSS model using manual thresholds setting-up methods, personalized sleep staging system can reach much better performance in terms of F-Measure.

## 8.2 Perspectives

In this thesis, a hybrid expert system conception to realize automatic sleep staging by combining symbolic fusion and feedback system control is proposed. Based on this conception, personalized sleep staging system is implemented and evaluated. While, in further work, the following points can be also researched:

- Based on the proposed hybrid expert system conception, generalized sleep staging system can be also researched to fully understand whether a generalized classification is sufficient, or whether a personalized classification is required to take individual variability into consideration.
- After obtaining the binary classification results of each stage, an effective method to generate Hypnogram can also be researched.
- To complete the existing SF-SSS, more sleep events and rules described by AASM can be involved.



# Publications

1. Chen, C., Liu, X., Ugon, A., Zhang, X., Amara, A., Garda, P., Ganascia, J. G., Philippe, C., & Pinna, A. (2016, May). Polysomnography Symbolic Fusion for Automatic Sleep Staging. In 5èmes Journées d'Etude sur la TéléSANTé (JETSAN).
2. Chen, C., Liu, X., Ugon, A., Zhang, X., Amara, A., Garda, P., Ganascia, J. G., Philippe, C., & Pinna, A. (2015, December). Symbolic Fusion: A Novel Decision Support Algorithm for Sleep Staging Application. In Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare (pp. 19-22). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
3. Chen, C., Ugon, A., Zhang, X., Amara, A., Garda, P., Ganascia, J. G., Philippe, C., & Pinna, A. (2016, October). Personalized sleep staging system using evolutionary algorithm and symbolic fusion. In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the (pp. 2266-2269). IEEE.
4. Chen, C., Ugon, A., Zhang, X., Amara, A., Garda, P., Ganascia, J. G., Kotti, A., Philippe, C., & Pinna, A. (2016, October). In Biomedical Circuits and Systems, 2016 12th IEEE BioCAS. IEEE.

# Appendix

## 8.3 Appendix 1

ID	Recall of Existing SF-SSS					Recall of SF-SSS (with modifications)				
	W	N1	N2	N3	R	W	N1	N2	N3	R
627	0.8374	0.0000	0.5483	0.9766	0.7245	0.8558	0.2037	0.8512	0.9766	0.7245
639	0.7500	0.0000	0.0380	0.9751	0.8133	0.7500	0.3013	0.1876	0.9751	0.8133
1818	0.6696	0.0000	0.0590	0.6854	0.8889	0.6696	0.0000	0.8984	0.6854	0.8889
3774	0.8087	0.0000	0.8949	0.9493	0.9175	0.8108	0.3636	0.9150	0.9493	0.9175
3870	0.7614	0.0000	0.1994	0.8058	0.9020	0.7614	0.4271	0.3263	0.8058	0.9020
3875	0.9058	0.0000	0.1416	0.7677	0.9380	0.9058	0.6250	0.6022	0.7677	0.9380
3928	0.4306	0.0000	0.5137	0.9074	0.4286	0.4306	0.1765	0.5167	0.9074	0.4286
4062	0.6019	0.0000	0.0510	0.8522	0.7269	0.6019	0.0915	0.0737	0.8522	0.7269
4064	0.7143	0.0000	0.5588	0.7992	0.1711	0.7143	0.0354	0.7974	0.7992	0.1711
45813	0.6680	0.0000	0.2290	0.6064	0.0000	0.7747	0.0000	0.2311	0.6064	0.0000
46618	0.9739	0.0000	0.1318	0.9352	0.7717	0.9739	0.1140	0.2273	0.9352	0.7717
48814	0.1781	0.0000	0.2625	0.8539	0.5867	0.2055	0.1667	0.3607	0.8539	0.5867
50062	0.6111	0.0000	0.3175	0.8203	0.7190	0.6111	0.1613	0.3175	0.8203	0.7190
50897	0.7500	0.0000	0.6621	0.8000	0.8056	0.7500	0.3043	0.6621	0.8000	0.8056
55341	0.8037	0.0000	0.3929	0.9339	0.9504	0.8037	0.2556	0.4660	0.9339	0.9504
56508	0.5000	0.0000	0.7286	0.8247	0.7835	0.5079	0.4103	0.7668	0.8247	0.7835

Table 8.1 – Recall Comparison between SF-SSS (with modifications) and existing SF-SSS on Database 1

## 8.4 Appendix 2

ID	Precision of Existing SF-SSS					Precision of SF-SSS (with modifications)				
	W	N1	N2	N3	R	W	N1	N2	N3	R
627	0.7822	0.0000	0.8015	0.7257	0.1655	0.7772	0.1146	0.7392	0.7257	0.1655
639	0.4737	0.0000	0.1928	0.7417	0.3971	0.4737	0.4395	0.3726	0.7417	0.3971
1818	0.9375	0.0000	0.4615	0.5379	0.1642	0.9146	0.0000	0.5475	0.5379	0.1642
3774	0.9695	0.0000	0.5457	0.5928	0.2373	0.9696	0.2857	0.4881	0.5928	0.2373
3870	0.7053	0.0000	0.4430	0.7698	0.3506	0.7053	0.2291	0.4779	0.7698	0.3506
3875	0.9870	0.0000	0.8630	0.7628	0.1709	0.9870	0.0181	0.4054	0.7628	0.1709
3928	0.7623	0.0000	0.5541	0.7101	0.2328	0.7623	0.2051	0.5363	0.7101	0.2328
4062	0.6217	0.0000	0.3396	0.6577	0.6157	0.6217	0.0787	0.3514	0.6577	0.6157
4064	0.6962	0.0000	0.3167	0.7089	0.7442	0.6962	0.1795	0.3284	0.7089	0.7442
45813	0.3627	0.0000	0.5266	0.2938	0.0000	0.3570	0.0000	0.5263	0.2938	0.0000
46618	0.3085	0.0000	0.3021	0.7108	0.4949	0.3085	0.1902	0.3497	0.7108	0.4949
48814	0.5000	0.0000	0.7617	0.8216	0.3235	0.5000	0.0884	0.7832	0.8216	0.3235
50062	0.8713	0.0000	0.4462	0.7653	0.6627	0.8713	0.1049	0.4439	0.7653	0.6627
50897	0.9634	0.0000	0.5344	0.6228	0.4158	0.9634	0.0729	0.5344	0.6228	0.4158
55341	0.8444	0.0000	0.4921	0.5963	0.4007	0.8444	0.1753	0.5096	0.5963	0.4007
56508	0.9130	0.0000	0.7207	0.6299	0.6084	0.9143	0.0424	0.7076	0.6299	0.6084

Table 8.2 – Precision Comparison between SF-SSS (with modifications) and existing SF-SSS on Database 1

## 8.5 Appendix 3

ID	AR W	AR N1	AR N2	AR N3	AR R	Total AR
627	0.8374	0	0.5692	0.8505	0.5816	0.6791
639	0.7500	0	0.2660	0.9626	0.7892	0.5283
1818	0.6696	0	0.0885	0.6262	0.8889	0.3629
3774	0.8087	0	0.9306	0.0000	0.8969	0.7093
3870	0.7614	0	0.2024	0.7698	0.9020	0.5052
3875	0.9058	0	0.1483	0.7613	0.9302	0.7071
3928	0.4306	0	0.7264	0.6296	0.3810	0.5055
4062	0.6019	0	0.6487	0.8348	0.6574	0.5699
4064	0.7143	0	0.9575	0.1120	0.1123	0.4103
45813	0.6680	0	0.4034	0.2553	0.0000	0.3819
46618	0.9739	0	0.3045	0.6883	0.7323	0.4506
48814	0.1781	0	0.7230	0.8371	0.5733	0.6336
50062	0.6111	0	0.6679	0.6702	0.6863	0.6095
50897	0.7500	0	0.6923	0.3911	0.7708	0.6421
55341	0.8037	0	0.6121	0.5248	0.9256	0.6010
56508	0.5000	0	0.7632	0.2629	0.7778	0.6402

Table 8.3 – Agreement Rate of existing SF-SSS on Database 1

## 8.6 Appendix 4

ID	AR W	AR N1	AR N2	AR N3	AR R	Total AR
627	0.9540	0.0000	0.8799	0.9252	0.5306	0.8353
639	0.7500	0.0000	0.1971	0.9875	0.8223	0.5207
1818	0.6071	0.0000	0.9721	0.3302	0.0000	0.6297
3774	0.8586	0.0000	0.9687	0.0000	0.9691	0.7492
3870	0.7727	0.0000	0.3414	0.7914	0.9608	0.5647
3875	0.9323	0.0000	0.6787	0.7355	0.9612	0.8519
3928	0.6296	0.0000	0.6991	0.7037	0.4286	0.5596
4062	0.9140	0.0000	0.2833	0.9304	0.8704	0.5934
4064	0.8571	0.0000	0.9804	0.0000	0.0481	0.4203
45813	0.9842	0.0000	0.2500	0.2660	0.0000	0.3899
46618	0.9739	0.0331	0.2864	0.7328	0.9213	0.4913
48814	0.2466	0.0000	0.6667	0.8989	0.7667	0.6427
50062	0.7847	0.0000	0.4818	0.8309	1.0000	0.6957
50897	0.8333	0.0000	0.7967	0.3511	0.8681	0.7077
55341	0.8185	0.0376	0.6398	0.6198	0.9504	0.6406
56508	0.6190	0.0256	0.8798	0.2113	0.9145	0.7339

Table 8.4 – Agreement Rates of SF-SSS (with modifications) on Database 1

## 8.7 Appendix 5

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7490	0.1973	0.6038	0.8007	0.8056	0.0338	0.0344	0.0194	0.0221	0.0313
3	0.7402	0.1052	0.7616	0.7277	0.8155	0.0191	0.0189	0.0104	0.0173	0.0262
4	0.5701	0.2972	0.8318	0.8149	0.6865	0.0391	0.0402	0.0219	0.0191	0.0255
5	0.4320	0.4581	0.5767	0.7795	0.6667	0.0478	0.0368	0.0281	0.0154	0.0434
7	0.7692	0.3123	0.7486	0.6813	0.8044	0.0262	0.0405	0.0078	0.0221	0.0237
8	0.7024	0.1121	0.7980	0.6905	0.6100	0.0318	0.0204	0.0183	0.0294	0.0381
9	0.8924	0.0946	0.6856	0.8327	0.8104	0.0205	0.0221	0.0199	0.0219	0.0172
10	0.5957	0.2560	0.6829	0.7979	0.7493	0.0346	0.0594	0.0189	0.0168	0.0227
11	0.7794	0.1984	0.7932	0.8673	0.6556	0.0496	0.0400	0.0204	0.0130	0.0357
12	0.8628	0.2032	0.7628	0.7569	0.7584	0.0292	0.0387	0.0165	0.0245	0.0196
13	0.7296	0.1684	0.7630	0.7821	0.7034	0.0234	0.0277	0.0136	0.0140	0.0262
18	0.5804	0.3370	0.7102	0.6932	0.4401	0.0223	0.0446	0.0108	0.0208	0.0347
29	0.7179	0.1154	0.7984	0.8488	0.8362	0.0254	0.0229	0.0175	0.0227	0.0225
41	0.8782	0.0691	0.6643	0.6376	0.3138	0.0051	0.0240	0.0200	0.0296	0.0514
49	0.8538	0.4418	0.6596	0.4281	0.1874	0.0129	0.0343	0.0155	0.0896	0.0475
63	0.5662	0.1684	0.7254	0.6419	0.6599	0.0120	0.0312	0.0077	0.0222	0.0469
Mean	<b>0.7137</b>	<b>0.2209</b>	<b>0.7229</b>	<b>0.7363</b>	<b>0.6565</b>	0.0271	0.0335	0.0167	0.0250	0.0320

Table 8.5 – F-Measures of Evaluation Set using 5% Training Set (DE-PSSS) on Database 2

## 8.8 Appendix 6

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7691	0.1924	0.6129	0.8127	0.8395	0.0116	0.0297	0.0124	0.0198	0.0204
3	0.7600	0.1099	0.7688	0.7435	0.8514	0.0153	0.0163	0.0054	0.0133	0.0126
4	0.6017	0.3226	0.8509	0.8159	0.7155	0.0136	0.0239	0.0112	0.0148	0.0176
5	0.4663	0.4787	0.6066	0.7949	0.7146	0.0296	0.0213	0.0199	0.0147	0.0214
7	0.7718	0.3232	0.7499	0.6973	0.8179	0.0239	0.0350	0.0059	0.0107	0.0215
8	0.7270	0.1050	0.8113	0.7044	0.6379	0.0213	0.0193	0.0147	0.0237	0.0186
9	0.9213	0.1150	0.7084	0.8464	0.8292	0.0122	0.0297	0.0054	0.0089	0.0114
10	0.6388	0.2884	0.6901	0.8134	0.7627	0.0566	0.0321	0.0204	0.0134	0.0173
11	0.8162	0.1984	0.8036	0.8802	0.7027	0.0209	0.0606	0.0124	0.0113	0.0245
12	0.8743	0.1971	0.7690	0.7633	0.7736	0.0271	0.0482	0.0191	0.0186	0.0214
13	0.7537	0.1280	0.7842	0.7924	0.7206	0.0133	0.0300	0.0100	0.0153	0.0286
18	0.5856	0.3507	0.7139	0.7168	0.4702	0.0199	0.0241	0.0060	0.0212	0.0305
29	0.7296	0.1428	0.8061	0.8775	0.8582	0.0319	0.0323	0.0141	0.0204	0.0146
41	0.8828	0.0804	0.6819	0.6228	0.3146	0.0030	0.0269	0.0136	0.0300	0.0576
49	0.8655	0.4673	0.6747	0.4363	0.2063	0.0074	0.0166	0.0143	0.0843	0.0556
63	0.5730	0.1790	0.7345	0.6614	0.6936	0.0155	0.0235	0.0034	0.0156	0.0169
Mean	<b>0.7335</b>	<b>0.2299</b>	<b>0.7354</b>	<b>0.7487</b>	<b>0.6818</b>	0.0202	0.0293	0.0118	0.0210	0.0244

Table 8.6 – F-Measures of Evaluation Set using 10% Training Set (DE-PSSS) on Database 2

## 8.9 Appendix 7

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7697	0.1954	0.6143	0.8147	0.8566	0.0120	0.0329	0.0105	0.0193	0.0078
3	0.7707	0.1258	0.7696	0.7515	0.8526	0.0161	0.0570	0.0043	0.0120	0.0155
4	0.6078	0.3250	0.8529	0.8225	0.7131	0.0176	0.0303	0.0129	0.0122	0.0145
5	0.4867	0.4959	0.6031	0.8044	0.7213	0.0247	0.0184	0.0186	0.0132	0.0217
7	0.7774	0.3402	0.7538	0.6969	0.8234	0.0209	0.0441	0.0039	0.0103	0.0176
8	0.7436	0.1171	0.8120	0.7102	0.6328	0.0148	0.0245	0.0172	0.0204	0.0245
9	0.9273	0.1096	0.7056	0.8489	0.8235	0.0112	0.0265	0.0082	0.0089	0.0142
10	0.6456	0.3046	0.6965	0.8171	0.7719	0.0518	0.0347	0.0184	0.0113	0.0186
11	0.8343	0.2227	0.8078	0.8764	0.7092	0.0198	0.0586	0.0114	0.0093	0.0224
12	0.8861	0.2248	0.7789	0.7746	0.7825	0.0183	0.0508	0.0088	0.0151	0.0097
13	0.7518	0.1304	0.7853	0.7964	0.7312	0.0141	0.0329	0.0122	0.0126	0.0267
18	0.5859	0.3775	0.7164	0.7130	0.4900	0.0151	0.0259	0.0072	0.0230	0.0239
29	0.7568	0.1373	0.8088	0.8828	0.8647	0.0298	0.0369	0.0158	0.0130	0.0107
41	0.8835	0.0957	0.6895	0.6451	0.3629	0.0039	0.0274	0.0110	0.0337	0.0589
49	0.8670	0.4651	0.6820	0.4810	0.2536	0.0053	0.0126	0.0121	0.0548	0.0700
63	0.5628	0.1895	0.7321	0.6737	0.7016	0.0188	0.0265	0.0057	0.0105	0.0149
Mean	<b>0.7411</b>	<b>0.2410</b>	<b>0.7380</b>	<b>0.7568</b>	<b>0.6932</b>	0.0184	0.0337	0.0111	0.0175	0.0232

Table 8.7 – F-Measures of Evaluation Set using 15% Training Set (DE-PSSS) on Database 2

## 8.10 Appendix 8

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7679	0.1996	0.6219	0.8206	0.8535	0.0132	0.0498	0.0132	0.0149	0.0150
3	0.7772	0.1494	0.7722	0.7490	0.8503	0.0139	0.0280	0.0036	0.0139	0.0122
4	0.6196	0.3023	0.8565	0.8222	0.7224	0.0112	0.0308	0.0083	0.0111	0.0159
5	0.4785	0.4901	0.6069	0.8041	0.7321	0.0166	0.0177	0.0223	0.0056	0.0222
7	0.7936	0.3453	0.7553	0.7066	0.8392	0.0207	0.0344	0.0049	0.0122	0.0205
8	0.7528	0.1248	0.8163	0.7171	0.6418	0.0133	0.0231	0.0145	0.0177	0.0141
9	0.9368	0.1095	0.7076	0.8496	0.8342	0.0086	0.0252	0.0079	0.0082	0.0080
10	0.6418	0.2949	0.7001	0.8167	0.7736	0.0488	0.0340	0.0147	0.0101	0.0150
11	0.8379	0.2376	0.8120	0.8818	0.7228	0.0114	0.0404	0.0120	0.0055	0.0193
12	0.8822	0.2258	0.7779	0.7910	0.7813	0.0124	0.0263	0.0113	0.0199	0.0155
13	0.7541	0.1306	0.7946	0.8019	0.7477	0.0153	0.0333	0.0086	0.0117	0.0143
18	0.5978	0.3834	0.7195	0.7342	0.4853	0.0149	0.0247	0.0060	0.0188	0.0284
29	0.7641	0.1502	0.8198	0.8888	0.8665	0.0225	0.0411	0.0130	0.0101	0.0087
41	0.8833	0.0877	0.6963	0.6591	0.3385	0.0035	0.0263	0.0074	0.0240	0.0504
49	0.8666	0.4790	0.6794	0.4827	0.2435	0.0079	0.0138	0.0102	0.0429	0.0569
63	0.5689	0.2140	0.7346	0.6718	0.6950	0.0151	0.0234	0.0047	0.0133	0.0145
Mean	<b>0.7452</b>	<b>0.2453</b>	<b>0.7419</b>	<b>0.7623</b>	<b>0.6955</b>	0.0156	0.0295	0.0102	0.0150	0.0207

Table 8.8 – F-Measures of Evaluation Set using 20% Training Set (DE-PSSS) on Database 2

## 8.11 Appendix 9

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7479	0.0795	0.6065	0.7942	0.8192	0.0105	0.0381	0.0219	0.0093	0.0278
3	0.6724	0.0547	0.7535	0.7389	0.8023	0.0496	0.0388	0.0090	0.0151	0.0314
4	0.5281	0.2398	0.8516	0.8076	0.6807	0.0218	0.0571	0.0107	0.0123	0.0184
5	0.3511	0.4123	0.5990	0.8046	0.5942	0.0215	0.0458	0.0253	0.0077	0.0426
7	0.7345	0.2055	0.7536	0.6816	0.7589	0.0095	0.0601	0.0066	0.0162	0.0247
8	0.6608	0.0404	0.7841	0.7052	0.5965	0.0180	0.0365	0.0135	0.0179	0.0508
9	0.8646	0.0645	0.6907	0.8304	0.8104	0.0288	0.0257	0.0135	0.0180	0.0155
10	0.5373	0.1555	0.6885	0.8014	0.7375	0.0354	0.0452	0.0161	0.0138	0.0217
11	0.6698	0.1582	0.7816	0.8701	0.6673	0.0497	0.0588	0.0109	0.0115	0.0279
12	0.7619	0.1135	0.7708	0.7547	0.7576	0.0344	0.0456	0.0106	0.0139	0.0220
13	0.6693	0.1038	0.7659	0.7923	0.6692	0.0281	0.0320	0.0123	0.0143	0.0410
18	0.5573	0.2224	0.6953	0.6959	0.4043	0.0153	0.0480	0.0114	0.0217	0.0491
29	0.6043	0.0375	0.7966	0.8497	0.7244	0.0562	0.0242	0.0119	0.0233	0.0584
41	0.8759	0.0337	0.6064	0.6045	0.2159	0.0047	0.0310	0.0383	0.0314	0.0730
49	0.8542	0.4479	0.6479	0.4238	0.2292	0.0073	0.0240	0.0223	0.0799	0.0605
63	0.5532	0.1364	0.7183	0.6609	0.6540	0.0275	0.0417	0.0107	0.0142	0.0292
Mean	<b>0.6652</b>	<b>0.1566</b>	<b>0.7194</b>	<b>0.7385</b>	<b>0.6326</b>	0.0261	0.0408	0.0153	0.0200	0.0371

Table 8.9 – F-Measures of Evaluation Set using 5% Training Set (CE-PSSS) on Database 2

## 8.12 Appendix 10

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7642	0.1229	0.6159	0.7942	0.8454	0.0173	0.0551	0.0158	0.0091	0.0121
3	0.7133	0.0368	0.7661	0.7430	0.8414	0.0287	0.0423	0.0055	0.0107	0.0114
4	0.5391	0.2753	0.8539	0.8083	0.7037	0.0194	0.0218	0.0061	0.0152	0.0274
5	0.3655	0.4385	0.6058	0.8062	0.6313	0.0279	0.0284	0.0189	0.0096	0.0316
7	0.7391	0.2567	0.7567	0.6889	0.7931	0.0120	0.0392	0.0045	0.0148	0.0185
8	0.6781	0.0146	0.7984	0.7163	0.6247	0.0223	0.0209	0.0131	0.0175	0.0177
9	0.8996	0.0450	0.6995	0.8487	0.8162	0.0201	0.0246	0.0085	0.0123	0.0142
10	0.5779	0.1414	0.6933	0.7967	0.7597	0.0342	0.0442	0.0150	0.0147	0.0225
11	0.7140	0.1550	0.7808	0.8748	0.7027	0.0207	0.0347	0.0085	0.0101	0.0258
12	0.8077	0.1228	0.7727	0.7606	0.7718	0.0262	0.0613	0.0078	0.0136	0.0128
13	0.7023	0.1116	0.7878	0.7958	0.7168	0.0236	0.0442	0.0113	0.0094	0.0287
18	0.5760	0.2569	0.7004	0.7077	0.4483	0.0228	0.0305	0.0109	0.0142	0.0267
29	0.6398	0.0365	0.8085	0.8685	0.7805	0.0345	0.0343	0.0116	0.0195	0.0413
41	0.8789	0.0687	0.6567	0.6129	0.2442	0.0033	0.0254	0.0179	0.0284	0.0678
49	0.8608	0.4574	0.6688	0.4511	0.2572	0.0076	0.0219	0.0114	0.0546	0.0650
63	0.5581	0.1529	0.7300	0.6685	0.6734	0.0123	0.0292	0.0070	0.0120	0.0215
Mean	<b>0.6884</b>	<b>0.1683</b>	<b>0.7309</b>	<b>0.7464</b>	<b>0.6632</b>	0.0208	0.0349	0.0109	0.0166	0.0278

Table 8.10 – F-Measures of Evaluation Set using 10% Training Set (CE-PSSS) on Database 2

## 8.13 Appendix 11

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7714	0.1223	0.6225	0.7963	0.8536	0.0086	0.0529	0.0137	0.0082	0.0121
3	0.7293	0.0605	0.7665	0.7541	0.8461	0.0142	0.0424	0.0061	0.0126	0.0178
4	0.5698	0.2698	0.8553	0.8239	0.7090	0.0144	0.0319	0.0078	0.0132	0.0214
5	0.3759	0.4497	0.6234	0.8089	0.6366	0.0260	0.0243	0.0162	0.0084	0.0384
7	0.7493	0.2612	0.7576	0.6924	0.8065	0.0104	0.0381	0.0036	0.0094	0.0224
8	0.6884	0.0281	0.7965	0.7191	0.6431	0.0210	0.0262	0.0127	0.0138	0.0153
9	0.9147	0.0700	0.7074	0.8503	0.8166	0.0110	0.0217	0.0063	0.0109	0.0125
10	0.5884	0.1440	0.7040	0.8081	0.7649	0.0349	0.0460	0.0125	0.0108	0.0137
11	0.7424	0.1527	0.7903	0.8806	0.7100	0.0190	0.0802	0.0110	0.0103	0.0233
12	0.8111	0.2109	0.7765	0.7592	0.7781	0.0272	0.0380	0.0078	0.0156	0.0144
13	0.7130	0.0968	0.7955	0.7929	0.7308	0.0242	0.0489	0.0091	0.0088	0.0233
18	0.5889	0.2899	0.7057	0.7113	0.4710	0.0194	0.0336	0.0071	0.0142	0.0282
29	0.6377	0.0341	0.8133	0.8801	0.8141	0.0380	0.0439	0.0090	0.0142	0.0361
41	0.8769	0.0773	0.6662	0.6290	0.3059	0.0045	0.0245	0.0202	0.0316	0.0478
49	0.8658	0.4578	0.6745	0.4568	0.2367	0.0056	0.0131	0.0089	0.0418	0.0538
63	0.5599	0.1618	0.7319	0.6708	0.6860	0.0175	0.0239	0.0056	0.0103	0.0139
Mean	<b>0.6989</b>	<b>0.1804</b>	<b>0.7367</b>	<b>0.7521</b>	<b>0.6756</b>	0.0185	0.0369	0.0098	0.0146	0.0246

Table 8.11 – F-Measures of Evaluation Set using 15% Training Set (CE-PSSS) on Database 2

## 8.14 Appendix 12

ID	Mean of F-Measure					Std of F-Measure				
	W	N1	N2	N3	R	W	N1	N2	N3	R
1	0.7698	0.0913	0.6216	0.7963	0.8509	0.0139	0.0522	0.0159	0.0124	0.0052
3	0.7436	0.0559	0.7721	0.7490	0.8508	0.0194	0.0421	0.0036	0.0118	0.0149
4	0.5739	0.2736	0.8607	0.8246	0.7267	0.0165	0.0214	0.0051	0.0155	0.0153
5	0.3754	0.4537	0.6338	0.8053	0.6564	0.0217	0.0247	0.0142	0.0096	0.0233
7	0.7458	0.2958	0.7590	0.6993	0.8104	0.0118	0.0340	0.0046	0.0115	0.0190
8	0.6938	0.0331	0.8001	0.7173	0.6391	0.0228	0.0313	0.0110	0.0182	0.0119
9	0.9179	0.0460	0.7087	0.8508	0.8325	0.0116	0.0332	0.0058	0.0076	0.0100
10	0.5890	0.1759	0.7080	0.8101	0.7767	0.0278	0.0421	0.0102	0.0113	0.0145
11	0.7298	0.1519	0.7974	0.8815	0.7089	0.0185	0.0685	0.0080	0.0071	0.0271
12	0.8432	0.2029	0.7750	0.7677	0.7825	0.0283	0.0463	0.0062	0.0151	0.0115
13	0.7308	0.1088	0.7968	0.8031	0.7459	0.0138	0.0429	0.0092	0.0075	0.0105
18	0.5896	0.3020	0.7072	0.7143	0.4708	0.0154	0.0405	0.0045	0.0146	0.0242
29	0.6728	0.0195	0.8097	0.8863	0.8282	0.0316	0.0334	0.0082	0.0128	0.0318
41	0.8834	0.0754	0.6811	0.6461	0.2944	0.0041	0.0197	0.0139	0.0267	0.0579
49	0.8662	0.4682	0.6794	0.4959	0.2510	0.0060	0.0117	0.0059	0.0355	0.0794
63	0.5559	0.1816	0.7320	0.6692	0.6932	0.0120	0.0150	0.0050	0.0166	0.0121
Mean	<b>0.7051</b>	<b>0.1835</b>	<b>0.7402</b>	<b>0.7573</b>	<b>0.6824</b>	0.0172	0.0349	0.0082	0.0146	0.0230

Table 8.12 – F-Measures of Evaluation Set using 20% Training Set (CE-PSSS) on Database 2

## 8.15 Appendix 13

ID	W	N1	N2	N3	R
1	0.5897	0.0909	0.2510	0.5670	0.1538
3	0.6403	0.0000	0.5181	0.5749	0.0000
4	0.4108	0.3457	0.5274	0.6079	0.4179
5	0.2356	0.0000	0.4587	0.6603	0.0222
7	0.6983	0.0000	0.4067	0.6335	0.0279
8	0.5063	0.0000	0.6233	0.5532	0.0116
9	0.6984	0.0000	0.4018	0.6248	0.0000
10	0.3516	0.0000	0.4299	0.7583	0.0000
11	0.4561	0.0000	0.1658	0.6848	0.0000
12	0.5976	0.0000	0.4855	0.6639	0.0233
13	0.6811	0.0000	0.1659	0.6681	0.0486
18	0.5856	0.2276	0.2769	0.5444	0.1404
29	0.3345	0.0000	0.6780	0.7444	0.0000
41	0.5374	0.0000	0.2089	0.3429	0.0000
49	0.7710	0.0291	0.1896	0.1647	0.0000
63	0.5416	0.0972	0.3872	0.4352	0.1533
Mean	<b>0.5397</b>	<b>0.0494</b>	<b>0.3859</b>	<b>0.5768</b>	<b>0.0624</b>

Table 8.13 – F-Measures of SF-SSS (with modifications) using MTSU on Database 2



## 8.16 Appendix 14

ID	Agreement Rate (DE-PSSS)	Cohen's Kappa (DE-PSSS)	Agreement Rate (CE-PSSS)	Cohen's Kappa (CE-PSSS)
1	0.8152	0.7527	0.7892	0.7189
2	0.8637	0.7956	0.8086	0.7149
3	0.8147	0.7395	0.7570	0.6602
4	0.7258	0.6367	0.6362	0.5115
5	0.7919	0.6906	0.7603	0.6541
6	0.8496	0.7668	0.7490	0.6130
7	0.8861	0.8444	0.8075	0.7398
8	0.8006	0.7162	0.7396	0.6297
9	0.8546	0.7902	0.7996	0.7087
10	0.8258	0.7458	0.7878	0.6842
11	0.8310	0.7534	0.7898	0.6954
12	0.6883	0.5549	0.6470	0.4968
13	0.8840	0.8233	0.7893	0.6730
14	0.8314	0.6848	0.8033	0.6259
15	0.6642	0.4834	0.6924	0.5306
16	0.7264	0.5884	0.6561	0.4899

Table 8.14 – Agreement Rate and Cohen's Kappa of PSSSs

# Bibliography

- [1] Khald AI Aboalayon and Miad Faezipour. “Multi-class SVM based on sleep stage identification using EEG signal”. In: *Healthcare Innovation Conference (HIC), 2014 IEEE*. IEEE. 2014, pp. 181–184.
- [2] Beena Ahmed, Amira Redissi, and Reza Tafreshi. “An automatic sleep spindle detector based on wavelets and the Teager energy operator”. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE. 2009, pp. 2596–2599.
- [3] G Alon et al. “Application of the cross-entropy method to the buffer allocation problem in a simulation-based environment”. In: *Annals of Operations Research* 134.1 (2005), pp. 137–151.
- [4] Thomas Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford, UK: Oxford University Press, 1996. ISBN: 0195099710.
- [5] Mohammed Bahoura and Jean Rouat. “Wavelet speech enhancement based on the teager energy operator”. In: *Signal Processing Letters, IEEE* 8.1 (2001), pp. 10–12.
- [6] Tim Benham et al. “CEoptim: Cross-Entropy R Package for Optimization”. In: *arXiv preprint arXiv:1503.01842* (2015).
- [7] Isabelle Bloch and HENRI MAiTRE. “Data fusion in 2D and 3D image processing: An overview”. In: *Computer Graphics and Image Processing, 1997. Proceedings., X Brazilian Symposium on*. IEEE. 1997, pp. 127–134.
- [8] Guillaume M JB Chaslot et al. “Cross-entropy for Monte-Carlo tree search”. In: (2008).
- [9] Alberto Colorni, Marco Dorigo, Vittorio Maniezzo, et al. “Distributed optimization by ant colonies”. In: *Proceedings of the first European conference on artificial life*. Vol. 142. Paris, France. 1991, pp. 134–142.
- [10] Heidi Danker-hopfe et al. “Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard”. In: *Journal of sleep research* 18.1 (2009), pp. 74–84.
- [11] Charles Darwin. “On the origin of species by means of natural selection. 1859”. In: *Murray, London* 502 (1991).
- [12] Belur V Dasarathy. “Sensor fusion potential exploitation-innovative architectures and illustrative applications”. In: *Proceedings of the IEEE* 85.1 (1997), pp. 24–38.
- [13] Gianni Di Caro and Marco Dorigo. “Ant colony optimization and its application to adaptive routing in telecommunication networks”. In: (2004).
- [14] Xianting Ding et al. “Cascade search for HSV-1 combinatorial drugs with high antiviral efficacy and low toxicity”. In: *International journal of nanomedicine* 7 (2012).
- [15] WB Dolan, PT Cummings, and MD LeVan. “Process optimization via simulated annealing: application to network design”. In: *AICHE Journal* 35.5 (1989), pp. 725–736.

- [16] Kathryn A Dowsland. “Nurse scheduling with tabu search and strategic oscillation”. In: *European journal of operational research* 106.2 (1998), pp. 393–407.
- [17] Zvi Drezner, George A Marcoulides, and Mark Hoven Stohs. “Financial applications of a tabu search variable selection model”. In: *Advances in Decision Sciences* 5.4 (2001), pp. 215–234.
- [18] Diego Álvarez Estévez. “Diagnosis of the sleep apnea-hypopnea syndrome: a comprehensive approach through an intelligent system to support medical decision”. PhD thesis. Universidade da Coruña, 2012.
- [19] Amr F Farag, Shereen M El-Metwally, and Ahmed A Morsy. “A Sleep Scoring System Using EEG Combined Spectral and Detrended Fluctuation Analysis Features”. In: *Journal of Biomedical Science and Engineering* 7.8 (2014), p. 584.
- [20] Luay Fraiwan et al. “Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier”. In: *Computer methods and programs in biomedicine* 108.1 (2012), pp. 10–19.
- [21] Fred Glover. “Future paths for integer programming and links to artificial intelligence”. In: *Computers & operations research* 13.5 (1986), pp. 533–549.
- [22] Fred Glover. “Tabu search-part I”. In: *ORSA Journal on computing* 1.3 (1989), pp. 190–206.
- [23] Steinn Gudmundsson, Thomas Philip Runarsson, and Sven Sigurdsson. “Automatic sleep staging using support vector machines with posterior probability estimates”. In: *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*. Vol. 2. IEEE. 2005, pp. 366–372.
- [24] David L Hall and Alan Steinberg. *Dirty secrets in multisensor data fusion*. Tech. rep. DTIC Document, 2001.
- [25] Masaaki Hanaoka, Masaki Kobayashi, and Haruaki Yamazaki. “Automatic sleep stage scoring based on waveform recognition method and decision-tree learning”. In: *Systems and Computers in Japan* 33.11 (2002), pp. 1–13.
- [26] Sari-Leena Himanen and Joel Hasan. “Limitations of rechtschaffen and kales”. In: *Sleep medicine reviews* 4.2 (2000), pp. 149–167.
- [27] Chih-Ming Ho. “Keynote: Personalized medicine enabled by FSC. X technology”. In: *Faible Tension Faible Consommation (FTFC), 2014 IEEE*. IEEE. 2014, pp. 1–2.
- [28] Jamil L Hossain and Colin M Shapiro. “The prevalence, cost implications, and management of sleep disorders: an overview”. In: *Sleep and Breathing* 6.2 (2002), pp. 85–102.
- [29] Jung Hu and Benjamin Knapp. “Electroencephalogram pattern recognition using fuzzy logic”. In: *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*. IEEE. 1991, pp. 805–807.
- [30] K-P Hui et al. “The cross-entropy method for network reliability estimation”. In: *Annals of Operations Research* 134.1 (2005), pp. 101–118.
- [31] E Huupponen et al. “EEG alpha activity detection by fuzzy reasoning”. In: *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*. Vol. 1. IEEE. 2001, pp. 411–416.
- [32] Hisao Ishibuchi, Shinta Misaki, and Hideo Tanaka. “Modified simulated annealing algorithms for the flow shop sequencing problem”. In: *European Journal of Operational Research* 81.2 (1995), pp. 388–398.
- [33] Firas Jabloun, A Enis Cetin, and Engin Erzin. “Teager energy based feature parameters for speech recognition in car noise”. In: *Signal Processing Letters, IEEE* 6.10 (1999), pp. 259–261.

- [34] Data Fusion Lexicon JDL. “Technical Panel For C3”. In: *FE White, San Diego, Calif, USA, Code 420* (1991).
- [35] James F Kaiser. “On a simple algorithm to calculate the energy of a signal”. In: *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*. IEEE. 1990, pp. 381–384.
- [36] Borhan Kazimipour, Xiaodong Li, and AK Qin. “Effects of population initialization on differential evolution for large scale optimization”. In: *Evolutionary Computation (CEC), 2014 IEEE Congress on*. IEEE. 2014, pp. 2404–2411.
- [37] Nizar Kerkeni et al. “Automatic classification of Sleep Stages on a EEG signal by Artificial Neural Networks”. In: *5th WSEAS International Conference on SIGNAL, SPEECH and IMAGE PROCESSING-WSEAS SSIP’05*. 2005.
- [38] Truong Quang Dang Khoa, Vo Van Toi, et al. “A tool for analysis and classification of sleep stages”. In: *Advanced Technologies for Communications (ATC), 2011 International Conference on*. IEEE. 2011, pp. 307–310.
- [39] Scott Kirkpatrick. “Optimization by simulated annealing: Quantitative studies”. In: *Journal of statistical physics* 34.5-6 (1984), pp. 975–986.
- [40] Tarek Lajnef et al. “Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines”. In: *Journal of neuroscience methods* 250 (2015), pp. 94–105.
- [41] Patrick Lambert and Thierry Carron. “Symbolic fusion of luminance-hue-chroma features for region segmentation”. In: *Pattern recognition* 32.11 (1999), pp. 1857–1872.
- [42] J Richard Landis and Gary G Koch. “The measurement of observer agreement for categorical data”. In: *biometrics* (1977), pp. 159–174.
- [43] Claire Laudy and Jean-Gabriel Ganascia. “Information fusion in a tv program recommendation system”. In: *Information Fusion, 2008 11th International Conference on*. IEEE. 2008, pp. 1–8.
- [44] Quoc Khai Le, Quang Dang Khoa Truong, et al. “A tool for analysis and classification of sleep stages”. In: *Advanced Technologies for Communications (ATC), 2011 International Conference on*. IEEE. 2011, pp. 307–310.
- [45] Sheng-Fu Liang et al. “A fuzzy inference system for sleep staging”. In: *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*. IEEE. 2011, pp. 2104–2107.
- [46] Sheng-Fu Liang et al. “A rule-based automatic sleep staging method”. In: *Journal of neuroscience methods* 205.1 (2012), pp. 169–176.
- [47] Xi Long et al. “Analyzing respiratory effort amplitude for automated sleep stage classification”. In: *Biomedical Signal Processing and Control* 14 (2014), pp. 197–205.
- [48] Haoyu Ma et al. “A hybrid classification method using artificial neural network based decision tree for automatic sleep scoring”. In: *World Academy of Science, Engineering and Technology* 79 (2011), pp. 279–284.
- [49] Holger R Maier et al. “Ant colony optimization for design of water distribution systems”. In: *Journal of water resources planning and management* 129.3 (2003), pp. 200–209.
- [50] Antonio Miguel Mora et al. “Sleeping with ants, SVMs, multilayer perceptrons and SOMs”. In: *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*. IEEE. 2010, pp. 126–131.
- [51] James N Morgan and John A Sonquist. “Problems in the analysis of survey data, and a proposal”. In: *Journal of the American statistical association* 58.302 (1963), pp. 415–434.

- [52] Sudipta Mukhopadhyay and GC Ray. “A new interpretation of nonlinear energy operator and its efficacy in spike detection”. In: *Biomedical Engineering, IEEE Transactions on* 45.2 (1998), pp. 180–187.
- [53] Godfrey Onwubolu and Donald Davendra. “Scheduling flow shops using differential evolution algorithm”. In: *European Journal of Operational Research* 171.2 (2006), pp. 674–692.
- [54] Kenneth Price, Rainer M Storn, and Jouni A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.
- [55] Allan Rechtschaffen and Anthony Kales. “A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects”. In: (1968).
- [56] M Janga Reddy and D Nagesh Kumar. “Multiobjective differential evolution with application to reservoir system optimization”. In: *Journal of Computing in Civil Engineering* 21.2 (2007), pp. 136–146.
- [57] Richard S Rosenberg, Steven Van Hout, et al. “The American Academy of Sleep Medicine inter-scoring reliability program: sleep stage scoring”. In: *J Clin Sleep Med* 9.1 (2013), pp. 81–87.
- [58] Reuven Rubinstein. “The cross-entropy method for combinatorial and continuous optimization”. In: *Methodology and computing in applied probability* 1.2 (1999), pp. 127–190.
- [59] Reuven Y Rubinstein. “Optimization of computer simulation models with rare events”. In: *European Journal of Operational Research* 99.1 (1997), pp. 89–112.
- [60] Sadiq M Sait, Mahmood R Minhas, and Junhaid A Khan. “Performance and low power driven VLSI standard cell placement using Tabu search”. In: *Evolutionary Computation, 2002. CEC’02. Proceedings of the 2002 Congress on*. Vol. 1. IEEE. 2002, pp. 372–377.
- [61] Ruhul A Sarker, Saber M Elsayed, and Tapabrata Ray. “Differential evolution with dynamic parameters selection for optimization problems”. In: *IEEE Transactions on Evolutionary Computation* 18.5 (2014), pp. 689–707.
- [62] Nicolas Schaltenbrand, Régis Lengelle, and Jean-Paul Macher. “Neural network model: application to automatic analysis of human sleep”. In: *Computers and Biomedical Research* 26.2 (1993), pp. 157–171.
- [63] C Muthamil Selvi and K Gnanambal. “Power system voltage stability analysis using Modified Differential Evolution”. In: *Computer, Communication and Electrical Technology (IC-CCET), 2011 International Conference on*. IEEE. 2011, pp. 382–387.
- [64] Daniel Shiffman, Shannon Fry, and Zannah Marsh. *The nature of code*. D. Shiffman, 2012.
- [65] Ibrahim Al-Shyoukh et al. “Systematic quantitative characterization of cellular responses induced by multiple signals”. In: *BMC systems biology* 5.1 (2011), p. 1.
- [66] Michael H Silber et al. “The visual scoring of sleep in adults”. In: *J Clin Sleep Med* 3.2 (2007), pp. 121–131.
- [67] American Academy of Sleep Medicine, Conrad Iber, et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine, 2007.
- [68] Stanislaw Solnik et al. “Teager–Kaiser energy operator signal conditioning improves EMG onset detection”. In: *European journal of applied physiology* 110.3 (2010), pp. 489–498.
- [69] MS Stachowicz and ME Kochanska. “Analysis of the application of fuzzy relations in modeling”. In: *Proc. North American Fuzzy Information Society*. Vol. 86. 1986.
- [70] Rainer Storn and Kenneth Price. “Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces”. In: *Journal of global optimization* 11.4 (1997), pp. 341–359.

- [71] M Emin Tagluk, Necmettin Sezgin, and Mehmet Akin. “Estimation of sleep stages by an artificial neural network employing EEG, EMG and EOG”. In: *Journal of medical systems* 34.4 (2010), pp. 717–725.
- [72] JY Tian and JQ Liu. “Automated sleep staging by a hybrid system comprising neural network and fuzzy rule-based reasoning”. In: *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE. 2006, pp. 4115–4118.
- [73] ML Tsetlin. “Finite automata and modeling the simplest forms of behavior”. PhD thesis. PhD thesis, VA Steklov Mathematical Institute, 1964.
- [74] Hideaki Tsutsui et al. “An optimized small molecule inhibitor cocktail supports long-term maintenance of human embryonic stem cells”. In: *Nature communications* 2 (2011), p. 167.
- [75] Brian Tung and Leonard Kleinrock. “Using finite state automata to produce self-optimization and self-control”. In: *Parallel and Distributed Systems, IEEE Transactions on* 7.4 (1996), pp. 439–448.
- [76] Adrien Ugon. “Fusion symbolique et données polysomnographiques”. PhD thesis. Université Pierre et Marie Curie (UPMC, Paris 6), 2013.
- [77] Adrien Ugon et al. “How to use symbolic fusion to support the sleep apnea syndrome diagnosis”. In: *Artificial Intelligence in Medicine*. Springer, 2011, pp. 45–54.
- [78] Vladimir Vapnik. “Pattern recognition using generalized portrait method”. In: *Automation and remote control* 24 (1963), pp. 774–780.
- [79] Vladimir N Vapnik. “An overview of statistical learning theory”. In: *Neural Networks, IEEE Transactions on* 10.5 (1999), pp. 988–999.
- [80] Jussi Virkkala et al. “Automatic sleep stage classification using two-channel electro-oculography”. In: *Journal of neuroscience methods* 166.1 (2007), pp. 109–115.
- [81] Feng-Sheng Wang and Horng-Jhy Jang. “Parameter estimation of a bioreaction model by hybrid differential evolution”. In: *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*. Vol. 1. IEEE. 2000, pp. 410–417.
- [82] Pak Kin Wong et al. “Closed-loop control of cellular functions using combinatory drugs guided by a stochastic search algorithm”. In: *Proceedings of the National Academy of Sciences* 105.13 (2008), pp. 5105–5110.
- [83] Jue Wu and Albert CS Chung. “Cross entropy: A new solver for markov random field modeling and applications to medical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*. Springer, 2005, pp. 229–237.
- [84] Meng Xiao et al. “Sleep stages classification based on heart rate variability and random forest”. In: *Biomedical Signal Processing and Control* 8.6 (2013), pp. 624–633.
- [85] Jihoon Yang and Vasant Honavar. “Feature subset selection using a genetic algorithm”. In: *Feature extraction, construction and selection*. Springer, 1998, pp. 117–136.