# Wearable in-ear electroencephalography for real world applications: Sleep monitoring and person authentication

Takashi Nakamura

Supervisor: Professor Danilo P. Mandic

A thesis submitted for the degree of
*Doctor of Philosophy (Ph.D)*

Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London
June 2019

*All my life needed was a sense of someplace to go. — Travis Bickle*

## Statement of Originality

I declare that the contents in this thesis are my own work under the supervision of Professor Danilo P. Mandic. All information derived from the work of others has been appropriately acknowledged in the text and a list of references is given. The material in this thesis have not been submitted in any form for another academic degree.

## Publications

### Journal articles in preparation

- Y. D. Alqurashi, **T. Nakamura**, J. Moss, M. I. Polkey, M. J. Morrell, and D. P. Mandic, "The efficacy of an in-ear electroencephalography sensor to measure non-rapid eye movement sleep in healthy participants overnight," (in preparation).

### Refereed journal articles

- **T. Nakamura**, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Hearables: Automatic overnight sleep monitoring with standardised in-ear EEG sensor," *IEEE Transactions on Biomedical Engineering* (in press).
- T. Adjei, W. v. Rosenberg, **T. Nakamura**, T. Chanwimalueang, and D. Mandic "The ClassA framework: HRV based assessment of SNS and PNS dynamics without LF-HF controversies," *Frontiers in Physiology*, vol. 10, Article no. 505, 2019.
- Y. Alqurashi, **T. Nakamura**, V. Goverdovsky, J. Moss, M. Polkey, D. Mandic, and M. Morrell, "A novel in-ear sensor to determine sleep latency during the Multiple Sleep Latency Test (MSLT) in healthy adults with and without sleep restriction," *Nature and Science of Sleep*, vol.10, pp. 385-395, 2018.
- **T. Nakamura**, V. Goverdovsky, and D. P. Mandic, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648-661, 2018.
- V. Goverdovsky, W. v. Rosenberg, **T. Nakamura**, D. Looney, D. J. Sharp, C. Papavassiliou, M. J. Morrell, and D. P. Mandic, "Hearables: Multimodal physiological in-ear sensing," *Scientific Reports*, vol. 7, no. 1, Article no. 6948, 2017.
- **T. Nakamura**, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, "Automatic sleep monitoring using ear-EEG," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, Article no. 2702558, 2017.

### Refereed conference papers

- **T. Nakamura**, H. J. Davies and D. P. Mandic, "Scalable automatic sleep staging in the era of Big Data," in *Proc. IEEE Int. Eng. in Med. and Biology Conference (EMBC)*, 2019 (to appear).
- H. J. Davies, **T. Nakamura** and D. P. Mandic, "A transition probability based classification model for enhanced N1 sleep stage identification during automatic sleep stage scoring," in *Proc. IEEE Int. Eng. in Med. and Biology Conference (EMBC)*, 2019 (to appear).

- **T. Nakamura**, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Automatic detection of drowsiness using in-ear EEG," in *Proc. IEEE Int. Joint Conf. on Neural Netw. (IJCNN)*, pp. 5569-5574, 2018.
- **T. Nakamura**, T. Adjei, Y. Alquarashi, D. Looney, M. J. Morrell, and D. P. Mandic, "Complexity science for sleep stage classification from EEG," in *Proc. IEEE Int. Joint Conf. on Neural Netw. (IJCNN)*, pp. 4387-4394, 2017.

**Conference abstracts**

- Y. Alqurashi, **T. Nakamura**, J. Moss, M. I. Polkey, D. P. Mandic, and M. J. Morrell, "The efficacy of a novel in-ear electroencephalography (EEG) sensor to measure overnight sleep in healthy participants," *The 2019 American Thoracic Society Inter. Conf. in Dallas*, 2019.
- Y. Alqurashi, J. Moss, **T. Nakamura**, V. Goverdovsky, M. I. Polkey, D. P. Mandic, and M. J. Morrell, "The efficacy of in-ear electroencephalography (EEG) to monitor sleep latency and the impact of sleep deprivation," *The 2017 American Thoracic Society Inter. Conf. in Washington*, 2017.

## Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

## Abstract

The electroencephalography (EEG) is a standard for studying the dynamics of brain activity, especially that related to cognition and disease. However, despite the EEG being extensively employed in both clinical settings and brain-computer interface (BCI), real world EEG-based applications are not yet well established. This is largely due to the cumbersome and obtrusive nature of current EEG sensing which has been prohibitive to the continuous acquisition of data over sufficiently long periods of time. To bring the power of EEG monitoring of neural activity out of clinic, this thesis explores the feasibility of EEG monitoring in the real world using the recent discovery of ear-canal based physiological sensing, termed in-ear EEG. The utility of wearable in-ear sensing technology, especially its capability to monitor brain activity for automatic sleep monitoring and person authentication, is investigated.

This thesis first demonstrates the utility of in-ear sensing for one of the cornerstones of well-being in modern society – sleep monitoring. Sleep is an essential process for human well-being, and its quality reflects both a person's lifestyle as well as various medical conditions. Current clinical sleep monitoring analysis reflects sleep profiles of individuals rigorously and comprehensively, however, the cumbersome nature of standard recordings can disturb patients' normal sleep, while the scoring process places enormous demands on a clinician's time. To mitigate these issues, feature extraction for automatic sleep monitoring is investigated. The combination of structural complexity analysis and non-linear frequency domain features is shown to be effective for classifying sleep stages, and the feature-fusion based classification model using those features is shown to outperform other existing studies using a publicly available overnight sleep EEG dataset. In addition, the efficacy of smoothing algorithms implemented by a hidden Markov model is investigated in this context. Following the investigation of automatic sleep staging algorithms, the feasibility of an in-ear EEG device for large scale sleep monitoring is scrutinised. To begin with, as proof-of-concepts, numerous nap recordings are conducted inside the laboratory with in-ear EEG. To provide wearable sleep monitoring in real-world, simultaneous conventional polysomnography (PSG) and in-ear EEG during the night is recorded in subjects' home; the agreement between automatic scoring based on in-ear EEG and manual scoring based on the PSG was 74.1 % and the corresponding kappa coefficients was 0.61 (Substantial Agreement). Through numerous recordings and analysis, a wearable in-ear EEG device for sleep monitoring is proposed and validated as a 24/7 continuous and unobtrusive technology for sleep quality assessment in the community.

The other major in-ear EEG application explored is person authentication, which refers to the process of confirming the claimed identity of an individual. The use of EEG as a biometrics modality has been investigated for the last 15 years; however, its feasibility in real world has not yet been established, mainly due to issues of reproducibility. To this end, a readily deployable EEG biometrics system based on in-ear EEG is proposed, that is user-friendly and discreet in nature and does not require supervision by skilled personnel or the use of cumbersome equipment. For rigour, recordings over two days

from 15 subjects are conducted; both frequency domain and autoregressive features are employed to identify an individual. The achieved classification accuracy was 95.7 % in the verification setup whereas the kappa coefficient was 0.65 (Substantial Agreement) in the identification setup.
Overall, the comprehensive analysis over a number of recordings and proof-of-concept studies in this thesis demonstrates the feasibility of the proposed in-ear EEG based applications. The so-verified wearable in-ear EEG paradigm promises faithful EEG applications in real-world environments, and resolves the critical issues with robustness associated with current use of EEG.

## Acknowledgements

First and foremost, I would like to thank my supervisor Professor Danilo P. Mandic who gave me the great opportunity to undertake this PhD. During my MSc and PhD, he has supported me during difficult time, guided me, and given me many opportunities to develop myself regarding to both academic and personal skills. I hope I was able to learn as much as possible over last four years.

Next, I am thankful to have been given the occasion to work with Professor Mary J. Morrell. She advised and taught me physiology and sleep, and gave radically different insights into my research from medical perspectives.

I would like to thank Dr Valentin Goverdovsky. He is a great mentor who taught me the research strategies and approaches, especially in the early days of my PhD.

I wish to acknowledge the financial support from the Nakajima Foundation for sponsoring during both MSc and PhD studies.

I would like to thank both Professor Jonathon Chambers and Dr Tobias Reichenbach for their insightful comments.

I am also grateful to my colleagues in the CSP group and collaborators. Fortunately, I have met countless incredible young researchers, but unfortunately I have met too many to write down in this part.

On top of this, I am extremely grateful to my friends I met in London, especially I have spent such a valuable time in Griffon Studios.

Last but not least, to my father and mother, I have been extremely fortunate to be in this family. I would like to thank my grandmothers who sponsored me during my studies. My brother has given me a lot of opportunities to open up myself and have deep conversations. Special thanks should go to my family for their unconditional love and support.

Takashi Nakamura
*June 2019*

# Contents

**Statement of Originality** v

**Publications** vii

**Copyright Declaration** ix

**Abstract** xi

**Acknowledgements** xiii

**Contents** xv

**List of Figures** xv

**List of Tables** xix

**List of Abbreviations** xxi

**Chapter 1. Introduction** 1

1.1 Brain monitoring: EEG . . . 1

1.1.1 Recent EEG research in biomedical engineering . . . 2

1.1.2 Portable EEG . . . 3

1.1.3 Mobile EEG . . . 4

1.1.4 Transparent EEG . . . 5

1.2 Ear-EEG system . . . 6

1.2.1 Original in-ear EEG system . . . 6

1.2.2 Sensor developments and around-ear EEG system . . . 7

1.2.3 Ear-EEG application . . . 8

1.2.4 Could ear-EEG be a 'transparent EEG'? . . . 8

1.3 Problem formulation . . . 10

1.3.1 In-ear EEG for sleep monitoring . . . 10

1.3.2 In-ear EEG for person authentication . . . 11

1.4 Contribution of the thesis . . . 12

1.5 Thesis organisation . . . 15

**Chapter 2. Algorithms and preliminaries** **19**
2.1 Feature extraction . . . 19
2.1.1 Structural complexity analyses for automatic sleep staging . . . 19
2.1.2 Frequency domain features for automatic sleep staging . . . 24
2.1.3 Other features for automatic sleep staging . . . 27
2.2 Classification . . . 28
2.2.1 Support vector machine (SVM) for classification . . . 28
2.2.2 Extreme learning machine (ELM) for classification . . . 34
2.3 Smoothing algorithm for automatic sleep staging . . . 36
2.3.1 Hidden Markov model (HMM) . . . 36
2.4 Evaluation . . . 39
2.5 Summary . . . 39

**Part I. State-of-the-art automatic sleep stage classification** **41**

**Chapter 3. Structural complexity analysis for automatic sleep staging** **43**
3.1 Introduction . . . 43
3.2 Methods . . . 46
3.2.1 Data acquisition . . . 46
3.2.2 Pre-processing . . . 47
3.2.3 Feature extraction . . . 48
3.2.4 Classification . . . 50
3.2.5 Performance evaluation . . . 50
3.3 Results . . . 50
3.4 Summary . . . 54

**Chapter 4. Automatic sleep staging for large scale data** **55**
4.1 Introduction . . . 55
4.2 Methods . . . 57
4.2.1 Data acquisition . . . 57
4.2.2 Pre-processing . . . 57
4.2.3 Feature extraction . . . 58
4.2.4 Classification . . . 58
4.2.5 Smoothing . . . 59
4.3 Results . . . 59
4.4 Discussion . . . 64
4.5 Summary . . . 66

**Part II. Proof-of-concept in-ear EEG applications** **67**

**Chapter 5. Automatic sleep staging using in-ear EEG for nap data** **69**
5.1 Introduction . . . 69
5.2 Methods . . . 71
5.2.1 Data acquisition . . . 71
5.2.2 Manual scoring . . . 72
5.2.3 Pre-processing for automatic sleep stage classification . . . 73
5.2.4 Feature extraction . . . 73
5.2.5 Classification . . . 74
5.2.6 Performance evaluation . . . 74
5.3 Results . . . 74
5.4 Summary . . . 78

**Chapter 6. Automatic detection of drowsiness using in-ear EEG** **79**
6.1 Introduction . . . 79
6.2 Methods . . . 80
6.2.1 Data acquisition . . . 80
6.2.2 Manual scoring . . . 82
6.2.3 Pre-processing . . . 83
6.2.4 Feature extraction . . . 83
6.2.5 Classification . . . 84
6.2.6 Performance evaluation . . . 84
6.2.7 Validation setup . . . 85
6.3 Results . . . 85
6.4 Summary . . . 86

**Chapter 7. Person authentication using in-ear EEG** **87**
7.1 Introduction . . . 87
7.2 Overview of EEG-based biometrics . . . 89
7.2.1 Biometric systems with verification/identification . . . 89
7.2.2 Feasible EEG biometrics design . . . 90
7.2.3 Previous protocols . . . 92
7.2.4 Biometrics based on collectable EEG systems . . . 94
7.2.5 Problem formulation . . . 94
7.3 Methods . . . 95
7.3.1 Data acquisition . . . 95
7.3.2 Pre-processing . . . 96
7.3.3 Feature extraction . . . 97
7.3.4 Validation scenarios . . . 98
7.3.5 Classification . . . 100
7.3.6 Performance evaluation . . . 102
7.4 Results . . . 103

7.4.1 Client-Imposter verification with different segment sizes . . . . . 103
7.4.2 Client-Imposter verification with different features . . . . . 104
7.4.3 Client-Imposter verification with different classifiers . . . . . 104
7.4.4 Validation including non-registered imposters . . . . . 105
7.4.5 Client-Imposter verification results per subject . . . . . 106
7.4.6 Biometrics identification scenarios . . . . . 106
7.5 Discussion . . . . . 108
7.6 Summary . . . . . 112

**Part III. In-ear EEG application towards 'transparent EEG' 115**

**Chapter 8. Automatic overnight sleep monitoring using in-ear EEG 117**
8.1 Introduction . . . . . 117
8.2 Methods . . . . . 120
8.2.1 Data acquisition . . . . . 120
8.2.2 Manual scoring . . . . . 122
8.2.3 Pre-processing . . . . . 122
8.2.4 Feature extraction . . . . . 123
8.2.5 Classification . . . . . 124
8.2.6 Performance evaluation . . . . . 124
8.2.7 Validation setup . . . . . 124
8.3 Results . . . . . 124
8.4 Discussion . . . . . 128
8.5 Summary . . . . . 129

**Chapter 9. Conclusions and future work 131**
9.1 Conclusions . . . . . 131
9.2 Future work . . . . . 132

**References 137**

# List of Figures


1.1 Examples of the *conventional EEG*, the *mobile EEG* and the *portable EEG*. The *transparent EEG* is categorised into in the middle of this diagram (Yellow circle). (1) Controlling a synthetic arm in laboratory [1, 2] and (2) controlling a wheelchair in a room [3] – BCI motor imagery tasks. (3) High-density EEG recording during dart games [4], and (4) high-density EEG recordings on the treadmill [5] – MoBI project. (5) Event related potential recordings on a fixed bike. (6) Visual evoked potential recordings during walking with a commercial headset [6]. (7) Emotiv EPOC+ [7] and (8) Neurosky Brainlink [8] – commercial wireless EEG headset. (9) The cEEGrid [9], and (10) the generic earpiece [10] – the ear-EEG technology. . . . . 6

1.2 The concept of in-ear physiological monitoring [11]. . . . . 7

1.3 An example of recent ear-EEG technology. (a, b) In-ear generic earpiece; (c, d) around-ear cEEGrid. Images are adapted from [9, 10, 12]. . . . . 8

1.4 Publications and thesis structure. . . . . 13

2.1 Extraction of motif for the permutation entropy with the embedding dimension $d = 3$. Figure is adapted and modified from [13]. . . . . 23

2.2 EEG recordings from a single subject in different sleep stages. (Red) The on-scalp C3-A2 channel; (Blue) the in-ear upper channel. . . . . 24

2.3 Averaged MSPE for the overnight sleep of one participant for an in-ear upper channel (Top) and an on-scalp C3-A2 channel (Bottom) over different sleep stages with scale $\tau = 1 - 20$. The error bars represent the standard error. . . . . 25

2.4 SEF at 50 % and 95 % of the total signal power between 4-16 Hz. . . . . 27

2.5 (Black) SEF$d$ feature between 0.5-16 Hz in Fpz-Cz channel of one subject (Subject ID: sc4001e0) from the Sleep-EDF dataset; (Blue) The manually scored hypnogram. . . . . 27

2.6 The geometry of a linear decision boundary in two dimensional space. . . 29

2.7 The geometry of a linear decision boundary in two dimensional space with slack variables. . . . . 31

2.8 The diagram of an ELM model. (a) Simple model; (b) matrix notation. . . 36

2.9 A diagram of a hidden Markov model. . . . . 37

2.10 A example of Viterbi decoding. Figure and example are based on Figure 17.13 in [14]. . . . . 38

2.11 Confusion matrix and performance criteria. . . . . 40

3.1 PSD for two EEG channels (Fpz-Cz, and Pz-Oz) from the *SEDFx-S* dataset with $M = 5$ sleep stages. . . . . 45

3.2 Flowchart of an automatic sleep staging system. . . . . 46

3.3 Epoch selection for EEG time series. (Top) No overlap method, the epoch length is 30 seconds; (Bottom) the epoch length enlarges to 90 seconds as combining the preceding and following epoch. . . . . 48

3.4 MSPE analyses for two EEG channels from the *SEDFx-S* dataset with $M = 5$ sleep stages, and scale $\tau = 1 - 20$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 20$). The error bars represent the standard error. . . . . 49

3.5 Hypnogram (Blue) and the predicted label (Red) of subject *sc4001e0* in the *SEDFx-S* dataset for the $M = 5$ class classification problem. . . . . 51

4.1 The flowchart of this study. . . . . 57

4.2 Kappa values for different numbers of training subjects in Setup-S and Setup-N ($L = 5000$). The error bar shows standard deviation of over ten independent iterations. . . . . 60

4.3 Kappa values for different numbers of hidden nodes in Setup-S and Setup-N ($N_{\text{train}} = 258$). The error bar shows standard deviation of over ten independent iterations. . . . . 61

4.4 The histogram of the kappa value of test subjects in Setup-S and Setup-N ($N_{\text{train}} = 258, L = 5000$). The dashed lines indicate the average for each setup. . . . . 61

4.5 The averaged confusion matrices over all 257 testing subjects for different scenarios. The title of each confusion matrix denotes: [Setup-{S,N}, ($N_{\text{train}}, L$), kappa value $\kappa$]. Observe that the sensitivities of each class improved after HMM smoothing. . . . . 63

4.6 (Black) Scored hypnogram of one subject; (Blue/Red) automatically predicted label before/after smoothing. . . . . 64

4.7 The distribution of kappa values for different classifiers. "**" represents the Bonferroni corrected p-value from *post hoc* Dunn's pairwise test is less than 0.01. . . . . 65

4.8 The confusion matrices for the fixed 257 testing subjects and for different classifiers. . . . . 65

5.1 Comparison with previous studies. (A) Evaluation of the agreement between the manually scored hypnograms based on on-scalp EEG channels and in-ear EEG channels [15]. (B) Our analysis framework for establishing the feasibility of in-ear EEG in sleep research. . . . . 71
5.2 Recording setup in this study. (Left) The electrodes were placed on scalp and ear; (Centre) the subject reclined in a comfortable chair; (Right) wearable in-ear sensor with two flexible electrodes. . . . . 72
5.3 Flowchart for the sleep stage prediction framework adopted in this study (*Scenario 2*). . . . . 72
5.4 PSD: (Top) On-scalp C3-A2 montage; (Bottom) in-ear channel EL1. . . . . 73
5.5 Structural complexity features for different sleep stages. Normalised MSFE from the in-ear EEG channel EL1 shows excellent separation between sleep stages. The error bars indicate the standard error. . . . . 74
5.6 Hypnogram for Subject Two scored based on on-scalp EEG channels (Blue) and the automatically predicted label based on in-ear EEG channel EL1 (Red). (Top) 2-class Wake v. Sleep; (Middle) 2-class W-N1 v. N2-N3; (Bottom) 4-class classification scenario. . . . . 77

6.1 Flowchart for the classification framework in this study. . . . . 81
6.2 Recording setup in this study. (Left) The electrodes were placed on scalp and ear; (Centre) the subject reclined in a comfortable chair; (Right) wearable in-ear sensor with two flexible electrodes [10]. . . . . 82
6.3 Averaged PSD: (Top) On-scalp C3 channel; (Bottom) in-ear Ch1. . . . . 83
6.4 The averaged MSFE for on-scalp and in-ear EEG. The error bars represent the standard error. . . . . 84
6.5 Cross validation approaches used in this study. . . . . 85

7.1 Person recognition systems. (Top) Verification system; (Bottom) identification system. . . . . 89
7.2 Feasible EEG biometrics verification framework. (Left) EEG recording registration. The registered EEG signal must have been recorded beforehand. (Right) Verification and identification system. . . . . 91
7.3 The in-ear sensor used in this study. (Left) Wearable in-ear sensor with two flexible electrodes; (Right) placement of the viscoelastic earpiece. . . . . 96
7.4 Flowchart for the biometrics analysis framework in this study. . . . . 97
7.5 PSD for the in-ear EEG Ch1 (Left) and the in-ear EEG Ch2 (Right) of Subject One (Top) and Subject Two (Bottom). The thick lines correspond to the averaged periodogram obtained by the all recordings from the first day (Red) and the second day (Blue), whereas the thin lines are the averaged periodogram obtained by a single trial. . . . . 98

7.6 Two validation scenarios (Setup-R and Setup-B), where $X_R^{(i,j,k)} \in \mathbb{R}^{N\times D}$ and $X_N^{(i,-,k)} \in \mathbb{R}^{N\times D}$ denote a feature matrix from a single trial of the subject $i$, recording day $j$, and trial $k$, from the $S_R$ recordings and the $S_N$ recordings, respectively. The number of segments per recording trial $N$ depends on the chosen segment lengths $L_{\text{seg}}$. The dimension $D$ is the number of features per EEG segment. . . . . 101

7.7 The identification rate with different segment size in Setup-R and Setup-B. The error bars indicate the standard error. . . . . 107

7.8 PSD for the in-ear EEG Ch1 (Left) and the in-ear EEG Ch2 (Right) of Subject Eight. The thick lines correspond to the averaged periodogram obtained by the all recordings from the first day (Red) and the second day (Blue), whereas the thin lines are the averaged periodogram obtained by a single trial. . . . . 110

7.9 PSD for the in-ear EEG Ch1 (Left), and the in-ear EEG Ch2 (Right) of one subject. The thick lines correspond to the averaged periodogram obtained by the recordings from a 'sleepy' trial (Red) and a 'normal' trial (Blue). Observe that the alpha power attenuated during the 'sleepy' trial. . . . . 111

7.10 Scatter plot for the maximum alpha power against total power ratio (Left) and for the frequency at maximum alpha power against total power ratio (Right) of the in-ear Ch1. . . . . 113

8.1 Flowchart for the automatic sleep stage classification in this study. . . . . 121

8.2 The in-ear sensor used in this study. (Left) Wearable in-ear sensor with two flexible electrodes; (Right) placement of the in-ear sensor. . . . . 122

8.3 Confusion matrix for the classification results: (Top) In-ear EEG; (Bottom) on-scalp EEG. The symbols SE/PR on the bottom right denote respectively sensitivity and precision. . . . . 125

8.4 Classification accuracy (Blue bars) and kappa values (Red dots) for individual participants: (Top) Based on in-ear EEG; (Bottom) based on on-scalp EEG). . . . . 126

8.5 Manually scored hypnograms of overnight sleep based on the full PSG (Blue) and the automatically predicted label by the proposed algorithm with the in-ear EEG and on-scalp EEG features (Red) from two participants. The black crosses above the W stage indicate the epochs removed from the analyses due to the amplitude thresholding. . . . . 127

8.6 Averaged PSD of in-ear EEG and on-scalp EEG during selected consecutive epochs of two different participants. The vertical lines denote the 2 and 4 Hz frequencies in order to designate the delta activity (0.5-4 Hz) and slow wave activity (0.5-2 Hz). . . . . 129

9.1 Future work. (Left) Sensor development; (Right) in-ear sensing for driver fatigue. . . . . 135

# List of Tables


1.1 Comparison of next generation EEG systems. . . . . 5
1.2 Various ear-EEG systems and their applications. . . . . 9

2.1 The details of feature extraction and classification methods in the chapters. 19
2.2 Sleep stage scoring guideline for EEG (Adapted from [16]). . . . . 24
2.3 Parameters for feature extraction. (Left) Setup-N; (Right) Setup-S. . . . . 26

3.1 Recent works for automatic sleep stage classification problem. . . . . 44
3.2 Comparison of datasets - after removing noisy epochs. . . . . 47
3.3 Sleep stage classes ($M$). . . . . 50
3.4 Classification accuracy with 10-fold CV for $M = 5$ class with two EEG channels. . . . . 51
3.5 *SEDF-W* - Confusion matrix for $M = 5$-stage sleep classification using MSFE and SEF features with overlapping windows extracted from the Fpz channel (a single EEG channel). . . . . 52
3.6 *SEDFx-S* - Confusion matrix for $M = 5$-stage sleep classification using MSPE and SEF features with overlapping windows extracted from the Fpz and Pz channels (two EEG channels). . . . . 52
3.7 *SEDFx-W* - Confusion matrix for $M = 5$-stage sleep classification using MSFE and SEF features with overlapping windows extracted from the Fpz channel (a single EEG channel). . . . . 52
3.8 *SEDFx-S* - Sensitivity and *Precision* (*in italic*) for sleep classification using MSPE and SEF with overlaps. . . . . 53
3.9 Performance comparison of S1 stage sensitivity (%) among existing methods for $M = 5$-stage sleep classification. . . . . 53
3.10 Performance comparison among existing methods for the Sleep-EDF database. . . . . 53

4.1 Proportion of sleep stages with the number of epochs and ratio. . . . . 57
4.2 Parameters and implementation details for the selected classifiers. . . . . 58
4.3 Computational time of different analyses. . . . . 62
4.4 Averaged kappa and computational time of different classifiers. . . . . 64

4.5 Recent works for automatic sleep stage classification with neural networks. 66

5.1 Confusion matrix for the 2-class Wake v. Sleep classification. . . . . 75
5.2 Confusion matrix for the 2-class Wake-N1 v. N2-N3 classification. . . . . 75
5.3 Confusion matrix for 4-class sleep stage classification. . . . . 75
5.4 Confusion matrix for the 2-class Wake v. Sleep classification. . . . . 76
5.5 Confusion matrix for the 2-class Wake-N1 v. N2-N3 classification. . . . . 76
5.6 Confusion matrix for the 4-class sleep stage classification. . . . . 76
5.7 Comparison between the manual scores and automatic predicted scores. Accuracy (%) / Kappa. . . . . 78

6.1 The configuration of EEG recordings. . . . . 82
6.2 Proportion of sleep stages over 23 subjects. . . . . 82
6.3 Confusion matrix for the 2-class Wake v. N1 classification (in-ear EEG, LOTOCV). . . . . 86
6.4 Confusion matrix for the 2-class Wake v. N1 classification (on-scalp EEG, LOTOCV). . . . . 86
6.5 Confusion matrix for the 2-class Wake v. N1 classification (in-ear EEG, 10-fold CV). . . . . 86
6.6 Confusion matrix for the 2-class Wake v. N1 classification (on-scalp EEG, 10-fold CV). . . . . 86

7.1 Recording and classification setup, and performance comparison among existing EEG biometrics. . . . . 91
7.2 Two EEG recordings and corresponding subset. . . . . 96
7.3 Dimensions of the training and validation matrix in Setup-R and Setup-B. 102
7.4 Hyper-parameters for SVM. . . . . 102
7.5 Summary of parameter choice in the proposed biometric application. . . . 103
7.6 Client-Impostor verification over different segment sizes. . . . . 104
7.7 Rigorous setup: Client-Imposter verification over different features in Setup-R. . . . . 104
7.8 Client-Impostor verification with different classifiers. . . . . 105
7.9 Confusion matrix of the Client-Imposter verification scenario from different datasets. . . . . 106
7.10 Recording details, accuracy and HTER in the verification problem, and sensitivity in the identification problem for each subject with different training and validation data in Setup-R. . . . . 107

8.1 Comparison of existing approaches to sleep research using in-ear/around-ear wearable sensing technology. . . . . 119
8.2 Proportion of scored sleep stages in different systems. . . . . 123

8.3 Sensitivity and *Precision* (*in italic*) for sleep classification in different class scenarios. . . . . 126

# List of Abbreviations

| | |
|---|---|
| 3D | Three dimensional |
| AASM | American Academy of Sleep Medicine |
| AC | Accuracy |
| AR | Autoregressive |
| BCI | Brain-computer interface |
| CCSHS | Cleveland Children's Sleep and Health Study |
| CNN | Convolutional neural networks |
| CRR | Correct recognition rate |
| CV | Cross-validation |
| ECG | Electrocardiography |
| EEG | Electroencephalography |
| EER | Equal error rate |
| ELM | Extreme learning machine |
| EMG | Electromyography |
| EOG | Electrooculography |
| ERP | Event related potential |
| FAR | False accept rate |
| FE | Fuzzy entropy |
| FN | False negative |
| FP | False positive |
| FRR | False reject rate |
| HMM | Hidden Markov model |
| HTER | Half total error rate |
| ICA | Independent component analysis |
| IR | Identification rate |
| KNN | K-nearest neighbours |
| (f)MRI | (functional) Magnetic resonance imaging |
| LDA | Linear discriminant analysis |
| LOTOCV | Leave-'one trial'-out cross-validation |
| MAP | Maximum a posteriori |
| MEG | Magnetoencephalography |
| ML | Machine learning |
| MoBI | Mobile brain/body imaging |
| MSE | Multi-scale entropy |

| | |
|---|---|
| MSFE | Multi-scale fuzzy entropy |
| MSPE | Multi-scale permutation entropy |
| MSSE | Multi-scale sample entropy |
| NaN | Not a number |
| PE | Permutation entropy |
| PR | Precision |
| PSD | Power spectral density |
| PSG | Polysomnography |
| R1IR | Rank-1 identification rate |
| R&K | Rechtschaffen and Kales |
| RBF | Radial basis function |
| REM | Rapid eye movement |
| RF | Random forests |
| RNN | Recurrent neural networks |
| SampE | Sample entropy |
| SE | Sensitivity |
| SEDF-W | Sleep-EDF whole |
| SEDFx-S | Sleep-EDF [expanded] sleep |
| SEDFx-W | Sleep-EDF [expanded] whole |
| SEF | Spectral edge frequency |
| SLFN | Single-hidden layer feedforward neural networks |
| SNR | Signal-to-noise ratio |
| SVM | Support vector machine |
| SWS | Slow wave sleep |
| TN | True negative |
| TP | True positive |

# Chapter 1

# Introduction

The rapidly expanding market for wearable devices is driven by an ever-increasing interest in physiological monitoring. According to the International Data Corporation, 222.3 million wearable devices, including watches, wristbands, eyewear, clothing, and earphones, will be shipped in 2021 (*cf.* 113.2 million shipments in 2017) [17]. Such wearable devices are typically designed to monitor the physiological responses of the user in real time, with the standard measurement being heart rate, which represents the number of contractions of the heart per minute. In 2018, latest consumer wearables were capable of monitoring the electrical activity of the heart, electrocardiogram (ECG or EKG), and accurately diagnosing heart conditions, with 95 % of success for the condition of atrial fibrillation [18].

In terms of user convenience, traditionally, physiological responses, such as the ECG and electroencephalogram (EEG: electrical brain signals), have been monitored in a hospital for diagnostic purposes. However, with recent advanced sensing technologies, it is now possible to monitor these responses using wearable devices. The main aim of this thesis is to explore the extent to which EEG sensing can be performed in out-of-clinic scenarios. Particular focus is placed on the wearable applications of our proof-of-concept in-ear EEG, which records brain electrical potentials from the ear canal.

## 1.1 Brain monitoring: EEG

#### Neural activity and EEG

The recording electrical neural activity of the brain became termed the electro- (electrical activities) encephalo- (emitting the signals for the brain) and graphy – EEG [19]. The EEG is a measurement of electrical activity from the surface of the scalp, in response to the general activity of the cerebral cortex. The EEG recordings are typically non-invasive and painless, whereby the electrodes are attached to the scalp with conductive gel (or paste) to ensure a low electrode-skin impedance. The electrical potential (typically of the order of tens of microvolts) is measured between selected pairs of electrodes and is produced by electrical fields generated by assemblies of cortical neurones synchronously firing – an EEG is a recording of the voltage generated by currents [20]. Many thousands

of underlying neurones must be activated together in order to generate an EEG signal big enough to be measured [20].

#### EEG and responses from non-invasive monitoring

Recording signals (or images) from human body has become a vital part of the diagnostic process in clinical settings [19]. Such digitised data are categorised as physiological responses, including the ECG from the heart, the electromyogram (EMG) from muscles, the EEG and magnetoencephalogram (MEG) from the brain, and the electrooculogram (EOG) from eye movements. The human body is imaged using radiography or sonography, including computerised tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), the functional MRI (fMRI). The EEG has a number of advantages over these other modalities as [21]:

- The EEG sensors (electrodes) are light enough to routinely wear day by day;
- The temporal resolution is high (*cf.* the fMRI image sequence is approximately two frames per second). Therefore the EEG is capable of monitoring brain activities in real-time;
- The recording setup is a relatively low cost.

In a clinical setting, EEG is used as a diagnostic tool, for instance, in the diagnosis of Alzheimers disease, and in the determination of lack of activity of the central nervous system (brain death) [22]. The EEG has also been used in the area of biomedical engineering, for example, in the development of brain-computer interface (BCI) systems [23] – where the interface typically translates measured electrophysiological signals from the brain (typically EEG) into control commands for prosthetics, games or facilitating software (e.g. word spellers for locked-in syndrome patients) [21], such that, the BCI decodes the user's brain signal.

### 1.1.1 Recent EEG research in biomedical engineering

In biomedical engineering research, the EEG has been recorded to study human cognition. Typically, the recordings have been conducted in controlled environments, namely static laboratory settings. The participants for the recordings sit still (or lie down), and are instructed to perform specific activities, for example attending to artificial stimuli in order to deliberately evoke specific responses from the brain. This research, herein termed *conventional EEG*, offers significant contributions in terms of understanding the human cognitive process. However, despite the fact that the cognitive processes are linked to the body's interaction with the environment [24], there is a limited number of studies, investigating how the brain is engaged with the cognitive activity in different real-world

recording environments or scenarios (e.g. the responses when the participants are walking). In other words, until recently most EEG research has only focused on scrutinising the cognitive process in a static setup.

Here, the recent EEG research in the field of engineering can be summarised into four categorises [9]:

- *Conventional EEG*: Research that approaches the analysis of cognition using the EEG in a static setup;
- *Portable EEG*: A recording system (i.e. electrodes, amplifiers) that is small enough to be carried around (Section 1.1.2);
- *Mobile EEG*: A technology that does not require the user to remain still; i.e., EEG studies during motion. This research addresses the cognitive process in different environments (Section 1.1.3);
- *Transparent EEG*: The *conventional EEG* $\wedge$ *mobile EEG* $\wedge$ *portable EEG*. A portable EEG recording system which is also capable of monitoring the user's brain activity in ambulatory settings, namely a feasible EEG system in the real world (Section 1.1.4).

### 1.1.2 Portable EEG

The term "*portable EEG*" refers to recording devices, including electrodes and amplifiers, that are small enough to be carried around – for example, devices that can fit inside the user's pocket. The *portable EEG* systems aim to operate in a 'stand-alone' fashion, whereby no additional equipment (like a computer) is needed to record and store the data, either through the use of an integrated data-logger, or a wireless transmission of data to an external device, such as a smartphone. The recorded signal can be sequentially analysed in real time to give feedback to the user from the device. Note that, unlike the *mobile EEG* system, the *portable EEG* does not require the system to be robust during any motion of the user.

An example of *portable EEG* devices are the commercial EEG headsets, such as Emotive EPOC+ (USA), Neurosky Brainlink (USA), CGX (USA), mBrainTrain SMARTING (Serbia), InteraXon Muse (Canada), and Neuroelectrics Starstim (Spain). Such headsets require either 1) application of a skin abrasion paste and electrolyte gel in order to obtain low impedance between the skin and electrodes (wet electrodes), 2) simple application of solution to electrodes or 3) no preparation (dry electrodes).

The advantage of wet electrodes is low skin-electrode impedance. The conventional high-density EEG headset with either 64, 128 or 256 channels is typically based on wet electrodes, and is capable of obtaining the high quality of signal required for research in cognitive neuroscience. However, the preparation of multiple electrodes and the consequent disassembly and cleaning of the patients scalp are both time-consuming and cumbersome for users. Compared to the traditional wet electrodes, the signal quality of

dry electrodes is typically lower; however the latest dry-electrode EEG headsets require shorter preparation time and have recently provided sufficient signal quality for BCI applications [25]. The main limitation of the dry EEG headset is comfort; the headsets need to be attached to the skin firmly, therefore applying constant pressure against the scalp and decreasing comfort.

Another example of a *portable EEG* system is ear-EEG – the development of both in-ear and around-ear EEG systems (discussed in detail in Section 1.2).

### 1.1.3 Mobile EEG

A "*mobile EEG*" system should enable the user to monitor their brain activity in a variety of real world environments in order to examine associated changes in motor cognition; this is termed 'active sensing' [26]. If the participants move during recordings, both 1) brain related electrical activity and 2) the electrical activity produced by muscles (e.g. eyes, neck movements) are recorded in the EEG channel. In other words, the recorded EEG signals are smeared by the artefacts. In order to minimise these motion artefacts, the EEG recordings are typically conducted in static environments which restrict the participant's movements.

Recently, both latest signal processing and advanced sensing technologies have enabled proof-of-concept *mobile EEG* studies, which are based on the replication of the conventional laboratory-based EEG recordings during motions [27]. One such *mobile EEG* project is called the mobile brain/body imaging (MoBI) approach [4, 5]. In this project, the authors aimed to monitor brain activity using a medical grade stationary EEG amplifier and a high-density on-scalp EEG headset while the participant was performing a gait motion on a treadmill. The recorded EEG signals during the movements were pre-processed and cleaned by applying signal processing algorithms, such as those based on independent component analysis (ICA) [28]; a source separation algorithm that can be used for identifying brain-related potential and separating these from motion related artefacts in the noisy measurements. These studies have also shown to be useful in identifying the characteristic artefacts arising from different real world scenarios, such as walking [29].

De Vos *et al.* utilised the latest sensing technology to conduct EEG recordings from subjects walking outside, and outdoor seated EEG recordings [30], while Zink *et al.* conducted similar EEG experiments on subjects during indoor cycling [31]. These two studies have demonstrated the feasibility of EEG monitoring in a dynamic recording setup, however, the utilised on-scalp EEG headsets cover the entire scalp and are not suited to long-term brain monitoring.

In a separate *mobile EEG* study, Liao *et al.* developed a different type of system, with dry electrodes attached to springs, with a view of increasing the headsets tolerance to motion and improving the skin-electrode contact impedance [32]. An extensive review of the *mobile EEG* applications can be found in [33].

### 1.1.4 Transparent EEG

The term "*transparent EEG*" was introduced by Bleichner and Debener [9] as "a portable, motion-tolerant, self-applicable, highly unobtrusive, near invisible, and comfortable to wear EEG system". The system should be wearable and offer ubiquity – "good technology will work seamlessly into our lives and workflows in such a way that we fail to even notice that it exists [34]" [35]. In other words, a *transparent EEG* system must:

1. Record an EEG signal-to-noise ratio (SNR) comparable to that of *conventional EEG*,
2. Utilise small sensors and amplifiers such that the EEG system is *portable*,
3. Perform robustly during motion, so that it is suited to *mobile EEG* applications.

Note that a *transparent EEG* should keep low skin-electrode impedance over a long period of time. Therefore, ideally, the wires and amplifiers should only be minimally exposed to the environment, and placed in such a way that does not impede the users everyday activity.

For example, an EEG system can be made small enough so that it can be integrated into head worn accessories such as spectacles, earphones or hearing aids. The recorded signals from these inconspicuous devices are wirelessly transmitted to a smartphone or are stored within the device. The transmitted data can be sequentially analysed in a smartphone or in a data logger, which gives feedback to the user. Figure 1.1 depicts examples of *conventional EEG*, *portable EEG*, and *mobile EEG*; Table 1.1 summarises the comparison of the different EEG systems.

**Table 1.1:** Comparison of next generation EEG systems.

| | *Conventional EEG* | *Portable EEG* | *Mobile EEG* | *Transparent EEG* |
|---|---|---|---|---|
| Properties | Clinical, high quality | Small system (e.g. electrodes, amplifier) | Robust for movement | High quality, small system, robust for movement |
| Objects | Diagnosis, cognitive, neuroimaging | EEG monitoring with small devices | Cognitive process during motion | EEG monitoring in the real world |
| Amplifier | Stationary | Small and/or wireless | - | Small and wireless |
| # of channels | Large (e.g. > 32 ch) | Small | - | Small |
| Environment | Controlled | (Normally) controlled | In the real world | In the real world |

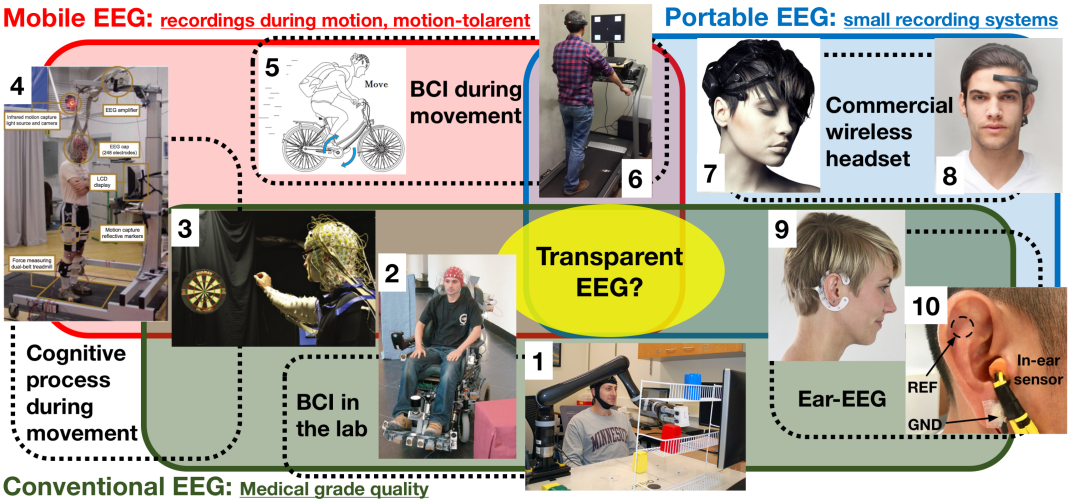



**Figure 1.1:** Examples of the *conventional EEG*, the *mobile EEG* and the *portable EEG*. The *transparent EEG* is categorised into in the middle of this diagram (Yellow circle). (1) Controlling a synthetic arm in laboratory [1, 2] and (2) controlling a wheelchair in a room [3] – BCI motor imagery tasks. (3) High-density EEG recording during dart games [4], and (4) high-density EEG recordings on the treadmill [5] – MoBI project. (5) Event related potential recordings on a fixed bike. (6) Visual evoked potential recordings during walking with a commercial headset [6]. (7) Emotiv EPOC+ [7] and (8) Neurosky Brainlink [8] – commercial wireless EEG headset. (9) The cEEGrid [9], and (10) the generic earpiece [10] – the ear-EEG technology.

## 1.2 Ear-EEG system

### 1.2.1 Original in-ear EEG system

A new concept that demonstrates the most advanced sensing technology has recently been introduced into the research community; the in-ear physiological monitor, shown in Figure 1.2 [11]. The objective of in-ear EEG is to record electrical potentials in the ear canal enabling a discreet, unobtrusive, and user-centred approach to brain monitoring. In-ear EEG promises a number of advantages over existing solutions, for example, fixed electrode locations, user comfort, and ease of use [36]. These features are all crucial in achieving long term brain monitoring in the real world. At the genesis of in-ear EEG research, the in-ear recording concept employed several standard EEG paradigms, benchmarked against conventional on-scalp EEG. The amplitude of in-ear EEG is smaller than that of on-scalp EEG, whereas the SNR of in-ear EEG is found to be similar to that of on-scalp EEG in terms of steady state responses [37, 38, 39]; therefore in-ear EEG has been validated to provide on-par signal quality, compared to conventional on-scalp EEG.

Since the initial validation studies [11, 36, 37, 38, 39], the in-ear EEG system has been further developed and validated, especially in 1) sensor developments and 2) applications.

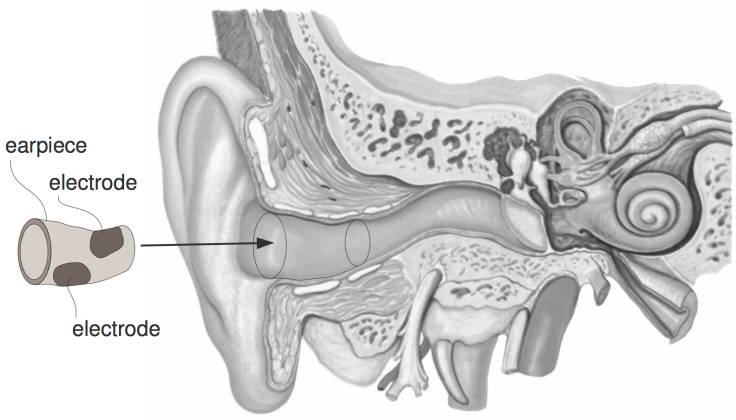



**Figure 1.2:** The concept of in-ear physiological monitoring [11].

### 1.2.2 Sensor developments and around-ear EEG system

The original in-ear sensor was in the form of a custom-made hard-shell earpiece, which is similar to an earplug used in a hearing-aid application. Even though the personalised earpiece fits tightly to the user's ear canal and enables good quality EEG recording, this custom-made earpiece, which is manufactured by 3D scanning, is not ideal in terms of the sensor production cost. In order to address the issues of cost, multiple earpieces have been introduced, such as those made by silicon [37] and using commercial earphone ear-tips [40].

Professor Mandic's group has introduced a one-size-fits-all viscoelastic earpiece; the sensor is made from an affordable/consumable standard foam earplug [41], shown in Figure 1.3 (a, b). The viscoelastic substrate expands after insertion and maintains a stable interface with the ear canal, so the electrodes fit firmly inside the ear canal. These viscoelastic earpieces are able to record both mechanical signals (e.g. pulsation in the ear) [42] and electrical signals (in-ear EEG) simultaneously. Recently, this viscoelastic in-ear sensor was also used for monitoring cardiac activity [10, 43].

In a different approach, around-ear EEG recordings have also been proposed. Norton *et al.* demonstrated mesh electrodes made by multilayers of gold electrodes, and conducted EEG recordings from the pinna [44]. Bleichner *et al.* proposed miniaturised EEG electrodes, which are able to record EEG from the concha – the around-ear EEG device is named cEEGrid [9, 12]. The sensor arrays are 3D printed, and placed around the ear. Figure 1.3 (c, d) show the around-ear EEG sensor (cEEGrid).

In this thesis, the term 'ear-EEG' refers to both in-ear EEG and around-ear EEG.

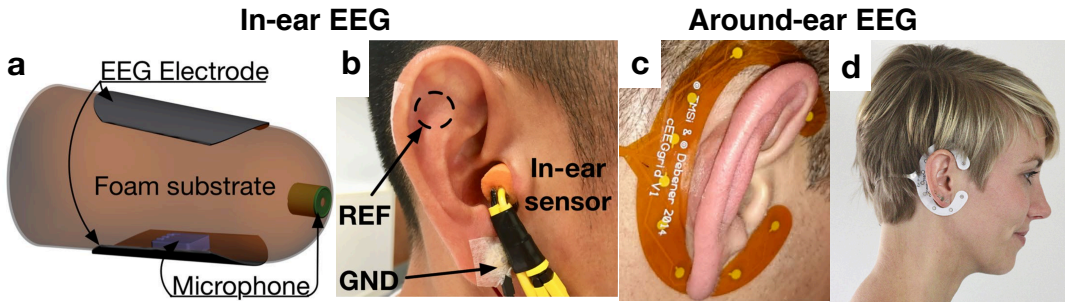



**Figure 1.3:** An example of recent ear-EEG technology. (a, b) In-ear generic earpiece; (c, d) around-ear cEEGrid. Images are adapted from [9, 10, 12].

### 1.2.3 Ear-EEG application

Before 2014, the ear-EEG research involved validation studies using several standard EEG paradigms investigating the quality of signals from the in-ear, compared to the recorded potentials from an on-scalp EEG headset. After these preliminary studies, ear-EEG application (i.e. use of ear-EEG) has been further investigated, especially by leading groups from Imperial College London (viscoelastic in-ear sensor), Aarhus University (personalised hard-shell in-ear sensor), and University of Oldenburg and University of Oxford (around-ear sensor). Several other groups have also produced ear-EEG work in the area of sensor development, investigating ear-EEG artefacts, sleep monitoring, fatigue evaluation, estimating hearing levels, biometric person authentication, and BCIs. Typically, these proposed ear-EEG applications have been validated by following the exact same or similar protocols from conventional on-scalp EEG based studies; the responses under a certain condition from the ear were compared to those from on-scalp EEG, for example, investigating the similarity of responses (correlation, coherence) in the two systems. Table 1.2 summarises the ear-EEG applications in the different categories and the research groups.

### 1.2.4 Could ear-EEG be a 'transparent EEG'?

As discussed in Section 1.1.4, the *transparent EEG* is "a portable, motion-tolerant, self-applicable, highly unobtrusive, near invisible, and comfortable to wear EEG system" [9]. Among recent EEG research in engineering (see Figure 1.1), the ear-EEG system pertains to a *portable EEG* – a technology with a recording system that is small enough to be carried around. The quality of ear-EEG (i.e. the SNR) is proven to be similar to that of conventional on-scalp EEG (*conventional EEG*). However, the ear-EEG system is only able to measure responses from the in-/around- ear, therefore designing certain application with ear-EEG is not feasible, for example an EEG-based motor imagery BCI is not feasible as the ear sensor would be too far from the motor cortex to translate motor signals. Here,

**Table 1.2:** Various ear-EEG systems and their applications.

| Before 2014 | Concept [11], overview [36], generic earpiece [37], steady-state responses [38, 39] | | | |
|---|---|---|---|---|
| From 2014 | In-ear viscoelastic (Imperial College) | In-ear hardshell (Aarhus) | Around-ear cEEGrid (Oldenburg, Oxford) | Others |
| Overview | [10] | [45], mutual information between ear and scalp [46], reference configuration [47] | [9, 12] | Behind-ear [44, 48] |
| Sensor development | Viscoelastic [41], co-located mechanical sensor [42] | High-density [49], dry electrode [50] | Before around-ear sensor [51] | Soft material [52], ear-EOG [53], heart and breathing rate monitor [54] |
| Artefacts | | Artefacts [55], epilepsy [56], seizures [57] | | Behind ear, epilepsy [58] |
| Sleep | Nap [15, 59, 60, 61], overnight [62] | Overnight [63, 64, 65] | Overnight [66, 67] | Overnight [68, 69, 70], driver drowsiness [71] |
| Fatigue | | | | Driver fatigue with Neurosky [72] |
| Hearing | | Hearing threshold [73, 74] | Auditory attention [75], speaker detection [76], recognition and ERP [77] | Hearing attention [78, 79] |
| Biometrics | [80] | | | With Neurosky [81] |
| BCI, cognitive | | P300 [82] | Visual task and ERP [83] | P300 [84, 85], with Neurosky [86] |

the properties of our in-ear EEG system, which is based on a one-size-fits-all viscoelastic earpiece (Figure 1.3 (a, b)) in outlined, in the context of *transparent EEG* applications:

- *Portability*: The size of a viscoelastic earpiece (in-ear EEG sensor) is approximately 25mm in length and 12mm in diameter, and the two electrodes are embedded on the surface of an earpiece. In terms of an amplifier, our research group has been developing and testing a custom-made portable bio-amplifier, which is the size of a matchbox and powered by a coin cell battery, called iAmp. The iAmp stores the recorded physiological data on a Micro SD card, on the other hand, the data cannot be analysed in real time.
- *Motion-tolerance*: Noise removal methods have been tested under jaw-clenching artefact conditions [10]. In terms of active sensing scenarios, for example recording during walking or cycling, the in-ear EEG system has not been validated yet. Due to the nature of the in-ear sensor, the viscoelastic substrate expands after insertion, so the electrodes fit firmly inside the ear canal; therefore, the sensor is not likely to be dislodged during subtle head movements associated with walking.
- *Ease of use*: Users are able to insert the sensor by themselves, as when wearing an earplug; however, initially, the users would need to be instructed by a trained person to apply conductive gel to the electrodes.
- *Discreet nature*: After the insertion, the sensor conforms to the shape of the ear canal, therefore the users can wear the viscoelastic in-ear sensor comfortably. The in-ear

sensor sits discreetly inside the ear canal, however the cables from the sensor to the amplifier are still visible.

- *Robustness*: For the recording time, Goverdovsky *et al.* investigated the skin-electrode impedance over a long period of time – after 8 hours, the median impedances were lower than 10 kΩ among four out of five subjects [10].

Overall, the latest in-ear system with a viscoelastic earpiece is capable of being a *transparent EEG* – brain monitoring in real world environments.

## 1.3 Problem formulation

In the area of BCIs, numerous EEG-based systems have been proposed for enhancing the user's quality of life – including controlling a prosthetic arm [2], controlling a wheelchair [3], and using a visual speller [87]. These experiments have been conducted in highly controlled environments, where the optimum quality of EEG has been recorded to faithfully investigate the relationship between the EEG responses and the participant's condition. These EEG-based applications are able to give neural feedback based on the recorded brain activity in real-time. Following these successes, EEG-based applications are investigated, emphasising the importance of the following factors:

- *Environment*: The recording setup can be applied in the real world (e.g. setup time, with/without trained person's assistance);
- *Unobtrusive*: The recording device should be comfortable to wear and should impede the user as little as possible [36];
- *Feasibility*: Maximising the tradeoff between the number of channels, the SNR, and unobtrusiveness.

To address these aspects, feasible brain monitoring systems and their applications, which are based on a readily wearable in-ear EEG sensor, are exhibited. Again, the recorded signal from the ear is spatially limited (ideally the entire cortex would be covered); therefore the feasibility of in-ear EEG applications, which can facilitate the user's quality of life, are evaluated in the two domains:

1. Automatic sleep monitoring with in-ear EEG;
2. Person authentication with in-ear EEG.

### 1.3.1 In-ear EEG for sleep monitoring

Sleep is an essential process for human well-being, and its quality reflects both a persons lifestyle as well as various medical conditions. Quality of sleep is considered one of the most important current topics in sleep medicine, and is typically investigated by recording

a persons sleep patterns in a sleep clinic. The current clinical sleep monitoring (polysomnography, PSG) represents sleep profiles of individuals rigorously and comprehensively [16], however, the expensive and cumbersome nature of PSG may even disturb patients normal sleep, thus affecting diagnosis and treatment. During the analysis, the recorded PSG data are scored manually by a trained sleep clinician; given the scale of sleep disorders in our modern society, this imposes unrealistic demands on their time and incurs economic costs. Therefore, from a point of view of continuous widely affordable healthcare in the community, conventional sleep monitoring based on the PSG is not realistic.

In the current research, to address these two issues, multiple potential solutions have been proposed, which fall under the two main categories:

- Employ standard PSG but replace the clinician with an automatic scoring system [88],
- Use wearable sensing, possibly with a reduced number of sensing modalities (e.g. only EEG or actigraphy – a wrist worn device), and perform the analysis either by the clinician or automatically based on machine learning (ML) [89].

In this thesis, with the one-size-fits-all in-ear sensing system, the feasibility of sleep monitoring in the community is validated in the following setups:

1. *Algorithms*: Develop automatic sleep staging algorithms based on EEG, and validate the performance using the publicly available sleep datasets;
2. *Wearable sensing*: Conduct sleep monitoring using viscoelastic in-ear sensors, which can be readily worn by users;
3. *Comparison with the current standard*: Record in-ear EEG and the PSG (or on-scalp EEG) simultaneously, and compare the so-obtained signals from two recording systems;
4. *Real world applications*: A 'real world' out-of-clinic overnight sleep scenario, namely in participants' homes to reflect their normal sleep patterns.

### 1.3.2 In-ear EEG for person authentication

Person authentication refers to the process of confirming the claimed identity of an individual, and is already present in many aspects of life, such as electronic banking and border control. Most extensively used recognition methods are based on knowledge (e.g. password, PIN) and tokens (e.g. passport, card) [90], however, these are also vulnerable to fraud, such as theft and forgery, and can be easily used by imposters. In contrast, biometric recognition methods (e.g. fingerprints, iris) rest upon unique physiological or behavioural characteristics of a person, thus largely overcome the above vulnerabilities.

One of the currently investigated biometric modalities is EEG; the EEG fulfils the requirements of *universality* (each person should possess the given biometric characteristic), as it can be recorded from anyone, together with the *uniqueness*. Specifically, the

individual differences of EEG alpha rhythms has been examined [91] and reported to exhibit a significant power in discriminating individuals [92] in the area of clinical neurophysiology. The brain activity is neither exposed to surroundings nor possible to be captured at a distance, therefore the brain patterns of an individual are robust to forgery, unlike face, iris, and fingerprints. The EEG is therefore more robust against imposters attacks than other biometrics. However, in order to utilise EEG signals in the real world, several key properties such as *permanence* (the biometric characteristic should neither change with time nor be alterable) and *collectability* (the characteristic should be readily measurable by a sensor and readily quantifiable) must be further addressed.

In this thesis, EEG-based biometrics are established and their feasibility is validated in the following setups:

1. *Collectability*: Biometrics verification is evaluated with a wearable and easy to set up in-ear sensor (in-ear EEG),
2. *Uniqueness* and *permanence*: These issues are addressed through subject-dependent EEG features which are recorded over temporally distinct recording days.

Note that the in-ear EEG biometrics is still in a proof-of-concept stage as the data collection was undertaken in the research lab, which is a controlled environment.

## 1.4 Contribution of the thesis

The contributions of this thesis can be divided into three parts, and these correspond to the organisation of this thesis, as shown in Figure 1.4 below.

#### Part I: State-of-the-art automatic sleep stage classification

For clinical diagnostic purposes, the PSG, which is based on a multitude of physiological responses, including the EEG, EOG, and EMG, has been extensively utilised to monitor sleep. While the PSG is able to faithfully reflect human sleep patterns, the scoring process is expensive as it involves time-consuming manual scoring by a trained clinician. To alleviate this problem, computer based sleep staging has been developed in order to minimise analysis time. Automatic sleep stage classification from EEGs using complexity science methods is conducted. The proposed method applies fuzzy entropy and permutation entropy for multi-scale entropy analysis. As a state-of-the-art, such classification methods are validated using the Sleep-EDF [93] dataset, and outperform methods proposed in previous studies. The same classification methods are applied for further validation in a 'big data' scenario – using the PSG recordings from 515 individuals [94].

*Relevant publications:*

**T. Nakamura**, T. Adjei, Y. Alquarashi, D. Looney, M. J. Morrell, and D. P. Mandic, "Complexity science for sleep stage classification from EEG," in *Proc. IEEE Int. Joint Conf.*

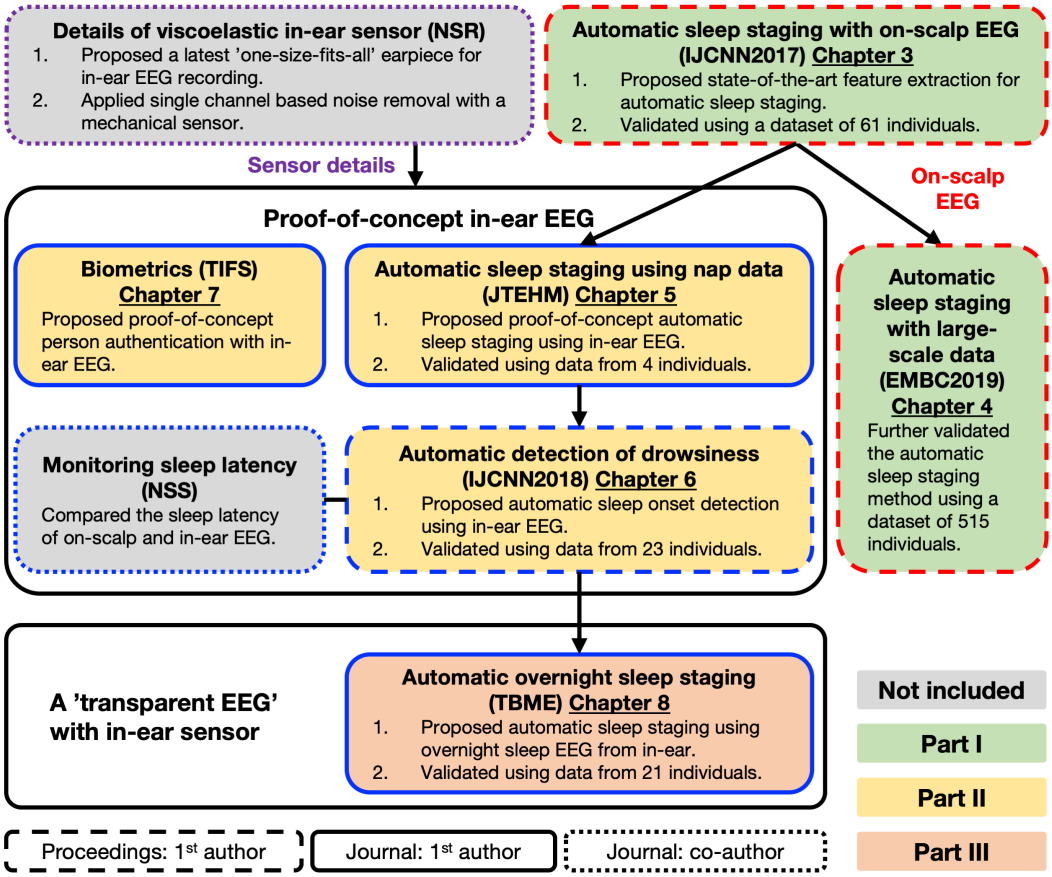



**Figure 1.4:** Publications and thesis structure.

*on Neural Netw. (IJCNN)*, pp. 4387-4394, 2017.

**T. Nakamura**, H. J. Davies and D. P. Mandic, "Scalable automatic sleep staging in the era of Big Data," in *Proc. IEEE Int. Eng. in Med. and Biology Conference (EMBC)*, 2019 (to appear).

### Part II: Proof-of-concept in-ear EEG applications

The concept of ear-EEG was originally proposed in 2011 [11]. This novel physiological sensing idea has been validated in several different ways, including comparing recorded steady-state responses from in-ear to those from on-scalp [36, 39]. After these validation studies, the quality of ear-EEG has been demonstrated to be on a par with that of conventional on-scalp EEG. Since then, multiple ear-EEG based applications have been proposed, and are shown in Table 1.2. The main challenge of ear-EEG application is the trade-off between the performance (e.g. classification, prediction, decoding accuracy) and the unobtrusiveness (impeding the user's activities as little as possible). In other words, the user's physiological condition is only examined from the in-ear responses. In this

thesis, in-ear EEG applications are developed, particularly in the area of sleep and person authentication. First of all, the exact same (or similar) experiments to previous on-scalp based EEG applications presented in the literature are designed to conduct experiments, which monitors EEG from in-ear locations (or from both on-scalp and in-ear). The aim of such proof-of-concept studies is to replicate and compare the results to the previous on-scalp EEG studies. Through comparison to the previous studies, the feasibility of our in-ear EEG application is evaluated in the area of sleep and person authentication.

*Relevant publications:*

**T. Nakamura**, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, "Automatic sleep monitoring using ear-EEG," *IEEE Journal of Translational Engineering in Health and Medicine (JTEHM)*, vol. 5, Article no. 2702558, 2017.

**T. Nakamura**, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Automatic detection of drowsiness using in-ear EEG," in *Proc. IEEE Int. Joint Conf. on Neural Netw. (IJCNN)*, pp. 5569-5574, 2018.

**T. Nakamura**, V. Goverdovsky, and D. P. Mandic, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 13, no. 3, pp. 648-661, 2018.

**Part III: In-ear EEG application towards 'transparent EEG'**

In Part I and II, 1) an automatic sleep stage classification with the state-of-the-art performance, and 2) a sleep monitoring with the in-ear EEG are proposed. Combining the two findings, the feasibility of sleep monitoring in the community is further investigated using an in-ear sensor. Overnight sleep monitoring with simultaneous standard PSG and in-ear EEG is conducted in participants' homes; such familiar environments are a key factor to achieving truly representative sleep monitoring, as this minimises the stress and inconvenience of the participants while maximising the likelihood of exhibiting usual sleep patterns. Through a rigorous comparative examination of performance metrics of an automatic sleep staging method based on in-ear EEG and the manually scored hypnogram (based on the standard PSG), the feasibility of automatic overnight sleep monitoring in out-of-clinic scenarios with readily deployable one-size-fits-all in-ear sensors is confirmed.

The overnight sleep study with viscoelastic earpieces mostly fulfils the requirements for *transparent EEG – Portable*: the sensor is only located the in-ear; *Motion-tolerant*: the sensor fits tightly and is not dislodged during the night; *Self-applicable*: the viscoelastic in-ear sensor can be easily applied by the user; *Highly unobtrusive*: the sensor does not disturb the subject's sleep; *Near invisible*: the in-ear sensor is much smaller than the conventional PSG; and *Comfortable to wear*: the sensor is capable of monitoring the quality of EEG during the entire night. Consequently, the proposed in-ear EEG based sleep monitoring can be

a proxy for a 24/7 continuously wearable alternative to conventional cumbersome and expensive sleep monitoring. As a part of *transparent EEG*, this in-ear EEG technology promises to be a viable method for readily wearable sleep monitoring in the community, a key to affordable and future healthcare.

*Relevant publications:*

**T. Nakamura**, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Hearables: Automatic overnight sleep monitoring with standardised in-ear EEG sensor," *IEEE Transactions on Biomedical Engineering (TBME)* (in press).

## 1.5 Thesis organisation

Figure 1.4 illustrates an overview of this thesis, which is organised as follows:

**Chapter 2: Algorithms and preliminaries**

This chapter outlines the signal processing and ML algorithms used in the automatic sleep staging for both on-scalp EEG and in-ear EEG. The combination of the selected feature extraction and classifier yields the state-of-the-art automatic sleep staging performance in benchmarking datasets. The structural complexity and frequency domain features are calculated in this study, whereas the classification method is based on supervised ML algorithms, including the support vector machine (SVM) and the extreme learning machine (ELM). In order to incorporate the transitional information in the automatic sleep staging, a smoothing algorithm based on a hidden Markov model (HMM) is added.

**Part I: State-of-the-art automatic sleep stage classification**

**Chapter 3: Structural complexity analysis for automatic sleep staging**

This chapter shows the possibility of automatic sleep stage classification based on structural complexity analyses. The classification performance is evaluated using a benchmarking publicly available sleep dataset, PhysioNet Sleep-EDF [93], which contains the overnight on-scalp EEG recordings from 61 individuals. The proposed methods outperform other existing studies in the literature, thus the combination of two features, 1) structural complexity features (multi-scale entropy) and 2) non-linear frequency domain features (spectral edge frequency), are shown to be effective for classifying sleep stages from a limited number of EEG channels.

**Chapter 4: Automatic sleep staging for large scale data**

Effective feature extraction for automatic sleep staging is proposed in Chapter 3; however, in the real world scenario, the scalability and computational cost of the classifier are as important as the prediction accuracy, especially for the large-scale data analysis. In this

chapter, the computationally efficient classifier, the ELM, is introduced. A publicly available large-scale PSG dataset, recorded from 515 individuals [94] is employed to evaluate the performance. The classification performance is compared with multiple validation setups, for example changing the number of training subjects. The chapter concludes by outlining the recent automatic sleep staging methods based on 'deep learning'.

**Part II: Proof-of-concept in-ear EEG applications**

**Chapter 5: Automatic sleep staging using in-ear EEG for nap data**

In Part I, the automatic sleep stage classification methods using structural complexity analyses is proposed, and their performance is evaluated using publicly available on-scalp sleep EEG datasets. Therefore, the automatic sleep staging can be a proxy for time-consuming, conventional, manual sleep scoring by clinician. Next, the issues related to sleep monitoring are resolved. Currently, the PSG is a standard methodology to monitor sleep in a clinical setting. The sleep profiles of individuals based on the PSG are rigorous and comprehensive, however, the patients' normal sleep can be impeded by the PSG recordings. To alleviate the problem, sleep monitoring with wearable devices, for example wrist worn actigraphy, has been examined in the research community. In this chapter, as a proof-of-concept, automatic wearable sleep monitoring in the community is investigated by analysing the agreement between the automatically predicted sleep stages by in-ear EEG and on-scalp EEG patterns.

**Chapter 6: Automatic detection of drowsiness using in-ear EEG**

In the previous chapter, the feasibility of the in-ear sensor in a sleep monitoring application is validated and shown that the in-ear sensor is capable of monitoring sleep-related physiological responses. In this chapter, the in-ear sensing application in the area of daytime sleepiness is further investigated. Sleepiness is a symptom and a subjective parameter (dependant on how the individual perceives the intensity of the urge to fall asleep). The state of being sleepy appears to be linked to drowsiness, which is the transition from wakefulness to light sleep. Daytime nap data using on-scalp EEG headset and in-ear EEG sensor is recorded simultaneously to validate whether in-ear EEG patterns can be used to discriminate states of wakefulness and light sleep.

**Chapter 7: Person authentication using in-ear EEG**

This chapter focuses on a different aspect of user well-being. The person authentication paradigm based on EEG (EEG biometrics) is examined. Biometric recognition methods (e.g. fingerprints, iris, face), which are based on unique physiological or behavioural characteristics of a person, have recently become more accessible with the advanced pattern recognition and sensing technologies. In the research community, there are multiple proof-of-concept biometrics based on on-scalp EEG; however, a technical problem still remains

for deployment in real world scenarios – the cumbersome and obtrusive nature of on-scalp EEG. The chapter first summarises the concept of EEG biometrics and the existing standard approaches, and formalises the research objective of readily applicable EEG biometrics using an in-ear sensor. Comprehensive analysis over a number of subjects, setups, and analysis features demonstrates the feasibility of the proposed in-ear EEG biometrics, and its potential in resolving the critical collectability, robustness, and reproducibility issues associated with current EEG biometrics.

**Part III: In-ear EEG application towards 'transparent EEG'**

**Chapter 8: Automatic overnight sleep monitoring using in-ear EEG**

In Part II, the feasibility of multiple in-ear EEG application in the area of sleep and person authentication is validated. These proof-of-concept in-ear EEG applications contribute to the *Portable EEG* research, whereby the recording devices are small enough to be worn around. The in-ear EEG recordings in Part II are conducted inside the research laboratory, which is a controlled environment. In this chapter, the feasibility of in-ear EEG sleep monitoring is further investigated in out-of-laboratory scenarios. To this end, both conventional PSG and in-ear EEG are recorded during the night, in the subjects' homes. The simultaneous EEG from on-scalp and in-ear are used for automatic sleep staging based on the proposed techniques in Chapter 3, whereas the corresponding hypnogram is manually scored based on the PSG recordings. From the agreement between automatic scoring based on in-ear EEG and manually scored hypnogram based on the PSG, it is found that the in-ear sensor is capable of monitoring overnight sleep outside the sleep laboratory and mitigates technical difficulties associated with PSG. This technology promises to be a viable method for readily wearable sleep monitoring in the community as an example of a *transparent EEG* system, a key to affordable and future healthcare.

**Chapter 9: Conclusions and future work**

This chapter outlines the research findings from this thesis and summarises future works.

# Chapter 2

# Algorithms and preliminaries

## Overview

In this chapter, algorithms for the automatic sleep staging are outlined, including 1) feature extractions, 2) classifiers, and 3) metrics for performance evaluation. Table 2.1 summarises the algorithms used in different chapters in this thesis. For a full list of acronyms and their meanings, please see List of Abbreviations in the preamble.

**Table 2.1:** The details of feature extraction and classification methods in the chapters.

| Chapter | Dataset | Details | | Feature extraction | Classifier |
|---|---|---|---|---|---|
| 3 | Sleep-EDF | Overnight | On-scalp | MSE + SEF | SVM |
| 4 | CCSHS | Overnight | On-scalp | MSE + SEF (Setup-N)<br>Time + Frequency (Setup-S) | ELM (SVM, RF, LDA)* |
| 5 | Private | Nap | In-ear | MSE + SEF | SVM |
| 6 | Private | Nap | In-ear | MSE | SVM |
| 8 | Private | Overnight | In-ear | MSE + SEF | SVM |
| 7 | Private | Biometrics | In-ear | AR + Frequency | Cosine distance (SVM, LDA) |

**Dataset**: Sleep-EDF [93], CCSHS (Cleveland Children's Sleep and Health Study) [95].
**Feature extraction**: MSE (multi-scale entropy), SEF (spectral edge frequency), AR (autoregressive model).
**Classifier**: SVM (support vector machine), ELM (extreme learning machine), RF (random forest), LDA (linear discriminant analysis).
*: HMM (hidden Markov model) is used as smoothing algorithm.

## 2.1 Feature extraction

### 2.1.1 Structural complexity analyses for automatic sleep staging

#### Multi-scale entropy (MSE)

The multi-scale entropy (MSE) method [96] measures the amount of structural complexity in a time series. The MSE can be calculated using different types of entropy, such as approximate entropy and sample entropy, from multiple coarse-grained time series. In the automatic sleep stage classification, the multi-scale sample entropy (MSSE) has previously been used [97]. In this thesis, the fuzzy entropy (FE) [98] and the permutation entropy (PE) [99] are used for MSE calculation.

Here $1:N$ denotes the sequence $1, 2, ..., N$. Given an EEG signal with $N$ data points, $x(i)$ where $i = 1:N$, a coarse-grained time series $y^{(\tau)}(j)$ is generated, where $\tau$ is the scale factor. This is achieved by dividing a given EEG signal into non-overlapping windows of length $\tau$, and by averaging within the divided window as

$$y^{(\tau)}(j) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x(i), \quad 1 \leq j \leq M_\tau, \tag{2.1}$$

where $M_\tau = \lfloor \frac{N}{\tau} \rfloor = |y^{(\tau)}|$ is the length of the obtained coarse-grained time series.

Therefore, $y^{(1)}(j)$ corresponds to the original EEG signal, while the length of the coarse-grained time series is simply the length of the original EEG signal $N$ divided by $\tau$. In this thesis, the scale parameter is selected between 1 to 30, $\tau = 1 - 30$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 30$) for multi-scale fuzzy entropy, whereas $\tau = 1 - 20$ for multi-scale permutation entropy.

**Multi-scale fuzzy entropy (MSFE)**

The FE [98] is used in this thesis as it is robust to noisy measurements and relatively consistent; this consistency makes it well suited to relatively short physiological signals, which in turn makes it appropriate for use with small embedding dimensions.

For the sample entropy (SampE) [100], the similarity between two vectors is calculated by a Heaviside function. Here, the similarity between a $m$-dimensional vector $Y_m(i) \in \mathbb{R}^m$ and another vector $Y_m(j) \in \mathbb{R}^m$ is evaluated by a given threshold $r$, whether the maximum norm distance between two vectors is smaller or larger than the threshold as

$$\mu(d_{ij}, r) = \begin{cases} 1 \ (d_{ij} \leq r) \\ 0 \ (d_{ij} > r), \end{cases} \tag{2.2}$$

where

$$d_{ij}^m = d[Y_m(i), Y_m(j)] = \max_{l=1,\ldots,m} (|y(i+l-1) - y(j+l-1)|) \tag{2.3}$$

denotes the maximum norm distance between two vectors with delays. $m$ is typically selected between 2 to 7.

For the FE, the degree of similarity is evaluated based on a fuzzy membership function, which is a 'smooth' curve with respect to the distance between the two delay vectors (*cf.* sharp discontinuity of a Heaviside function in the SampE). In other words, as the distance between $Y_m(i)$ and $Y_m(j)$ increases, the degree of similarity between the two vectors decreases gradually to zero. The multi-scale fuzzy entropy (MSFE) is calculated in the similar way to the MSSE; an entropy is calculated for each coarse-grained time series.

First, the coarse-grained time series is calculated from an input EEG signal $x(i)$ with

$N$ data points by (2.1) with a given scale parameter $\tau$. For the coarse-grained time series, $y^{(\tau)}(i)$ where $i = 1 : M_\tau$, with the length of $M_\tau$ and a given parameter $m$, a delay vector is computed by removing the baseline (local mean) as follows

$$Y_m^{(\tau)}(i) = \left[ y^{(\tau)}(i) - \overline{y^{(\tau)}(i)}, y^{(\tau)}(i+1) - \overline{y^{(\tau)}(i)}, ..., y^{(\tau)}(i+m-1) - \overline{y^{(\tau)}(i)} \right], \tag{2.4}$$

where

$$\overline{y^{(\tau)}(i)} = \frac{1}{m} \sum_{j=0}^{m-1} y^{(\tau)}(i+j). \tag{2.5}$$

Repeat $M_\tau - m + 1$ times to compute all delay vectors. The distance (the maximum norm) between any two delay vectors with the length of $m$, $Y_m(i)$ and $Y_m(j)$, is computed by (2.3) as

$$\begin{aligned} d_{ij}^{(\tau),m} &= d[Y_m^{(\tau)}(i), Y_m^{(\tau)}(j)] \\ &= \max_{l=1,...,m} \left[ |y^{(\tau)}(i+l-1) - \overline{y^{(\tau)}(i)} - y^{(\tau)}(j+l-1) + \overline{y^{(\tau)}(j)}| \right]. \end{aligned} \tag{2.6}$$

With a given delay vector $Y_m^{(\tau)}(i)$ and a tolerance parameter (threshold) $r$, the degree of similarity $D_{ij}^{(\tau),m}$ between the another delay vector $Y_m^{(\tau)}(j)$ is calculated using a fuzzy membership function as

$$D_{ij}^{(\tau),m}(r) = \mu(d_{ij}^{(\tau),m}, r) \tag{2.7}$$

$$= \exp\left( -\frac{\left(d_{ij}^{(\tau),m}\right)^2}{2r^2} \right). \tag{2.8}$$

In this thesis, the Gaussian function is used as a fuzzy membership function. Here $\phi^{(\tau),m}(r)$ define the function to be,

$$\phi^{(\tau),m}(r) = \frac{1}{M_\tau - m} \sum_{i=1}^{M_\tau - m} \left( \frac{1}{M_\tau - m - 1} \sum_{j=1, j\neq i}^{M_\tau - m - 1} D_{ij}^{(\tau),m}(r) \right). \tag{2.9}$$

The $\phi^{(\tau),m}(r)$ represents the probability that any two delay vectors are similar in the dimension $m$. (Namely, the $\phi^{(\tau),m+1}(r)$ is the probability that any two delay vectors will be similar in the dimension $m+1$.) Finally, the original MSSE is calculated as the negative of the logarithm of the conditional probability; the probability, that the two delay vectors close to each other in an $m$-dimensional space, will also be close to each other when the

dimensionality is increased by one. Then the MSFE is computed as

$$MSFE(\tau, m, r, N) = -\ln\left[\frac{\phi^{(\tau),m+1}(r)}{\phi^{(\tau),m}(r)}\right]. \tag{2.10}$$

**Multi-scale permutation entropy (MSPE)**

In this thesis, PE [99] is calculated for each coarse-grained time series, which is called the multi-scale permutation entropy (MSPE) [13]. The PE is a metric for detecting dynamical changes and for estimating the information contained in a time series based on comparing consecutive values of a time series. Compared to other entropy metrics, the PE requires less computational time and is robust to noise in the measurements; hence, the method is suited to time series with poor stationarity characteristics, such as physiological signals [101].

First, the coarse-grained time series is calculated from an input time series $x(i)$ with $N$ data points by (2.1) with a given scale parameter $\tau$. For the coarse-grained time series, $y^{(\tau)}(i)$ where $i = 1 : M_\tau$, with the length of $M_\tau$, the series of vectors are obtained with given embedding dimension $d$ and time lag $L$

$$v_d^{(\tau)}(i) = \left[y^{(\tau)}(i), y^{(\tau)}(i+L), ..., y^{(\tau)}(i+(d-1)L)\right]. \tag{2.11}$$

The so-obtained vector $v_d^{(\tau)}(i)$ is arranged in increasing order; in other words, sorted in ascending order as

$$v_d^{(\tau)}(i)' = \left[y^{(\tau)}(i+(j_1-1)L) \leq y^{(\tau)}(i+(j_2-1)L) \leq ... \leq y^{(\tau)}(i+(j_d-1)L)\right]. \tag{2.12}$$

From the sorted time series $v_d^{(\tau)}(i)'$, the symbolic sequence $[j_1, j_2, ..., j_d]$ is obtained. Note that for any sequence of length $d$, there are $d!$ possible symbolic sequences – let the symbolic sequences call motif $s_k$ where $k = 1 : d!$. Figure 2.1 illustrates the example of motif extraction from a time series with given embedding dimension $d = 3$. For each motif $s_k$, the frequency of occurrence, $f(s_k)$, in the time series with the length of $M_\tau$ is calculated. When the time lag is $L = 1$, the relative frequency of each motif is therefore,

$$p(s_k) = \frac{f(s_k)}{M_\tau - d + 1}. \tag{2.13}$$

The PE for the time series is defined as the Shannon entropy for the $d!$ motifs, which can be expressed as

$$H(d) = -\sum_{k=1}^{d!} p(s_k) \ln p(s_k), \tag{2.14}$$

while the normalised MSPE entropy is given by

$$0 \leq MSPE(\tau, d, L = 1, N) = \frac{H(d)}{\ln d!} \leq 1. \tag{2.15}$$

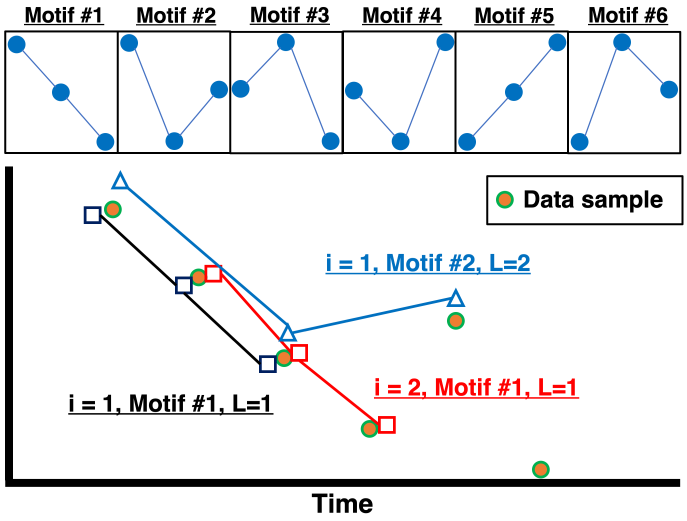



**Figure 2.1:** Extraction of motif for the permutation entropy with the embedding dimension $d = 3$. Figure is adapted and modified from [13].

### MSE for sleep analysis

The MSE method is employed to quantify the amount of structure (correlation) in time series, often referred to as the underlying system complexity. The complexity is small for both deterministic (predictable) and uncorrelated random (unpredictable) signals evaluated over larger scales $\tau$, whereas it is large for correlated (linear and/or nonlinear) stochastic processes [102]. Therefore, multi-scale sample entropy values for uncorrelated white noise monotonically decrease as scale increases, whilst the entropy values for $1/f$ noise remain constant over multiple scales [96].

The motivation of MSE analysis for automatic sleep staging is to measure the degree of complexity of an EEG signal, which represents the depth of sleep (sleep stages). As standard clinical methodology, the recorded data are scored manually according to a sleep staging criteria, briefly summarised in Table 2.2. From the table, the sleep stage is determined by multiple factors, and is not based solely on one frequency or time domain component. Figure 2.2 illustrates different EEG sleep features, including alpha, theta, K-complex, and delta activities, from an on-scalp and in-ear EEG channel. For instance, the high frequency components ($\alpha$ activities) are observed in wakefulness, whereas the dominant $\delta$ activities are observed in the N3 sleep stage. Consequently, the lower complexity in the N3 sleep stage than other sleep stages is observed in the MSPE analysis, shown in

Figure 2.3.

**Table 2.2:** Sleep stage scoring guideline for EEG (Adapted from [16]).

| Stage | Details |
|---|---|
| W | $\alpha$ = 8-13 Hz activity. low-voltage and mixed-frequency activity (4-7 Hz). |
| N1 | Low-voltage and mixed-frequency activity. |
| N2 | Appearance of sleep spindles and/or K complexes. Sleep spindles and K complexes each must last 0.5 s. < 20 % of the epoch may contain high-voltage activity (> 75 $\mu$V, < 2 Hz). |
| N3 | > 20 % of the epoch consists of high-voltage (> 75 $\mu$V) and low-frequency (< 2 Hz) activity. |
| REM | Relatively low-voltage and mixed-frequency. |

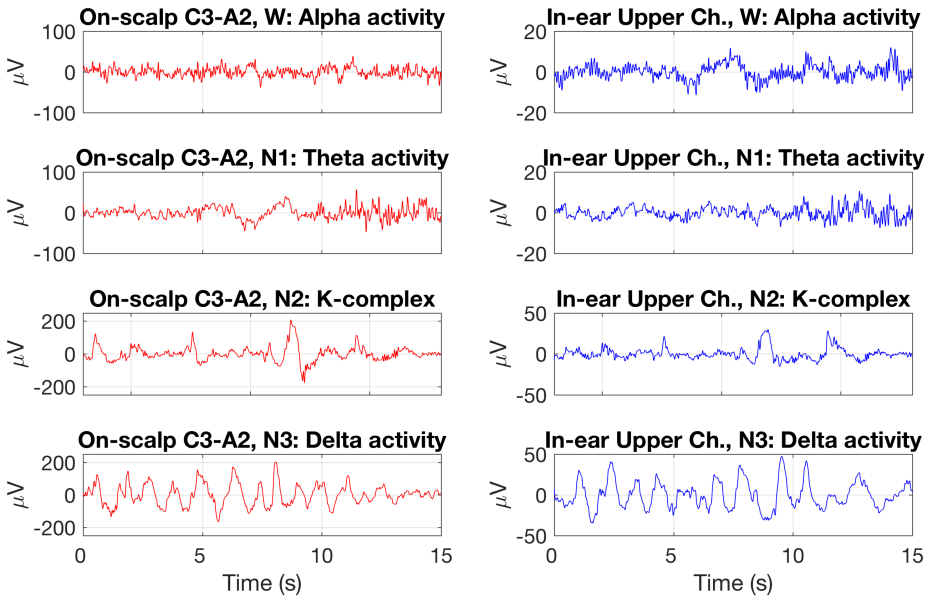



**Figure 2.2:** EEG recordings from a single subject in different sleep stages. (Red) The on-scalp C3-A2 channel; (Blue) the in-ear upper channel.

### 2.1.2 Frequency domain features for automatic sleep staging

#### Spectral edge frequency (SEF)

The $r\%$ spectral edge frequency (SEF) is the frequency below which $r\%$ of the signal power is contained; in other words, SEF$r$ corresponds to $r\%$ of the signal power and is given by

$$\sum_{f=f_{\text{low}}}^{f_{\text{high}}} |\text{magnitude}(f)|^2 \times \frac{r}{100} = \sum_{f=f_{\text{low}}}^{\text{SEF}r} |\text{magnitude}(f)|^2.$$

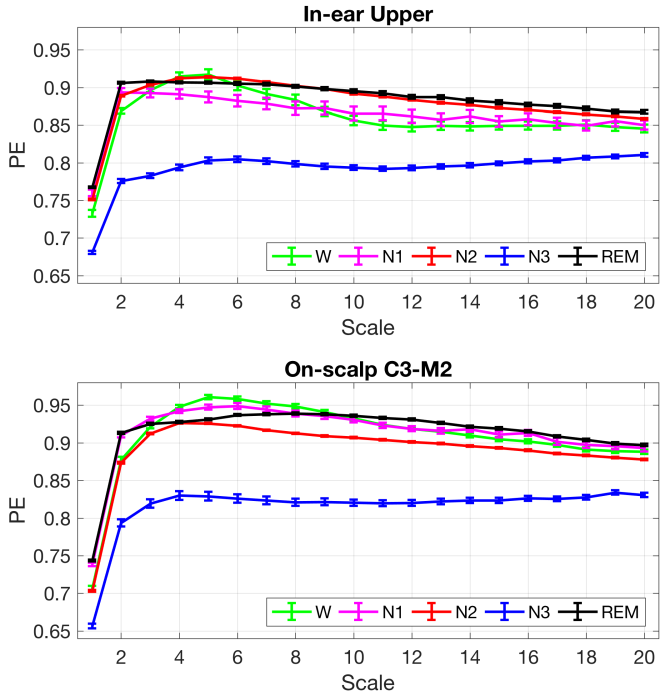



**Figure 2.3:** Averaged MSPE for the overnight sleep of one participant for an in-ear upper channel (Top) and an on-scalp C3-A2 channel (Bottom) over different sleep stages with scale $\tau = 1 - 20$. The error bars represent the standard error.

Imtiaz *et al.* [103] used the difference between SEF95 and SEF50, called SEF$d$, in order to detect the REM stage effectively; SEF$d$ = SEF95 - SEF50. Figure 2.4 shows the SEF analysis between 4-16 Hz.

The SEF has been employed to analyse physiological responses, specifically for sleep EEG [104]. Figure 2.5 shows SEF$d$ analysis between 0.5-16 Hz in the Fpz-Cz channel of one subject (Subject ID: sc4001e0) from the Sleep-EDF dataset, and the corresponding manually scored overnigth hypnogram. The SEF$d$ decreases when the subject is in a deep sleep (i.e. SWS or sleep stage 3) whereas it increases during REM sleep. Compared to the normal spectral features, the SEF analysis is more robust for characterising different sleep stages for two reasons:

- The window size (epoch size) of sleep stage analysis is 30 s according to the current scoring criteria [105]. Abrupt electrode noise is occasionally recorded in 30 s epochs, which can be ignored and does not affect the manual scoring; however, the abrupt and high amplitude noise smears the spectral features, and especially changes the spectral power value. In contrast, the SEF values are calculated from the frequency value, not the spectral power. Therefore, the SEF feature is robust to noisy recording

**Table 2.3:** Parameters for feature extraction. (Left) Setup-N; (Right) Setup-S.

*Setup-N*

| Domain | Details | $D$ | Frequency band |
|---|---|---|---|
| Structural complexity | MSE* | 20/30* | $\delta-\beta$ |
| | | | $\delta-\beta$ |
| Frequency | SEF | 3 | $\delta-\beta$ |
| | | 3 | $\delta-\alpha$ |
| | | 3 | $\theta$ |
| | | 3 | $\alpha$ |
| | | 3 | $\alpha_{\text{low}}$ |
| | | 3 | $\alpha_{\text{high}}$ |
| | | 3 | $\beta$ |
| Total | | 41/51* | (1 channel) |
| | | 82/102* | (2 channels) |

MSE*: MSPE ($D = 20$), 41 features in total
MSE*: MSFE ($D = 30$), 51 features in total

*Setup-S*

| Domain | Details | $D$ | Frequency band |
|---|---|---|---|
| Time | Line length | 2 | $\delta-\beta$ |
| | Kurtosis | 2 | |
| | SampE | 2 | |
| Frequency | Power features | 4 | $\delta/(\delta-\beta)$ |
| | | 4 | $\theta/(\delta-\beta)$ |
| | | 4 | $\alpha/(\delta-\beta)$ |
| | | 4 | $\delta/\theta$ |
| | | 4 | $\delta/\alpha$ |
| | | 4 | $\theta/\alpha$ |
| | Kurtosis features | 1 | $\delta$ |
| | | 1 | $\theta$ |
| | | 1 | $\alpha$ |
| | | 1 | $\sigma$ |
| Total | | 34 | |

conditions.

- The EEG amplitude becomes larger when the skin-electrode impedance is higher. For overnight recordings from a single subject, the skin-electrode impedance is not same over multiple recordings; In addition, the skin-electrode impedance may change across the course of a recording. Consequently, the SEF feature is robust against the skin-electrode impedance problem.

In this thesis, the frequency ranges for SEF50, SEF95, and SEF$d$ were chosen in the following bands; $\delta-\beta$ = 0.5-30 Hz, $\delta-\alpha$ = 0.5-16 Hz, $\theta$ = 4-8 Hz, $\alpha$ = 8-15 Hz, $\alpha_{\text{low}}$ = 8-11 Hz, $\alpha_{\text{high}}$ = 11-15 Hz, and $\beta$ = 16-30 Hz.

The above frequency bands for sleep analysis were selected in two steps.

- First, the target frequency ranges were chosen based on sleep staging criteria [105], summarised in Table 2.2. The frequency bands given in the table can be categorised into 1) low-frequency ($\delta$), 2) mixed-frequency ($\theta$), and 3) $\alpha$ band. In [19], the dominant typical brain rhythms are given as 1) $\delta$ = 0.5-4 Hz, 2) $\theta$ = 4-8 Hz, 3) $\alpha$ = 8-13 Hz, and 4) $\beta$ = 13-30 Hz. In addition, the frequency of sleep spindles, a signature response of N2 sleep, is 11-15 Hz. Consequently, without overlaps, the five frequency ranges are selected as 1) $\delta$ = 0.5-4 Hz, 2) $\theta$ = 4-8 Hz, 3) $\alpha$ = 8-15 Hz for sleep spindle, and 4) $\beta$ = 15-30 Hz.

- Second, the automatic classification performance is compared using the SEF features of the selected frequency bands, extracted from the publicly available sleep-EDF dataset [93]. There are three findings from this analysis: 1) the SEF features of $\delta$ bands negatively influence the classification model (lower accuracy), 2) adding the SEF features of whole frequency bands $\delta-\beta$ and $\delta-\alpha$ increases the classification accuracy, and 3) in addition to the whole $\alpha$ band, utilising two additional narrow alpha bands, $\alpha_{\text{low}}$ = 8-11 Hz and $\alpha_{\text{high}}$ = 11-15 Hz, improves the prediction performance. Through the analysis, the frequency ranges for automatic sleep staging

are selected; the same SEF features are used in the course of automatic sleep stage experiments in this thesis.

Table 2.3 summarises the details of feature extraction for automatic sleep staging in this thesis. *Setup-N* denotes the combination of MSE and SEF features, whereas *Setup-S* denotes the same feature extraction technique in [106]. In Chapter 4, the classification performance of automatic sleep staging based on the two methods, *Setup-N* and *Setup-S*, are compared.

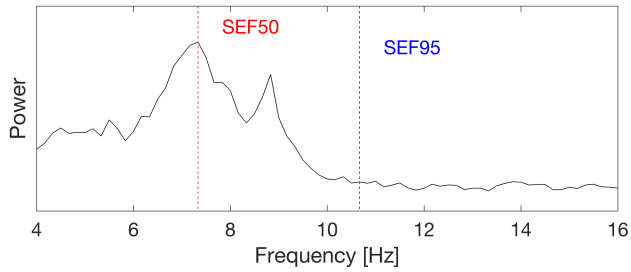



**Figure 2.4:** SEF at 50 % and 95 % of the total signal power between 4-16 Hz.

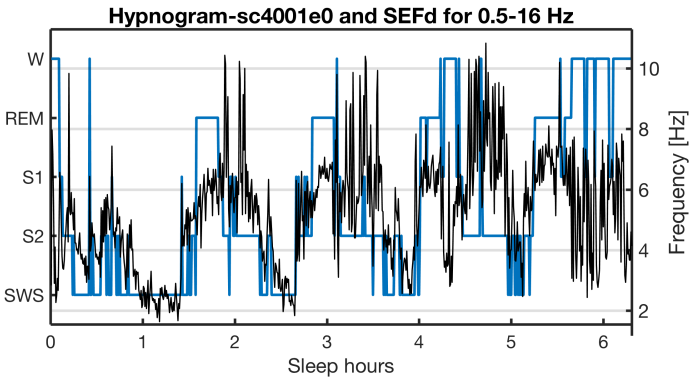



**Figure 2.5:** (Black) SEF$d$ feature between 0.5-16 Hz in Fpz-Cz channel of one subject (Subject ID: sc4001e0) from the Sleep-EDF dataset; (Blue) The manually scored hypnogram.

### 2.1.3 Other features for automatic sleep staging

#### Time and frequency domain features

In Chapter 4, two channels of EEG (i.e. C3-A2 and C4-A1 montage) are used for automatic sleep staging classification. For *Setup-S*, $D = 34$ features are extracted from each 30 s EEG epoch. As the time domain features, line length, kurtosis, and SampE are calculated; six features (three features $\times$ two EEG channels) are extracted. Given a single channel of EEG data with $N = 30 \times f_s$ data points with the sampling frequency $f_s$, $x(i)$ where $i = 1 : N$,

line length is calculated as $\frac{1}{N-1}\sum_{i=1}^{N-1}|x(i+1)-x(i)|$. As the frequency domain features, the following bands are considered: 1) $\delta$ = 0.5-4 Hz, 2) $\theta$ = 4-8 Hz, 3) $\alpha$ = 8-12 Hz, and 4) $\delta-\beta$ = 0.5-20 Hz. First, the spectrograms of two EEG channels are calculated using multitaper method with the window length of 2 s and 50 % window overlap. Therefore, for each 30 s epoch, there are 29 bins in the time domain ((30-2)/1 + 1 = 29). The so-obtained spectrograms are averaged from two EEG channels. For each averaged spectrogram, four features (95th percentile, minimum, mean, and standard deviation of band power) are culculated in given frequency bands: 1) $\delta/(\delta-\beta)$, 2) $\theta/(\delta-\beta)$, 3) $\alpha/(\delta-\beta)$, 4) $\delta/\theta$, 5) $\delta/\alpha$, and 6) $\theta/\alpha$. In total, 24 features (four features $\times$ six bands) are extracted as frequency power. Additionally, kurtosis of a given band in the spectrogram is calculated: 1) $\delta$, 2) $\theta$, 3) $\alpha$, and 4) $\sigma$ = 12-20 Hz. In total, four features are extracted as frequency kurtosis. Overall, 34 features (6 + 24 + 4) are extracted for *Setup-S* in Chapter 4.

## 2.2 Classification

### 2.2.1 Support vector machine (SVM) for classification

Support vector machine (SVM) is a popular ML method for classification [107]. The SVM is a generalisation of a simple classifier called the maximal margin classifier, which requires that the classes are separable by a linear boundary [108]. As an extension of the maximal margin classifier, the SVM can accommodate non-linear class boundaries. The SVM is originally designed by a binary classification problem, but extended to multi-class classification.

#### Maximal margin classifier

The linear discriminant function is defined to be

$$f(\mathbf{x}) = \text{sgn}(g(\mathbf{x})), \text{ where } g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0. \tag{2.16}$$

The $\mathbf{w}$ and $w_0$ are the parameters to determine the discriminant function. Given $d$-dimensional input vectors with $N$ samples, $\{\mathbf{x}_i \in \mathbb{R}^d, i = 1:N\}$, and the corresponding binary class label $\{\mathbf{y}_i \in \{1,-1\}, i = 1:N\}$, the decision boundary (hyperplane) between two classes is

$$g(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + w_0 = 0. \tag{2.17}$$

Suppose an input $\mathbf{x}_i$ does not satisfy (2.17); hence,

$$g(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + w_0 > 0. \tag{2.18}$$

Then, (2.18) means that $\mathbf{x}_i$ lies to one side of the hyperplane. On the other hand, if

$$g(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + w_0 < 0, \tag{2.19}$$

then, $\mathbf{x}_i$ lies on the other side of the hyperplane, as shown in Figure 2.6. the observations from the blue class are denotes by $y_i = 1$, whereas those from the orange class are written as $y_i = -1$. In summary,

$$\begin{cases} g(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + w_0 > 0 \ (\text{if } y_i = 1), \\ g(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + w_0 < 0 \ (\text{if } y_i = -1), \end{cases} \tag{2.20}$$

the two equations can be written in a simple form as

$$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) > 0. \tag{2.21}$$

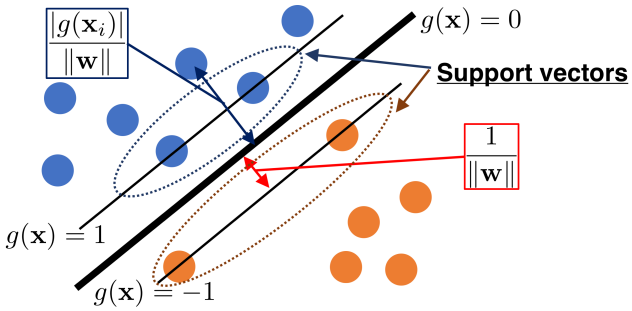



**Figure 2.6:** The geometry of a linear decision boundary in two dimensional space.

The main object of maximal margin classifier is to maximise the margin between the decision boundary ($g(\mathbf{x}) = 0$) and the nearest data to the decision boundary, called the support vectors, as illustrated in Figure 2.6. The distance between the decision boundary and the input $\mathbf{x}_i$ is $\frac{|g(\mathbf{x}_i)|}{\|\mathbf{w}\|}$, therefore the distance between the support vectors and the decision boundary is $\frac{1}{\|\mathbf{w}\|}$. It is assumed that there is no data between hyperplane $g(\mathbf{x}) = \pm 1$ and the decision boundary $g(\mathbf{x}) = 0$, so the following condition is imposed,

$$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1 \quad i = 1, ...., N. \tag{2.22}$$

Here, the margin between the decision boundary and the nearest data is maximised by finding the optimum weight parameters $\mathbf{w}, w_0$, i.e.

$$\arg\max_{\mathbf{w}, w_0} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i \left[ y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \right] \right\}. \tag{2.23}$$

In equation (2.23), in order to maximise the margin, $\frac{1}{\|\mathbf{w}\|}$, it is required to minimise $\|\mathbf{w}\|$ [14]. Therefore, the aim is to minimise $\|\mathbf{w}\|$ with the condition (2.22); in other words, the

new object of this maximal margin classifier is written as

$$\begin{aligned} &\min_{\mathbf{w},w_0} \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to} \quad y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1 \quad i = 1, ...., N. \end{aligned} \tag{2.24}$$

For convenience, $\frac{1}{2}\|\mathbf{w}\|^2$ can be solved instead of $\|\mathbf{w}\|$. The equation (2.24) is a constrained optimisation problem, and can be solved by the Lagrangian multiplier method [109]. The Lagrange (primal) function is

$$\mathcal{L}_P(\mathbf{w}, w_0, \lambda) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \lambda_i (y_i(\mathbf{w}^T\mathbf{x}_i + w_0) - 1), \tag{2.25}$$

which is minimised with respect to $\mathbf{w}, w_0$. $\lambda$ represents $\{\lambda_i\}_{i=1}^N$. It follows

$$\frac{\partial \mathcal{L}_P(\mathbf{w}, w_0, \lambda)}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i, \tag{2.26}$$

$$\frac{\partial \mathcal{L}_P(\mathbf{w}, w_0, \lambda)}{\partial w_0} = 0 \quad \Rightarrow \quad 0 = \sum_{i=1}^{N} \lambda_i y_i. \tag{2.27}$$

By substituting (2.26) – (2.27) into (2.25), the Lagrangian (dual) objective function is obtained

$$\mathcal{L}_D(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \tag{2.28}$$

which gives a lower bound on the objective function (2.24). It is required to maximise $\mathcal{L}_D(\lambda)$ subject to

$$\sum_{i=1}^{N} \lambda_i y_i = 0, \quad \lambda \geq 0. \tag{2.29}$$

In order to maximise the dual objective function (2.28), the sequential minimal optimisation (SMO) [110] can be used. The SMO is an efficient algorithm for solving the quadratic programming problem [111] (i.e. to find $\hat{\lambda}$, which is the optimal estimate of parameter $\lambda$.) From (2.26), the solution for $\mathbf{w}$ has the form

$$\hat{\mathbf{w}} = \sum_{i=1}^{N} \hat{\lambda}_i y_i \mathbf{x}_i, \tag{2.30}$$

with nonzero coefficients $\hat{\lambda}_i$ only for those observations $i$ for which the constraints in (2.24) are met. These observations are the support vectors. Consequently, the solution for $\mathbf{w}$ and

$w_0$ has the form

$$\hat{\mathbf{w}} = \sum_{i \in \mathcal{S}} \hat{\lambda}_i y_i \mathbf{x}_i, \tag{2.31}$$

$$\hat{w_0} = \hat{\mathbf{w}}^T \mathbf{x}_{\mathcal{S}} - y_{\mathcal{S}}, \tag{2.32}$$

where $\mathcal{S}$ denotes the subset of the input data, namely the support vectors. Therefore, the linear discriminant function for the new input sample, $\mathbf{x}^*$, is written as

$$f(\mathbf{x}^*) = \text{sgn}(\hat{\mathbf{w}}^T \mathbf{x}^* + \hat{w_0}) \tag{2.33}$$

$$= \text{sgn}\left( \sum_{i \in \mathcal{S}} \hat{\lambda}_i y_i \mathbf{x}_i \mathbf{x}^* + \hat{w_0} \right). \tag{2.34}$$

This is the maximal margin classifier, specifically called the hard margin classifier.

### Soft margin classifier

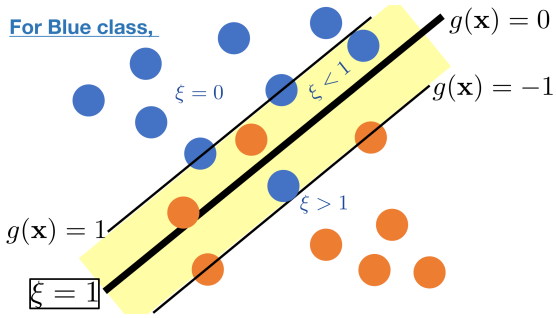



**Figure 2.7:** The geometry of a linear decision boundary in two dimensional space with slack variables.

If the data is not linearly separable (two classes are not necessarily separable by a hyperplane [108]), there is no feasible solution in which satisfies the condition (2.22), $y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1$ for all $i = 1, ...., N$ [14]. Now, the concept of a separating hyperplane is introduced to almost separate the classes – it is called a soft margin classifier. The object of the soft margin classifier is maximising the margin between two classes, but also allow for some training data to be on the wrong side of margin. To this end, the slack variables $\xi = (\xi_1, ..., \xi_N)$ is introduced to modify the objective function and the constraint in (2.24). The slack variables $\xi \geq 0$ satisfies

- $\xi = 0$, if the point is on or inside the correct margin boundary,
- $1 > \xi > 0$, if the point lies inside the margin, but on the correct side of the decision boundary,
- $\xi = 1$, if the point is on the decision boundary,

- $\xi > 1$, if the point lies on the wrong side of the decision boundary.

Figure 2.7 illustrates the geometry of a linear boundary in two dimensional space. the hard constraints in (2.22), such that $y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1$, are replaced with the soft margin constraints as

$$\begin{aligned} y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1 \quad &\Rightarrow \quad 1 - y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \leq 0 && (2.35)\\ &\Rightarrow \quad 1 - y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \leq \xi_i && (2.36)\\ &\Rightarrow \quad 1 - y_i(\mathbf{w}^T\mathbf{x}_i + w_0) - \xi_i \leq 0 \quad \text{where} \quad \xi_i \geq 0. && (2.37) \end{aligned}$$

The new object of the classifier becomes

$$\begin{aligned} \min_{\mathbf{w},w_0,\xi} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \right\} \\ \text{subject to} \quad \xi_i \geq 0,\ y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1 - \xi_i,\ i = 1:N, \end{aligned} \tag{2.38}$$

where the cost parameter $C$ is a regularisation that controls the number of errors tolerated on the training set. Since $\xi_i > 1$ means the point $i$ is misclassified, $\sum_i \xi_i$ can be interpreted as an upper bound on the number of misclassified points [14]. When $C$ is smaller, the constraint becomes 'softer' and more data points are allowed to be on the wrong side of boundary; i.e. the values of slack variables $\xi_i$ become larger. Conversely, the constraint becomes 'harder' with the larger $C$ parameter, and the value of slack variables become smaller. When $C = \infty$, the soft margin classifier is equivalent to the hard margin classifier. As well as the hard margin classifier, the function in (2.38) is solved by the Lagrangian multiplier method. By introducing the two Lagrangian multipliers $\lambda$ and $\mu$, the constrained optimisation problem in (2.38) is solved by

$$\mathcal{L}_P(\mathbf{w}, w_0, \lambda, \mu, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\lambda_i(y_i(\mathbf{w}^T\mathbf{x}_i + w_0) + \xi_i - 1) - \sum_{i=1}^{N}\mu_i\xi_i, \tag{2.39}$$

which is again minimised with respect to $\mathbf{w}, w_0, \xi$. $\mu$ represents $\{\mu_i\}_{i=1}^N$. It follows

$$\frac{\partial\mathcal{L}_P(\mathbf{w}, w_0, \lambda, \mu, \xi)}{\partial\mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{N}\lambda_i y_i \mathbf{x}_i, \tag{2.40}$$

$$\frac{\partial\mathcal{L}_P(\mathbf{w}, w_0, \lambda, \mu, \xi)}{\partial w_0} = 0 \quad \Rightarrow \quad 0 = \sum_{i=1}^{N}\lambda_i y_i, \tag{2.41}$$

$$\frac{\partial\mathcal{L}_P(\mathbf{w}, w_0, \lambda, \mu, \xi)}{\partial\xi_i} = 0 \quad \Rightarrow \quad 0 = C - \lambda_i - \mu_i. \tag{2.42}$$

By substituting (2.40) – (2.42) into (2.39), the same Lagrangian (dual) objective function to

the function in the hard margin classifier (2.28) is obtained as

$$\mathcal{L}_D(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j, \tag{2.43}$$

however, the constraints are different to (2.29) as

$$\sum_{i=1}^{N} \lambda_i y_i = 0, \quad 0 \leq \lambda_i \leq C. \tag{2.44}$$

Again, same to the hard margin classifier in (2.28), the sequential minimal optimisation (SMO) can be used to find the parameter $\hat{\lambda}$ to maximise the dual objective function in (2.43).

**The SVM as a non-linear soft margin classifier**

The support vector classifier (soft margin classifier) described above finds linear boundaries in the input feature space. The SVM is an extension of the soft margin classifier with enlarging the feature space using kernels, which translate to nonlinear boundaries in the original space. Let us denote the basis functions by $h_m(x)$, $m = 1, ..., d$. The input to the soft margin classifier is written as $h(\mathbf{x}_i) = (h_1(x_{i1}), ..., h_d(x_{id}))$, $i = 1, ..., N$. The soft margin classifier using the input as $g(\mathbf{x}) = \mathbf{w}^T h(\mathbf{x}) + w_0$ is fitted; this is similar to (2.16), but a non-linear function. The class is then assigned by the sign function $f(\mathbf{x}) = \text{sgn}(g(\mathbf{x}))$.

The SVM utilises the above-mentioned idea. The linear boundaries in the enlarged space achieve better training-class separation, and translate to nonlinear boundaries in the original space [109]. Previously, the optimisation problem using Lagrangian multiplier method in (2.39) is represented, and the solution in (2.43) is given by the inner products of the input data. Here, the inner products with the transformed feature vectors $h(\mathbf{x}_i)$ can be computed as

$$\mathcal{L}_D(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \langle h(\mathbf{x}_i)^T, h(\mathbf{x}_j) \rangle. \tag{2.45}$$

By substituting (2.40), the function $g(\mathbf{x})$ can be written as

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{w}^T h(\mathbf{x}) + w_0 \\ &= \sum_{i=1}^{N} \lambda_i y_i \langle h(\mathbf{x}_i), h(\mathbf{x}_j) \rangle + w_0. \end{aligned} \tag{2.46}$$

As well as the soft margin classifier, given $\lambda_i, w_0$ can be determined by solving $y_i g(\mathbf{x}_i) = 1$ in (2.46), for any $\mathbf{x}_i$ for which $0 < \lambda_i < C$. In both (2.45) and (2.46), $h(\mathbf{x})$ is only involved in the inner products. Actually, the details of the specific non-linear transformation $h(\mathbf{x})$

is not required, but the kernel function

$$K(\mathbf{x}, \mathbf{x}') = \langle h(\mathbf{x}), h(\mathbf{x}') \rangle \tag{2.47}$$

that computes the inner product in the transformed space, is needed. In this thesis, the radial basis function (RBF) kernel

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2). \tag{2.48}$$

is used with selected hyper-parameters $\gamma$.

#### SVM for multi-class problem

There are two approaches to extending the SVM to multi-class classification problems [107]: 1) one-against-rest and 2) one-against-one classification approach. This thesis applies the one-against-one approach. If $M$ is the number of classes (categories), then $M(M-1)/2$ classifiers are constructed and each classifier is trained by data from two classes. For the training data from the $j$th and the $k$th classes, the following two-class classification problem is solved as

$$\min_{\mathbf{w}, w_0, \xi} \quad \left\{ \frac{1}{2} \|\mathbf{w}^{jk}\|^2 + C \sum_{i=1}^{N} \xi_i^{jk} \right\} \tag{2.49}$$

$$\text{subject to} \quad \xi_i^{jk} \geq 0, \ i = 1 : N \tag{2.50}$$

$$y_i((\mathbf{w}^{jk})^T \mathbf{x}_i + w_0^{jk}) \geq 1 - \xi_i^{jk}, \text{ if } \mathbf{x}_i \text{ in } j\text{th class}, \tag{2.51}$$

$$y_i((\mathbf{w}^{jk})^T \mathbf{x}_i + w_0^{jk}) \leq -1 + \xi_i^{jk}, \text{ if } \mathbf{x}_i \text{ in } k\text{th class}. \tag{2.52}$$

In this thesis, a voting strategy is utilised – each binary classification is considered to be a voting where votes can be cast for all data points $\mathbf{x}$.

### 2.2.2 Extreme learning machine (ELM) for classification

Among multiple regression/classification algorithms, neural networks have been popular due to their capability of nonlinear mapping from the input data. The extreme learning machine (ELM) has been originally proposed by Huang *et al.* [112] as single-hidden layer feedforward neural networks (SLFNs) with 1) randomly chosen weights in the hidden nodes and 2) analytically determined output weights [113]. Conventional implementations for feedforward neural networks are slower in general due to 1) gradient-based learning algorithm and 2) all the parameters of neural networks are adaptively tuned by such gradients. Besides such disadvantages, the gradient descent based learning methods may converge to the local minima [113]. The SLFNs with $L$ random hidden nodes (i.e. the input weights and hidden layer biases are randomly chosen) can learn $L$ distinct ob-

servations with negligible errors. In other words, unlike traditional learning algorithms, the SLFNs with random hidden nodes does not require tuning in weight parameters, and these parameters can be independent to the input data (i.e. training data for classification). The proof can be found in [114].

The ELM is a fast algorithm for SLFNs [115] with randomly selected input weights $\mathbf{W}$ and biases $\mathbf{b}$. Unlike the neural networks with back-propagation, the output weights $\boldsymbol{\beta}$ are independent to $\mathbf{W}$ and $\mathbf{b}$, and the learning does not require iteration. Consider a set of $N$ inputs $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i \in \mathbb{R}^d, \mathbf{t}_i \in \mathbb{R}^c$ $(i = 1, ..., N)$. $d$ is the dimension of the inputs whereas $c$ is the number of classes (categories). A SLFN with $L$ hidden nodes has outputs given by

$$\sum_{j=1}^{L} \boldsymbol{\beta}_j g(\mathbf{w}_j \mathbf{x}_i + b_j), \quad i = 1, ..., N \tag{2.53}$$

where $\boldsymbol{\beta}_j, g(\cdot), \mathbf{w}_j, b_j$ are the output weights, an activation function (typically a sigmoid function), the input weights, and the biases, respectively, as shown in Figure 2.8 (a). In other words, as typical neural networks, the input data are transformed by the hidden layer into a different representation in the ELM framework. First, the input samples $\mathbf{x}_i$ are projected into the hidden layer with the input weights $\mathbf{w}_j$ and biases $b_j$. Then, such projected data is transformed by an activation function $g(\cdot)$. After the transformation, $\mathbf{h}$ denotes the data in the hidden layer. The estimated outputs $\mathbf{y}_i$ is obtained with the given input $\mathbf{x}_i$ and the target output (class) $\mathbf{t}_i$ as

$$\mathbf{y}_i = \sum_{j=1}^{L} \boldsymbol{\beta}_j g(\mathbf{w}_j \mathbf{x}_i + b_j) = \mathbf{t}_i + \epsilon_i, \quad i = 1, ..., N, \tag{2.54}$$

where $\epsilon_i$ is the noise.

The advantage of ELM is that the solution is obtained in a closed form. In a matrix form, the ELM can be written as

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_L \mathbf{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_L \mathbf{x}_N + b_L) \end{bmatrix},$$

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^T ... \boldsymbol{\beta}_L^T\right)^T, \mathbf{T} = \left(\mathbf{t}_1^T ... \mathbf{t}_N^T\right)^T. \tag{2.55}$$

Figure 2.8 (b) illustrates the diagram of an ELM in a matrix form. In summary, the ELM requires two projections: 1) $\mathbf{XW}$ for the inputs and 2) $\mathbf{H}\boldsymbol{\beta}$ for the outputs, where $\mathbf{H} = g(\mathbf{XW} + \mathbf{b})$.

When the number of training data samples is larger than the number of hidden nodes ($N > L$), which means the ELM problem is over-determined, the unique solution is given by minimising the sum of square errors. As the matrix $\mathbf{H}$ is a full rank, the pseudoinverse

of matrix $\mathbf{H}$ is computed by

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{2.56}$$

$$\boldsymbol{\beta} = \mathbf{H}^{\dagger}\mathbf{T}, \tag{2.57}$$

$$\mathbf{H}^{\dagger} = (\mathbf{H}^{T}\mathbf{H})^{-1}\mathbf{H}^{T}. \tag{2.58}$$

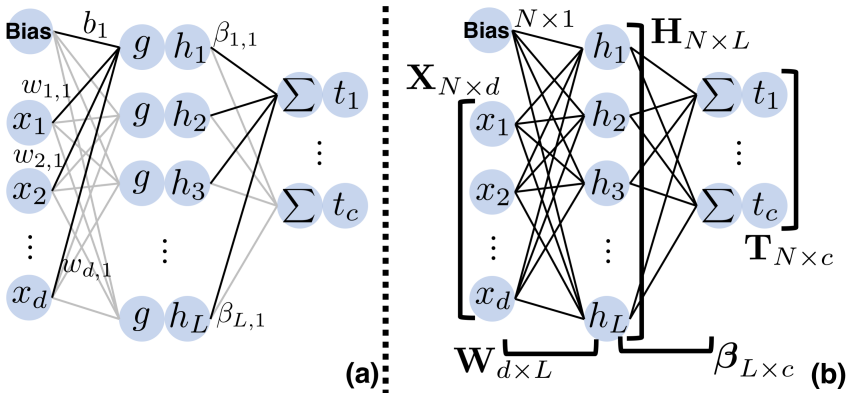



**Figure 2.8:** The diagram of an ELM model. (a) Simple model; (b) matrix notation.

## 2.3 Smoothing algorithm for automatic sleep staging

### 2.3.1 Hidden Markov model (HMM)

Hidden Markov model (HMM) [116] is a ML method for modelling time series. A key property of HMM is to represent the probability distributions over the sequence of observations. The HMM assumes:

1. the observation at time $t$, $x_t \in \{O_1, ..., O_K\}$, generated by some process; the state $z_t \in \{S_1, ..., S_J\}$ is not observed (i.e. *hidden*),
2. the state of this *hidden* process satisfies the *Markov property* – the current state $z_t$ is independent from all the states prior to $t-1$ (first-order Markov property).

In other words, the HMM consists of a discrete-time and a discrete-state Markov chain with hidden state $z_t$. The observed sequence is written as $\{x_t\}_{t=1}^{T}$, whereas $\{z_t\}_{t=1}^{T}$ denotes the hidden sequence. The observation model can be written as $p(x_t|z_t)$. The joint

distribution of this model is therefore

$$\begin{aligned} p(\{z_t\}_{t=1}^T, \{x_t\}_{t=1}^T) &= p(z_1, ..., z_T, x_1, ..., x_T) \\ &= p(z_1, ..., z_T) p(x_1, ..., x_T | z_1, ..., z_T) \\ &= \left[ p(z_1) \prod_{t=2}^{T} p(z_t | z_{t-1}) \right] \left[ \prod_{t=1}^{T} p(x_t | z_t) \right]. \end{aligned} \tag{2.59}$$

Figure 2.9 illustrates the graphic of this factorisation of the joint probability in (2.59).

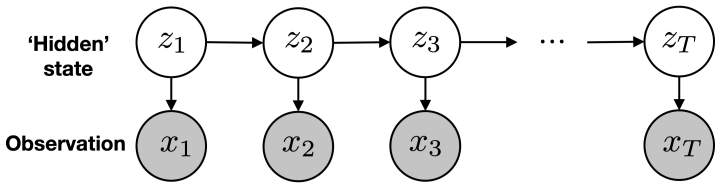



**Figure 2.9:** A diagram of a hidden Markov model.

There are multiple techniques to infer the hidden state sequence of an HMM with given parameters. In this thesis, a maximum a posteriori (MAP) estimation (Viterbi decoding algorithm) is used. The decoding algorithm finds a most probable sequence as computing

$$\begin{aligned} \underset{\{z_t\}_{t=1}^T}{\operatorname{argmax}} p(\{z_t\}_{t=1}^T | \{x_t\}_{t=1}^T) &= \underset{\{z_t\}_{t=1}^T}{\operatorname{argmax}} p(\{z_t\}_{t=1}^T, \{x_t\}_{t=1}^T) \\ &= \underset{j}{\operatorname{argmax}} \max_{\{z_t\}_{t=1}^{T-1}} p(z_T = j, \{z_t\}_{t=1}^{T-1}, \{x_t\}_{t=1}^T) \\ &= \underset{j}{\operatorname{argmax}}\, \delta_T^j \end{aligned} \tag{2.60}$$

where $\delta_T^j$ is the probability of ending up in the state $j$ at time $T$. The most probable path to the state $j$ at time $t$ must consist of the most probable path to some other state $i$ at time $t-1$. Therefore, the probability $\delta_t^j$ can be recursively calculated over $t$,

$$\begin{aligned} \delta_t^j &\triangleq \max_{z_1, ..., z_{t-1}} p(z_t = j, z_1, ..., z_{t-1}, x_1, ..., x_t) \\ &= p(x_t | z_t = j) \max_i p(z_t = j | z_{t-1} = i) \delta_{t-1}^i \\ &= \phi_t^j \max_i \psi_{ij} \delta_{t-1}^i. \end{aligned} \tag{2.61}$$

The three factors that are multiplied in (2.61) for extending the previous paths to compute

the decoding probaility at time $t$ are

$$\begin{aligned} \phi_t^j &: \text{the likelihood of the observation } x_t \text{ given the current state } j \\ \psi_{ij} &: \text{the transition probability from previous state } i \text{ to current state } j \\ \delta_{t-1}^i &: \text{the previous decoding probability from the previous time step } t-1 \end{aligned}$$

Consequently, (2.60) tells us the most likely previous state on the most probable path to $z_t = j$.

In summary, the Viterbi decoding algorithm computes $\delta_t^j$ for all time $t$ and state $j$ using dynamic programming. It is required to initialise the algorithm by setting

$$\begin{aligned} \delta_1^j &= p(x_1|z_1 = j)p(z_1 = j) \\ &= \phi_1^j \pi_j \end{aligned} \tag{2.62}$$

and to terminate the algorithm by computing the most probable final state

$$\max_{\{z_t\}_{t=1}^T} p(\{z_t\}_{t=1}^T, \{x_t\}_{t=1}^T) = \max_j \delta_T^j. \tag{2.63}$$

The most probable sequence of states can be computed using traceback

$$z_T^* = \operatorname*{argmax}_j \delta_T^j \tag{2.64}$$

$$z_{t-1}^* = \operatorname*{argmax}_i p(z_t^*|z_{t-1} = i)\delta_{t-1}^i. \tag{2.65}$$




**Figure 2.10:** A example of Viterbi decoding. Figure and example are based on Figure 17.13 in [14].

Figure 2.10 depicts a toy example of the Viterbi algorithm, based on Figure 17.13 in [14]. Suppose we observed the discrete sequence of observation $\{x_t\}_{t=1}^4 = \{O_1, O_3, O_4, O_6\}$. The model starts in state $z_1 = S_1$. The transition matrix and emission matrix are given

in the left side of Figure 2.10. First, the probability of generating $O_1$ in state $z_1 = S_1$ is 0.5, which means $\delta_1^1 = 0.5$, whereas $\delta_1^i = 0$ for all other states $i$. Next, the probability of translation from $S_1$ to $S_1$ (self-translation) is 0.3, whilst that of translation from $S_1$ to $S_2$ is 0.7. When the next state is $S_1$, the probability of generating $O_3$ is 0.3; or if the next state is $S_2$, the probability of generating $O_3$ is 0.2. Hence, the most probable state in $t = 2$ is,

$$\delta_2^1 = \delta_1^1 \psi_{11} \phi_2^1 = 0.5 \cdot 0.3 \cdot 0.3 = 0.045 \tag{2.66}$$

$$\delta_2^2 = \delta_1^1 \psi_{12} \phi_2^2 = 0.5 \cdot 0.7 \cdot 0.2 = 0.07 \tag{2.67}$$

therefore $S_2$. The algorithm continues with such calculation until the end of the sequence.

## 2.4 Evaluation

The performance indicators to evaluate the proposed automatic sleep staging methods are class-specific sensitivity (SE) and precision (PR), as well as overall accuracy (AC) and kappa coefficient ($\kappa$), defined as follows:

$$\mathrm{SE}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i}, \ \mathrm{PR}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i}, \ \mathrm{AC} = \frac{\sum_{i=1}^{M} \mathrm{TP}_i}{N},$$
$$\pi_e = \frac{\sum_{i=1}^{M} \{(\mathrm{TP}_i + \mathrm{FP}_i)(\mathrm{TP}_i + \mathrm{FN}_i)\}}{N^2}, \ \kappa = \frac{\mathrm{AC} - \pi_e}{1 - \pi_e}.$$

The parameter TP (true positive) represents the number of positive (target) epochs correctly predicted, TN (true negative) is the number of negative (non-target) epochs correctly predicted, FP (false positive) is the number of negative epochs incorrectly predicted as positive class, FN (false negative) is the number of positive epochs incorrectly predicted as negative class, $M$ is the number of classes, and $N$ is the total number of epochs. Figure 2.11 shows a $M$ class confusion matrix and corresponding sensitivity and precision. The Cohen's kappa ($\kappa$) determines the agreement between scorers, with chance agreement removed. The values between $0.00-0.20$, $0.21-0.40$, $0.41-0.60$, $0.61-0.80$, and $0.81-1.00$ respectively correspond to Slight, Fair, Moderate, Substantial, and Almost Perfect Agreement [117].

The performance metrics for the biometrics is discussed later, in Section 7.3.6.

## 2.5 Summary

Several feature extraction methods and classifiers for automatic sleep stage classification have been reviewed. The frequency response varies in different sleep stages (i.e. depth of sleep). The MSE methods calculate the complexity of time series, whereas the SEF is a non-linear feature in frequency domain. The SVM is one of the traditional classifiers in ML, which is a flexible and powerful algorithm with the kernel methods. As a com-

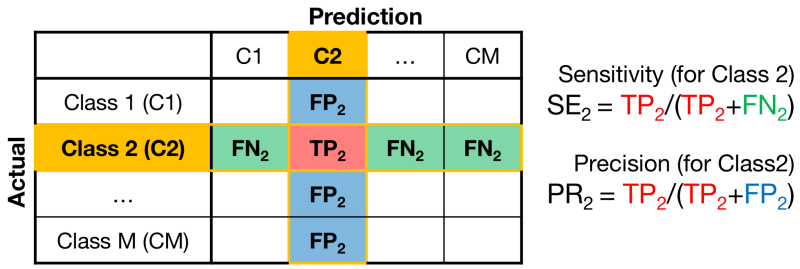



**Figure 2.11:** Confusion matrix and performance criteria.

putationally efficient classifier, the ELM is employed in the automatic sleep staging for large-scale data. The HMM is utilised for a smoothing which aims to incorporate the transitional information in automatic sleep staging. The performance of these algorithms is validated using publicly available sleep EEG datasets in Part I of this thesis.

# Part I

# State-of-the-art automatic sleep stage classification

# Chapter 3

# Structural complexity analysis for automatic sleep staging

## Overview

This chapter provides the details of the automatic sleep stage classification using both structural complexity feature and non-linear frequency domain feature, which are summarised in Chapter 2. The classification performance of the proposed method is validated using a publicly available dataset, the Sleep-EDF [93], which has been commonly used in the research community.

## 3.1 Introduction

Sleep can be defined as a reversible behavioural state of perceptual disengagement and unresponsiveness to the environment. Sleep is also a complex amalgam of physiologic and behavioural processes [118]. The function of sleep is not fully understood yet, and among the many hypotheses proposed, the most widely accepted ones are brain thermoregulation, brain detoxification, tissue restoration and metabolic homeostasis [119]. Quality of sleep also reflects the state of body and mind, and in addition, numerous sleep disorders are common [120]; these disorders include insomnia, breathing disturbances during sleep (i.e. sleep apnea), and narcolepsy, which are diagnosed using polysomnography (PSG).

The PSG recording requires multiple electrodes which includes at least two EEG channels, two EOG channels to observe eye movements, at least one chin EMG channel. The PSG recording is usually conducted within a hospital or at a sleep centre. After the recording, the PSG is analysed by experts in order to identify individual sleep patterns. The classification of sleep stages has been classically performed based on the visual interpretation of each 30-second PSG epoch, according to the Rechtschaffen and Kales (R&K) sleep scoring manual [121], or the manual of the American Academy of Sleep Medicine (AASM) [105]. The sleep stages include: wake (W), stage1 (S1), stage2 (S2), stage3 (S3), stage4 (S4), and rapid eye movement (REM). The stages S3 and S4 are called slow wave

**Table 3.1:** Recent works for automatic sleep stage classification problem.

| Authors | Dataset | Channel | Features | Classifier | $M^*$ | AC (%) |
|---|---|---|---|---|---|---|
| [122] | Sleep-EDF | Pz-Oz | Band power | ANN | 4 | 81.6 |
| [104] | | Fpz-Cz<br>Pz-Oz | Band power, power ratio, SEF | Decision tree | 5 | 82.2 |
| [123] | | Pz-Oz | VG, difference VG | SVM | 5 | 88.9 |
| [124] | | Pz-Oz | EMD, 1st-4th statistical moment | Bagging decision tree | 5 | 90.7 |
| [125] | | Pz-Oz | DWT, 2nd-4th statistical moment | RF | 5 | 91.5 |
| [97] | Own | C3-A2 | MSSE, AR coefficient | LDA | 5 | 88.1 |
| [126] | | C4-A1 | Time, frequency, non-linear domain | SVM | 5 | 89.4 |

$M^*$: Number of class. ANN = artificial neural network, SEF = spectral edge frequency, VG = visibility graph, SVM = support vector machine, EMD = empirical mode decomposition, DWT = discrete wavelet transform, RF = random forests, MSSE = multi-scale sample entropy, AR = autoregressive, LDA = linear discriminant analysis.

sleep (SWS) and these two stages are merged into one condition. For the ASSM manual, the stages are noted as non-REM stage 1-3 (N1, N2, N3) and REM.

The main limitation of epoch based visual sleep scoring is that it is time-consuming. The scoring of eight hours overnight PSG takes more than an hour for an expert [122]. In addition, the multiple electrode montage disturbs patients' sleep, also, with PSG recordings typically occurring in hospitals, or other unfamiliar environments, some patients find themselves unable to sleep as usual. In the last two decades, computer based sleep staging has been developed in order to minimise analysis time.

To alleviate the aforementioned problems, a large number of automatic sleep stage algorithms, based on a small number of electrodes have been reported, as summarised in Table 3.1. These methods are typically based on feature extraction techniques and pattern recognition algorithms. In terms of the analysis data, the PhysioNet Sleep-EDF database [93] has been available for more than ten years. For datasets extracted from Sleep-EDF database, Imtiaz *et al.* [104] calculated spectral band power and spectral edge frequency (SEF) from two EEG channels. These methods correctly classified 82.2 % of epochs for a 5-stage task. Zhu *et al.* [123] classified the sleep stages based on graph domain features, and achieved an accuracy of 89.0 %. Hassan *et al.* [124] applied empirical mode decomposition (EMD) analysis for a single lead EEG montage, and calculated first to fourth statistical moments for each intrinsic mode function (IMF). Their proposed method achieved 90.7 % accuracy by bootstrap aggregating decision tree. Recently, Silveira *et al.* [125] reported 91.5 % accuracy for 5-stage classification problems when using discrete wavelet transform and random forests classifiers (RF).

Among research groups who recorded PSG by themselves, Liang *et al.* [97] recorded PSG from twenty healthy subjects, and calculated multi-scale sample entropy (MSSE) and autoregressive coefficients from a single EEG channel. The LDA yielded 88.1 % accuracy for testing data.

For the sleep stage classification, distinguishing between S1 sleep and REM is the most challenging with state-of-the-art papers. In the R&K sleep scoring guideline, S1 sleep is defined as 50 % of the epoch consists of relatively low voltage mixed (2-7 Hz)

activity, and <50 % of the epoch contains alpha activity, whereas REM is represented by a relatively low amplitude and mixed frequency (2-7 Hz) EEG with episodic rapid eye movements and the absence of (or reduced) chin EMG activity [16]. Figure 3.1 shows the power spectral density (PSD) for two EEG channels (Fpz-Cz and Pz-Oz) of the *SEDFx-S* dataset (see details in Section 3.2.1) with $M = 5$ sleep stages. For the wake (W) condition, observe a peak in the alpha band (8-13 Hz). Since the alpha rhythm is predominantly observed from the occipital lobe, the Pz channel has stronger alpha than the Fpz. The alpha power reduced in the S1 stage, and the strong peak around 13-14 Hz was present for the S2 stage. During the SWS stage, the delta activity become larger, however, the spectrum of REM mostly overlapped with that of S1 except for the beta band. In order to distinguish between REM and S1, it is considered that an automatic sleep stage classification problem based on structural complexity analyses by means of entropy.

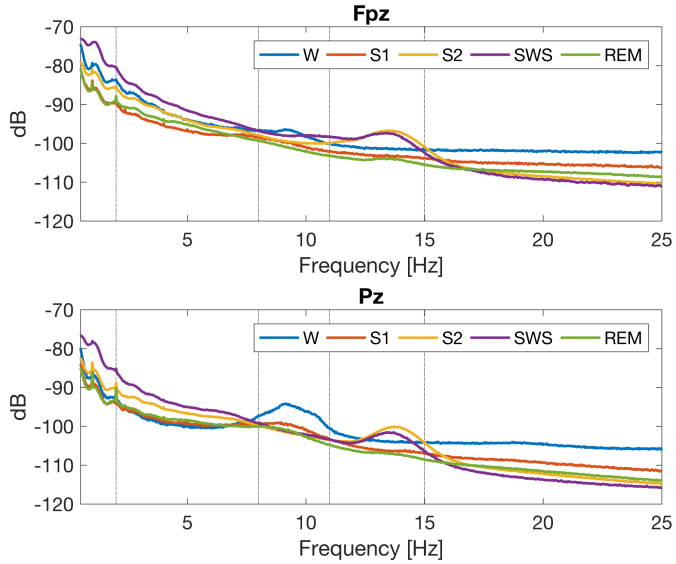



**Figure 3.1:** PSD for two EEG channels (Fpz-Cz, and Pz-Oz) from the *SEDFx-S* dataset with $M = 5$ sleep stages.

Multi-scale entropy (MSE) [96] was introduced as a non-parametric method for estimating dynamical complexity over multiple scales of a time series. The MSE are calculated for evaluating multi-scale complexity using fuzzy entropy (FE) and permutation entropy (PE).

Here, the structural complexity is utilised to perform sleep stage classification which yields the following advantages:

- The required number of EEG channels is reduced to one or two channels of EEG, which is a prerequisite for wearable devices.
- Structural complexity analyses as well as spectral edge frequency (SEF) are able to

distinguish between the S1 stage and REM sleep, which has been notoriously difficult with the state-of-the-art methods using only EEG.

## 3.2 Methods

To evaluate the performance of the proposed method, the EEG recording was obtained from the Sleep-EDF database [expanded] [93]. The extracted features from each EEG channel are used in the classification algorithm. Figure 3.2 shows the flowchart of the proposed analysis.

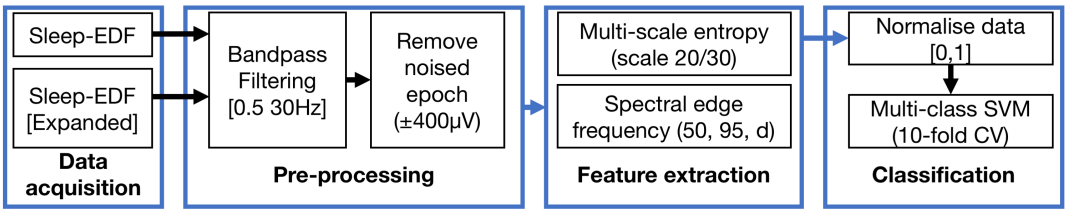



**Figure 3.2:** Flowchart of an automatic sleep staging system.

### 3.2.1 Data acquisition

The Sleep-EDF database [127] consists PSG data of eight subjects, which include two EEG, Fpz-Cz and Pz-Oz (for simplicity, denote Fpz and Pz later), one EOG (horizontal), and one submental chin EMG. Each PSG is sampled at 100 Hz and manually scored by well-trained technicians based on the R&K manual. Recently, the Sleep-EDF database [expanded] [93] has been made public, and contains in total PSG recordings of 61 subjects, from healthy Caucasians aged 25-101, without any sleep-related medication.

Although the Sleep-EDF database is extensively used for automatic sleep stage scorings [97, 104, 122, 123, 124, 125, 128], the usage of the database varies for research groups. One problem is that the database contains two types of PSG data: some of them are recorded over 24 hours (titled as $sc*$), and the others contain only overnight data (denoted by $st*$). In terms of $sc*$ data, the majority of epochs were labeled as the wake condition, because the data were obtained both over day and night periods; in other words, the data contains a large number of pre-sleep wake epochs (from afternoon to evening) and post-sleep wake epochs (from morning to afternoon). Therefore, most current analyses are biased towards the wake condition which dominates the 24-hour data.

The lack of consistency in the data structure makes the direct comparison between existing methods and the proposed idea difficult. Here, the dataset from the Sleep-EDF database were partitioned in three different ways in order to enable comparisons with existing methods.

1. *Sleep-EDF Whole (SEDF-W)* - The number of subjects is eight. The data are downloaded from [127], and extracted whole scored data (from *hypnogram start time* to last scored data) and excluded epochs labeled '?'. This dataset has a similar form to the dataset used in [123, 124, 128].
2. *Sleep-EDF [expanded] Whole (SEDFx-W)* - The number of subjects is 61. The data are downloaded from [93], and extracted whole scored data (from *hypnogram start time* to last scored data) and excluded epochs labeled '?'. This dataset has a similar form to the dataset used in [125]
3. *Sleep-EDF [expanded] Sleep (SEDFx-S)* - The number of subjects is 61. The data are downloaded from [93]. Imtiaz *et al.* [129] proposed an open source toolbox for Matlab to extract only overnight data in order to create a standardised dataset. The toolbox truncates the data in the following way:

   - From *light off* time to *light off* time;
   - if *light off/on* time is not available, from fifteen minutes before the first scored sleep epoch to fifteen minutes after the last scored sleep epoch.

   This dataset has a similar form to the dataset used in [104].

**Table 3.2:** Comparison of datasets - after removing noisy epochs.

| Dataset | | Wake | Sleep Stage | | | | |
|---|---|---|---|---|---|---|---|
| | | | S1 | S2 | S3 | S4 | REM |
| *SEDF-W* | # of epochs | 7943 | 581 | 3596 | 667 | 619 | 1589 |
| | Ratio (%) | 53.0 | 8.2 | 51.0 | 9.5 | 8.8 | 22.5 |
| *SEDFx-W* | # of epochs | 74346 | 4715 | 27070 | 5056 | 3757 | 11755 |
| | Ratio (%) | 58.7 | 9.0 | 51.7 | 9.7 | 7.2 | 22.5 |
| *SEDFx-S* | # of epochs | 6448 | 4676 | 26849 | 4996 | 3705 | 11749 |
| | Ratio (%) | 11.0 | 9.0 | 51.7 | 9.6 | 7.1 | 22.6 |

Table 3.2 summarises the number of epochs and the ratio of epochs with respect to each sleep stage for datasets extracted from the Sleep-EDF database after the pre-processing. The ratio for the wake condition, and the ratios for S1, S2, S3, S4, and REM are calculated as follows:

$$\begin{aligned} \text{Wake Ratio} &= \frac{\text{Number of wake epochs}}{\text{Number of whole epochs}}, \\ \text{Sleep Ratio} &= \frac{\text{Number of each stage epochs}}{\text{Number of whole sleep epochs}}. \end{aligned}$$

### 3.2.2 Pre-processing

A fourth-order Butterworth filter with passband from 0.5-30 Hz was applied to the two EEG channels. Afterwards, the epochs which contained amplitudes of more than $\pm 400\,\mu$V

were removed, since the amplitude of the K-complex is almost always less than 400 $\mu$V [130]. Epochs containing `NaN` values were removed from the analyses. In total, there were 58423 epochs for the *SEDFx-S* dataset; 11.0 % of the data were scored as awake, and 9.0 %, 51.7 %, 9.6 %, 7.1 %, and 22.6 % of sleep data were scored as S1, S2, S3, S4, and REM, respectively.

### 3.2.3 Feature extraction

After the pre-processing, multi-scale fuzzy entropy (MSFE), multi-scale permutation entropy (MSPE), and spectral edge frequency (SEF) were calculated from each epoch of the EEG channels. The details of the algorithms are given in Chapter 2.

#### Epoch length

The length of epochs for manual sleep stage scoring was 30 seconds, and the epoch-based criteria are often dependent on the preceding and following epochs. Here, epochs were extracted in two different ways: 1) use only the current 30 seconds epoch to calculate features (No overlap), or 2) enlarge the epoch size from 30 seconds to 90 seconds (With overlap); in other words, utilise both the preceding and following 30 seconds of epoch data as well as the current epoch to take account of sleep transition, as illustrated in Figure 3.3. Since the sampling frequency was 100 Hz, the length of each epoch for the no overlap method is $N = 3000$, and $N = 9000$ data points were used for the proposed overlap methods. For the overlap method, the first and last epoch of EEG recordings were excluded from analyses.

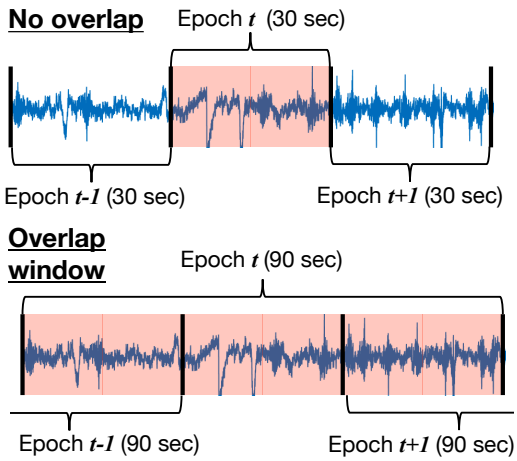



**Figure 3.3:** Epoch selection for EEG time series. (Top) No overlap method, the epoch length is 30 seconds; (Bottom) the epoch length enlarges to 90 seconds as combining the preceding and following epoch.

### Multi-scale entropy

The parameters used to calculate the MSFE were $\tau = 1 - 30$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 30$), $m = 2$, $r = 0.15\times$(*standard deviation of each epoch*). For the MSPE, The scales were chosen as $\tau = 1-20$. When the window contains less than $d!$ samples, the PE can be calculated with the dimension equal to or less than the $d-1$. The length of coarse-grained time series for no overlapping epochs with scale $\tau_{\text{max}} = 20$ is $\frac{N=3000}{\tau=20} = 150$ which is larger than $5! = 120$, so the PE with dimension $d = 5$ can be calculated with $L = 1$. Figure 3.4 illustrates MSPE analysis for two EEG channels (Fpz, and Pz) of *SEDFx-S* dataset with $M = 5$ sleep stages.

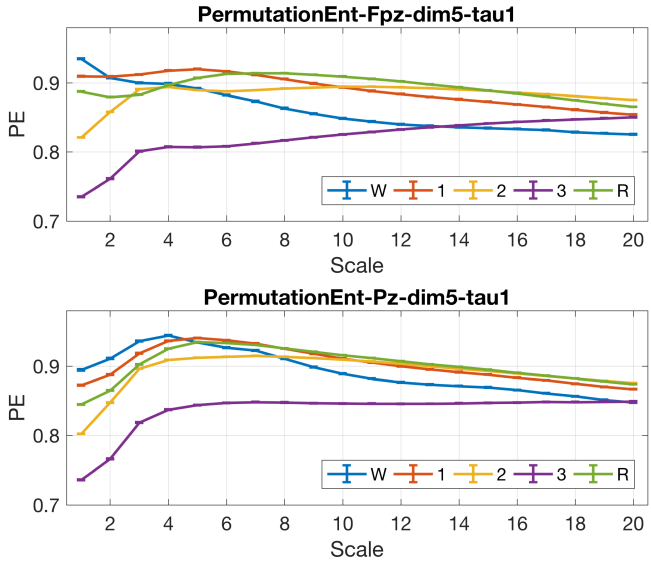



**Figure 3.4:** MSPE analyses for two EEG channels from the *SEDFx-S* dataset with $M = 5$ sleep stages, and scale $\tau = 1 - 20$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 20$). The error bars represent the standard error.

### Spectral edge frequency (SEF)

The frequency ranges for SEF50, SEF95, and SEF$d$ were selected in the same bands given in Table 2.3 (Setup-N).

### Other frequency domain features

In order to compare the classification performance, a simple frequency domain feature was used. The absolute spectral band power was calculated by the sum of the power spectral density within the given frequency range $f_{\text{low}}$ and $f_{\text{high}}$.

$$\text{Abs. Band power} = \sum_{f=f_{\text{low}}}^{f_{\text{high}}} \text{Power}(f).$$

As well as SEF, the selected frequency ranges for absolute spectral band power were $\delta - \beta = 0.5\text{-}30\,\text{Hz}$, $\delta - \alpha = 0.5\text{-}16\,\text{Hz}$, $\theta = 4\text{-}8\,\text{Hz}$, $\alpha = 8\text{-}15\,\text{Hz}$, $\alpha_{\text{low}} = 8\text{-}11\,\text{Hz}$, $\alpha_{\text{high}} = 11\text{-}15\,\text{Hz}$, and $\beta = 16\text{-}30\,\text{Hz}$.

### 3.2.4 Classification

The extracted features were normalised to between $[0, 1]$ before classification. A 10-fold cross-validation (CV) approach was utilised for this study. As a classifier, the multi-class SVM (Section 2.2.1) with the RBF kernel was employed; the hyper-parameters were tuned based on the training accuracy.

### 3.2.5 Performance evaluation

The performance indicators to evaluate the proposed methods were class-specific sensitivity (SE) and precision (PR), as well as overall accuracy and kappa coefficient ($\kappa$); the details are given in Section 2.4.

## 3.3 Results

Feature extraction parts was undertaken using Matlab 2015a, and the classification was conducted in Python 2.7.11, Anaconda 2.3.0, operated on a MacBook Pro with 2.2GHz Intel Core i7, 16GB of RAM. First, the classification results using different features with no overlap and overlap blocks were compared. Second, the classification performance for $M = 2-6$ class sleep stages (See, Table 3.3) was evaluated. Finally, the proposed methods were compared to the existing methods using the Sleep-EDF database.

**Table 3.3:** Sleep stage classes ($M$).

| $M$ | Sleep stages |
|---|---|
| 6 | Wake, S1, S2, S3, S4, REM |
| 5 | Wake, S1, S2, SWS(S3-4), REM |
| 4 | Wake, S1-2, SWS, REM |
| 3 | Wake, NREM(S1-4), REM |
| 2 | Wake, Sleep(REM & NREM) |

Table 3.4 shows the classification accuracy for the *SEDFx-S* dataset in the 5-stage discrimination, using different features extracted from EEG. The overlap epoch based method performed better than the conventional 30-second epoch method. In terms of classification performance with a single feature, the SEF feature over seven frequency bands performed the best, which achieved 85 % of classification accuracy with overlapping epochs, which was approximately 14 % higher than the classification results using a simple band power feature. The performance using multi-scale entropy (MSFE or MSPE) was slightly lower than that using SEF feature. The classification accuracy using both

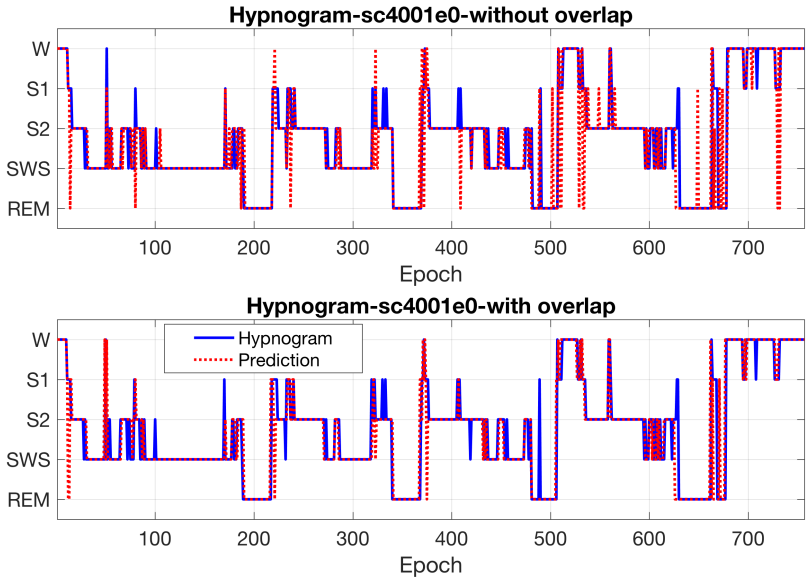



**Figure 3.5:** Hypnogram (Blue) and the predicted label (Red) of subject *sc4001e0* in the *SEDFx-S* dataset for the $M = 5$ class classification problem.

MSPE and SEF features was the highest; 88.6 % of epochs were correctly predicted using the overlapping window. Figure 3.5 illustrates the labeled hypnogram and the predicted label for subject *sc4001e0* with no overlapping epochs (Top panel) and with overlapping epochs (Bottom panel). Since the overlapping epochs method extracted EEG not only from the current 30-second epoch but also from the preceding and following epochs, the predicted class was more smooth, e.g. 280-320 SWS epoch; on the other hand, the result without overlapping epochs predicted quick transitions, such epochs were not correctly predicted by the overlapping method, e.g. transition from SWS to S2 at 100. Overall, the overlap method performed slightly better than no overlap epoch based feature extraction.

**Table 3.4:** Classification accuracy with 10-fold CV for $M = 5$ class with two EEG channels.

| Features | # of features | *SEDFx-S* (Fpz + Pz) | |
|---|---|---|---|
| | | No overlapping | With overlapping |
| Band power | 14 (7 frequency bands) | 74.5 | 70.7 |
| SEF | 42 (7 frequency bands, 3 features each) | 82.5 | 85.0 |
| MSFE | 60 ($\tau = 1 - 30$) | 81.2 | 84.7 |
| MSPE | 40 ($\tau = 1 - 20$) | 81.2 | 83.5 |
| MSFE + SEF | 102 | 85.5 | 87.9 |
| MSPE + SEF | 82 | 86.2 | **88.6** |

Table 3.5, 3.6, and 3.7 depict the confusion matrix obtained from the classification results with multi-scale entropy and SEF features by overlapped epoch obtained from the *SEDF-W* dataset, the *SEDFx-S* dataset and the *SEDFx-W* dataset, respectively. Since

the majority of epochs were labeled as the wake (W) condition in the *SEDFx-W* dataset, the sensitivity of the W stages was approximately 10 % higher than that of the *SEDFx-S* dataset. Therefore, the accuracy for the *SEDFx-W* dataset became higher.

Table 3.8 shows the sensitivity and precision of the classification for different sleep stages $M = 2 - 6$ with MSPE and SEF features by overlapped epoch obtained from the *SEDFx-S* dataset. For the all 2- to 6-stages classification, the kappa values achieved Almost Perfect Agreement, as 0.86, 0.88, 0.86, 0.84, and 0.81, respectively.

**Table 3.5:** *SEDF-W* - Confusion matrix for $M = 5$-stage sleep classification using MSFE and SEF features with overlapping windows extracted from the Fpz channel (a single EEG channel).

| | | Algorithm | | | | | SE(%)/*PR(%)* |
|---|---|---|---|---|---|---|---|
| | | W | S1 | S2 | SWS | REM | |
| | W | 7841 | 60 | 18 | 1 | 10 | 98.9/*98.5* |
| Manually | S1 | 65 | 300 | 127 | 4 | 84 | **51.7**/*61.1* |
| scored | S2 | 27 | 75 | 3330 | 115 | 47 | 92.7/*90.3* |
| hypnogram | SWS | 7 | 1 | 150 | 1128 | 0 | 87.7/*90.4* |
| | REM | 18 | 55 | 62 | 0 | 1454 | 91.5/*91.2* |
| Accuracy = 93.8%, $\kappa$ = 0.90 (SE:Sensitivity, *PR:Precision)* | | | | | | | |

**Table 3.6:** *SEDFx-S* - Confusion matrix for $M = 5$-stage sleep classification using MSPE and SEF features with overlapping windows extracted from the Fpz and Pz channels (two EEG channels).

| | | Algorithm | | | | | SE(%)/*PR(%)* |
|---|---|---|---|---|---|---|---|
| | | W | S1 | S2 | SWS | REM | |
| | W | 5559 | 489 | 150 | 21 | 124 | 87.6/*87.0* |
| Manually | S1 | 545 | 2699 | 926 | 13 | 487 | **57.8**/*68.6* |
| scored | S2 | 149 | 490 | 24951 | 810 | 442 | 93.0/*90.3* |
| hypnogram | SWS | 35 | 7 | 1161 | 7495 | 2 | 86.2/*89.9* |
| | REM | 104 | 248 | 452 | 3 | 10939 | 93.1/*91.2* |
| Accuracy = 88.6%, $\kappa$ = 0.84 (SE:Sensitivity, *PR:Precision)* | | | | | | | |

**Table 3.7:** *SEDFx-W* - Confusion matrix for $M = 5$-stage sleep classification using MSFE and SEF features with overlapping windows extracted from the Fpz channel (a single EEG channel).

| | | Algorithm | | | | | SE(%)/*PR(%)* |
|---|---|---|---|---|---|---|---|
| | | W | S1 | S2 | SWS | REM | |
| | W | 73419 | 448 | 199 | 24 | 150 | 98.9/*98.3* |
| Manually | S1 | 758 | 2312 | 1009 | 10 | 620 | **49.1**/*65.4* |
| scored | S2 | 275 | 452 | 24929 | 845 | 563 | 92.1/*89.0* |
| hypnogram | SWS | 55 | 7 | 1213 | 7530 | 7 | 85.5/*89.5* |
| | REM | 187 | 319 | 652 | 2 | 10592 | 90.1/*88.8* |
| Accuracy = 93.8%, $\kappa$ = 0.90 (SE:Sensitivity, *PR:Precision)* | | | | | | | |

Table 3.9 shows the S1 sleep stage sensitivity of the proposed methods. In order to evaluate the performance, the *SEDF-W*, the *SEDFx-S*, and the *SEDFx-W* were compared to the results in [123, 124, 128], [104], and [125], respectively. The sensitivities of the S1 stage are 51.7 %, 57.8 %, 49.1 % for the *SEDF-W*, *SEDFx-S*, and *SEDFx-W* datasets, respectively. The proposed method was able to distinguish the S1 stage better than other methods;

**Table 3.8:** *SEDFx-S* - Sensitivity and *Precision* (*in italic*) for sleep classification using MSPE and SEF with overlaps.

| SE(%) *PR(%)* | $M = 2$ | $M = 3$ | $M = 4$ | $M = 5$ | $M = 6$ |
|---|---|---|---|---|---|
| W | 85.8 *89.4* | 86.9 *88.4* | 86.7 *87.8* | 87.6 *87.0* | 87.6 *86.8* |
| S1 | | | | 57.8 *68.6* | 57.9 *68.1* |
| S2 | | | 92.8 *91.6* | 93.0 *90.3* | 92.9 *90.2* |
| S3 | 98.8 *98.3* | 96.3 *96.3* | | | 64.9 *69.6* |
| S4 | | | 86.1 *90.1* | 86.2 *89.9* | 83.7 *85.2* |
| REM | | 92.4 *91.7* | 92.2 *91.8* | 93.1 *91.2* | 93.1 *91.2* |
| Accuracy | 97.4 | 94.5 | 91.0 | 88.6 | 86.6 |
| $\kappa$ | 0.86 | 0.88 | 0.86 | 0.84 | 0.81 |

**Table 3.9:** Performance comparison of S1 stage sensitivity (%) among existing methods for $M = 5$-stage sleep classification.

| *SEDF-W* | | *SEDFx-S* | | *SEDFx-W* | |
|---|---|---|---|---|---|
| Zhu *et al.* [123] | 15.8 | | | | |
| Hassan *et al.* [124] | 47.0 | Imtiaz *et al.* [104] | 29.8 | Silveira *et al.* [125] | 6.1 |
| Hassan *et al.* [128] | 37.4 | | | | |
| **Proposed** | **51.7** | **Proposed** | **57.8** | **Proposed** | **49.1** |

however, discriminating the S1 stage still remains challenging. The sensitivity of the S1 is smaller than the sensitivities for the other sleep conditions.

Table 3.10 summarises the accuracies of various methods utilised for a different Sleep-EDF database. The accuracies were highest for the proposed method in $M = 2 - 6$ sleep state classification, except $M = 2$ stage classification in the *SEDF-W* dataset. Overall, the proposed method outperformed almost all other methods, regardless of the dataset extracted from the Sleep-EDF database.

**Table 3.10:** Performance comparison among existing methods for the Sleep-EDF database.

| Dataset | Methods | # of epochs | EEG ch. | $M = 2$ | $M = 3$ | $M = 4$ | $M = 5$ | $M = 6$ |
|---|---|---|---|---|---|---|---|---|
| *SEDF-W* | [123] | 14963 | Pz | 97.9/0.96 | 92.6/0.87 | 89.3/0.83 | 88.9/0.83 | 87.5/0.81 |
| | [124] | 15188 | Pz | **99.5**/- | 94.1/- | 92.1/- | 90.7/- | 86.9/- |
| | [128] | 15188 | Pz | 97.5/0.95 | 94.8/0.91 | 92.1/0.87 | 91.5/0.86 | 90.4/0.84 |
| | **MSFE + SEF** | **14979** | **Fpz** | 98.8/**0.98** | **97.1**/**0.95** | **95.3**/**0.93** | **93.8**/**0.90** | **93.0**/**0.89** |
| | | | **Pz** | 98.7/0.97 | **97.1**/**0.95** | 94.8/0.92 | 93.6/**0.90** | 92.8/**0.89** |
| | **MSPE + SEF** | | **Fpz** | 98.6/0.97 | 97.0/**0.95** | 94.8/0.92 | 93.4/**0.90** | 92.6/0.88 |
| | | | **Pz** | 98.5/0.97 | 96.9/**0.95** | 94.4/0.91 | 93.3/0.89 | 92.1/0.88 |
| *SEDFx-S* | [104] | 59316 | Fpz, Pz | -/- | -/- | -/- | 82.2/- | -/- |
| | **MSFE + SEF** | **58301** | **Fpz, Pz** | 97.2/0.85 | **94.5**/**0.88** | 90.6/0.85 | 87.9/0.83 | 85.8/0.80 |
| | **MSPE + SEF** | | | **97.4**/**0.86** | **94.5**/**0.88** | **91.0**/**0.86** | **88.6**/**0.84** | **86.6**/**0.81**6 |
| *SEDFx-W* | [125] | 106376 | Pz | 97.3/0.94 | 93.9/0.87 | 92.3/0.84 | 91.5/0.83 | 90.5/0.80 |
| | **MSFE + SEF** | **126577** | **Fpz** | **98.5**/**0.97** | **96.7**/**0.94** | **95.1**/**0.91** | **93.8**/**0.90** | **92.9**/**0.88** |
| | | | **Pz** | **98.5**/**0.97** | 96.6/**0.94** | 94.7/**0.91** | 93.6/0.89 | 92.6/**0.88** |
| | **MSPE + SEF** | | **Fpz** | 98.3/**0.97** | **96.7**/**0.94** | 94.7/**0.91** | 93.5/0.89 | 92.4/0.87 |
| | | | **Pz** | **98.5**/**0.97** | 96.6/**0.94** | 94.5/0.90 | 93.3/0.89 | 92.2/0.87 |

Accuracy(%)/$\kappa$. $M$: the number of sleep stage (Table 3.3).

## 3.4 Summary

The efficacy of structural complexity analyses of EEG signals in sleep stage classification has been investigated. This has been achieved by using the MSFE and MSPE, instead of using only spectral indices for sleep stage analysis. From the analyses using the Sleep-EDF database, It has been found that the proposed features could be effectively utilised for classifying sleep stages from a limited number of EEG channels. Additionally, it has been found that MSE is able to improve the performance of discriminating the S1 sleep stages, which has been the most challenging for automatic sleep stage classification tasks. In the next chapter, the feasibility of the proposed automatic sleep staging is further examined using a large-scale dataset.

# Chapter 4

# Automatic sleep staging for large scale data

## Overview

In the previous chapter, the automatic sleep staging based on the structural complexity analysis has been presented, and its performance has been evaluated using the Sleep-EDF dataset, which contains overnight sleep recordings from 61 subjects. The results have been based on data of limited scale, which compromises both the validation and the reproducibility of such automatic sleep staging; i.e., the computational issues have been neglected. This chapter further addresses the feasibility of the automatic sleep staging using a large collection of publicly accessible PSG data, recorded over 515 subjects.

## 4.1 Introduction

PSG is the gold standard for monitoring sleep and evaluating patients' sleep disorders in a clinic. The recorded physiological data are visually reviewed and manually scored by a human expert, according to some given guidelines, such as the AASM guideline [105]. Recently, automatic sleep staging approaches have been proposed in the research community, aiming to replicate human scoring based on PSG data [97, 123], which both provides economic savings and gives a way for future healthcare.

Automatic sleep staging approaches normally utilise supervised ML techniques with fine-tuned feature extraction for PSG data, and such studies typically propose novel combinations of features and classifiers [131, 132]. Prediction performances of such ML techniques are evaluated using either publicly available datasets (such as [93, 133, 134, 135, 136]) or proprietary data recorded as part of research projects [126]. However, only very few studies have demonstrated the feasibility of automatic classification, while catering rigorously for the generalisation ability for new PSG data.

The obstacles which prevent more widespread generalisation of findings in automatic sleep staging include numerous issues starting with the recording configuration during data acquisition; there are not always the same as the current standard – the AASM. For example, one of the most widely used datasets in automatic sleep staging research is PhysioNet Sleep-EDF [93], however, their EEG montage is Fpz-Cz and Pz-Oz (*cf.* C3-A2

and C4-A1 in AASM). Next, the size of data is usually small, typically recorded from less than 100 subjects. In other words, to become a robust substitution for labour-intensive human scoring, an ideal automatic scoring system needs to accurately capture not only the sleep patterns in its training data, but also to generalise well for unseen sleep data coming from different acquisition montages, different nights of the same subjects and different subjects. In addition, insufficient data for training leads to poor generalisation and impacts robustness in a real world setup.

Recently, some large-scale and heterogeneous sleep PSG datasets have been made publicly available [94]. Such 'Big Data' were created with the aim to validate the performance of ML methods when predicting sleep stages based on EEG from diverse patients; some automatic sleep staging approaches have also been rigorously validated using large-scale data [106, 137, 138]. For analyses with large-scale data, the level of scalability of the algorithm becomes a critical issue, together with the computational cost (e.g. execution time, resources), which increase with the size of data. Therefore, for automatic sleep staging in the real world to become a reality, it is not only the performance of an algorithm but also the computational costs that need to be scrutinised among different methods.

Here, automatic sleep staging is further examined from the viewpoints of feasibility and compatibility with the Big Data paradigms. This is achieved over the following aspects, which are fundamental for working systems in healthcare but almost ignored in the open literature:

1. Utilise a large-scale sleep data (from 515 individuals) to further validate the feasibility of classification methods;

2. Conduct comparative classification over multiple feature extraction methods and classifiers, including the technique proposed in Chapter 3, which was validated using the Sleep-EDF dataset (from 61 individuals, the channel configuration was not C3-A2 and C4-A1);

3. Investigate the variability of sleep staging performance based on the number of training subjects and model complexity, following initial suggestions by the previous study in [106];

4. Compare the performance not only in terms of the accuracy but also the computation time, both of which are critical for Big Data analyses.

Through these comprehensive and fair analyses, the viability of automatic sleep staging with both structural complexity and frequency domain features has been confirmed.

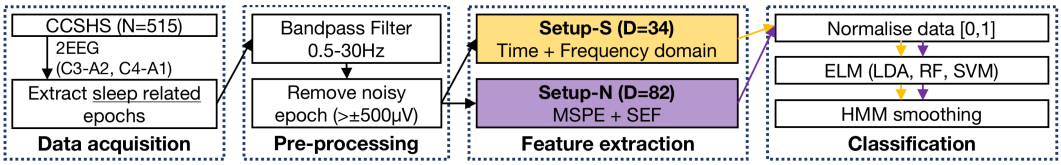



**Figure 4.1:** The flowchart of this study.

## 4.2 Methods

### 4.2.1 Data acquisition

The dataset from the Cleveland Children's Sleep and Health Study (CCSHS) [95] was used. It contains PSG recordings from 515 individuals, which were recorded from healthy subjects aged between 16 to 19 ($17.7 \pm 0.4$), and each file contains PSG data in the .edf format and a hypnogram scored by a clinician. Two EEG channels (C3-A2 and C4-A1) are used for this study. The data were ordered according to their ID numbers; the first 258 PSGs (from ID1800001 to ID1800460) were assigned as training data whereas the rest of 257 PSGs (from ID1800464 to from ID1800906) were used for testing.

### 4.2.2 Pre-processing

Some recordings in the CCSHS dataset contain both a large number of pre-sleep wake epochs and post-sleep wake epochs since some of the recordings were conducted over both a day and night period. First, 'sleep-related' data were only extracted that periods between twenty minutes before the first scored sleep epoch, and twenty minutes after the last scored sleep epoch, were considered. The so-extracted two channels of EEG data were filtered using a fourth-order Butterworth filter with the passband 0.5-30 Hz. Another EEG rejection method, based on the signal amplitude, was deployed. Any epoch with absolute voltage of more than 500 $\mu$V was removed from further analyses, which was the same analysis to [106] (*cf.* 400 $\mu$V in Chapter 3). In total, 1.4 % of the epochs were removed from further analyses. Table 4.1 shows the number of epochs and the ratio of sleep epochs in both training and testing data.

**Table 4.1:** Proportion of sleep stages with the number of epochs and ratio.

| | | | W | N1 | N2 | N3 | REM |
|---|---|---|---|---|---|---|---|
| Training | $N_{\text{train}} = 258$ | # of epochs | 32557 | 9694 | 125703 | 55737 | 50531 |
| | | Ratio (%) | 11.9 | 3.5 | 45.8 | 20.3 | 18.4 |
| Testing | $N_{\text{test}} = 257$ | # of epochs | 35095 | 9278 | 122423 | 54133 | 48765 |
| | | Ratio (%) | 13.0 | 3.4 | 45.4 | 20.1 | 18.1 |

### 4.2.3 Feature extraction

For a fair comparison with latest automatic sleep staging results with large scale data [106] and our previous idea given in Chapter 3, the features from each 30-second epoch were extracted in the same way as in [106] (*Setup-S*) and Chapter 3 (*Setup-N*). The details of feature extraction for this study is given in Table 2.3. Each feature value, $x$, was log-transformed as $\text{sign}(x)\log_{10}(|x+1|)$, and then normalised to $[0, 1]$, for each subject.

#### Setup-S

In Setup-S (the same feature extraction technique as in [106]), $D = 34$ features were extracted from each 30-seconds epoch; the details are discussed in Section 2.1.3 and Table 2.3.

#### Setup-N

In Setup-N (the same feature extraction technique as in [139] and in Chapter 3), $D = 82$ features were extracted from each 30-seconds epoch. As a structural complexity feature, MSPE was computed.The parameters for MSPE were: 1) the scale, $\tau = 1 - 20$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 20$), 2) the embedding dimension, $d = 5$, and 3) the time delay, $L = 1$. In total, 40 features (twenty features $\times$ two EEG channels) were extracted as structural complexity domain features. As a frequency domain feature, the SEF was used. In total, 42 features (three features $\times$ seven frequency bands $\times$ two EEG channels) were extracted as structural complexity domain features. Overall, 82 features (40 + 42) were extracted for Setup-N (see Table 2.3).

### 4.2.4 Classification

The classification implemented in Python. The ELM [113], LDA, RF, and SVM were used for classifiers, which were compared the prediction performance and computational time. The details of the ELM is given in Section 2.2.2. The ELM was implemented by `hpelm` package [115], whereas LDA, RF, and SVM were implemented by `sklearn` in Python. Table 4.2 summarises the packages and parameters for different classifiers.

**Table 4.2:** Parameters and implementation details for the selected classifiers.

| Algorithm | Framework | Package |
|---|---|---|
| LDA | `sklearn` | `LinearDiscriminantAnalysis` |
| RF | `sklearn` | `RandomForestClassifier (n_estimators=300)` |
| SVM | `sklearn` | `LinearSVC(C=1)` |
| ELM | `hpelm` | `ELM(classification='c')` |

### 4.2.5 Smoothing

The classifiers and their weight parameters were adjusted by the features from training samples, and the prediction was performed independently, based on the given feature vectors. In other words, this supervised classification approach only takes into account the current epochs, whereas manual scoring by human takes account of both neighbouring epochs and sleep transition. In Chapter 3, the overlapping window was applied (details can be found in Section 3.2.3), whereas the HMM was reported as being effective in improving the prediction accuracy using such sleep transitional information [106]. The details of the HMM are discussed in Section 2.3.1. It is assumed that the hidden states were hypnogram scored by a human in the training data, whilst the observations were predicted labels of the same training data. The transition matrix was given by counting the transition of the hypnogram in the training data, and the emission matrix was the obtained confusion matrix of training data (hypnogram v. prediction). The Viterbi algorithm (implemented by the Matlab build-in function `hmmviterbi`) was utilised to find the sequence of hidden states, which corresponds to 'smoothed' prediction by the algorithm.

## 4.3 Results

The pre-processing, feature extraction, and smoothing analyses were performed in Matlab 2016b, and the classifier was implemented in Python 2.7.12, operated on an iMac with 2.8GHz Intel Core i5, and 16GB of RAM. The versions of `sklearn` and `hpelm` were respectively 0.19.1 and 1.0.10.

#### ELM and the number of training subjects

First, the experiments were conducted using the ELM classifier and by changing the number of training subjects ($N_{\text{train}}$), ranging from $10$ to $258$. For each experiment with the selected number of training subjects, $N_{\text{train}}$, the simulation was repeated ten times; this was achieved independently and for randomly chosen subsets from training data (except for the experiment of $N_{\text{train}} = 258$, since all training data were used). The initial hidden weights and bias of ELM were randomly generated. For each simulation, the ELM model was trained to compute the Cohen's kappa value [117] (also see Chapter 2.4 in detail) using the confusion matrix derived from the same training subset (training performance); and then, the testing performance was evaluated by computing the kappa values of the confusion matrix, which were derived from all testing data (257 subjects). The statistical significance was evaluated by the Wilcoxon signed-rank test. The $p$-values were computed by comparing to the kappa values for each testing subjects from ten independent simulations (2570 samples).

Figure 4.2 shows the kappa values for different numbers of training subjects $N_{\text{train}}$, ranging from $10$ to $258$ subjects, using different features based on [106] (Setup-S) and [139]

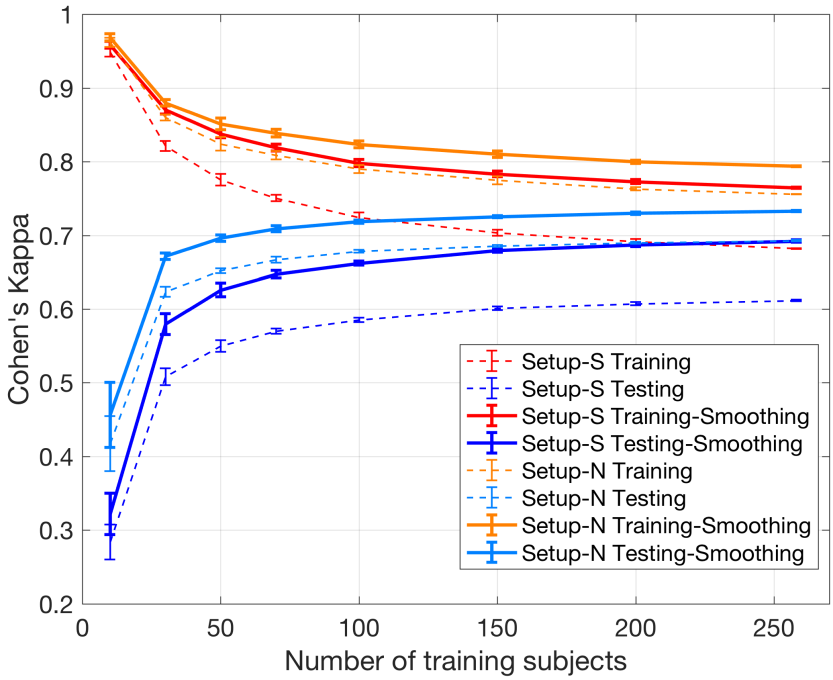



**Figure 4.2:** Kappa values for different numbers of training subjects in Setup-S and Setup-N ($L = 5000$). The error bar shows standard deviation of over ten independent iterations.

(Setup-N). The number of hidden nodes within ELM, $L = 5000$, was fixed. The highest averaged testing kappa values were obtained with $N_{\text{train}} = 258$, the values of which were 0.61 and 0.69 in Setup-S and Setup-N, respectively. To evaluate the statistical significance of increasing the number of training subjects, the testing performances were compared. In Setup-S, increasing the number of subjects until 200 led to the improvement of the performance: ($p$-values: 100 v. 150 – $4.6 \times 10^{-06}$, 150 v. 200 – 0.032, 200 v. 258 – 0.14), whereas there was no statistical significance between 150 and 200 in Setup-N: ($p$-values: 100 v. 150 – 0.011, 150 v. 200 – 0.14, 200 v. 258 – 0.31). The computational time of training increased with the number of training subjects ($N_{\text{train}} = 10$: 8 s, $N_{\text{train}} = 258$: 165 s in Setup-N), whereas the times for testing were the same regardless of $N_{\text{train}}$, as shown in Table 4.3.

These learning curves have similar shapes to the results in [106]; the training errors (kappa) are smaller than the testing errors when there are fewer training subjects. In [106], the training errors and testing errors converged when the number of training subjects was $N_{\text{train}} = 400$. The overfilling problem can be alleviated by regularisation (adding a regularisation term with a parameter). The regularisation term penalises functions that are overfitted. In the ELM framework, both $L^1$ (Lasso) and $L^2$ regularisation are available; the $L^2$ regularisation is called Tikhonov regularisation (ridge regression in statistics, or

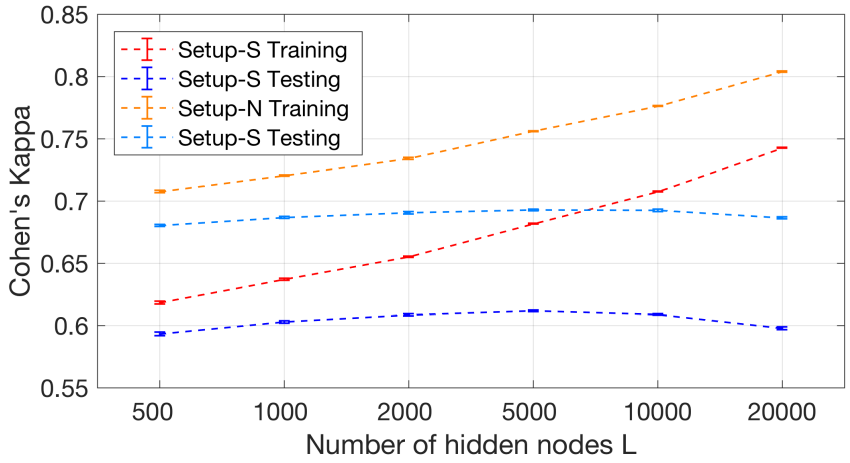



**Figure 4.3:** Kappa values for different numbers of hidden nodes in Setup-S and Setup-N ($N_{\text{train}} = 258$). The error bar shows standard deviation of over ten independent iterations.

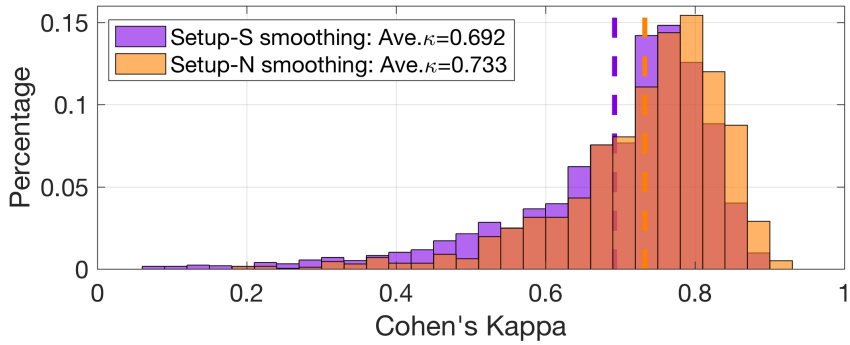



**Figure 4.4:** The histogram of the kappa value of test subjects in Setup-S and Setup-N ($N_{\text{train}} = 258, L = 5000$). The dashed lines indicate the average for each setup.

weight decay in machine learning). The advantage of $L^1$ regularisation over $L^2$ regularisation is that it produces simpler and sparse models which only involve a subset of the features. In other words, $L^1$ regularisation can be used for selecting the most 'effective' features for classification. On the other hand, $L^2$ regularisation keeps all features of the model. In general, one might expect the $L^1$ regularisation to perform better in a setting where a relatively small number of features has substantial coefficients, and the remaining features have coefficients which are very small or that are equal to zero [108], whereas $L^2$ regularisation will perform better when the prediction is a function of many features, all with coefficients of roughly equal size. In addition, it is feasible for ELM to handle imbalanced class prediction with weighted version of ELM. As sleep data typically consist of a large number of N2 sleep epochs, but, fewer N1 epochs, testing the efficacy of weighted ELM will be the subject of future work.

**Table 4.3:** Computational time of different analyses.

| Training/ *Testing* (s) | Setup-N $L = 500$ | 5000 | 20000 | Setup-S 5000 |
|---|---|---|---|---|
| $N_{\text{train}} = 10$ | -/- | 8/*51* | -/- | 8/*50* |
| 150 | -/- | 98/*51* | -/- | 97/*50* |
| 258 | 7/*3* | 165/*51* | 1679/*202* | 164/*50* |

### ELM and the number of hidden nodes

The experiments were conducted by changing the number of hidden nodes, $L$, of ELM, ranging from 500 to 20000, in order to evaluate the performance against model complexity. The number of training subjects was fixed to $N_{\text{train}} = 258$ (all training data were used). Figure 4.3 illustrates the kappa values for different numbers of hidden nodes, $L$, in Setup-S and Setup-N. In both setups, the highest averaged kappa values were achieved with $L = 5000$, while increasing the number of hidden nodes to more than $5000$ negatively affected the performance. The number of hidden nodes, $L$, also affects computational time, especially for training, as shown in Table 4.3; the averaged training time of $L = 20000$ was 1679 s ($\approx$ 28 mins), which was approximately ten times larger than simulation, with $L = 5000$ (165 s).

### HMM smoothing

The predictions after HMM smoothing are given in Figure 4.2 (bold solid lines). Using all training data ($N_{\text{train}} = 258$), the achieved kappa values were 0.69 (*cf.* 0.61 before smoothing) in Setup-S, and 0.73 (*cf.* 0.69 before smoothing) in Setup-N. Among 257 testing subjects, the prediction performance was improved by smoothing for 96 % of subjects in Setup-S, whereas for 93 % of subjects in Setup-N. Figure 4.4 shows the histogram of kappa values for all testing subjects in both Setup-S and Setup-N. The prediction was made with fixed parameters of $N_{\text{train}} = 258$, $L = 5000$, and a smoothing algorithm was applied.

Figure 4.5 depicts the confusion matrices obtained with different settings of ELM, upon changing the number of training subjects $N_{\text{train}}$, the number of hidden nodes $L$, and applying smoothing. The upper confusion matrices were results for Setup-S whilst lower ones were for Setup-N. Comparing to the matrices in the second column (i.e. $N_{\text{train}} = 258$, $L = 5000$, before smoothing) and the third column (the same $N_{\text{train}}$ and $L$, but after smoothing), the sensitivities of each class improved with smoothing, especially for the N1 stage (from 4.0 % to 20.6 % in Setup-S, and from 0.8 % to 20.3 % in Setup-N). In the hypnogram plot, the advantage of smoothing was also confirmed. Figure 4.6 depicts an overnight hypnogram of one subject; the upper graph shows the manually scored hypnogram based on the PSG recordings (Black) whereas the bottom panel shows the automatically predicted label based on the ELM ($N_{\text{train}} = 258$, $L = 5000$) before smoothing (Blue) and after smoothing (Red). The smoothing successfully took account of neighbouring epochs and eliminated fragmentation in prediction. The kappa value for this subject

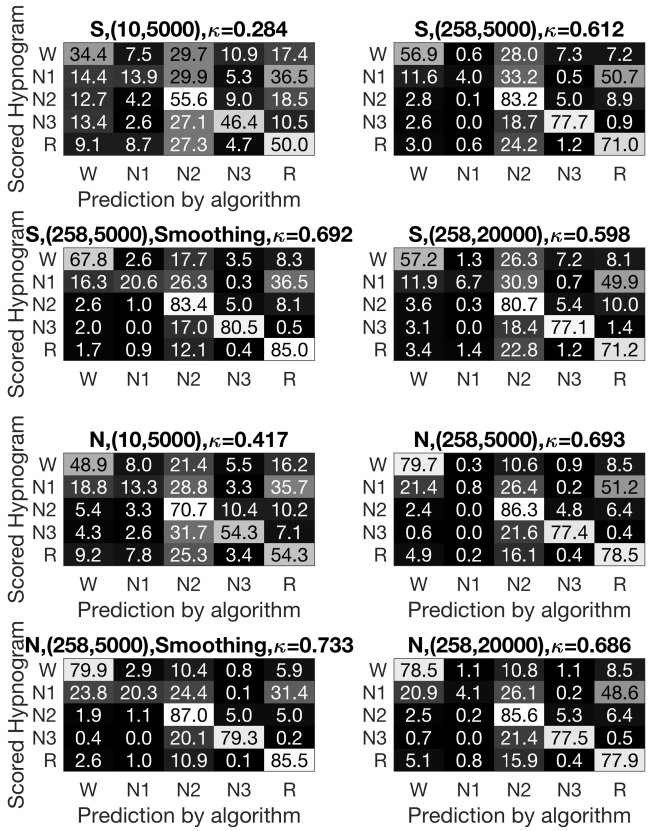



**Figure 4.5:** The averaged confusion matrices over all 257 testing subjects for different scenarios. The title of each confusion matrix denotes: [Setup-{S,N}, ($N_{\text{train}}$, $L$), kappa value $\kappa$]. Observe that the sensitivities of each class improved after HMM smoothing.

was 0.81 after smoothing, improved from 0.67 before.

### Performance with different classifiers

Finally, the performance of different classifiers, the LDA, RF and SVM, was evaluated. Table 4.4 summarises the prediction performance and computational time (Training time/*Testing time*) for different classifiers in Setup-N. The highest kappa value, $\kappa$, was obtained by RF (0.695), whereas the RF required the longest training time (878 s).

Figure 4.7 shows the distribution of kappa values using different classifiers, LDA, RF, SVM, and ELM. For the classification, the number of training subjects was fixed to $N_{\text{train}} = 258$ (all training data were used). The boxplot was obtained from all 257 testing subjects. For RF and ELM, the simulation was repeated ten times to average over such ten independent simulations. For evaluating statistical significance between different classi-

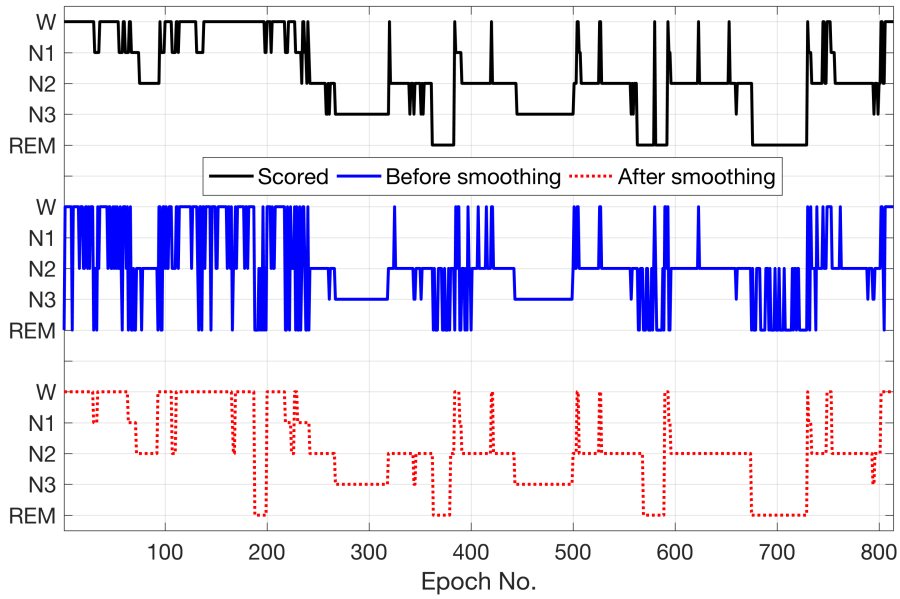



**Figure 4.6:** (Black) Scored hypnogram of one subject; (Blue/Red) automatically predicted label before/after smoothing.

fication methods, the non-parametric ANOVA (i.e. Kruskal-Wallis H test), followed by *post hoc* Dunn's pairwise test with Bonferroni correction was calculated. The plot shows that the ELM has significantly better testing kappa than both LDA and SVM, whereas the performance of RF is similar to that of ELM. The similar results are reported in [106].

**Table 4.4:** Averaged kappa and computational time of different classifiers.

| | LDA | RF | SVM | ELM |
|---|---|---|---|---|
| $\kappa$ | 0.67±0.11 | 0.69±0.11 | 0.67±0.11 | 0.69±0.10 |
| Training/*Testing* time (s) | 2/*0.1* | 878/*23* | 308/*0.1* | 165/*51* |

## 4.4 Discussion

The automatic sleep staging using a publicly available large-scale dataset, CCSHS ($N = 515$), was evaluated in this chapter. The selected structural complexity and frequency domain features have been classified by the ELM classifier, and this achieved average kappa value was 0.73 with the HMM smoothing. This classification approach is 'conventional' ML based on the fine-tuned feature engineerings and supervised classifier.

Since late 2016, a large number of automatic sleep stage classification based on 'deep learning' has been proposed and published in the research community. The different neural networks, such as recurrent neural networks (RNN) and convolutional neural net-

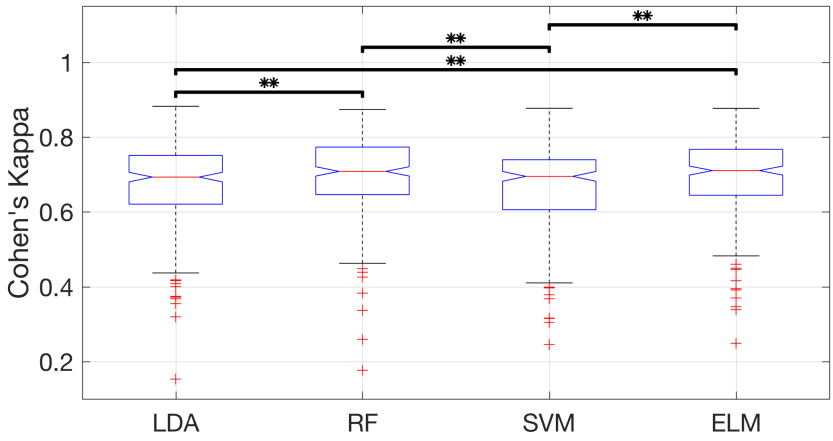



**Figure 4.7:** The distribution of kappa values for different classifiers. "**" represents the Bonferroni corrected p-value from *post hoc* Dunn's pairwise test is less than 0.01.

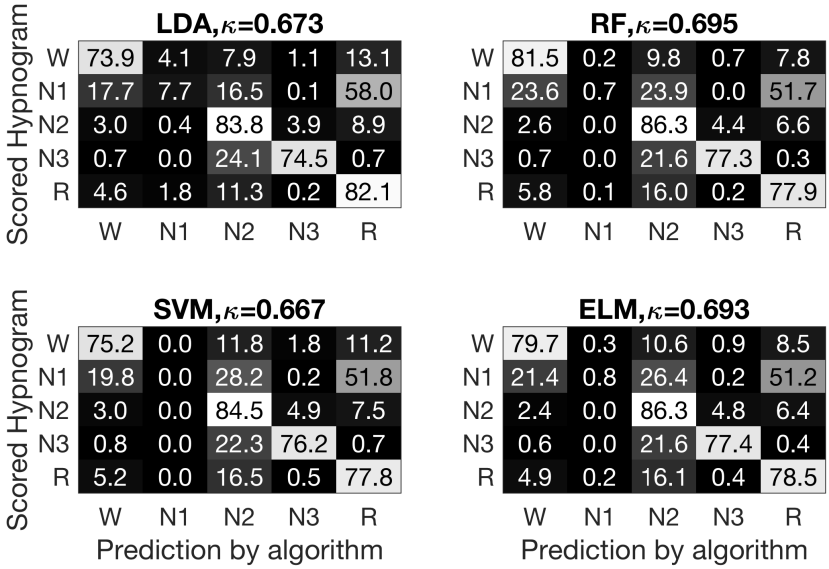



**Figure 4.8:** The confusion matrices for the fixed 257 testing subjects and for different classifiers.

**Table 4.5:** Recent works for automatic sleep stage classification with neural networks.

| Authors | Dataset* | # of subjects | Channel | E2E** | Performance |
|---|---|---|---|---|---|
| [140] | Sleep-EDF | 20 | 1EEG | ◯[A] | AC=80.3 % |
| [141] | Sleep-EDF | 20 | 1EEG | ◯ | AC=74.8 % |
| [142] | Sleep-EDF | 20 | 1EEG | ◯ | AC=82.0 %, $\kappa$=0.76 |
| [143] | UCD | 25 | 1EEG | ◯[B] | AC=92 %, $\kappa$=0.84 |
| [144] | Sleep-EDF | 61 | 1EEG | ◯ | AC=87.9 % |
| [145] | MASS | 62 | 6EEG, 2EOG, 3EMG | ◯ | -[C] |
| [146] | MASS | 62 | 1EEG-EOG | × | AC=85.9 % |
| [142] | MASS | 62 | 1EEG-EOG | ◯ | AC=86.2 %, $\kappa$=0.80 |
| [147] | MASS | 200 | 1EEG, 1EOG, 1EMG | ◯[A] | AC=87.1 %, $\kappa$=0.815 |
| [148] | SHHS | 5793 | 1EEG | ◯ | AC=87 %, $\kappa$=0.81 |
| [138] | Own | 1617[D] | 2EEG, 2EOG | ◯ | AC=81.4 %, $\kappa$=0.74 |
| [137] | Own | 10000 | 6EEG | × | AC=85.8 %, $\kappa$=0.795 |
| [149] | CCSHS+CFS | 198 | 1EEG | × | AC=85.6 %, $\kappa$=0.80 |
| [139] – Ch. 3 | Sleep-EDF | 61 | 1EEG | × | AC=93.8 %, $\kappa$=0.90 |
| [150] – Ch. 4 | CCSHS | 515 | 2EEG | × | $\kappa$=0.73 |

Dataset* – UCD: University College Dublin, MASS: Montreal Archive of Sleep Studies, SHHS: Sleep Heart Health Study, CFS: Cleveland Family Study.
E2E**: End-to-end. [A]: The input is spectrogram of EEG signal.
[B]: The input is complex value of EEG signal. [C]: Overall performance is not given.
[D]: Data from healthy subjects for training, from patients for testing.

works (CNN) with deep architecture, have been used in automatic sleep staging. Table 4.5 summarises the examples of 'deep learning' based automatic sleep staging articles, some recent studies using large-scale datasets, and our study proposed in both Chapter 3 and this chapter. The table outlines differences include 1) the benchmarking dataset, 2) the use of data (e.g. number of the channels), and 3) the performance metrics (for convenience, only summarised the accuracy and corresponding kappa value). Due to these differences, the comparison among these studies is not possible; however, the 'deep learning' based approaches do not always outperform than the conventional ML approaches in automatic sleep stage classification.

## 4.5 Summary

In this chapter, the feasibility of automatic sleep staging has been further examined in Big Data scenarios using a publicly available large-scale dataset, CCSHS ($N$=515). Extensive experiments have been conducted to evaluate the scalability of the proposed methods by changing the number of training subjects, model complexity, and choice of classifiers; these analyses were inspired by [106]. Another virtue of this current work is the use of an HMM state machine for physically meaningful discrimination between 'state transition' associated with the evolution of the scored sleep stages, resulting in much smoother and more accurate automatic sleep scores. The extracted features in both structural complexity and frequency domains have been classified by a computationally efficient ELM algorithm, and the achieved average kappa value was 0.73 with the HMM smoothing.

# Part II

# Proof-of-concept in-ear EEG applications

# Chapter 5

# Automatic sleep staging using in-ear EEG for nap data

## Overview

In Part I, automatic sleep stage classification methods based on the on-scalp EEG have been evaluated, using publicly available overnight sleep datasets. Such automatic sleep staging are shown to be a proxy for time-consuming conventional manual sleep scoring by a clinician. In terms of recordings, the PSG is a standard clinical methodology to monitor sleep in current medicine. The so-acquired sleep profiles of individuals are rigorous and comprehensive, however, the expensive and cumbersome nature of PSG may even disturb patients normal sleep, thus affecting diagnosis and treatment. In this chapter, automatic wearable sleep monitoring in the community is examined, by analysing the agreement between the automatically predicted sleep stages by in-ear EEG and on-scalp EEG patterns.

## 5.1 Introduction

Sleep is an essential process in the internal control of the state of body and mind and its quality is strongly linked with a number of cognitive and health issues, such as stress, depression and memory [119]. For clinical diagnostic purposes, PSG has been extensively utilised which is based on a multitude of physiological responses, including the EEG, EOG, and EMG. While the PSG is able to faithfully reflect human sleep patterns, both the recording and scoring process are expensive as this involves an overnight stay in a specialised clinic and time-consuming manual scoring by a medically trained person. In addition, hospitals are unfamiliar environments for patients, which compromises the reliability of the observed sleep patterns. In other words, the conventional recording process is not user-centred and not ideal for long-term sleep monitoring.

With the advance in wearable physiological monitoring devices, it has become possible to monitor some of sleep-related physiological responses out of the clinic. The next step towards sleep care in the community is therefore to monitor sleep-related physiological signals in an affordable way, at home, and over long periods of time, together with automatic detection of sleep patterns (sleep scoring) without the need for a trained med-

ical expert. Indeed, consumer technologies are becoming increasingly popular for the self-monitoring of sleep [151], and include both smartphone applications (softwares) and wearable devices. While such technologies aim to assess 'sleep quality' and are affordable, these are typically not direct measures of neural activity, and instead measure indirect surrogates of sleep such as limb movement [152].

Another fast developing aspect of sleep research is automatic sleep scoring, with the aim to replace the time-consuming manual scoring of sleep patterns from full PSG with computer software, as already discussed in Part I of this thesis. Automatic sleep stage scoring employs ML algorithms, and it is now possible to achieve up to 90 % accuracy of classification between the W, N1, N2, N3 and REM sleep stages from a single channel on-scalp EEG [125, 126].

More recent approaches for sleep monitoring aim to move beyond actigraphy and develop advanced multimodal sensors and wearable devices. In this direction, Le *et al.* introduced a wireless wearable sensor to monitor vectorcardiography (VCG), ECG, and respiration for detecting obstructive sleep apnea in real time [153]. Using a wearable in-ear EEG sensor [36], Looney *et al.* evaluated sleep stages during nap episodes from a viscoelastic in-ear EEG sensor [15]. They performed simultaneous sleep monitoring from four subjects, using both on-scalp and in-ear EEG data channels, and reported Substantial Agreement between the corresponding hypnograms, manually and blindly scored by a trained clinician, as shown in Figure 5.1A. The in-ear EEG data were recorded from one-size-fits-all viscoelastic earpieces [41].

The in-ear sensing technology has been proven to provide sufficiently good EEG signal. Even though amplitude of in-ear EEG is smaller than that of on-scalp EEG, the signal-to-noise ratio (SNR) was found to be similar [36, 39, 41]. In a sleep monitoring scenario, in-ear wearable sensors have the following advantages:

- *Affordability and unobtrusiveness*: The in-ear sensor is made from viscoelastic material [41], such as those used in standard earplugs.
- *User-centred nature*: Users are able to insert the sensor by themselves as when wearing earplugs. The device is comfortable to wear and does not disturb sleep.
- *Robustness*: The sensor expands after the insertion and maintains a stable interface with the ear canal, and is thus not likely to dislodge during sleep.

In this chapter, automatic wearable sleep monitoring in the community is examined, by analysing the agreement between the automatically predicted sleep stages by in-ear EEG and on-scalp EEG patterns. To this end, the sleep-related EEG-patterns were obtained from both scalp and inside the ear simultaneously, using a stationary data acquisition unit. For rigour, the in-ear EEG automatic scoring procedures were validated for the following scenarios:

1. Agreement between automatically predicted sleep stages based on in-ear EEG patterns and the manually scored hypnogram from in-ear EEG (*Scenario 1*).

2. Agreement between automatically predicted sleep stages based on in-ear EEG patterns and the manually scored hypnogram from on-scalp EEG (*Scenario 2*).

Figure 5.1B illustrates the proposed analysis framework. The results are benchmarked against the results in [15] where both the on-scalp and in-ear EEG hypnograms were scored manually, corresponds to Figure 5.1A. In this way, a proof-of-concept for the feasibility of in-ear EEG in automatic scoring of sleep patterns out-of-clinic and in the community is established.

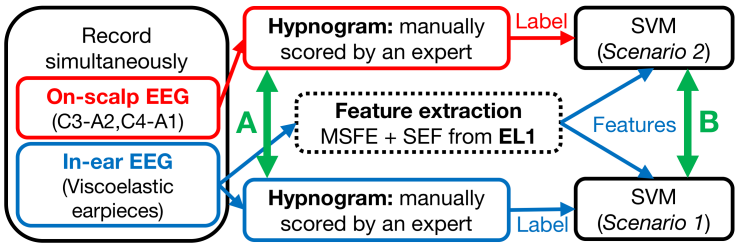



**Figure 5.1:** Comparison with previous studies. (A) Evaluation of the agreement between the manually scored hypnograms based on on-scalp EEG channels and in-ear EEG channels [15]. (B) Our analysis framework for establishing the feasibility of in-ear EEG in sleep research.

## 5.2 Methods

### 5.2.1 Data acquisition

The EEG recordings were conducted at Imperial College London between May 2014 and March 2015 under the ethics approval, ICREC 12_1_1, Joint Research Office at Imperial College London. Four healthy male subjects (age: 25-36 years) without history of sleep disorders participated in the recordings. All participants were instructed to reduce their sleep to less than five hours the night before, and agreed to refrain from consuming caffeine and napping on the recording day. The four on-scalp EEG channels C3, C4, A1 and A2 (according to international 10-20 system), were recorded using standard gold-cup electrodes. The forehead was used for the ground, and the standard configurations for sleep scoring were utilised (i.e. C3-A2 and C4-A1). The in-ear EEG was recorded from both the left and right ear, and the in-ear EEG sensor was made based on a viscoelastic earplug with two cloth electrodes [41]. Earwax was removed from the ear canals, and the sensor expanded after the insertion, to conform to the shape of the ear canal. The reference gold-cup standard electrodes were attached behind the ipsilateral mastoid and the ground electrodes were placed on the ipsilateral helix, as illustrated in Figure 5.2. Both on-scalp EEG and in-ear EEG were recorded simultaneously using the g.tec g.USBamp amplifier with 24-bit resolution, at a sampling frequency $f_s = 1200$ Hz.

The participants seated in a comfortable chair in a dark and quiet room. The duration




**Figure 5.2:** Recording setup in this study. (Left) The electrodes were placed on scalp and ear; (Centre) the subject reclined in a comfortable chair; (Right) wearable in-ear sensor with two flexible electrodes.

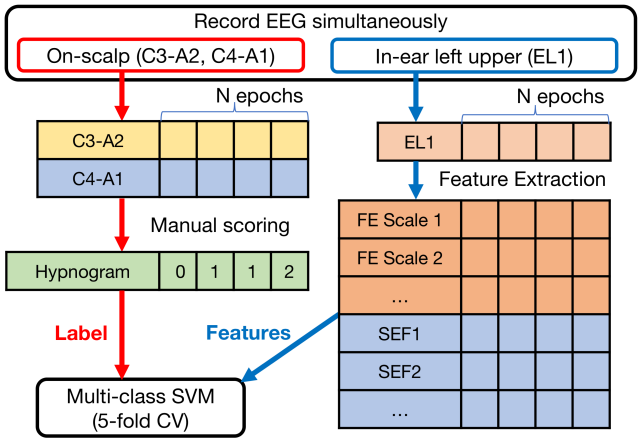



**Figure 5.3:** Flowchart for the sleep stage prediction framework adopted in this study (*Scenario 2*).

of recording was 45 minutes, while to increase the number of transitions between the wake and sleep stage, a loudspeaker played 10 s abrupt noise at random intervals.

### 5.2.2 Manual scoring

For on-scalp EEG, a fourth-order Butterworth bandpass filter with passband 1-20 Hz was applied to two bipolar EEG configurations (C3-A2 and C4-A1). Due to low-frequency interference in in-ear EEG channels, the low cutoff frequency was set to 1 Hz for the Subject One and Three, and 2 Hz for the Subject Two and Four. Next, the in-ear EEG amplitudes were normalised to the same range as those of on-scalp EEG, and both on-scalp EEG and in-ear EEG were manually scored by a clinical expert, who had six years of experience in EEG-based sleep stage scoring. The processed EEG data was blinded and the epoch-based manual sleep scoring was performed according to the AASM criteria [105]. The epoch size was set to 30 s, therefore 90 epochs were scored in each recording.

### 5.2.3 Pre-processing for automatic sleep stage classification

For automatic sleep stage classification, the recorded EEG from the left ear channel 1 (EL1) is used, for a fair comparison with automatic scoring algorithms for a single EEG channel montage in the literature. First, the data was downsampled to 200 Hz, and the epochs with the amplitudes of more than $\pm 400\,\mu$V were removed from subsequent analyses. The data were then bandpass filtered with the passband of 0.5-30 Hz. The pre-processing resulted in a loss of approximately 20 % of the data, and eventually 293 (hypnogram based on on-scalp EEG, W:67, N1:46, N2:140, N3:40, and hypnogram based on in-ear EEG, W:52, N1:49, N2:162, N3:30) epochs were used for the classification. Figure 5.4 shows the averaged PSD for on-scalp and in-ear EEG channels with 4-class sleep stages.

### 5.2.4 Feature extraction

After the pre-processing, two types of features, MSFE and SEF, were extracted from each epoch of the EEG. The details of analysis is found in Section 2.1.1 and 2.1.2. The parameters used to calculate the MSFE were $\tau = 1 - 30$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 30$), $m = 2$, $r = 0.15\times$(*standard deviation of each epoch*). Figure 5.5 illustrates MSFE analysis for in-ear EEG (EL1) channel with 4-class sleep stages. The frequency ranges for SEF50, SEF95, and SEF$d$ were chosen in the same bands given in Table 2.3 (Setup-N).

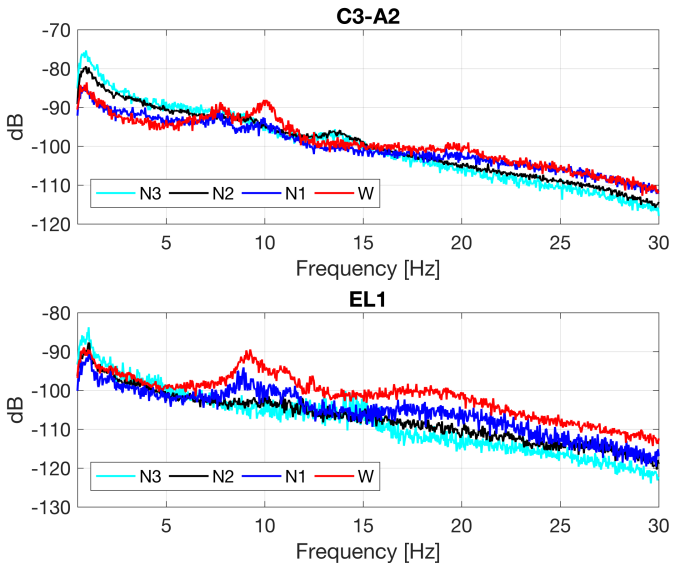



**Figure 5.4:** PSD: (Top) On-scalp C3-A2 montage; (Bottom) in-ear channel EL1.

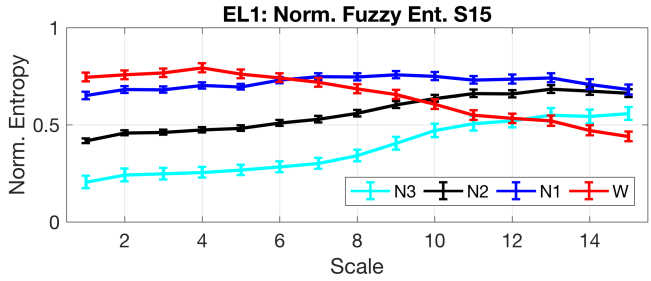



**Figure 5.5:** Structural complexity features for different sleep stages. Normalised MSFE from the in-ear EEG channel EL1 shows excellent separation between sleep stages. The error bars indicate the standard error.

### 5.2.5 Classification

Classification was performed based on the SEF and MSFE features, which were normalised to the range $[0, 1]$. The one-against-one multi-class SVM with a RBF kernel was employed as a classifier.

### 5.2.6 Performance evaluation

Feature extraction was performed using Matlab 2016b, and the classification was conducted in Python 2.7.12, Anaconda 4.2.0 operated on an iMac with 2.8GHz Intel Core i5, 16GB of RAM. A 5-fold CV was performed to evaluate the automatic sleep stage classification. The performance indicators to evaluate the proposed methods were class-specific sensitivity (SE) and precision (PR), as well as overall accuracy and kappa coefficient ($\kappa$); the details are given in Section 2.4.

## 5.3 Results

**Scenario 1: Sleep stage classification from in-ear EEG against the manually scored hypnogram based on in-ear EEG**

First, the agreement between the hypnogram scored based on in-ear EEG channels and the predicted label based on extracted features from the in-ear EEG channel EL1 was evaluated. Tables 5.1, 5.2, and 5.3 show the confusion matrices obtained from the classification results based on the SEF and MSFE features for the 2-class scenarios Wake v. Sleep and W-N1 v. N2-N3, and the 4-class (W, N1, N2, N3) scenario. For the 2-class classification scenarios, the overall classification accuracies were respectively 95.2 % and 86.0 %, with an Almost Perfect ($\kappa = 0.83$) to Substantial ($\kappa = 0.68$) Agreement of kappa coefficients [117], as shown in Table 5.1 and 5.2.

The accuracy for the more difficult 4-class sleep stage classification was 78.5 % with the kappa coefficient $\kappa = 0.64$, which indicates a Substantial Agreement, as shown in

**Table 5.1:** Confusion matrix for the 2-class Wake v. Sleep classification.

| | | Algorithm based on in-ear EEG | | |
|---|---|---|---|---|
| | | Wake | Sleep | SE(%)/*PR(%)* |
| Score based on in-ear EEG | Wake | 42 | 10 | 80.8 / *91.3* |
| | Sleep | 4 | 237 | 98.3 / *96.0* |
| Accuracy = 95.2 %, $\kappa$ = 0.83 (SE:Sensitivity, *PR:Precision)* | | | | |

**Table 5.2:** Confusion matrix for the 2-class Wake-N1 v. N2-N3 classification.

| | | Algorithm based on in-ear EEG | | |
|---|---|---|---|---|
| | | W-N1 | N2-N3 | SE(%)/*PR(%)* |
| Score based on in-ear EEG | W-N1 | 76 | 25 | 75.3 / *82.6* |
| | N2-N3 | 16 | 176 | 91.7 / *87.6* |
| Accuracy = 86.0 %, $\kappa$ = 0.68 (SE:Sensitivity, *PR:Precision)* | | | | |

Table 5.3.

**Table 5.3:** Confusion matrix for 4-class sleep stage classification.

| | | Algorithm based on in-ear EEG | | | | |
|---|---|---|---|---|---|---|
| | | W | N1 | N2 | N3 | SE(%)/*PR(%)* |
| Score based on in-ear EEG | W | 46 | 3 | 3 | 0 | 88.5 / *92.0* |
| | N1 | 3 | 17 | 27 | 2 | 34.7 / *56.7* |
| | N2 | 1 | 10 | 146 | 5 | 90.1 / *78.9* |
| | N3 | 0 | 0 | 9 | 21 | 70.0 / *75.0* |
| Accuracy = 78.5 %, $\kappa$ = 0.64 (SE:Sensitivity, *PR:Precision)* | | | | | | |

**Scenario 2: Sleep stage classification from in-ear EEG against the manually scored hypnogram based on on-scalp EEG**

Second, the agreement between the hypnogram scored based on on-scalp EEG channels and the predicted label based on extracted features from the in-ear EEG channel EL1 was evaluated. Tables 5.4, 5.5, and 5.6 show the corresponding confusion matrices, obtained from the classification based on the SEF and MSFE features for the 2-class Wake v. Sleep and W-N1 v. N2-N3 scenarios, and the 4-class (W, N1, N2, N3) scenario. For the 2-class classification problems, the achieved classification accuracies were more than 90 %, with the Substantial Agreements ($\kappa = 0.75$ and $\kappa = 0.80$) [117].

The achieved accuracy for the 4-class sleep stage classification was 76.8 %, with the kappa coefficient $\kappa = 0.65$, which indicates a Substantial Agreement.

Figure 5.6 depicts the hypnograms scored manually based on on-scalp EEG channels (Blue) and the automatically predicted label based on the in-ear EL1 channel (Red) for the 2-class Wake v. Sleep (Top) and W-N1 v. N2-N3 (Middle) scenarios, and the 4-class (Bottom) scenario, for the Subject Two. Only the first epoch was removed because of the electrical noise when the amplifier starts, therefore the hypnogram was scored based on 89 epochs, which corresponds to 44 minutes of 30 s recording. For the 4-class problems, even

**Table 5.4:** Confusion matrix for the 2-class Wake v. Sleep classification.

| | | Algorithm based on in-ear EEG | | |
|---|---|---|---|---|
| | | Wake | Sleep | SE(%)/*PR(%)* |
| Score based on on-scalp EEG | Wake | 50 | 17 | 74.6 / *87.7* |
| | Sleep | 7 | 219 | 96.9 / *92.8* |
| Accuracy = 91.8 %, $\kappa$ = 0.75 (SE:Sensitivity, *PR:Precision)* | | | | |

**Table 5.5:** Confusion matrix for the 2-class Wake-N1 v. N2-N3 classification.

| | | Algorithm based on in-ear EEG | | |
|---|---|---|---|---|
| | | W-N1 | N2-N3 | SE(%)/*PR(%)* |
| Score based on on-scalp EEG | W-N1 | 96 | 17 | 85.0 / *90.0* |
| | N2-N3 | 11 | 169 | 89.7 /*90.9* |
| Accuracy = 90.4 %, $\kappa$ = 0.80 (SE:Sensitivity, *PR:Precision)* | | | | |

though some epochs were predicted incorrectly, for example epoch 62 (hypnogram:N3, prediction:N2), the majority of epochs were correctly classified. This confirms that the features extracted from the in-ear EEG data were effectively used for the automatic sleep stage classification, and provided a substantial match to the on-scalp EEG patterns scored manually by an expert. Therefore, it is empirically confirmed that the recorded in-ear EEG carried a sufficient amount of information to evaluate human sleep robustly.

### Agreement between the predicted and manual sleep scores

Upon establishing the feasibility of predicting on-scalp EEG sleep stages from in-ear EEG features, these findings were benchmarked against our recent results based on manual scoring of both on-scalp and in-ear EEG [15]. To this end, Table 5.7 compares the manual and automatic labels for the following scenarios:

- *Scenario 1*: The manually scored hypnogram based on in-ear EEG channels v. the predicted label based on the in-ear EL1 channel (Table 5.1, 5.2, and 5.3).
- *Scenario 2*: The manually scored hypnogram based on on-scalp EEG channels v. the predicted label based on the in-ear EL1 channel (Table 5.4, 5.5, and 5.6).
- The hypnogram manually scored based on on-scalp EEG channels v. that scored based on in-ear EEG channels.

**Table 5.6:** Confusion matrix for the 4-class sleep stage classification.

| | | Algorithm based on in-ear EEG | | | | |
|---|---|---|---|---|---|---|
| | | W | N1 | N2 | N3 | SE(%)/*PR(%)* |
| Score based on on-scalp EEG | W | 53 | 5 | 8 | 1 | 79.1 / *77.9* |
| | N1 | 12 | 23 | 11 | 0 | 50.0 / *65.7* |
| | N2 | 3 | 5 | 125 | 7 | 89.3 / *79.1* |
| | N3 | 0 | 2 | 14 | 24 | 60.0 / *75.0* |
| Accuracy = 76.8 %, $\kappa$ = 0.65 (SE:Sensitivity, *PR:Precision)* | | | | | | |

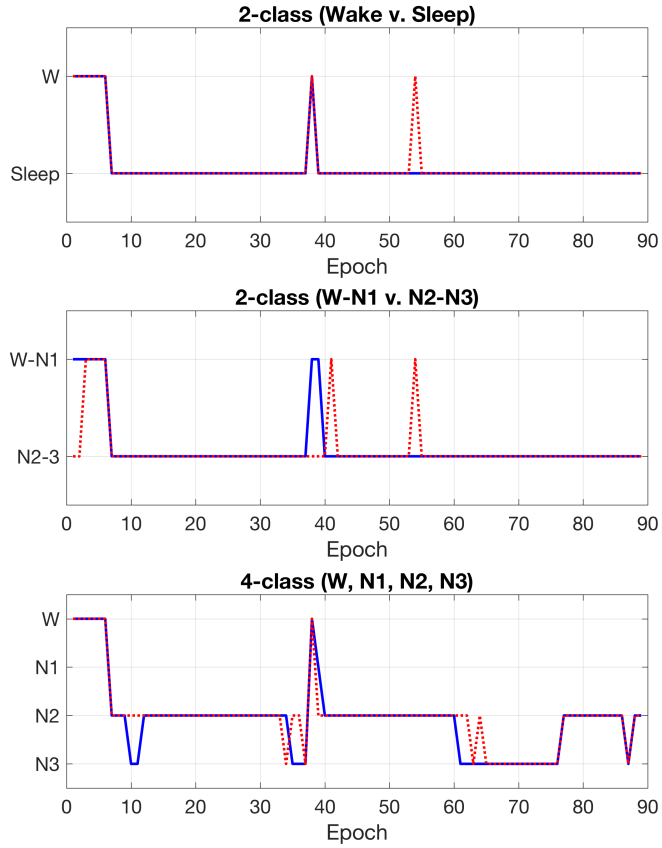



**Figure 5.6:** Hypnogram for Subject Two scored based on on-scalp EEG channels (Blue) and the automatically predicted label based on in-ear EEG channel EL1 (Red). (Top) 2-class Wake v. Sleep; (Middle) 2-class W-N1 v. N2-N3; (Bottom) 4-class classification scenario.

In all cases, the proposed automatically scored labels were a significant match to the corresponding labels scored manually in [15].

**Table 5.7:** Comparison between the manual scores and automatic predicted scores. Accuracy (%) / Kappa.

| Methods | *Scenario 1* | *Scenario 2* | Looney *et al.* [15] |
|---|---|---|---|
| Agreement | Score (in-ear EEG) v. Prediction (in-ear EEG) | Score (on-scalp EEG) v. Prediction (in-ear EEG) | Score (on-scalp EEG) v. Score (in-ear EEG) |
| Wake v. Sleep | 95.2 / 0.83 | 91.8 / 0.75 | 84.0 / 0.60 |
| W-N1 v. N2-N3 | 86.0 / 0.68 | 90.4 / 0.80 | 83.0 / 0.65 |
| 4 class | 78.5 / 0.64 | 76.8 / 0.65 | - / - |

## 5.4 Summary

In this chapter, an automatic sleep stage monitoring system using in-ear EEG, which is capable of minimising both patient's inconvenience and the involvement of a medical specialist, has been evaluated. For rigour, the experiments have been conducted in two scenarios: *Scenario 1* examined automatic scores for in-ear EEG against manual scores for in-ear EEG, while *Scenario 2* examined automatic scores for in-ear EEG against manual scores for on-scalp EEG. This has provided a proof-of-concept for the feasibility of in-ear EEG in automatic scoring of sleep patterns out-of-clinic and in the community. Therefore, it has been empirically confirmed that the in-ear EEG carries sufficient amount of information to faithfully represent human sleep patterns. For this pilot study the number of subjects was four; the studies with a larger cohort of subjects, overnight sleep, and other aspects of fully wearable scenarios are given in Part III of this thesis.

# Chapter 6

# Automatic detection of drowsiness using in-ear EEG

## Overview

In the previous chapter, the feasibility of in-ear EEG for sleep monitoring using nap data has been examined. In this chapter, a further proof-of-concept of the utility of in-ear EEG in day time drowsiness monitoring is investigated from validating whether in-ear EEG patterns can classify between wakefulness and light sleep.

## 6.1 Introduction

Although the function of sleep is not yet fully understood, sleep is strongly linked to quality of life. For example, the lack of sleep during the night contributes to the feeling of tiredness, and poor memory and work performance. The concept of 'sleepiness' can be interpreted as a tendency towards sleep; the inability to concentrate or to keep eyes open may indicate a higher level of sleepiness. For several decades, multiple researchers have investigated sleepiness and performance capacities related to physiological patterns, especially EEG and EOG for shift workers [154], and for healthy subjects [155]. These studies have shown that the alpha and theta activity, in particular, exhibit significant correlations between both subjective and objective measures of sleepiness. Physiological patterns have been investigated and linked to mental workload using a driving simulator for healthy subjects [156], while a recent study of obstructive sleep apnea (OSA) patients elucidated that auditory event related potentials, appear to be a significant parameter to predict poor driving performance [157].

In order to uncover physiological characteristics of certain states of body and mind, such as sleepiness, a multitude of physiological electrodes are required. In terms of sleep monitoring, PSG recordings have been traditionally used in clinic to diagnose various sleep disorders. Typically, PSG recordings require at least two EEG channels, two EOG channels to observe eye movements, and at least one chin EMG channel. However, this conventional setup is both a cumbersome and expensive procedure as it requires a trained clinician; this is therefore prohibitive to the physiological states in the real world.

Automatic sleep staging systems have been proposed based on PSG recordings; this was originally motivated to replace the time-consuming sleep stage scoring process by clinicians. The state-of-the-art approaches include single-channel on-scalp EEG based classification has been validated using various benchmark datasets [126, 139], which were also discussed this in Part I of this thesis. For wearable sensing, in-ear EEG technology [36] has been recently used in sleep monitoring [15, 69]. The technology can vastly improve the technically cumbersome setup for sleep monitoring, as the latest in-ear EEG sensor is readily wearable by users, and does not disturb their daily activities. In other words, the unobtrusive and wearable in-ear sensors are capable of monitoring physiological responses out-of-clinic. Here, the in-ear sensing capability in the area of day time sleepiness is further investigated.

Sleepiness is a symptom and a subjective parameter, which depends on how an individual perceives the degree of urge to fall a sleep. The exact cause of sleepiness is still under investigation, but one cause could be a lack of sleep during the night. The state of being sleepy appears to be linked to 'drowsiness', which is the transition from wakefulness to light sleep. For the 'drowsiness' classification, Patrick *et al.* conducted automatic sleep staging using a publicly available dataset, and only extracted wake and Non-REM1 (N1) [158]. Their selected features from a single EEG channel and computationally inexpensive classification algorithm can detect sleep onset effectively, and in the real world scenarios.

Previously, daytime nap recordings with conventional on-scalp EEG headsets and our in-ear EEG sensor simultaneously were conducted, and it was found the feasibility of in-ear EEG for measuring sleep latency [159] in healthy volunteers [59]. With the same day time nap EEG data, the aim of this chapter is to validate whether in-ear EEG patterns can distinguish between wakefulness and light sleep. The results provide a evidence that, due to the unobtrusive and wearable nature of our in-ear sensor, the light sleep classification application with in-ear sensing is capable of achieving 'drowsiness' monitoring.

## 6.2 Methods

Figure 6.1 summarises the analysis framework for this study. The recorded on-scalp and in-ear EEG data were manually scored by a clinician. The data were also pre-processed and supervised classification was performed.

### 6.2.1 Data acquisition

The EEG data were recorded at Imperial College London between December 2015 and April 2016 under the ethics approval, ICREC 12_1_1, Joint Research Office at Imperial College London. 36 healthy subjects were screened, and 25 healthy subjects (28.5$\pm$5.3 years) participated in the recordings; the details of recruiting processing can be found

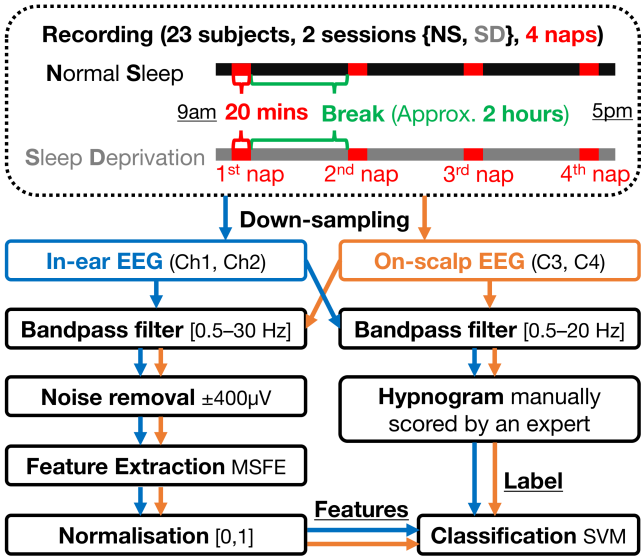



**Figure 6.1:** Flowchart for the classification framework in this study.

in [59]. 23 out of 25 subjects were instructed to participate in two recording sessions. Each session was executed between 9 am to approximately 5 pm. The subjects attended the recordings after either their normal sleep (average $\geq$ 7 hours) or sleep deprivation ($\leq$ 5 hours) in the night before, and refrained from consuming caffeine. The subjects performed four naps (trial) with the length of twenty minutes on each recording day, and the interval between trials were approximately two hours. If needed, participants were allowed to leave the lab in between naps (e.g. for toilet break or bringing in the lunch). The sensors were kept in place throughout the trials, unless requested by participants, or to reapply the electrodes due to poor electrode impedance. Overall, 21 out of 23 subjects participated in both recording sessions. Table 6.1 summarises the recording details.

On-scalp and in-ear EEG were recorded simultaneously using the g.tec g.USBamp amplifier with 24-bit resolution, at a sampling frequency $f_s = 1200$ Hz. For on-scalp EEG, standard gold-cup electrodes located at the C3 and C4 position (according to the international 10-20 system) were used for the recordings, a montage extensively used in sleep medicine for visual sleep scoring, and the ground electrode for on-scalp EEG recordings was attached to the forehead.

The in-ear sensor was made based on a one-size-fits-all viscoelastic earplug with two flexible electrodes. The size of in-ear sensors was the same for all subjects. The in-ear sensor was inserted into either subject's left or right of ear, according to their preference. After the insertion, the sensor expanded to conform firmly to the shape of the ear canal. The standard gold-cup electrodes were attached behind the ipsilateral helix and the ipsilateral earlobe for reference and ground, respectively. After the electrodes were

**Table 6.1:** The configuration of EEG recordings.

| | |
|---|---|
| # of subjects | 23 |
| # of sessions | 2 days (Normal sleep, Sleep deprivation) |
| # of trials per session | 4 trials |
| Length of trial | 20 minutes |
| Channels | On-scalp EEG (C3, C4)<br>In-ear EEG (Ch1, Ch2) |




**Figure 6.2:** Recording setup in this study. (Left) The electrodes were placed on scalp and ear; (Centre) the subject reclined in a comfortable chair; (Right) wearable in-ear sensor with two flexible electrodes [10].

attached, the subjects were instructed to seat and recline in a comfortable chair in a dark and quiet room. Figure 6.2 illustrates the recording setup (Left and Centre panel) and the viscoelastic in-ear sensor (Right panel).

### 6.2.2 Manual scoring

The recorded signal was first downsampled to 200 Hz, and a fourth-order Butterworth bandpass filter with passband 1-20 Hz was applied to both on-scalp and in-ear EEG. The pre-processed on-scalp and in-ear EEG data were blinded and separately scored based on the AASM criteria [105]. The epoch size was set to 30 s, therefore 40 epochs were scored in each recording trial ($30\,\text{s} \times 40\,\text{epochs} = 1200\,\text{s}$). Some recording trials were too noisy to score; one typical issue was when a subject put the side of their head on the pillow, the earpiece and cables could become trapped underneath of their heads. Overall, 161 trials were used for further analyses. Note that some subjects did not sleep at all during the twenty minutes trials, so in this scenario, the hypnograms from such trials were all scored as Wake.

### 6.2.3 Pre-processing

For classification analyses, the downsampled EEG signals were bandpassed with a fourth-order Butterworth filter with passband from 0.5-30 Hz. Since this study focus on a detection of drowsiness, which is the boundary between Wakefulness and N1, the epochs

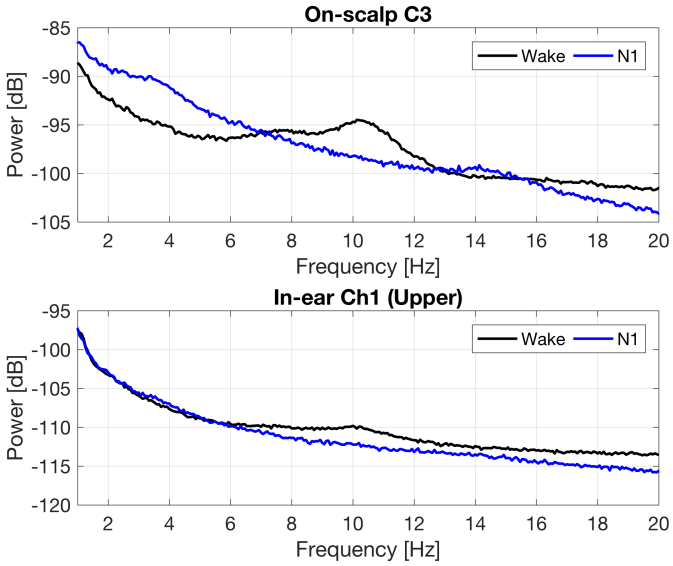



**Figure 6.3:** Averaged PSD: (Top) On-scalp C3 channel; (Bottom) in-ear Ch1.

**Table 6.2:** Proportion of sleep stages over 23 subjects.

| Dataset | | Wake | N1 | Total |
|---|---|---|---|---|
| In-ear EEG | # of epochs | 3997 | 2002 | 5999 |
| | Ratio (%) | 66.6 | 33.4 | 100.0 |
| On-scalp EEG | # of epochs | 3728 | 2071 | 5799 |
| | Ratio (%) | 64.3 | 35.7 | 100.0 |

labeled as Wake and N1 were analysed. In order to remove noisy epochs from further analyses, for each on-scalp and in-ear EEG, the epochs (i.e. 30 s segment of recordings) which contained amplitudes of more than $\pm 400\,\mu$V were removed. Therefore, the total number of on-scalp and in-ear EEG epochs for further analyses was different, with their proportion given in Table 6.2.

Figure 6.3 illustrates the averaged PSD over 23 subjects for on-scalp EEG channel (C3) and in-ear EEG channel (Ch1, upper electrode), corresponding to Wake and N1 conditions. Even though the amplitude of on-scalp EEG was different to that of in-ear EEG, the trends of PSD were similar. For example, for both on-scalp and in-ear EEG channels, there is a close match between the peaks in the alpha band (8-13 Hz) for Wake, and the alpha power attenuation in the N1.

### 6.2.4 Feature extraction

After the pre-processing, the MSFE were calculated for each epoch of both on-scalp and in-ear EEG data. The maximum scale for the MSFE was $\tau = 1 - 30$ ($\tau_{\min} = 1$ and $\tau_{\max} = 30$).

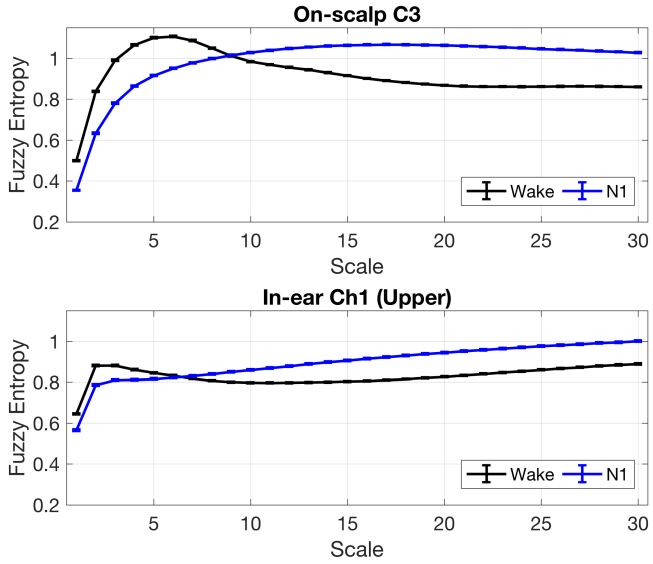



**Figure 6.4:** The averaged MSFE for on-scalp and in-ear EEG. The error bars represent the standard error.

Since there are two channels of EEGs for both on-scalp and in-ear EEG, $30 \times 2 = 60$ features were extracted from each epoch for the classification. The other parameters used to calculate the MSFE were $m = 2$, $r = 0.15\times$(*standard deviation of each epoch*). Figure 6.4 depicts MSFE analysis for on-scalp and in-ear EEG channels (C3 and in-ear Ch1) of the 161 trials over 23 subjects with different sleep stages. Observe that for both on-scalp and in-ear EEG, the Wake and N1 were visually well separated.

### 6.2.5 Classification

The binary-class SVM with a RBF kernel was employed as a classifier. The regularisation parameter was set to $C = 10$, and the hyper-parameters for classification were set to $\gamma = 1$; the same hyper-parameters were used throughout the analysis.

### 6.2.6 Performance evaluation

Feature extraction was performed using Matlab 2016b, and the classification was conducted in Python 2.7.12, Anaconda 4.2.0 operated on an iMac with 2.8GHz Intel Core i5, 16GB of RAM. The performance indicators to evaluate the proposed methods were class-specific sensitivity (SE) and precision (PR), as well as overall accuracy and kappa coefficient ($\kappa$); the details are given in Section 2.4. The parameter $M = 2$ is the number of classes (i.e. Wake or N1).

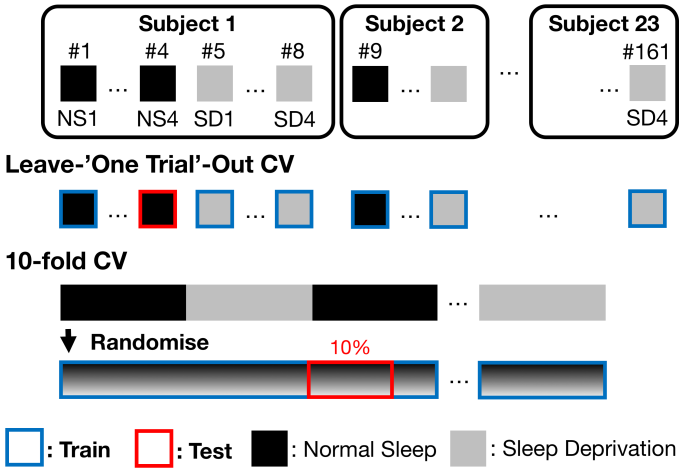



**Figure 6.5:** Cross validation approaches used in this study.

### 6.2.7 Validation setup

Figure 6.5 illustrates the two different validation methods employed in this study. First, the Leave-'One Trial'-Out Cross-Validation (LOTOCV) approach was applied; a single nap trial was assigned as test data and the rest of 160 trials was used as training data (as mentioned in Section 6.2.2, the number of epochs was 161). The validation trials were therefore repeated 161 times. Although this approach is computationally expensive, the test and training data do not come from the same recording trial, therefore, this approach is applicable for real world setups. Second, a 10-fold CV approach was utilised, whereby EEG recordings from all the 161 trials with scored hypnogram were concatenated into one large matrix, and then randomly split into the training data (90 %) and test data (10 %).

## 6.3 Results

#### Setup 1: LOTOCV

Tables 6.3 and 6.4 show the classification results using LOTOCV methods for in-ear EEG and on-scalp EEG recordings, respectively. For in-ear EEG, the classification accuracy of two stages (i.e. Wake or N1) was 80.0 % with the kappa coefficient of $\kappa = 0.53$. The classification and corresponding kappa for on-scalp EEG were 86.8 % and $\kappa = 0.72$ (Substantial Agreement), respectively.

#### Setup 2: 10-fold CV

Tables 6.5 and 6.6 show the confusion matrices for 10-fold CV methods using in-ear EEG and on-scalp EEG, respectively. The classification accuracy for in-ear EEG was 82.9 % and

**Table 6.3:** Confusion matrix for the 2-class Wake v. N1 classification (in-ear EEG, LOTOCV).

| | | Algorithm based on in-ear EEG | | |
|---|---|---|---|---|
| | | Wake | N1 | SE(%) / PR(%) |
| Hypnogram based on in-ear EEG | Wake | 3546 | 451 | 88.7 / 82.5 |
| | N1 | 750 | 1252 | 62.5 / 73.5 |
| Accuracy = 80.0 %, $\kappa$ = 0.53 | | | | |

**Table 6.4:** Confusion matrix for the 2-class Wake v. N1 classification (on-scalp EEG, LOTOCV).

| | | Algorithm based on on-scalp EEG | | |
|---|---|---|---|---|
| | | Wake | N1 | SE(%) / PR(%) |
| Hypnogram based on on-scalp EEG | Wake | 3316 | 412 | 88.9 / 90.4 |
| | N1 | 353 | 1718 | 83.0 / 80.7 |
| Accuracy = 86.8 %, $\kappa$ = 0.72 | | | | |

the corresponding kappa was $\kappa = 0.60$. For on-scalp EEG, the classification accuracy and kappa were 88.8 % and $\kappa = 0.76$ (Substantial Agreement), respectively.

**Table 6.5:** Confusion matrix for the 2-class Wake v. N1 classification (in-ear EEG, 10-fold CV).

| | | Algorithm based on in-ear EEG | | |
|---|---|---|---|---|
| | | Wake | N1 | SE(%) / PR(%) |
| Hypnogram based on in-ear EEG | Wake | 3645 | 352 | 91.2 / 84.4 |
| | N1 | 673 | 1329 | 66.4 / 79.1 |
| Accuracy = 82.9 %, $\kappa$ = 0.60 | | | | |

**Table 6.6:** Confusion matrix for the 2-class Wake v. N1 classification (on-scalp EEG, 10-fold CV).

| | | Algorithm based on on-scalp EEG | | |
|---|---|---|---|---|
| | | Wake | N1 | SE(%) / PR(%) |
| Hypnogram based on on-scalp EEG | Wake | 3374 | 354 | 90.5 / 92.0 |
| | N1 | 295 | 1776 | 85.8 / 83.4 |
| Accuracy = 88.8 %, $\kappa$ = 0.76 | | | | |

## 6.4 Summary

This chapter aims to classify 'drowsiness', which is an important condition on the boundary between wakefulness and light sleep. Day time nap recordings from 23 healthy subjects were conducted, and their sleep stages were automatically classified based on EEG signals obtained from both on-scalp and in-ear sensors. Our previous study scored and compared the sleep onset based on both on-scalp and in-ear EEG [59], where the results achieved in this study have confirmed the feasibility of in-ear EEG for automatic light sleep classification, which is applicable for wearable out-of-clinic drowsiness monitoring.

# Chapter 7

# Person authentication using in-ear EEG

## Overview

This chapter focuses on a different aspect of user well-being. The use of EEG as a biometric has been investigated for the last fifteen years, however its feasibility in the real world is not yet established, mainly due to the issues with collectability and reproducibility. To this end, a wearable biometric system based on in-ear EEG is evaluated.

## 7.1 Introduction

Person authentication refers to the process of confirming the claimed identity of an individual, and is already present in many aspects of life, such as electronic banking and border control. The existing authentication strategies can be categorised into: 1) knowledge-based (password, PIN), 2) token-based (passport, card), 3) biometric (fingerprints, iris) [90]. Most extensively used recognition methods are based on knowledge and tokens, however, these are also most vulnerable to fraud, such as theft and forgery, and can be straightforwardly used by imposters. In contrast, biometric recognition methods rest upon unique physiological or behavioural characteristics of a person, which then serve as 'biomarkers' of an individual, and thus largely overcome the above vulnerabilities. However, at present, biometric authentication systems are cumbersome to administer and require considerable computational and man-power overloads, such as special recording devices and the corresponding classification software.

With the current issues in global security, there is a rapid growth in biometrics applications based on various modalities, including palm patterns with high spectral wave [160], patterns of eye movement [161], patterns in the ECG [162], and otoacoustic emissions [163]. Each such biometric modality has its strengths and weaknesses, and typically suits only a chosen type of application and its corresponding scenarios [164]. A robust biometric system in the real world should satisfy the following requirements [90]:

- *Universality*: each person should possess the given biometric characteristic,
- *Uniqueness*: not any two people should share the given characteristic,

- *Permanence*: the biometric characteristic should neither change with time nor be alterable,
- *Collectability*: the characteristic should be readily measurable by a sensor and readily quantifiable.

In addition, a practical biometric system must be harmless to the users, and should maximise the trade-off between performance, acceptability, and circumvention; in other words, it should be designed with the accuracy, speed, and resource requirements in mind [164].

One of the currently investigated biometric modalities is the EEG, an electrical potential between specific locations on the scalp which arises as a result of the electrical field generated by assemblies of cortical neurones, and reflects brain activity of an individual, such as intent [21]. From a biometrics perspective, the EEG fulfils the above requirements of *universality*, as it can be recorded from anyone, together with the *uniqueness*. Specifically, the individual differences of EEG alpha rhythms has been examined [91] and reported to exhibit a significant power in discriminating individuals [92] in the area of clinical neurophysiology. The brain activity is neither exposed to surroundings nor possible to be captured at a distance, therefore the brain patterns of an individual are robust to forgery, unlike face, iris, and fingerprints. The EEG is therefore more robust against imposters' attacks than other biometrics and among different technologies to monitor brain function. However, in order to utilise EEG signals in the real world, several key properties such as *permanence* and *collectability* must be further addressed.

Most of proof-of-concept EEG biometrics were conducted over only one day (or even over one single trial) recording with EEG channels covering the entire head, while in the classification stage, the training and validation datasets were randomly selected from the same recording day (or the same trial). Apart from its usefulness as a proof-of-concept, this setup does not satisfy the feasibility requirement for a real world biometric application, since:

- Recording on-scalp EEG with multiple electrodes is time-consuming to set up and cumbersome to wear. Such a sensor therefore does not meet *collectability* requirement.
- EEG recordings from one day (or a single trial) cannot truly evaluate the performance in identifying features of an individual, as this scenario does not satisfy the *permanence* requirement either.
- The training and validation data within this scenario are typically mixed, thereby introducing a *performance bias* in classification. The classification results from such studies are therefore unrealistically high – ***biased scenario***.

Therefore, for feasible EEG biometrics, the EEG sensor should be wearable, easy to administer, and fast to set up, while in order to evaluate the performance, the recorded signal should be split in a rigorous way – the training and validation datasets in the classification

stage should be created so as *not to share the same recording days*. The setting is referred as the ***rigorous scenario***. While a considerable body of research has been undertaken to explore the EEG biometrics and to find the most informative subject-specific characteristic of EEG (*uniqueness*), most studies either focused on reducing the number of electrodes (*collectability*) or on evaluating whether the traits are temporally robust (*permanence*) by using EEG data obtained over multiple recording days.

In summary, EEG-based biometrics in real world resolve the following critical issues:

1. *Collectability*. Biometrics verification is evaluated with a wearable and easy to set up in-ear sensor [36],
2. *Uniqueness* and *permanence*. These issues are addressed through subject-dependent EEG features which are recorded over temporally distinct recording days,
3. *Reproducibility*. The recorded data are split into the training and validation data in two different setups, *biased* and *rigorous* setup,
4. *Fast response*. The classification is performed by both a fast non-parametric (cosine distance) and standard parametric approaches (linear discriminant analysis and support vector machine).

Through these distinctive features, a proof-of-concept for a wearable in-ear EEG biometrics in the community is introduced.

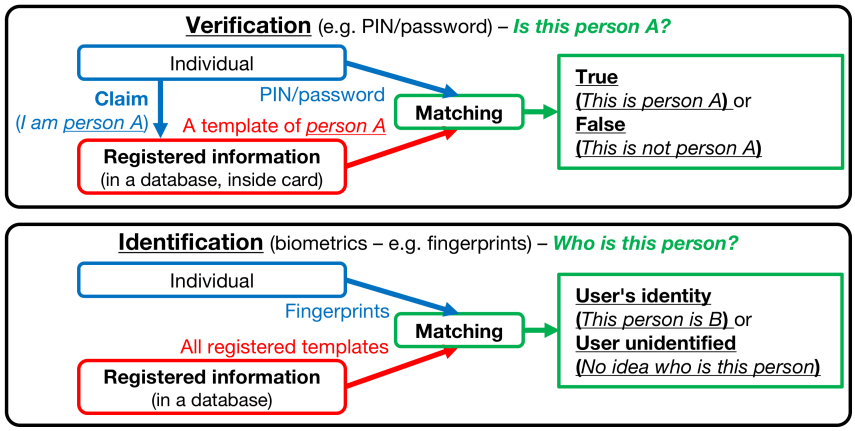



**Figure 7.1:** Person recognition systems. (Top) Verification system; (Bottom) identification system.

## 7.2 Overview of EEG-based biometrics

### 7.2.1 Biometric systems with verification/identification

Depending on the context, the two categories of biometric systems are: 1) verification systems and 2) identification systems, summarised in Figure 7.1 [164]. Verification refers to

validating a person's identity based on their individual characteristics, which are stored/registered on a server. In technical terms, this type of a biometric system performs a *one-to-one matching* between the 'claimed' and 'registered' data, in order to determine whether the claim is true. In other words, the question asked for this application is 'Is this person A?' as illustrated in Figure 7.1 (Top panel). In contrast, an identification system confirms the identify of an individual from cross-pattern matching of the all available information, that is, based on *one-to-many template matching*. The underlining question for this application is 'Who is this person?' as illustrated in Figure 7.1 (Bottom panel).

### 7.2.2 Feasible EEG biometrics design

Traditionally, EEG-based biometrics research has been undertaken based on both publicly available datasets [165] and custom recordings as part of research efforts [166]. However, most of existing studies failed to rigorously address the key criterion, *collectability*, which is also related to repeatability. A large number of studies, especially those conducted at the dawn of EEG biometrics research, employed classification of the clients based on supervised learning with the training and validation data coming from the same recording trial. However, this experimental setup cannot truly evaluate the performance in identifying individual features, since such classification does not take into account the varying characteristic among multiple recording trials and recording days. In addition, EEG is prone to contamination by artefacts from subjects' movements (e.g. eye blinks, chewing), while the sources of external noise include electrode noise, power line noise, and electromagnetic interference from the surroundings. This opens the possibility to additionally incorrectly associate 'EEG patterns' with either trial-dependent features or so-called noise-related features – in other words, this setup is *biased* in favour of high classification rate. Therefore, given the notorious variability of EEG patterns across days, biometrics studies based on a single recording day (even for a single subject) can only validate very limited scenarios, without any notion of repeatability and long-term feasibility [166].

Figure 7.2 shows the concept of a *rigorous* EEG biometrics verification/identification system in the real world. Individuals participate in EEG recordings and their EEG signals are registered and stored on a server or in a database. The client is granted access to their account by providing new EEG data in verification scenarios, whereby the algorithm discriminates the identify of an individual is in identification scenarios through new EEG recordings. Recall that the registered EEG must be recorded beforehand.

In order to fulfil the feasibility requirement, several studies performed successful EEG-based biometrics from multiple recording trials conducted on multiple distinct days, thus satisfying the *collectability* requirement. However, the majority of these studies were still conducted in an unrealistic scenario, whereby the training and validation data in the classification process are split into segments, with all the segments coming from multiple trials but on the same recording day being randomly assigned to the training and valida-

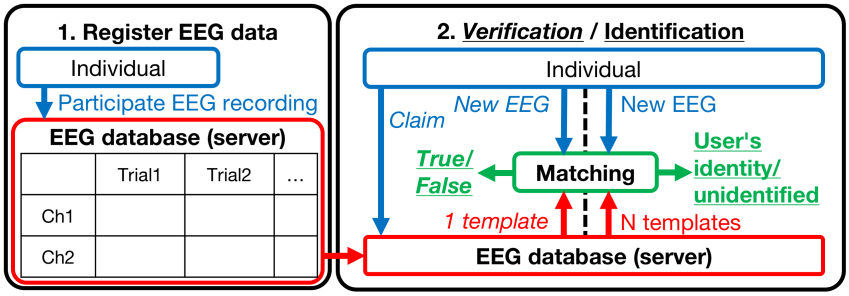



**Figure 7.2:** Feasible EEG biometrics verification framework. (Left) EEG recording registration. The registered EEG signal must have been recorded beforehand. (Right) Verification and identification system.

tion datasets. Therefore, this *biased* setup, despite being based on the classification from multiple recording trials, mixes the training and test recordings from the same recording day and thus cannot truly evaluate the performance in the identification of individual features.

In order to truly validate the robustness of a EEG biometrics application, within a *rigorous* setup, it is therefore necessary to both: 1) conduct multiple recordings over multiple days, and 2) to assign recordings on one day as the training data and use the recordings from the other days as the validation data. In other words, the training and validation datasets should be created *so as not to share the same recording days* (as illustrated later in Figure 7.6, Setup-R). As emphasised by Rocca *et al.*, the issue of the repeatability of EEG biometrics in different recording sessions is still a critical open issue [167].

**Table 7.1:** Recording and classification setup, and performance comparison among existing EEG biometrics.

| Author | Sub | Days | Interval | Ch. | Task* | Seg. | Features | Classifier** | Performance |
|---|---|---|---|---|---|---|---|---|---|
| ***Biased validation setup*** | | | | | | | | | |
| [168] | 10 | 5 | 2 weeks | 4 | EC<br>EO | 5 s | AR | ANN | CRR = 97.0 %<br>CRR = 96.0 % |
| [169] | 51 | 4 | 34 ± 74 days | 2 | EC | 4 s | AR, PSD, MuI, COH, CC | FDA | EER = 3.4 % |
| [170] | 40 | 2 | - | 1 | EC | 180 s | AR, PSD | KNN, LDA | CRR = 97.5 % |
| ***Rigorous validation setup*** | | | | | | | | | |
| [171] | 9 | 3 | 3 days | 8 | MI | 1 s | PSD | MAP model | HTER = 19.3 % |
| [172] | 4 | 2 | 10 days - 5 months | 1 | EC | 50 s | PSD | LDA | AC = 100 % |
| [173] | 9 | 2 | 1 - 3 weeks | 3<br>5 | EC | 1 s | AR | Linear classier | CRR = 100 %<br>CRR = 100 % |
| [174] | 15<br>9 | 2<br>3 | 5 - 40 days<br>134 - 188 days | 1 | ERP | 1.1 s | Time-series | CC | CRR = 89.0 %<br>CRR = 93.0 % |
| [175] | 50 | 3 | Ave. 34 days | 19 | EC<br>EO | 45 s | AR, PSD, COH | L1, L2, Cos.dist. | R1IR = 90.8 %<br>R1IR = 85.6 % |
| Task* | EC: resting state with eyes closed, EO: resting state with eyes open, MI: motor imagery<br>ERP: event related potential | | | | | | | | |
| Classifier** | ANN: artificial neural networks, FDA: Fisher discriminant analysis, KNN: k-nearest neighbours,<br>LDA: linear discriminant analysis, MAP: maximum a posteriori, CC: cross correlation,<br>L1 (Manhattan) distance, L2 (Euclidean) distance, Cos. (cosine) distance | | | | | | | | |

### 7.2.3 Previous protocols

Table 7.1 summarises the state-of-the-art of the existing EEG biometrics applications based on multiple data acquisition days.

#### Biased setup

In the first category (*Biased* setup) is the studies where the training and validation features were randomly selected regardless of the data acquisition days. Abdullah *et al.* [168] collected four channels of EEG data from ten male subjects during the resting state, in the both eyes open (EO) and eyes closed (EC) scenarios, in five separate recording days over a course of two weeks. In each recording day, five trials of 30 s recordings were recorded, and the recorded data were split into 5 s segments with an overlap of 50 %. The autoregressive (AR) coefficients of the order $p = 21$ were extracted from each segment, and the extracted features were *randomly* divided into the training (90 %) and validation (10 %) sets, namely 10-fold CV. An artificial neural network (ANN) yielded 97.0 % of correct recognition rate (CRR) for the EC task, and 96.0 % of CRR for the EO task. Riera *et al.* [169] recorded two forehead channels of EEG from 51 subjects over four separate recording days. The average temporal distance between the first and the fourth recording was $34 \pm 74$ days. The participants performed the EC task, and the duration of recordings was between two and four minutes. The recorded EEG was split into 4 s segments, and five types of features were calculated for each segment, namely AR coefficients of order $p = 100$, PSD in the 1-40 Hz band, mutual information (MuI), coherence (COH), and cross-correlation (CC). The authors trained various classifiers and identified the five best classifiers. The Fisher's discriminant analysis (FDA) was then employed and was first trained with different types of discriminant functions using the first to third recording trials; then the fourth recording trials were used for testing. Next, the best five classifiers from the training process were utilised for authentication tests, using the first and the second minutes of recordings from each trial; therefore, *the training data for the classifiers and test data for the validation were not disjoint* (biased setup). The discriminant analysis with the selected discriminant function achieved 3.4 % of equal error rate (EER). Su *et al.* [170] analysed five minutes of the EC task from 40 subjects, with six recording trials performed in two separate recording days for each subject. The recorded EEG data from the FP1 channel were split into segments of multiple lengths. The PSD in the 5-32 Hz band and AR coefficients of the order $p = 19$ were chosen as features. The extracted features were *randomly* divided into the training (50 %) and validation (50 %) sets. As a result of 100 iterations, the classifier combining LDA and k-nearest neighbours (KNN) achieved an average CRR = 97.5 %, for a segment length of 180 s.

### Rigorous setup

Multiple research groups considered EEG biometrics based on splitting the training and validation data in a *rigorous* way, *so as not to share the data from the same recording days* to highlight the feasibility of their system (*Rigorous* setup). Marcel *et al.* [171] analysed eight channels of EEG from nine subjects, with four recording trials over three consecutive days. The 15 s trials consisted of two different motor imagery (MI) mental tasks, the imagination of hand movements. The recorded data were split into 1 s segments, and PSD in the 8-30 Hz band was calculated for each segment. The Gaussian mixture model (GMM) was chosen as a classifier, and maximum a posteriori (MAP) estimation was used for adapting a model for client data. By combining recordings over two days as training data, the authors achieved 19.3 % of half total error rate (HTER), which is a performance criterion widely used in biometrics; the details can be found in Section 7.3.6. Lee *et al.* [172] conducted an experiment of 300 s in duration from four subjects over two days, based on single channel of EEG in the EC scenario. The data were segmented into multiple window sizes, and to extract frequency domain features, PSD was calculated only for the $\alpha$ band (8-12 Hz). Even though the dataset size was relatively small, with 50 s of segment length, the LDA achieved 100 % classification accuracy. Rocca *et al.* [173] recorded two resting state EEGs, in both the EO and EC scenarios, from nine subjects over two different recording days, which were spanned one to three weeks. The recording length was 60 s, and the recorded data were split into 1 s segments with an overlap of 50 %. The AR model (Burg algorithm) of the order $p = 10$ was employed for feature extraction, and the training and validation data were split without mixing the trials from different recording days. The recognition results, with features from selected three or five channels of on-scalp EEG, were obtained by linear classification based on minimising the mean square error, to achieve CRR = 100 %. Armstrong *et al.* [174] recorded event related potentials (ERPs) and constructed two datasets. One dataset included EEG recorded from fifteen subjects in two separate days, with a 5 to 40 inter-day interval, and the other one contained EEG from nine subjects obtained in three separate days, with the average interval between the first and third recordings of 156 days. The recorded data from the O2 channel were split into 1.1 s long segments, which contained an ERP and started from a 100 ms pre-stimulus. The CC between the training and validation data was used as a feature for classification, and CRR = 89.0 % was achieved for validating the second day recordings whereas CRR = 93.0 % for classifying the third day recordings. Maiorana *et al.* [175] analysed nineteen channels of EEG from 50 subjects during both EC and EO tasks in three different recording days, with the average interval between the first and the third recording of 34 days. Each recording trial consisted of 240 s of data, segmented into 5 s windows, with an overlap of 40 %. Three types of features were extracted, including a channel-wise AR model (using the Burg algorithm) of the order $p = 12$, channel-wise PSD, and the COH between the EEG channels. The L1, L2, and cosine distances were calculated for the

extracted features, and the rank-1 identification rate (R1IR) achieved 90.8 % accuracy in the EC task and 85.6 % in the EO task.

### 7.2.4 Biometrics based on collectable EEG systems

With a perspective of *collectability*, a biometrics application with dry EEG electrodes was recently introduced [176]. While conventional wet EEG headsets require the application of a conductive gel which is generally time-consuming, the dry headset with sixteen scalp channels took on the average two minutes to be operational. The BCI based biometrics application with rapid serial visual presentation paradigm achieved CRR = 100 % with 27 s window size over all 29 subjects. Although the recordings were performed over a single recording day per subject, the application with a dry headset was a step forward towards establishing collectable EEG biometrics in real world.

In a recent effort to enable collectable EEG, the in-ear sensing technology [36] was introduced into the research community. The in-ear EEG has been proven to provide on-par signal quality, compared to conventional on-scalp EEG, in terms of steady state responses [39], monitoring sleep stages [60], and also for monitoring cardiac activity [10]. The advantages of the in-ear EEG sensing for a potential biometrics application in the real world are:

- *Unobtrusiveness*: The viscoelastic in-ear EEG sensor is made from affordable/consumable standard earplugs [41],
- *Robustness*: The viscoelastic substrate expands after the insertion, so the electrodes fit firmly inside the ear canal [10], where the position of electrodes remains the same in different recording sessions,
- *User-friendliness*: The sensor can be applied straightforwardly by the user, without the need for a trained person.

Therefore, biometrics with in-ear EEG offers a high degree of *collectability*, a critical issue in real world applications.

### 7.2.5 Problem formulation

Here, the possibility of biometrics verification with a wearable in-ear sensor, which is capable of fulfilling the *collectability* requirement, is investigated. The data were recorded over temporally distinct recording days, in order to additionally highlight the *uniqueness* and *permanence* aspects. Although the changes in EEG rhythms may well depend on the time period of years rather than days, the alpha band features during the resting state with eyes closed were reported as the most stable EEG feature over two years [177]. Since EEG alpha rhythms predominantly originate during wakeful relaxation with eyes closed, this recording task was chosen to be the resting state with eyes closed. This task

was used in multiple previous studies [168, 169, 170, 172, 173, 175]. In order to design a feasible biometrics application in the real world, the two different types of imposters were considered: 1) registered subjects in a database, and 2) subjects not belonging to a database. Previously, Riera *et al.* [169] also used a single trial of EEG recording from multiple subjects as 'intruders', while the 'imposters' data were EEG recordings available from multiple other experiments. For rigour, two types of data were collected: 1) based on multiple recordings from fifteen subjects over two days, and 2) multiple recordings from five subjects, which were only used for imposters' data. The classification was performed by both a non-parametric and parametric approach. The non-parametric classifier, minimum cosine distance, is a simplest way for evaluating the similarity between the training and validation matrix. Through the binary client-imposter classification, the feasibility of our in-ear EEG biometrics was evaluated.

## 7.3 Methods

### 7.3.1 Data acquisition

The recordings were conducted at Imperial College London, for two different groups of subjects and under the ethics approval, Joint Research Office at Imperial College London ICREC12_1_1. One set of data were the recordings used as both clients and imposters data, denoted by $S_R$, and the other subset were the recordings for only imposters' data, denoted by $S_N$. Table 7.2 summarises the two recording configurations of $S_R$ and $S_N$.

For the $S_R$ subset of recordings, fifteen healthy male subjects (aged 22-38 years) participated in two temporally separate sessions, with the interval between two recording sessions between five and fifteen days, depending on the subject. The participants were seated in a comfortable chair during the experiment, and were asked to rest with eyes closed. The length of each recording was 190 s, and the recording was undertaken three times (trials) per one day. The interval between each recording trial was approximately five to ten minutes. In total, six trials were recorded per subject.

The in-ear EEG sensor is made of a memory-foam substrate and two conductive flexible electrodes, as shown in Figure 7.3. The size of earpiece was the same for over twenty subjects (both the $S_R$ and $S_N$ subjects). The in-ear sensor was inserted in the subject's left ear canal after earwax was removed; it then expanded to conform to the shape of the ear canal. The reference gold-cup standard electrodes were attached behind the ipsilateral earlobe and the ground electrodes were placed on the ipsilateral helix. For simplicity, the upper electrode is denoted by Ch1, while Ch2 refers to the bottom electrode. The two EEG signals from flexible electrodes were recorded using the g.tec g.USBamp amplifier with a 24-bit resolution, at a sampling frequency $f_s$ = 1200 Hz.

For the $S_N$ subset of recordings, five healthy subjects (aged 22-29 years) participated in three recording trials. Similar to the $S_R$ subset of recordings, the participants were

seated in a comfortable chair, and were resting with eyes closed. The duration of recording was also 190 s. A viscoelastic earpiece with two flexible electrodes [41] was inserted in the subject's left ear canal and the same reference and ground configuration was utilised for the $S_R$ subset of recordings.

Similar to the setup in [175], there was no restriction on the activities that the subjects performed, and no health test such as their diet and sleep, was carried out neither before or between an EEG acquisition and the following one, nor during the days of the recordings. This lack of restrictions allowed us to acquire data in conditions close to real life.

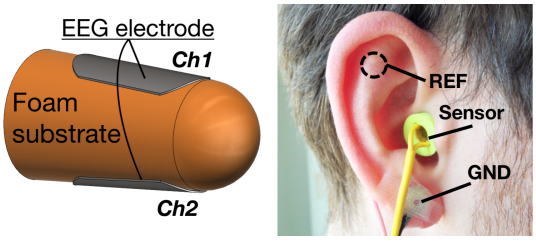



**Figure 7.3:** The in-ear sensor used in this study. (Left) Wearable in-ear sensor with two flexible electrodes; (Right) placement of the viscoelastic earpiece.

**Table 7.2:** Two EEG recordings and corresponding subset.

| Subset | Client/Imposter $S_R$ | Imposter only $S_N$ |
|---|---|---|
| Task | Resting state with eyes closed | |
| # of Subjects | 15 | 5 |
| # of Trials | 6 (2 days, 3 trials) | 3 (1 day, 3 trials) |
| Recording duration | 190 s | |

### 7.3.2 Pre-processing

The two channels of the so-obtained in-ear EEG were analysed based on the framework illustrated in Figure 7.4. In each recording, for both the $S_R$ and $S_N$ recordings, the first 5 s of recording data were removed from the analyses, in order to omit noisy recordings arising at the beginning of the acquisition. The two recorded channels of EEG were band-pass filtered with the fourth-order Butterworth filter with the pass-band 0.5-30 Hz. The bandpass filtered signals were split into segments. The symbol $N$ denotes the number of segments per recording trial from both the $S_R$ and $S_N$ subsets. The lengths of segments were chosen as $L_{\text{seg}} = 10, 20, 30, 60, 90$ s. Therefore, when the segment length was $L_{\text{seg}} = 60$ s, $N = 3$ ($190 - 5 = 185$ s, $\lfloor 185/60 \rfloor = 3$) segments were extracted from every recording trial of the $S_R$ and $S_N$. Within each segment, the data was split into epochs of 2 s length. The epochs with the amplitudes of greater than 50 $\mu$V for either Ch1 or Ch2 were considered corrupted by artefacts and removed from the analyses. This method resulted in a

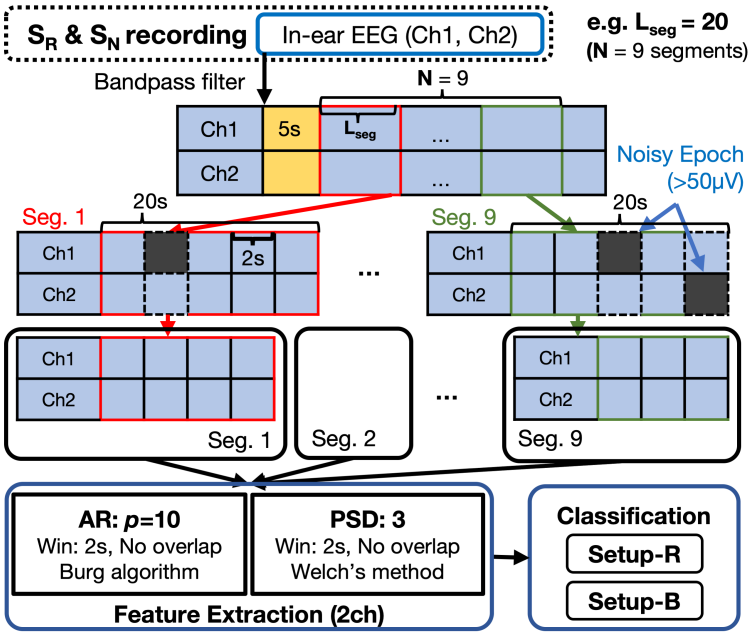



**Figure 7.4:** Flowchart for the biometrics analysis framework in this study.

loss of 4.3 % of the data, namely approximately 7.7 s out of 190 s per recording trial.

### 7.3.3 Feature extraction

After the pre-processing, two types of features were extracted from each segment of the in-ear EEG. For a fair comparison with the state-of-the-art, these features were selected to be the same or similar to those used in the recent studies based on the resting state with eyes closed [172, 173], and included: 1) a frequency domain feature – PSD, and 2) coefficients of an AR model.

#### The PSD features

Figure 7.5 shows power spectral density for the in-ear EEG Ch1 (Left) and for the in-ear EEG Ch2 (Right) of two subjects. For this analysis, the recorded signals were conditioned with the fourth-order Butterworth filter with the pass-band 0.5-30 Hz. The PSD were obtained using Welch's averaged periodogram method, the window length was 20 s with 50 % of overlap. The PSDs are overlaid between different recording days (Red: Day1, Blue: Day2), as well as among different recording trials with the same recording days, especially visible from 3 to 20 Hz. Previously, Maiorana *et al.* utilised PSD features for EEG biometrics based on the resting state eyes closed and achieved the best performance between the PSD features from theta to beta band, which was classified by the minimum cosine approach [175]; the inclusion of the delta band decreased their identification perform-

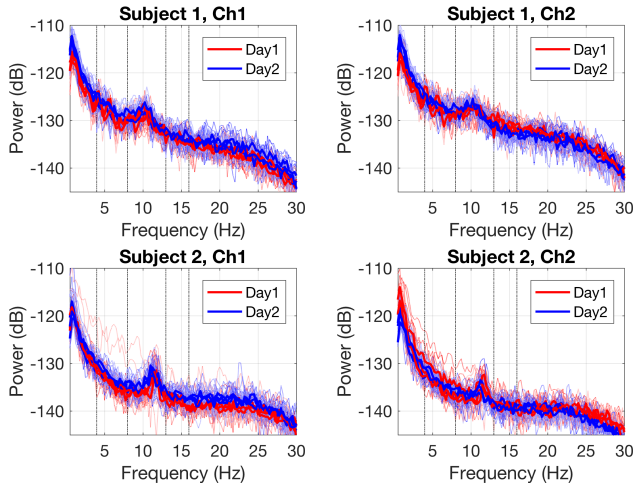



**Figure 7.5:** PSD for the in-ear EEG Ch1 (Left) and the in-ear EEG Ch2 (Right) of Subject One (Top) and Subject Two (Bottom). The thick lines correspond to the averaged periodogram obtained by the all recordings from the first day (Red) and the second day (Blue), whereas the thin lines are the averaged periodogram obtained by a single trial.

ance. In our in-ear EEG biometrics approach, the obtained PSDs were visually examined and being found that the ratio between the the total $\alpha$ band (8-13 Hz) power and the total $\theta - \alpha_{\text{high}}$ band (4-16 Hz) power is a relatively more significant individual factor for biometrics, rather than the total $\alpha$ band (8-13 Hz) power, which is proposed in [172]. Therefore, in each segment of length $L_{\text{seg}}$, univariate PSD was calculated by Welch's method with 2 s of the window length and no overlap. Three features were obtained for each PSD: 1) The ratio between the total $\alpha$ band (8-13 Hz) power and the total $\theta - \alpha_{\text{high}}$ band (4-16 Hz) power, 2) the maximum power in $\alpha$ band, and 3) the frequency corresponding to the maximum of $\alpha$ band power. In total, $D = 6$ (three features $\times$ two channels) frequency domain features were extracted from each segment.

#### The AR features

The Burg algorithm of order $p = 10$ was used to estimate the AR coefficients. For each segment, univariate AR parameter estimation of its $\alpha$ band (8-13 Hz) with a window length of 2 s and no overlap was applied. The AR model was chosen as a feature, because it was used in a previous successful study on EEG biometrics based on the resting state with eyes closed [173]. A total of $D = 20$ features (ten coefficients $\times$ two channels) were therefore extracted for each in-ear EEG segment.

### 7.3.4 Validation scenarios

**Setup-R:** Select the training and validation data **without mixing** segments from the two recording days, e.g. $[i, j, k]$ = [1,1,1].

1: $VC$: The matrix of the selected-subject, the selected-day and the selected-trial, [1,1,1]
$Y_{VC} = X_R^{(1,1,1)}$
2: $VI$: The matrices of the selected-trial and the selected-day from the non-selected-subjects, [(2:15),1,1]
$Y_{VI} = [X_R^{(2,1,1)^T}, X_R^{(3,1,1)^T}, ..., X_R^{(15,1,1)^T}]^T$
3: $TC$: The matrices of all recording trials recorded at the non-selected-day from the selected-subject, [1,2,(1:3)]
$Y_{TC} = [X_R^{(1,2,1)^T}, X_R^{(1,2,2)^T}, X_R^{(1,2,3)^T}]^T$
4: $TI$: The matrices of all recording data recorded at the non-selected-day from the non-selected-subjects, [(2:15),2,(1:3)]
$Y_{TI} = [X_R^{(2,2,1)^T}, ..., X_R^{(2,2,3)^T}, X_R^{(3,2,1)^T}, ..., X_R^{(15,2,3)^T}]^T$.
5: $VI_N$: The matrices of the selected-trial from the $S_N$ recording subjects, [(1:5),-,1]
$Y_{VI_N} = [X_N^{(1,-,1)^T}, X_N^{(2,-,1)^T}, ..., X_N^{(5,-,1)^T}]^T$.

With the extraction of the both univariate AR and PSD features from two channels, the dimension $D$ of features per EEG segment was twenty six. Recall that the first 5 s of recording data were removed from the analyses. For each trial of the $S_R$ and $S_N$ recordings, the data with the duration of 190 s was split into segments of length $L_{\text{seg}} = 10, 20, 30, 60, 90$ s. Therefore, $N = 18, 9, 6, 3, 2$ segments were respectively obtained. Each recording trial was represented by the feature matrix $X_R$ for the $S_R$ recordings and $X_N$ for the $S_N$ recordings, such matrices have $N \times D$ elements. In this way, a set of six feature matrices was obtained from one subject for the $S_R$ recordings (three recording trials per one day, over two different days), whilst a set of three feature matrices was obtained from one subject for the $S_N$ recordings (three recording trials per one day, one recording day).

As emphasised in Introduction, the aim of study is to establish a feasible EEG biometrics which satisfies the *collectability* requirements, which are also related to repeatability. Therefore, for rigour, all feature matrices $X_R$ from the first day of recordings were used as the training data, and feature matrices from the second day of recordings were used as the validation data, and vice versa (Setup-R) – Setup-*R*: *Rigorous* setup, Setup-*B*: *Biased* setup. Our goal was to examine the robustness of the proposed approach over the two different time periods in Setup-R. For the second setup, Setup-B, training feature matrices were also selected from the trials which were recorded at the same recording day as the validation matrices. Namely, the training and validation data are split by mixing the data from the same recording days. *Notice that, although used in most available EEG biometrics studies [168, 169, 170], Setup-B could not evaluate the repeatability/reproducibility of the application, because the training and validation data were both from the same recording days.* In other words, such an approach benefits from the recording-day-dependent EEG characteristic in the classification. However, as the number of feasible biometric modalities with in-ear EEG sensor is limited and for comparison with other studies, for convenience, the results for Setup-B are also provided.

Figure 7.6 summarises the two validation scenarios, Setup-R and Setup-B. For clarity, write $VC$ the validation feature matrix for the client, by $VI$ the validation feature matrix

**Setup-B:** Select the training and validation data **with mixing** segments from the two recording days, e.g. $[i, j, k]$ = [2,2,2].

1: $VC$: The matrix of the selected-subject, the selected-day and the selected-trial, [2,2,2]
$Y_{VC} = X_R^{(2,2,2)}$
2: $VI$: The matrices of the selected-trial and the selected-day from the non-selected-subjects, [(1,3:15),2,2]
$Y_{VI} = [X_R^{(1,2,2)^T}, X_R^{(3,2,2)^T}, ..., X_R^{(15,2,2)^T}]^T$.
3: $TC$: The matrices of all recording trials recorded at the non-selected day from the selected-subject, [2,1,(1:3)], and the matrices of non-selected trials recorded at the selected-day from the selected-subject, [2,2,(1,3)]
$Y_{TC} = [X_R^{(2,1,1)^T}, X_R^{(2,1,2)^T}, X_R^{(2,1,3)^T}, X_R^{(2,2,1)^T}, X_R^{(2,2,3)^T}]^T$.
4: $TI$: The matrices of all recording data recorded at the non-selected-day from the non-selected-subjects, [(1,3:15),1,(1:3)],
and the matrices of non-selected trials recorded at the selected-day from the non-selected-subjects, [(1,3:15),2,(1,3)]
$Y_{TI} = [X_R^{(1,1,1)^T}, X_R^{(1,1,2)^T}, X_R^{(1,1,3)^T}, X_R^{(1,2,1)^T}, X_R^{(1,2,3)^T}, X_R^{(3,1,1)^T}, ..., X_R^{(15,2,3)^T}]^T$.
5: $VI_N$: The matrices of the selected-trial from the $S_N$ recording subjects, [(1:5),-,2]
$Y_{VI_N} = [X_N^{(1,-,2)^T}, X_N^{(2,-,2)^T}, ..., X_N^{(5,-,2)^T}]^T$.

for the imposters, while $TC$ is the training feature matrix for the client, and $TI$ as the training feature matrix for the imposters. The feature matrix from a single trial of the subject $i$, recording day $j$, and trial $k$, from the $S_R$ recordings is denoted by $X_R^{(i,j,k)}$. Then, the training feature matrix $Y_T$ and the validation matrix $Y_V$ are given as

$$\begin{aligned} Y_T &= [Y_{TC}^T, Y_{TI}^T]^T, \\ Y_V &= Y_{V_R} = [Y_{VC}^T, Y_{VI}^T]^T. \end{aligned}$$

Besides, in order to evaluate feasibility in the real world, $S_N$ recordings is used, which are EEG recordings only used for imposters; Riera *et al.* termed the imposter only data as 'intruders' [169]. For an additional scenario in both Setup-R and Setup-B (see Section 7.4.4), $S_N$ recordings were used as the validation data for imposters, $VI_N$. The feature matrix from a single trial of the subject $i$, and trial $k$, from the $S_N$ recordings is denoted by $X_N^{(i,-,k)}$. Therefore, the validation matrix is given by

$$Y_V = Y_{V_R} + Y_{VI_N} = [Y_{VC}^T, Y_{VI}^T, Y_{VI_N}^T]^T.$$

Table 7.3 summarises the properties of matrices for the training matrix and the validation matrix in the both Setup-R and Setup-B.

### 7.3.5 Classification

For both the Setup-R and Setup-B, every trial from every subject were selected for the validation of client data, so as to have validated 90 times (three trials $\times$ two days $\times$ fifteen subjects). For each validation, both the largest and smallest values were found for each feature (column-wise) from the training matrix, then the validation matrix was normalised to the range $[0, 1]$ based on these largest and smallest values. Three classification

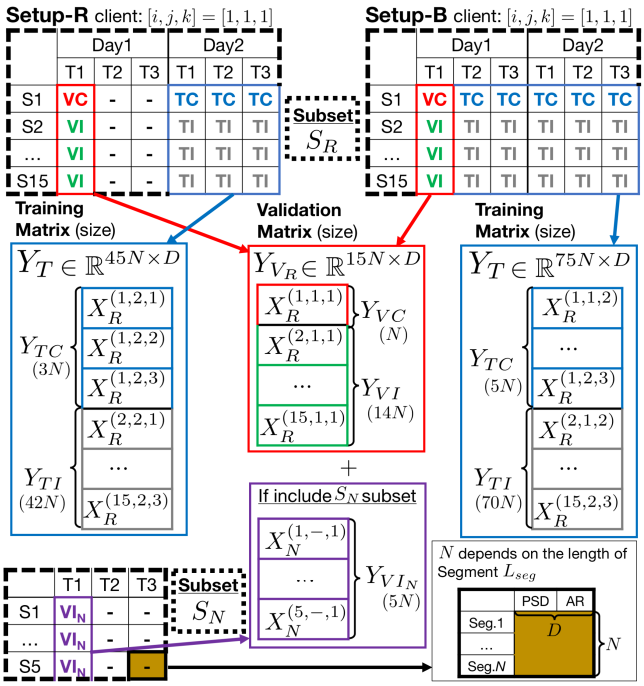



**Figure 7.6:** Two validation scenarios (Setup-R and Setup-B), where $X_R^{(i,j,k)} \in \mathbb{R}^{N \times D}$ and $X_N^{(i,-,k)} \in \mathbb{R}^{N \times D}$ denote a feature matrix from a single trial of the subject $i$, recording day $j$, and trial $k$, from the $S_R$ recordings and the $S_N$ recordings, respectively. The number of segments per recording trial $N$ depends on the chosen segment lengths $L_{\text{seg}}$. The dimension $D$ is the number of features per EEG segment.

algorithms were employed: 1) a non-parametric approach – minimum cosine distance [175], 2) LDA and 3) SVM.

### Non-parametric approach: Cosine distance

The cosine distance is the simplest way for evaluating the similarity between the rows of the validation matrix, $Y_{V_{(l,:)}}$, where $l = 1, ..., 15N$ for $Y_V = Y_{V_R}$ and $l = 1, ..., 20N$ for $Y_V = Y_{V_R} + Y_{VI_N}$, and the training matrix, $Y_T$, and is given by

$$d\left(Y_{V_{(l,:)}}, Y_T\right) = \min_n \frac{\sum_{m=1}^{D} Y_{V_{(l,m)}} Y_{T_{(n,m)}}}{\sqrt{\sum_{m=1}^{D} (Y_{V_{(l,m)}})^2} \sqrt{\sum_{m=1}^{D} (Y_{T_{(n,m)}})^2}}.$$

In other words, the cosine distance is used for evaluating the similarity between a given test sample (e.g $l$th row of the validation matrix, $Y_{V_{(l,:)}}$) and a template (training) feature matrix, $Y_T$. The distances between the $l$th row of the validation matrix, $Y_{V_{(l,:)}}$, and the each row of training matrix $Y_T$ were first computed, then the minimum among the computed distances was selected.

**Table 7.3:** Dimensions of the training and validation matrix in Setup-R and Setup-B.

| | | Setup-R | Setup-B |
|---|---|---|---|
| Train | $Y_{TC}$ | $3N \times D$ | $5N \times D$ |
| | $Y_{TI}$ | $42N \times D$ | $70N \times D$ |
| | $Y_T$ | $45N \times D$ | $75N \times D$ |
| Validation | $Y_{VC}$ | $N \times D$ | |
| | $Y_{VI}$ | $14N \times D$ | |
| | $Y_{V_R}$ | $15N \times D$ | |
| | $Y_{VI_N}$ | $5N \times D$ | |
| Total validation elements | | $Y_V = Y_{V_R}$ : 90(15$N$)<br>$Y_V = Y_{V_R} + Y_{VI_N}$ : 90(15$N$ + 5$N$)<br>$i$ = 1:15 (Sub), $j$ = 1:2 (Day), $k$ = 1:3 (Trial) | |

**Parametric approach**

The binary-class LDA and SVM were employed as a parametric classifier. Regarding SVM, for both Setup-R and Setup-B, four hyper-parameters: type of kernel, regularisation constant for loss function $C$, inverse of bandwidth $\gamma$ of kernel function, and order of polynomial $d$, were tuned by 5-fold CV within the training matrix. Then, the same hyper-parameters were used in order to obtain the optimal weight parameters within the training matrix. The same hyper-parameters and weight parameters as in the training were used for validation. Table 7.4 summarises the hyper-parameters for SVM.

**Table 7.4:** Hyper-parameters for SVM.

| Type of kernel | $K(\mathbf{x}, \mathbf{x}')$ | Hyper-parameters |
|---|---|---|
| Linear | $\mathbf{x}^T\mathbf{x}'$ | - |
| Sigmoid | $\tanh(\gamma\mathbf{x}^T\mathbf{x}' + r)$ | $\gamma, (r = 0)$ |
| RBF | $\exp(-\gamma\lvert\mathbf{x} - \mathbf{x}'\rvert^2)$ | $\gamma$ |
| Polynomial | $(\gamma\mathbf{x}^T\mathbf{x}' + r)^d$ | $\gamma, d, (r = 0)$ |

### 7.3.6 Performance evaluation

Feature extraction and classification with minimum cosine distance and with LDA was performed using Matlab 2016b, and the classification with SVM was conducted in Python 2.7.12 Anaconda 4.2.0 operated on an iMac with 2.8GHz Intel Core i5, 16GB of RAM.

For the verification setup (the number of classes $M = 2$, client-imposter classification), the performance was evaluated through the false accept rate (FAR), false reject rate (FRR), half total error rate (HTER), accuracy (AC), and true positive rate (TPR), defined as:

$$\begin{gathered} \mathrm{FAR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN}),\ \ \mathrm{FRR} = \mathrm{FN}/(\mathrm{TP} + \mathrm{FN}), \\ \mathrm{HTER} = \frac{\mathrm{FAR} + \mathrm{FRR}}{2},\ \ \mathrm{AC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}, \\ \mathrm{TPR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}). \end{gathered}$$

For the identification setup ($M = 15$), the performance was evaluated by subject-wise sensitivity (SE), identification rate (IR) and Cohen's kappa ($\kappa$) coefficient as:

$$\mathrm{SE}_i = \mathrm{TP}_i/(\mathrm{TP}_i + \mathrm{FN}_i),\ \mathrm{IR} = \frac{\sum_{i=1}^{15} \mathrm{TP}_i}{N_{\text{segment}}}$$
$$\pi_e = \frac{\sum_{i=1}^{15} \{(\mathrm{TP}_i + \mathrm{FP}_i)(\mathrm{TP}_i + \mathrm{FN}_i)\}}{{N_{\text{segment}}}^2},\ \kappa = \frac{\mathrm{IR} - \pi_e}{1 - \pi_e},$$

where $N_{\text{segment}}$ is the total number of segments.

## 7.4 Results

The biometric verification results within a one-to-one client-imposter classification problem are next summarised. In terms of the verification, the following scenarios were considered:

- Client-imposter verification based on varying segment lengths $L_{\text{seg}}$ (Section 7.4.1),
- Verification with various combinations of features (Section 7.4.2),
- Verification across different classifiers, both non-parametric and parametric ones (Section 7.4.3),
- Verification of registered clients and imposters ($S_R$), and of non-registered-imposters ($S_N$) (Section 7.4.4),
- Subject-wise verification (Section 7.4.5).

In addition, biometric identification is considered, that is, a one-to-many subject-to-subject classification problem (Section 7.4.6). Table 7.5 summarises the details of the considered scenarios.

**Table 7.5:** Summary of parameter choice in the proposed biometric application.

| Section | # of subjects | $L_{\text{seg}}$ | Features | Classifier | Setup | System |
|---|---|---|---|---|---|---|
| 7.4.1 | 15 | 10, 20, 30, 60, 90 | PSD+AR | Cos. Dist. | R&B | Verification |
| 7.4.2 | 15 | 60 | PSD, AR, PSD+AR | Cos. Dist. | R | |
| 7.4.3 | 15 | 60 | PSD+AR | Cos. Dist., LDA, SVM | R&B | |
| 7.4.4 | 15+5 | 60 | PSD+AR | Cos. Dist., LDA, SVM | R&B | |
| 7.4.5 | 15 | 60 | PSD+AR | Cos. Dist. | R | |
| 7.4.6 | 15 | 10, 20, 30, 60, 90 | PSD+AR | Cos. Dist. | R&B | Identification |

### 7.4.1 Client-Imposter verification with different segment sizes

Table 7.6 summarises validation results for both Setup-R and Setup-B, over different segment sizes $L_{\text{seg}} = 10, 20, 30, 60, 90$ s. Both the PSD and AR features were used. The elements in TP, FN, FP and TN columns denote the number of segments classified by a validation stage. The chance level in this scenario is 14/15 = 93.3 %; this is because every

subject is once used for client data, $V_{VC} \in \mathbb{R}^{N \times D}$ in Table 7.3, and imposter data are selected from the other 'non-client' subjects, $V_{VI} \in \mathbb{R}^{14N \times D}$ in Table 7.3. The ratio between $V_{VC}$ and $V_{VI}$ is therefore $1 : 14$, and thus the chance level is 14/15. In Setup-R, the results with $L_{\text{seg}} = 60\,\text{s}$ achieved both the best HTER score, 17.2 %, and the best accuracy (AC), 95.7 %. Notice that, the number of TP (=183) is larger than FN + FP (=174) with $L_{\text{seg}} = 60\,\text{s}$, therefore, the likelihood of making a true positive verification is higher than making a false verification. In Setup-B, the results with $L_{\text{seg}} = 90\,\text{s}$ obtained both the best HTER score, 6.9 %, and the best accuracy (AC), 98.3 %.

**Table 7.6:** Client-Impostor verification over different segment sizes.

| | $L_{\text{seg}}$ | TP | FN | FP | TN | FAR (%) | FRR (%) | HTER (%) | AC (%) |
|---|---|---|---|---|---|---|---|---|---|
| ***Setup-R*** | 10 s | 622 | 996 | 996 | 21656 | 4.4 | 61.6 | 33.0 | 91.8 |
| | 20 s | 371 | 439 | 439 | 10901 | 3.9 | 54.2 | 29.1 | 92.8 |
| | 30 s | 271 | 269 | 269 | 7291 | 3.6 | 49.8 | 26.7 | 93.4 |
| | 60 s | 183 | 87 | 87 | 3693 | **2.3** | **32.2** | **17.2** | **95.7** |
| | 90 s | 120 | 60 | 60 | 2460 | 2.4 | 33.3 | 17.8 | 95.6 |
| ***Setup-B*** | 10 s | 1097 | 523 | 523 | 22157 | 2.3 | 32.3 | 17.3 | 95.7 |
| | 20 s | 611 | 199 | 199 | 11141 | 1.8 | 24.6 | 13.2 | 96.7 |
| | 30 s | 425 | 115 | 115 | 7445 | 1.5 | 21.3 | 11.4 | 97.2 |
| | 60 s | 233 | 37 | 37 | 3743 | 1.0 | 13.7 | 7.3 | 98.2 |
| | 90 s | 157 | 23 | 23 | 2497 | **0.9** | **12.8** | **6.9** | **98.3** |

TP: True Positive, FN: False Negative, FP: False Positive, TN: True Negative, FAR: False Accept Rate, FRR: False Reject Rate, HTER: Half Total Error Rate, AC: Accuracy

### 7.4.2 Client-Imposter verification with different features

Table 7.7 shows the validation results in Setup-R, and over a range of different selections of features, such as AR coefficients, frequency band power, and the combination of AR and band power features for the segment length of $L_{\text{seg}} = 60\,\text{s}$. The classification results using both AR features and PSD features were the highest in terms of both HTER and AC, which corresponds to Table 7.6 (Fourth row), for $L_{\text{seg}} = 60\,\text{s}$.

**Table 7.7:** Rigorous setup: Client-Imposter verification over different features in Setup-R.

| $L_{\text{seg}} = 60$ | # of features $D$ | FAR (%) | FRR (%) | HTER (%) | AC (%) |
|---|---|---|---|---|---|
| AR | 20 | 4.8 | 66.7 | 35.8 | 91.1 |
| PSD | 6 | 4.4 | 61.1 | 32.8 | 91.9 |
| AR + PSD | 26 | **2.3** | **32.2** | **17.2** | **95.7** |

### 7.4.3 Client-Imposter verification with different classifiers

Table 7.8 shows the imposter-client verification accuracy based on the minimum cosine distance, LDA, and SVM, for both Setup-R and Setup-B, with a segment size of $L_{\text{seg}} = 60\,\text{s}$. Both the PSD and AR features were used. In Setup-R, the results with cosine distance were

the best in terms of both HTER score, 17.2 % and AC, 95.7 %. In Setup-B, the results of both HTER and AC were the best based on the SVM classifier, 5.5 % and 99.0 %, respectively.

**Table 7.8:** Client-Impostor verification with different classifiers.

| $L_{\text{seg}} = 60$ s | TP | FN | FP | TN | FAR (%) | FRR (%) | HTER (%) | AC (%) |
|---|---|---|---|---|---|---|---|---|
| ***Setup-R: rigorous*** | | | | | | | | |
| Cos. Dist | 183 | 87 | 87 | 3693 | 2.3 | **32.2** | **17.2** | **95.7** |
| LDA | 149 | 121 | 139 | 3641 | 3.7 | 44.8 | 24.2 | 93.6 |
| SVM | 135 | 135 | 54 | 3726 | **1.4** | 50.0 | 25.7 | 95.3 |
| ***Setup-B: biased*** | | | | | | | | |
| Cos. Dist | 233 | 37 | 37 | 3743 | 1.0 | 13.7 | 7.3 | 98.2 |
| LDA | 200 | 70 | 87 | 3693 | 2.3 | 25.9 | 14.1 | 96.1 |
| SVM | 241 | 29 | 11 | 3769 | **0.3** | **10.7** | **5.5** | **99.0** |

### 7.4.4 Validation including non-registered imposters

Table 7.9 summarises the confusion matrices of both Setup-R and Setup-B with segment sizes $L_{\text{seg}} = 60$ s, classified by the minimum cosine distance, LDA, and SVM; these correspond to Table 7.8, panels Setup-R and Setup-B. The confusion matrices were categorised into:

- Client matrix $Y_{VC}$ from dataset $S_R$,
- Imposter matrix $Y_{VI}$ from dataset $S_R$
- Imposter matrix $Y_{VI_N}$ from dataset $S_N$.

Notice that the minimum cosine distance approach assigns the class (client or imposter) of the nearest data from the training matrix. In this study, every trial from every subject was selected for the validation of client data, so as to have validated ninety times; therefore, *the nearest data (from dataset $S_R$) for each imposter data from dataset $S_N$, also always become the 'client' once.* Hence, when the nearest data for an $S_N$ data become the 'client' data in the training matrix, the imposter data from dataset $S_N$ are straightforwardly classified as 'client'. Therefore, regardless of data, the TPR for Imposter matrix $Y_{VI_N}$ is 93.3 % for the minimum cosine distance approach; however for comparison among different classifiers, these results are also included.

In Setup-R, the TPR of client $Y_{VC}$, achieved by the minimum cosine distance was the highest, with respective value of 67.8 %. However, the TPRs obtained by SVM for imposters $Y_{VI}$ and $Y_{VI_N}$ were 98.6 % and 96.2 %, respectively, which was higher than those achieved by LDA. In Setup-B, both the TPR of client $Y_{VC}$ and that of imposters $Y_{VI}$ and $Y_{VI_N}$ by SVM were the highest, with respective values of 89.3 %, 99.7 % and 96.3 %.

**Table 7.9:** Confusion matrix of the Client-Imposter verification scenario from different datasets.

| | | $L_{\text{seg}} = 60$ s | Prediction | | | TPR |
|---|---|---|---|---|---|---|
| | | | Client | Imposter | Total | (%) |
| ***Setup-R*** Label | Cos. | Client $Y_{VC}$ | 183 | 87 | 270 | **67.8** |
| | | Imposter $Y_{VI}$ | 87 | 3693 | 3780 | 97.7 |
| | | Imposter $Y_{VI_N}$ * | 90* | 1260* | 1350* | 93.3* |
| | LDA | Client $Y_{VC}$ | 149 | 121 | 270 | 55.2 |
| | | Imposter $Y_{VI}$ | 139 | 3641 | 3780 | 96.3 |
| | | Imposter $Y_{VI_N}$ | 136 | 1214 | 1350 | 89.9 |
| | SVM | Client $Y_{VC}$ | 135 | 135 | 270 | 50.0 |
| | | Imposter $Y_{VI}$ | 54 | 3726 | 3780 | **98.6** |
| | | Imposter $Y_{VI_N}$ | 52 | 1298 | 1350 | **96.2** |
| ***Setup-B*** Label | Cos. | Client $Y_{VC}$ | 233 | 37 | 270 | 86.3 |
| | | Imposter $Y_{VI}$ | 37 | 3743 | 3780 | 99.0 |
| | | Imposter $Y_{VI_N}$ * | 90* | 1260* | 1350* | 93.3* |
| | LDA | Client $Y_{VC}$ | 200 | 70 | 270 | 74.1 |
| | | Imposter $Y_{VI}$ | 87 | 3693 | 3780 | 97.7 |
| | | Imposter $Y_{VI_N}$ | 152 | 1198 | 1350 | 88.7 |
| | SVM | Client $Y_{VC}$ | 241 | 29 | 270 | **89.3** |
| | | Imposter $Y_{VI}$ | 11 | 3769 | 3780 | **99.7** |
| | | Imposter $Y_{VI_N}$ | 50 | 1300 | 1350 | **96.3** |

* Always the same result regardless of the data

### 7.4.5 Client-Imposter verification results per subject

Table 7.10 (Middle columns) summarises the subject- and day-wise validation results with PSD and AR features from $L_{\text{seg}} = 60$ s segments in Setup-R, which corresponds to Table 7.6 (Fourth row), for $L_{\text{seg}} = 60$ s. The first and second columns in the verification part show respectively subject-wise HTER and AC, with the training matrix $Y_T$ selected from all the first day recordings and classified based on the second day recordings. The third and fourth columns in the verification part show classification results obtained based on the training matrix $Y_T$, which was selected from all the second day recordings in order to classify the first day recordings.

### 7.4.6 Biometrics identification scenarios

Table 7.10 (Right column) summarises the subject-wise identification rate obtained by the minimum cosine distance classifier with the PSD and AR features from $L_{\text{seg}} = 60$ s segments in Setup-R. Previously, a binary client-imposter classification problem (e.g. $M = 2$) was considered for each subject, each day, and each trial, however, the classification algorithm used in this study was the simple minimum cosine distance between the validation matrix and training matrix. For the prediction of $l$th row of the validation matrix, $Y_{V_{(l,:)}}$, the minimum distance between the training matrix, $Y_T$, was found, e.g. the $n$th row of the training matrix, and the same label was assigned to the $n$th row of the training matrix as the prediction label for $l$th row of the validation matrix, $Y_{V_{(l,:)}}$. Notice that *the minimum distance approach is applicable for biometrics identification problems, which is a one-*

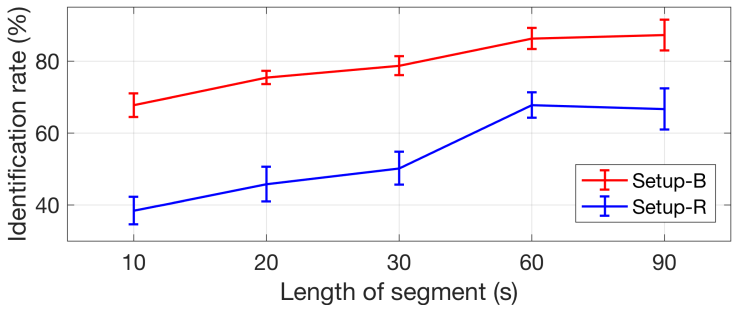



**Figure 7.7:** The identification rate with different segment size in Setup-R and Setup-B. The error bars indicate the standard error.

*to-many classification.* The number of classes was $M = 15$, which corresponds to the the number of subjects in the $S_R$ recordings, therefore the chance level was $1/15 = 6.7\,\%$. The achieved identification rate was $67.8\,\%$ with $L_{\text{seg}} = 60\,\text{s}$ segments in Setup-R, while the achieved kappa coefficient was $\kappa = 0.65$ (Substantial Agreement) [117].

Figure 7.7 shows identification rate of both Setup-R and Setup-B, with different segment sizes $L_{\text{seg}} = 10, 20, 30, 60, 90$ s. In Setup-B, the identification rate with $L_{\text{seg}} = 10$ s was $67.7\,\%$, which was almost the same as the result with $L_{\text{seg}} = 60$ s in Setup-R. The highest identification rate, $87.2\,\%$, was achieved with $L_{\text{seg}} = 90$ s, where corresponding kappa was $\kappa = 0.86$ (Almost Perfect Agreement) in Setup-B.

**Table 7.10:** Recording details, accuracy and HTER in the verification problem, and sensitivity in the identification problem for each subject with different training and validation data in Setup-R.

| | | Impedance | | | | **Verification** (Client-Imposter, $M = 2$) | | | | **Identification** |
|---|---|---|---|---|---|---|---|---|---|---|
| $S_R$ | Interval | (kΩ) | | Time* | | $Y_T$:Day1 → $Y_V$:Day2 | | $Y_T$:Day2 → $Y_V$:Day1 | | ($M = 15$) |
| Sub | day | Day1 | Day2 | Day1 | Day2 | AC(%) | HTER(%) | AC(%) | HTER(%) | SE(%) |
| 1 | 5 | $< 9$ | $< 10$ | M | A | **98.5** | 6.0 | **100** | **0.0** | **94.4** |
| 2 | 6 | $< 6$ | $< 6$ | A | A | 94.8 | 28.6 | 97.0 | 11.9 | 61.1 |
| 3 | 8 | $< 5$ | $< 8$ | A | A | **98.5** | **0.8** | 94.8 | 33.8 | 66.7 |
| 4 | 7 | $< 9$ | $< 10$ | A | M | 95.6 | 28.2 | 96.3 | 7.2 | 66.7 |
| 5 | 7 | $< 8$ | $< 4$ | A | A | 96.3 | 17.5 | 97.0 | 22.2 | 61.1 |
| 6 | 7 | $< 12$ | $< 14$ | A | A | *91.1* | 20.2 | 97.8 | 6.4 | 77.8 |
| 7 | 15 | $< 12$ | $< 13$ | M | M | **98.5** | 11.1 | 96.3 | 22.6 | 66.7 |
| 8 | 6 | $< 11$ | $< 11$ | M | M | *91.1* | *35.8* | *91.9* | *35.4* | *33.3* |
| 9 | 7 | $< 10$ | $< 13$ | A | A | 96.3 | 12.3 | 95.6 | 12.7 | 77.8 |
| 10 | 8 | $< 9$ | $< 9$ | M | M | 91.9 | 35.4 | 93.3 | 8.7 | 61.1 |
| 11 | 6 | $< 11$ | $< 13$ | A | A | 95.6 | 17.8 | 94.8 | 13.1 | 72.2 |
| 12 | 5 | $< 13$ | $< 11$ | M | M | 96.3 | 17.5 | 99.3 | 0.4 | 83.3 |
| 13 | 7 | $< 13$ | $< 9$ | M | M | 95.6 | 17.8 | 97.8 | 11.5 | 72.2 |
| 14 | 6 | $< 8$ | $< 9$ | M | A | 97.8 | 11.5 | 95.6 | 33.4 | 55.6 |
| 15 | 5 | $< 9$ | $< 13$ | A | A | 94.1 | 13.5 | *91.9* | 25.0 | 66.7 |
| Ave. | 7 | - | - | - | - | 95.5 | 18.2 | 96.0 | 16.3 | IR = **67.8 %** |
| ***Setup-R: rigorous***, $L_{\text{seg}} = 60$ | | | | | | Overall AC=95.7 %, HTER=17.2 % | | | | $\kappa = 0.65$ |
| *M: Morning (9-12), A: Afternoon (12-18) | | | | | | **bold**: best result, *italic*: worst result, out of fifteen subjects. | | | | |

## 7.5 Discussion

**Verification with different segment sizes**

Firstly, the classification results were compared for different segment lengths $L_{\text{seg}}$, shown in Table 7.6. Within the same setup, i.e. Setup-R or Setup-B, the performance of HTER and AC increased with the segment length, although the results with $L_{\text{seg}} = 60\,\text{s}$ and $L_{\text{seg}} = 90\,\text{s}$ are almost the same. Longer segments allowed for more data epochs to be averaged over, hence the EEG noise inference for classification diminished and the inherent EEG characteristic were able to be captured by averaging. However, a longer segment length also implies a longer recording time, which is not ideal for feasible EEG biometrics. Compared to the results in Setup-R and Setup-B for the same segment length, both the HTERs and accuracy (ACs) of Setup-B were clearly better than those of Setup-R. In terms of client discrimination, the decrease in FRR was significant from Setup-R to Setup-B. In Setup-B, a larger number of client segments was correctly classified (see TP) than in Setup-R. In setup-R, only the result with $L_{\text{seg}} = 60\,\text{s}$ achieved TP > (FN + FP), which indicates that the likelihood of making a true positive verification is higher than making a false verification; in Setup-B, the shortest segment size, $L_{\text{seg}} = 10\,\text{s}$, achieved TP > (FN + FP).

The difference between Setup-R and Setup-B was that the training matrices $Y_T$ in Setup-B included the trials which were recorded on the same day as the validation trial. In other words, the assigned validation data (trial) and the part of assigned training data (trials) were recorded within 5 - 10 minutes in the same environment. Therefore, the training matrix contains significantly similar EEG recordings to the validation matrix in Setup-B, and this leads to a higher classification performance than the classification in Setup-B.

With an increase in the segment size $L_{\text{seg}}$, the number of segments per recording trial, $N$, became smaller, especially for $N = 2$ with the segment size $L_{\text{seg}} = 90\,\text{s}$; therefore, the training matrix only contains six (two segments $\times$ three trials) examples of client data in Setup-R (*cf.* ten examples in Setup-B), which might not be enough for training client data. Hence, the performance with $L_{\text{seg}} = 90\,\text{s}$ was slightly lower than that with $L_{\text{seg}} = 60\,\text{s}$ in Setup-R.

**Verification with different classifiers**

Table 7.8 shows the classification comparison among the minimum cosine distance methods, LDA and SVM. The SVM was used as a parametric classifier; firstly, the optimal hyper-parameters (see details in Table 7.4) were selected from 5-fold CV within the training matrix, and then weight parameters based on these chosen hyper-parameters were obtained. Notice that the classifier could have been tuned in different ways, for example in order to minimise false acceptance or minimise false rejection. *The optimal tuning in this*

*study was performed so as to maximise class sensitivities, i.e. maximise the number of TP and TN elements, which resulted in minimum HTERs.* In both Setup-R and Setup-B, the FARs by SVM were smaller than those achieved by both the minimum cosine distance and the LDA, because the tuning was performed for maximising TN elements. Since the number of imposter elements was fourteen times bigger than the number of clients in both Setup-R and Setup-B (i.e. chance level was 14/15), the SVM parameters were tuned for higher sensitivity to imposters. As a result, the FRR by SVM, which were related to client sensitivity given in Table 7.9, were higher than those achieved by both minimum cosine distance and LDA in Setup-R.

In Setup-B, as mentioned above, the training matrix contains the data from the same recording day, which are more similar EEG patterns than the data obtained from a different recording day. Therefore, the SVM model chose hyper-parameters and weight parameters from the training matrix, so as to better the validation data in Setup-B, which led to higher performance than by both the minimum cosine distance and LDA.

Notice that, as described before, tuning of the hyper-parameters was performed within the training matrix, then the so-obtained hyper-parameters were used for finding the optimal weight parameters within the training matrix. The same hyper-parameters and weight parameters were used for classifying the validation matrix. This setup is applicable for feasible EEG biometrics scenarios in the real world.

**Validation including non-registered imposters**

In Table 7.9, the confusion matrices for the client matrix $Y_{VC}$ and imposter matrix $Y_{VI}$ from dataset $S_R$ and imposter matrix $Y_{VI_N}$, are given, which were then used for a comparison between Setup-R and Setup-B. Compared to the results obtained within the same classifier (minimum cosine distance, LDA, SVM) in Setup-R and Setup-B, the TPR of clients $Y_{VC}$ and the sensitivity of imposter $Y_{VI}$ from dataset $S_R$ in Setup-B were higher than those in Setup-R. As described before, in Setup-B, the two client trials from the same recording day as the validation trial were included into the training matrix, and therefore more segments were correctly classified. In contrast, the TPRs of imposter $Y_{VI_N}$ from dataset $S_N$ by both the LDA and SVM in Setup-B were almost the same to those in Setup-R; 89.9 % and 88.7 % for LDA, 96.2 % and 96.3 % for SVM. Compared to the TPR between two imposter data ($Y_{VI}$ and $Y_{VI_N}$), regardless of the classifiers, the TPR of $Y_{VI}$ were higher than those of $Y_{VI_N}$. Since the imposter data from $S_N$ were not included in the training matrix, the more $S_N$ data were misclassified as 'client' than $S_R$ data misclassified as 'client'.

However, in the real world scenarios for biometrics, imposters are not always 'registered'. The lower TPR for $Y_{VI_N}$ means that the application is inadequate for attack from non-registered subjects. One potential way to overcome the vulnerability of the minimum cosine distance classifier is by introducing threshold for classification. If the nearest distance is larger than the given distance parameter, the segment is excluded from the classification or is classified as imposter.

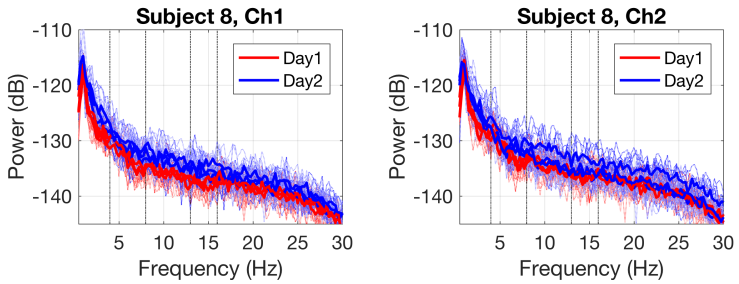



**Figure 7.8:** PSD for the in-ear EEG Ch1 (Left) and the in-ear EEG Ch2 (Right) of Subject Eight. The thick lines correspond to the averaged periodogram obtained by the all recordings from the first day (Red) and the second day (Blue), whereas the thin lines are the averaged periodogram obtained by a single trial.

### Client-Imposter verification results per subject

For subject-wise classification, Table 7.10 summarises classification results obtained by the minimum cosine distance in Setup-R, for different training-validation scenarios. The results varied across subjects and for training-validation configurations between 91.1 % to 100 % of AC and between 0.0 % to 35.8 % of HTER.

The size of viscoelastic earpiece was the same for twenty subjects (both $S_R$ and $S_N$ subjects), therefore all the subjects were able to wear it comfortably. The upper bounds of the electrode impedance over three recordings per day of each participant are given in Table 7.10 (Impedance part). The highest performance was achieved by Subject One, with maximum impedances of 9 kΩ and 10 kΩ for the first and second recording, respectively. Even though the impedances for Subject Two and Subject Five were smaller than those for Subject One, the corresponding performance was below average over fifteen subjects. Besides, the lowest performance was exhibited by Subject Eight, for whom the impedances were smaller than 11 kΩ for Day1 and 11 kΩ for Day2. Figure 7.8 shows average PSDs for Subject Eight – observe that the PSDs for the EEG recorded on Day2 (Blue) is slightly larger than those on Day1 (Red).

### Biometrics identification

In terms of biometrics identification results, a one-to-many subject-to-subject classification problem, the average sensitivity over fifteen subjects, i.e. the identification rate, was 67.8 % in Setup-R with $L_{\text{seg}} = 60$ s, as shown in Table 7.10 (Right column). Figure 7.7 illustrates the identification rates of both Setup-R and Setup-B, with different segment sizes $L_{\text{seg}} = 10, 20, 30, 60, 90$ s. The identification rate increased with segment length, although the results with $L_{\text{seg}} = 60$ s and $L_{\text{seg}} = 90$ s are almost the same.

Notice that the performances with $L_{\text{seg}} = 10$ s in Setup-B and $L_{\text{seg}} = 60$ s in Setup-R were almost the same, 67.7 % and 67.8 %, respectively. The highest identification rate in Setup-B was 87.2 % with $L_{\text{seg}} = 90$ s. Indeed, the training matrix for Setup-B included the

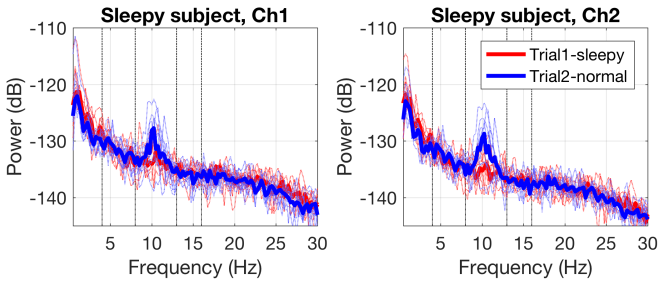



**Figure 7.9:** PSD for the in-ear EEG Ch1 (Left), and the in-ear EEG Ch2 (Right) of one subject. The thick lines correspond to the averaged periodogram obtained by the recordings from a 'sleepy' trial (Red) and a 'normal' trial (Blue). Observe that the alpha power attenuated during the 'sleepy' trial.

trials which were recorded at the same day as the validation trial; therefore the performance was better than in Setup-R.

In a previous biometrics identification study, Maiorana *et al.* [175] analysed nineteen channels of EEG during EC tasks in three different recording days, and achieved the rank-1 identification rate (R1IR) of 90.8 % for a segment length 45 s. Notice that it is hard to compare the performance with our approach, because the number of channels was very different, as nineteen scalp EEG channels covered the entire head vs our two in-ear EEG channels embedded on an earplug. Therefore, although our results were lower, the proof-of-concept in-ear biometrics emphasised the *collectability* aspect in fully wearable scenarios.

#### Alpha attenuation in the real world scenarios

One limitation of using the alpha band, is the sensitivity to drowsiness, a state where the alpha band power is naturally elevated. For illustration, Figure 7.9 shows the PSD obtained from a subject, calculated by Welch's averaging periodogram method. The subject slept during one recording, then the subject was woken up and another recording started less than ten minutes after the first recording. The PSD graphs in Figure 7.9 are overlapped except for the alpha band; the alpha power observed during the 'sleepy' recording trial was smaller than that at the 'normal' recording, thus demonstrating the alpha attenuation due to fatigue, sleepiness, and drowsiness. The alpha attenuation is well known in the research in sleep medicine [16], where it is particularly used to monitor sleep onset.

#### Inter-trial and inter-subject variability

Another concern of this application is inter-trial and inter-subject variability. Figure 7.10 shows scatter plot for two selected features with a segment size $L_{\text{seg}} = 60$ s. The numbers represent the subject IDs. The error bar represents the standard deviation, separately calculated by recording sessions (i.e. Day1 trials and Day2 trials). For some subjects, e.g. Subject Six, the plots obtained from Day1 and Day2 were closely located with small inter-

trial variabilities (the shorter length of error bar). Overall, the figure shows that the inter-trial variability was smaller than inter-subject variability in this study, when the number of subjects was $N = 15$. However, the inter-subject variability reduces when the number of subjects increases, whereas the inter-trial variability may not change significantly, regardless of the number of subjects. As the state-of-the-art, Maiorana [178] conducted recordings over five different sessions (days) over 16 months from 45 subjects using 19 channels of an on-scalp EEG headset, and achieved an error rate of less than 5 %. The feasible authentication approach for a larger number of subjects will remain as a future work.

## 7.6 Summary

This chapter aims to establish a repeatable and highly collectable EEG biometrics using a wearable in-ear sensor. A biometric verification problem has been considered, which was cast into a one-to-one client-imposter classification setting. A proof-of-concept has been introduced for a feasible, collectable and reproducible EEG biometrics in the community by virtue of an unobtrusive, discreet, and convenient to use in-ear EEG device. Robust PSD and AR features have been introduced to identify an individual, and unlike most of the existing studies, the classification has been performed rigorously, without mixing the training and validation data from the same recording days.

The aspects that need to be further addressed in order to fulfil the requirements for 'truly wearable biometrics' in the real world will focus on extensions and generalisations of this proof-of-concept to cater for:

- Intra-subject variability with respect to the circadian cycle and the mental state, such as fatigue, sleepiness, and drowsiness;
- Truly wearable scenarios with portable and affordable amplifiers;
- Inter- and intra-subject variability over the period of months and years.

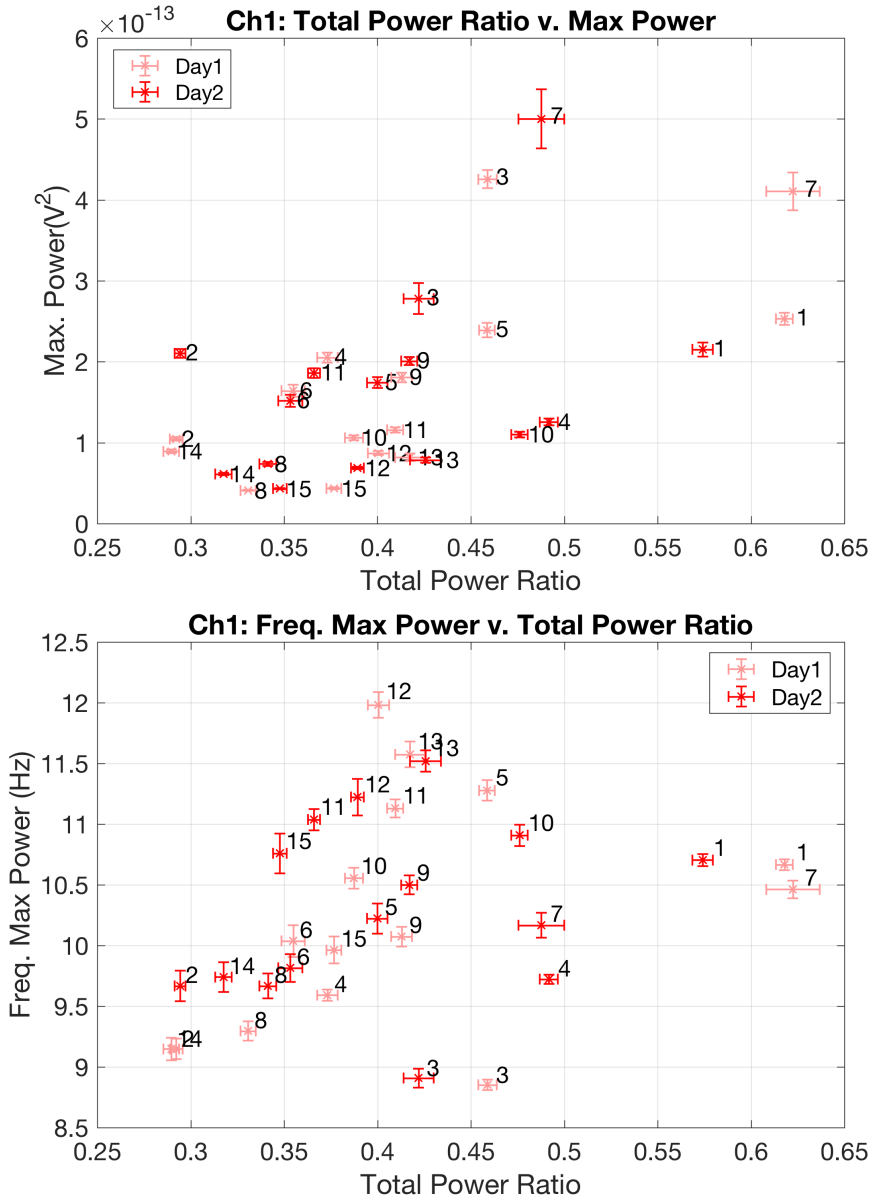



**Figure 7.10:** Scatter plot for the maximum alpha power against total power ratio (Left) and for the frequency at maximum alpha power against total power ratio (Right) of the in-ear Ch1.

# Part III

# In-ear EEG application towards 'transparent EEG'

# Chapter 8

# Automatic overnight sleep monitoring using in-ear EEG

## Overview

In Part I, the automatic sleep stage classification methods based on on-scalp EEG have been evaluated, using publicly available overnight sleep data; the classification model can be a proxy for time-consuming conventional manual sleep scoring by a clinician. In Part II, proof-of-concept in-ear EEG applications have been developed in the area of sleep and person authentication. Combining the two concepts, the feasibility of sleep monitoring in the community using in-ear EEG have been investigated.

## 8.1 Introduction

In our modern 24/7 society, sleep quality has become a major issue affecting the state of body and mind, with implications on both general health and economy. Consequently, quality of sleep is considered one of the most important current topics in sleep medicine [179], and is typically investigated by recording a person's sleep patterns in sleep clinics. However, current sleep monitoring in a clinic is expensive, cumbersome to administer, and prohibitive to perform recordings continuously over days and weeks; this also affects the subsequent diagnosis and treatment. For example, Beebe *et al.* designed and conducted sleep monitoring for three weeks in home using a wrist-worn actigraphy, and found that the daytime behavioural problems for adolescents may have been caused by inadequate sleep [180]. The issues and previous findings have spurred the development of unobtrusive, wearable sensors capable of long-term monitoring of physiological variables related to sleep [181]. On the other hand, such miniaturisation of sensors inevitably affects the quality of recorded data, which calls for advanced signal processing and ML tools, throughout the process, from data conditioning to automatic sleep staging.

The PSG is a standard clinical methodology to diagnose sleep disorders [16]. The so-acquired sleep profiles of individuals are rigorous and comprehensive, however, the expensive and cumbersome nature of PSG may even disturb patients' normal sleep, thus affecting diagnosis and treatment. During the analysis, the recorded PSG data are scored manually by a trained technician; given the scale of sleep disorders in our modern soci-

ety, this imposes unrealistic demands on clinical time. Therefore, from a point of view of continuous widely affordable healthcare in the community and the future healthcare, conventional sleep monitoring using the PSG is not realistic. To address this issue, multiple potential solutions have been proposed and fall under the two main categories:

1. Employ standard PSG but replace the clinician with an automatic scoring system [88],
2. Use wearable sensing, possibly with a reduced number of sensing modalities (e.g. only EEG or actigraphy), and perform the analysis with the clinician or automatically [89].

To address the first issue of time-consuming sleep stage scoring process performed by clinicians, automatic sleep staging systems have been proposed based on full PSG recordings [106, 138] or more recently based on only EEG recordings.

Regarding the second issue of a wider deployment of wearable devices, commercial product already exist based on wrist activity (i.e. actigraphy) [182, 183]. Information from such wearable devices is proven to be sufficient to distinguish between wakefulness and sleep, and the agreement between the actigraphy based wake/sleep stages and those manually scored based on standard PSG recordings can be as high as 91 % [184]. More recently, data from commercial wrist worn devices were found to allow for identifying multiple sleep parameters, such as 'sleep efficiency' and 'total sleep time' [185]. The smartphone accelerometers (i.e. off-body sensing) have also indicated the possibility of monitoring 'sleep duration', although the study did not simultaneously record standard PSG as a 'ground-truth' [181].

Despite relative success, these proof-of-concept achievements based on 'wearables' are not yet capable of faithfully providing the much more complex information regarding the clinically valid sleep stages, that is, to discriminate between wakefulness, Non-REM sleep, and REM sleep.

With the development in sensor technology, one of the most convenient wearable solutions for physiological monitoring introduced in the research community is sensing from the ear. Multiple strategies have also been proposed for ear-EEG based sleep research, as summarised in Table 8.1.

The original study by Looney *et al.* [15] recorded daytime naps with simultaneous in-ear EEG and on-scalp EEG systems, from four healthy participants. The corresponding manually scored hypnograms conclusively validated the feasibility of in-ear EEG for sleep monitoring. The same data were also used for automatic sleep stage classification in [60], which further demonstrated the possibility of out-of-clinic sleep monitoring with in-ear EEG (Chapter 5 in this thesis). After the initial proof-of-concept stage, Alqurashi *et al.* [59] conducted comprehensive multiple daytime nap recordings to establish the matching of the corresponding sleep latencies based on in-ear EEG and on-scalp EEG under two conditions: 1) after normal sleep and 2) after sleep restriction. The same nap data over twenty

**Table 8.1:** Comparison of existing approaches to sleep research using in-ear/around-ear wearable sensing technology.

| Author | # | Earpiece | Setup | Details |
|---|---|---|---|---|
| [15] | 4 | In-ear soft | DT | Comparison of manually scored hypnograms based on in-ear and on-scalp |
| Ch. 5 | | | | Comparison of automatic sleep scoring based on in-ear and on-scalp |
| [59] | 23 | | | Comparison of sleep latency based on in-ear and on-scalp |
| Ch. 6 | | | | Automatic detection of drowsiness based on in-ear EEG |
| **Ch. 8** | **21** | **In-ear soft** | **ON** | **Automatic overnight sleep staging using in-ear EEG** |
| [69] | 8 | In-ear soft | ON | Automatic sleep scoring based on in-ear EEG (accuracy not given) |
| [63] | 1 | In-ear hard | | Comparison of manually scored hypnograms based on in-ear and on-scalp |
| [65] | 9 | | | Automatic overnight sleep staging using in-ear EEG |
| [9] | 1 | Around-ear | | Demonstration of EEG patterns in different sleep stages (activities) |

#: the number of participants, soft: viscoelastic, hard: hardshell, DT: day time, ON: overnight.

three participants were used by Nakamura *et al.* [61] to investigate automatic detection of drowsiness (i.e. to distinguish between wakefulness and light sleep) using in-ear EEG (Chapter 6 in this thesis). Nguyen *et al.* [69] conducted overnight sleep recordings over eight participants to evaluate their in-ear sensing system; their sensors were able to record the EEG, EOG, and EMG, key physiological variables for sleep monitoring. It is important to highlight that the sleep studies in [15, 59, 60, 61, 69] were conducted using one-size-fits-all viscoelastic in-ear sensors, which are not optimised for a particular user but promise affordable and widely deployable solution. Owing to their flexibility and stress-strain properties (memory foam), these viscoelastic earpieces can be squeezed and shaped to fit comfortably any ear; such an in-ear sensor is readily applicable to a large population, a pre-requisite for the future healthcare in the community. With custom-made hardshell earpieces, a technology derived from hearing aids earpieces, Zibrandtsen *et al.* [63] monitored overnight sleep EEG activities from a single participant, and confirmed the similarities in temporal and spectral features between in-ear EEG and conventional on-scalp EEG. Recently, Mikkelsen *et al.* [65] validated an automatic overnight sleep staging using hardshell binaural in-ear EEG recordings over nine participants. The latest around-ear EEG device (cEEGrid) [9], which strictly speaking records on-scalp EEG behind the ear, has been utilised for overnight sleep recordings. It has been shown to be capable of monitoring specific sleep patterns, such as the K-complex, theta activity, and delta activity in NREM 3 stage sleep.

With one-size-fits-all in-ear sensing system [10], the feasibility of sleep monitoring using in-ear EEG is further validated in the following setups:

1. Viscoelastic in-ear sensors [41] for off-the-shelf sleep monitoring of a relatively large sample of healthy young adults;
2. A 'real world' out-of-clinic overnight sleep scenario, namely in participants' homes to reflect their normal sleep patterns;
3. In conjunction with the conventional PSG for hypnogram generation, which then serves as the 'ground-truth' for further analyses.

For rigour and feasibility considerations, the exact same shape of standardised earpieces

(size, materials) was used throughout the recordings on multiple participants. The recordings were undertaken in participants' homes; such familiar environments are a key to representative sleep monitoring, as this minimises the stress and inconvenience of the participants and the likelihood of unusual sleep patterns. Our comprehensive setup involved simultaneous in-ear EEG and PSG recordings; the in-ear EEG data were used for automatic sleep stage classification, whilst the PSG data were scored manually by a clinician. Through a rigorous performance metrics comparison of an automatic sleep staging method based on in-ear EEG and the manually scored hypnogram (based on the standard PSG), the feasibility of automatic overnight sleep monitoring in out-of-clinic scenarios with readily deployable one-size-fits-all in-ear sensors have been confirmed.

## 8.2 Methods

Figure 8.1 presents the flowchart of this study, whereby after simultaneous in-ear EEG and PSG recording in the first step, two channels of on-scalp EEG were extracted from the PSG data. The in-ear EEG and on-scalp EEG data were preprocessed through down-sampling, bandpass filtering, and removal of noisy epochs. Then, both structural complexity and frequency domain features for classification were extracted. The full PSG recordings were manually scored, and the so-obtained hypnogram was used as the 'ground-truth' of sleep stages for automatic classification.

### 8.2.1 Data acquisition

The in-ear EEG and PSG data were simultaneously recorded between October 2017 and June 2018 under the ethics approval, ICREC 17IC4150, Joint Research Office at Imperial College London. In total, 22 healthy participants (aged 23.8 $\pm$ 4.8 years) were recorded, after an informed consent was obtained. Only two participants had worn our in-ear sensor before, whilst none of participants had participated in overnight PSG recordings before.

Participants were visited in their own home at night (approximately two hours before their usual bedtime) to setup the in-ear EEG and PSG sensors; after the sensor setup, the clinician started the recording and left the participant's home. The participant went to bed as per usual and had their normal overnight sleep. The next morning, the clinician visited to detach the sensors; the participants were instructed to detach the sensors at any point during the night should they feel any discomfort.

The in-ear EEG and PSG were recorded simultaneously from two data acquisition systems, and these two amplifiers were manually controlled to start and to stop each recording. To ensure data alignment, the agreements between their time stamps were checked before every recording. For the in-ear EEG, the g.tec g.USBamp amplifier with 24-bit resolution at a sampling frequency $f_s$ = 1200 Hz was used for the recording. The in-ear sensor was in the form of a one-size-fits-all viscoelastic earplug with two flexible

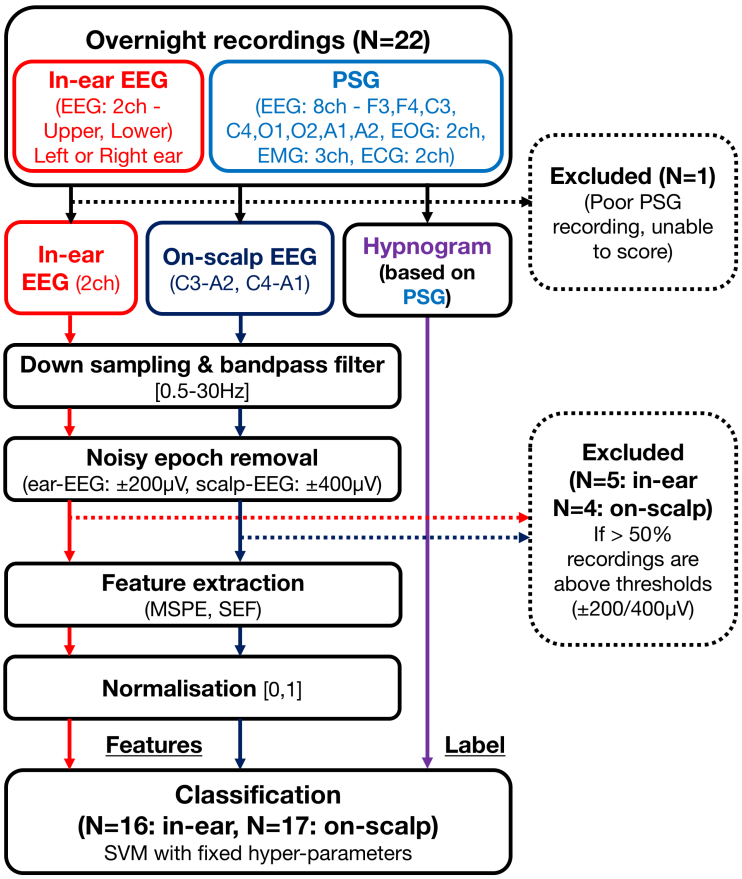



**Figure 8.1:** Flowchart for the automatic sleep stage classification in this study.

electrodes, the details can be found in [10, 41]. The size of in-ear sensors was the same for all participants, approximately 25 mm in length and 12 mm in diameter. Before insertion, a participant's ear canal was cleaned with a cotton bud to remove ear wax; then conductive gel was applied to the electrode. The in-ear sensor was inserted into either participant's left or right ear, according to their preference, within a monaural setup. After the insertion, the sensor adapted firmly to the shape of the ear canal. Standard gold-cup electrodes were used as a reference (behind the ipsilateral helix) and ground (the ipsilateral earlobe). Figure 8.2 shows the in-ear sensors with two electrodes (Left panel) and the placement of earpiece (Right panel). Before the overnight recordings, the quality of in-ear EEG responses were inspected during the participants' resting state with their eyes closed, and the electrode impedance was checked.

The PSGs were recorded by the SOMNO Screen device, by SOMNO medics. The electro-physiological sensors were placed onto participants' scalp (including eight channels of EEG: F3, F4, C3, C4, O1, O2, A1, and A2, according to the international 10-20 system), face (including two channels of EOG), chin (three channels of EMG), chest, ab-

domen, and legs (including two channels of ECG). The ground and reference electrodes were attached on the forehead. Multiple other signals were monitored, such as participants' movement, body position, pulse rate and the pulse waveform (pulse oximeter), Naso/Oral flow (thermistor), and snoring sound. The data were recorded at a sampling frequency of 256 Hz, and transmitted to a laptop wirelessly.

### 8.2.2 Manual scoring

The recorded PSG data, including EEG, EOG, EMG, and respiration, were analysed by the Domino Plus system, by SOMNO medics. The obtained PSGs were filtered with the passband frequencies from 0.2-35 Hz, and such processed PSG data were manually scored by a clinician based on the AASM criteria [105]. The labeled sleep stages were Wakefulness (W), NREM1 (N1), NREM2 (N2), NREM3 (N3), REM, and Movement. One PSG recording was not scored due to high frequency artefacts; therefore, overall, 21 out of 22 participants' overnight recordings were used for further analyses.

### 8.2.3 Pre-processing

The recorded signals were first aligned in accordance with the time stamps from the separate in-ear EEG and PSG amplifiers. In total, approximately 165 hours (i.e. approximately eight hours per participant) of in-ear EEG and PSG data were used for further analyses. For the classification, the following channel configuration was used: 1) two in-ear EEG channels (upper and lower channel of the earpiece), and 2) two on-scalp EEG channels (C3-A2 and C4-A1) from the PSG recordings. The in-ear EEG and on-scalp EEG signals were downsampled to 120 Hz and 128 Hz, respectively; the downsampled frequencies of two systems were different due to the recording sampling rates (1200 Hz for in-ear EEG, and 256 Hz for on-scalp EEG). The downsampled EEG signals were bandpass filtered with a fourth-order Butterworth filter with the passband from 0.5-30 Hz. An example of the recorded signal from an on-scalp and the in-ear EEG channel is illustrated in Figure 2.2.

Epochs scored as 'Movement' were removed. The size of epochs was considered as

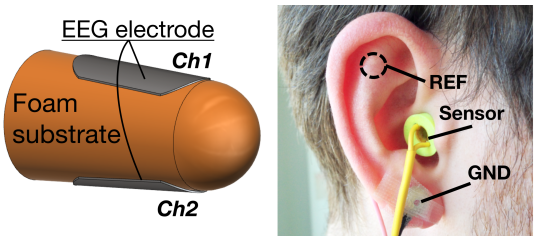



**Figure 8.2:** The in-ear sensor used in this study. (Left) Wearable in-ear sensor with two flexible electrodes; (Right) placement of the in-ear sensor.

a standard (i.e. 30 s segment of recordings) while epochs which contained amplitudes of more than $\pm 200\,\mu$V for in-ear EEG, and $\pm 400\,\mu$V for on-scalp EEG, were deemed to be contaminated by noise. This is because, for the on-scalp EEG, the amplitude of the K-complex, a signature response of NREM 2 sleep, is normally less than $\pm 400\,\mu$V [130]; $\pm 200\,\mu$V was assumed applicable for such amplitude thresholding in in-ear EEG, since the amplitude of in-ear EEG is smaller than that of on-scalp EEG [36], as also seen in Figure 8.6. In order to remove noisy epochs from further analyses, the epochs for classification were selected using the criteria below:

1. Find all epochs which contain amplitudes larger than the threshold (i.e. $\pm 200\,\mu$V for in-ear EEG, and $\pm 400\,\mu$V for on-scalp EEG) in each channel;
2. Remove an epoch if *at least* one channel of in-ear EEG (or on-scalp EEG) has amplitudes above the threshold;
3. Count the number of removed epochs (noisy epochs);
4. Do not consider a trial if the number of noisy epochs is more than 50 % of the entire recording from a participant.

This epoch/participant rejection strategy was applied to in-ear EEG and on-scalp EEG separately, therefore, the total numbers of epochs in the in-ear EEG and on-scalp EEG systems for further analyses were different, with details of their proportion given in Table 8.2. The number of participants and epochs for further analysis of in-ear EEG were respectively $N = 16$ and 11610 (out of 15120 – 76.8 % remaining epochs), while $N = 17$ and 13040 (out of 15970 – 81.7 % remaining epochs) for on-scalp EEG. This, in turn, means that the remaining participants for in-ear EEG and on-scalp EEG were not necessarily the same, since the removed participants were chosen based on amplitude thresholding.

**Table 8.2:** Proportion of scored sleep stages in different systems.

| In-ear EEG (N=16) | | | | | | |
|---|---|---|---|---|---|---|
| | Wake | N1 | N2 | N3 | REM | All |
| # of epochs | 2568 | 183 | 5332 | 2048 | 1479 | 11610 |
| Ratio (%) | 22.1 | 1.6 | 45.9 | 17.6 | 12.7 | 100 |
| On-scalp EEG (N=17) | | | | | | |
| | Wake | N1 | N2 | N3 | REM | All |
| # of epochs | 2385 | 163 | 6510 | 1902 | 2080 | 13040 |
| Ratio (%) | 18.3 | 1.2 | 49.9 | 14.6 | 16.0 | 100 |

### 8.2.4 Feature extraction

After the pre-processing stage, the MSPE and SEF metrics were calculated for each epoch of both in-ear EEG and on-scalp EEG data. The details of MSPE can be found in Section 2.1.1; the same parameters (i.e. the scale: $\tau = 1 - 20$ ($\tau_{\text{min}} = 1$ and $\tau_{\text{max}} = 20$) , the embedding dimension: $d = 5$, and the time delay: $L = 1$) were used for this study.

The MSPE analysis for in-ear EEG and on-scalp EEG channels of overnight data for one participant is shown in Figure 2.3. As a frequency domain feature, the SEF was used. For this study, the PSD of each epoch (30 s) was obtained using Welch's averaged periodogram method with the window length of 4 s and 50 % window overlap. The frequency ranges for SEF50, SEF95, and SEF$d$ were chosen in the same bands given in Table 2.3 (Setup-N).

### 8.2.5 Classification

The extracted features were normalised to between $[0, 1]$ participant-wise before performing classification. The multi-class SVM with the RBF kernel was employed as a classifier.

### 8.2.6 Performance evaluation

The pre-processing and feature extraction analyses were undertaken in Matlab 2016b, and the classification was implemented in Python 2.7.12, Anaconda 4.2.0 operated on an iMac with 2.8GHz Intel Core i5, and 16GB of RAM. The performance indicators to evaluate the proposed methods were class-specific sensitivity (SE) and precision (PR), as well as overall accuracy (AC) and kappa coefficient ($\kappa$); the details are given in Section 2.4.

### 8.2.7 Validation setup

A 10-fold CV approach was utilised; EEG recordings from all subjects were concatenated into one large matrix, which was then randomly split into the training data (90 %) and the test data (10 %). The validation process was repeated ten times with changing the selection of training and test data.

## 8.3 Results

The feature matrices based on two in-ear EEG channels and two on-scalp EEG channels were classified by a multi-class SVM with fixed hyper-parameters. A 10-fold CV approach was utilised.

Figure 8.3 shows the classification results for $M = 5$ class (W, N1, N2, N3, and REM) using the in-ear EEG and on-scalp EEG. In the in-ear EEG setup, the overall accuracy was 74.1 % and the corresponding $\kappa$ value was 0.61, which indicates Substantial Agreement, whereas the accuracy and $\kappa$ of on-scalp EEG were respectively 85.9 % and 0.79 (Substantial Agreement). The sensitivities to each sleep stage of in-ear EEG were approximately 10 % lower than those of on-scalp EEG, except for the REM condition. The sensitivities to REM of in-ear EEG and on-scalp EEG were 45.8 % and 81.8 %, respectively. Notice that more epochs labeled REM were misclassified as N2 (679 epochs) than correctly classified as REM (677 epochs) in the in-ear EEG setup.

Figure 8.4 depicts the participant-wise classification accuracy (Blue bar plot) and the corresponding $\kappa$ (Red dot plot) for both the in-ear EEG and on-scalp EEG scenarios. Although not without variations, overall, data from all participants were amenable to be classified using the proposed methods.

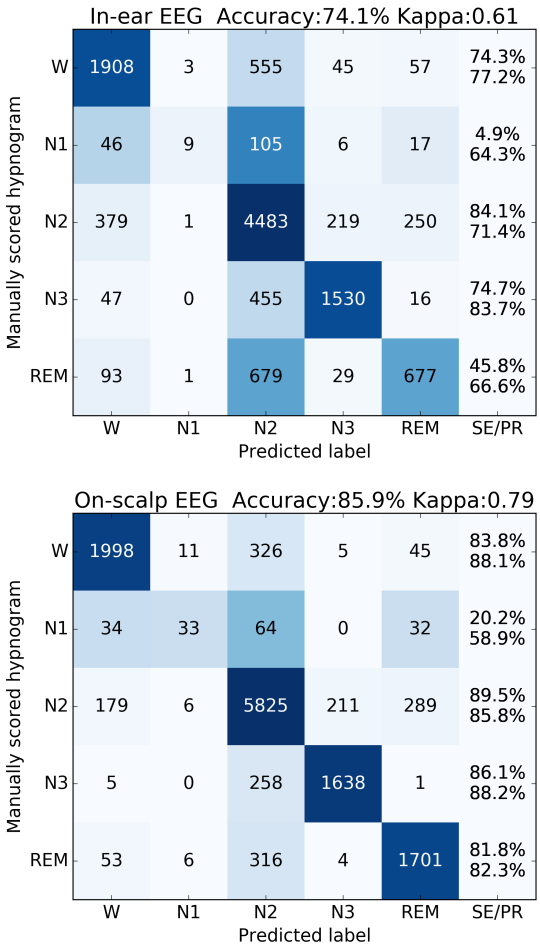



**Figure 8.3:** Confusion matrix for the classification results: (Top) In-ear EEG; (Bottom) on-scalp EEG. The symbols SE/PR on the bottom right denote respectively sensitivity and precision.

Figure 8.5 illustrates the overnight hypnograms of two participants for both the in-ear EEG and on-scalp EEG system; the graphs show the manually scored hypnogram based on the full PSG recordings (Blue) and the automatically predicted label based on the proposed algorithm (Red). The black crosses denote the epochs removed from the analyses due to amplitude thresholding. In Figure 8.5 (A), notice the REM condition between time stamps 03:30 and 04:30; regarding the predicted label based on in-ear EEG, a large portion of epochs was misclassified as N2 sleep, whereas the majority of epochs was correctly classified as REM in the prediction based on on-scalp EEG. For further illustration, in a

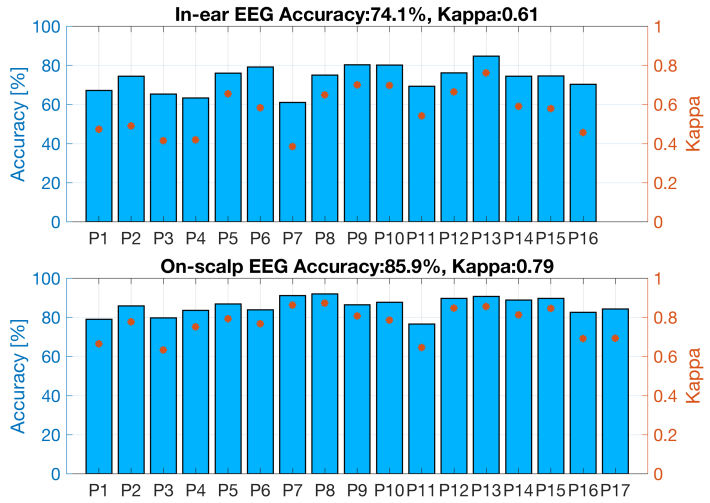



**Figure 8.4:** Classification accuracy (Blue bars) and kappa values (Red dots) for individual participants: (Top) Based on in-ear EEG; (Bottom) based on on-scalp EEG).

hypnogram of another participant in Figure 8.5 (B), multiple sleep cycles, for example three REM conditions, can be observed.

Table 8.3 displays the classification results over different number of classes ($M = 2, 3, 5$). The class-wise sensitivity and precision (*in italic*), and overall accuracy and $\kappa$ are also provided. The notation $M = 2$ means the 2-stage classification (Wake v. Sleep), whereas $M = 3$ denotes 3-stage classification (Wake v. NREM sleep v. REM sleep). The accuracy of in-ear EEG ranged from 74.1 % to 89.9 % with the corresponding $\kappa$ from 0.60 to 0.68. The accuracies and kappa coefficients ($\kappa$) of on-scalp EEG were higher than those of in-ear EEG.

**Table 8.3:** Sensitivity and *Precision* (*in italic*) for sleep classification in different class scenarios.

| In-ear EEG | $M = 2$ | $M = 3$ | $M = 5$ |
|---|---|---|---|
| Wake | 66.7 *84.1* | 70.4 *80.6* | 74.3 *77.2* |
| N1 | | | 4.9 *64.3* |
| N2 | 96.4 *91.1* | 92.0 *82.3* | 84.1 *71.4* |
| N3 | | | 74.7 *83.7* |
| REM | | 42.1 *68.1* | 45.8 *66.6* |
| Accuracy | 89.9 | 80.9 | 74.1 |
| $\kappa$ | 0.68 | 0.60 | 0.61 |
| **On-scalp EEG** | $M = 2$ | $M = 3$ | $M = 5$ |
| Wake | 81.6 *91.3* | 82.9 *89.3* | 83.8 *88.1* |
| N1 | | | 20.2 *58.9* |
| N2 | 98.3 *96.0* | 94.3 *91.7* | 89.5 *85.8* |
| N3 | | | 86.1 *88.2* |
| REM | | 80.3 *83.1* | 81.8 *82.3* |
| Accuracy | 95.2 | 90.0 | 85.9 |
| $\kappa$ | 0.83 | 0.80 | 0.79 |

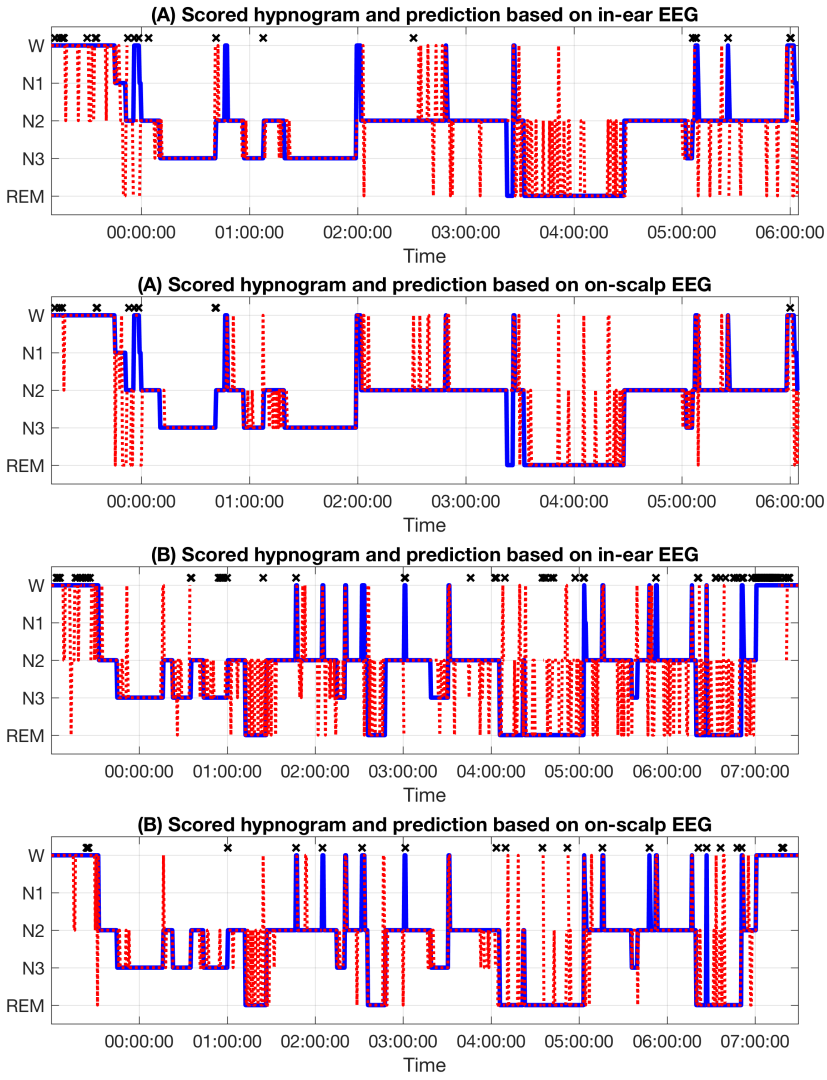



**Figure 8.5:** Manually scored hypnograms of overnight sleep based on the full PSG (Blue) and the automatically predicted label by the proposed algorithm with the in-ear EEG and on-scalp EEG features (Red) from two participants. The black crosses above the W stage indicate the epochs removed from the analyses due to the amplitude thresholding.

## 8.4 Discussion

This study has proposed an overnight sleep monitoring system using a viscoelastic in-ear sensor, and has validated the feasibility of automatic sleep staging based on in-ear EEG. Compared to the gold standard – manually scored hypnogram based on a standard PSG recording – the obtained classification accuracy using in-ear EEG features was 74.1 % with the corresponding $\kappa$ value of 0.61, which indicates Substantial Agreement.

Compared to the classification performance based on in-ear EEG, the results based on on-scalp EEG were better, especially regarding the sensitivity to REM stage. As seen in Figure 8.3, the majority of manually labeled REM epochs were misclassified as N2 in the in-ear EEG setup. Figure 8.6 illustrates the averaged power spectral density for the in-ear EEG (Left) and the on-scalp EEG (Right) of two participants. For this analysis, the recorded signals were manually selected in order to compare N2 v. N3 v. REM; for the top panel (Participant One), a single consecutive 90 minutes of sleep data were selected, whereas two consecutive 60 minutes of sleep data were selected (i.e. 120 minutes in total) for the bottom panel (Participant Two). The PSDs were obtained using Welch's averaged periodogram method, the window length was 4 s with 50 % of window overlap. The trends in in-ear EEG and on-scalp EEG analyses were similar and included: 1) high-alpha (12-15 Hz) activities in N2 sleep, 2) prominent delta activities (0.5-4 Hz), especially slow wave activity (0.5-2 Hz) in N3 sleep, and 3) relatively lower voltage amplitude in REM. However, the slow wave activities of N2 and REM sleep in the in-ear EEG were similar, as evidenced by an overlap in their spectrum, whereas clear visual separation was present in the on-scalp EEG setup. This overlap might have caused the lower discrimination performance for N2 and REM in the in-ear EEG scenario. Note that the on-scalp EEG montage (C3-A2 and C4-A1) is the gold standard for sleep medicine, and has been studied and validated over decades. Also the algorithm applied in this study was originally tested and developed on a publicly available dataset of on-scalp EEG, as discussed in Chapter 3. Future work will consider introducing a fine-tuned classifier, specifically designed for in-ear EEG. Collecting a very large cohort of in-ear sleep EEG data will allow us to examine more practical validation setups, such as leave-one-subject-out CV in addition to K-fold CV.

This study leaves room for improvement, for example, some noisy epochs were removed by amplitude thresholding, and some participants were removed from the analyses. According to visual inspection of the shape of the recorded signal, the noise was categorised into: 1) abrupt electrode noise and 2) constant noise from respiration. The first type of noise might have been caused by participants' movement. Note that the in-ear sensor has been extensively validated in a research lab when the participants were in the resting state, but only a few studies were conducted with participants' movements (e.g. jaw movement [10]). These issues can be resolved with an advanced sensor design. The second type of noise was due to the placement of the in-ear sensor on the participant's

head. In certain recordings, the recorded signal was overlaid by a large amplitude of slow oscillation, which represents an artefact from respiration. This study utilised a monaural setup in order to minimise both the time for technical setup and the participants' inconvenience, however, the locations on the participants' head and earpiece might have interfered with the quality of recordings.

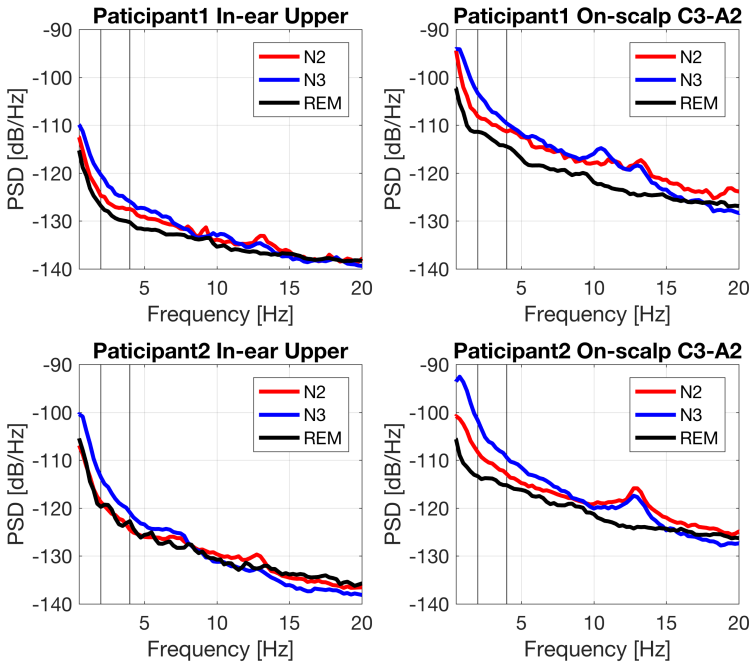



**Figure 8.6:** Averaged PSD of in-ear EEG and on-scalp EEG during selected consecutive epochs of two different participants. The vertical lines denote the 2 and 4 Hz frequencies in order to designate the delta activity (0.5-4 Hz) and slow wave activity (0.5-2 Hz).

## 8.5 Summary

The aim of the chapter was to propose and to validate an automatic overnight sleep monitoring system with readily deployable one-size-fits-all viscoelastic in-ear sensors. Full standard PSG and in-ear EEG were simultaneously recorded for twenty-two healthy participants, who have participated in overnight sleep recordings at their own home which both minimised participants' inconvenience and provided a 'real world' out-of-clinic scenario. The on-scalp EEG and in-ear EEG have showed a high degree of similarity in both the structural complexity and the spectral domains. This study has demonstrated that a single in-ear sensor is capable of monitoring overnight sleep in an unobtrusive and inexpensive way, and that one-size-fits-all viscoelastic sensor promises to become a viable proxy to conventional sleep monitoring in a clinic.

# Chapter 9

# Conclusions and future work

## 9.1 Conclusions

This thesis addresses the feasibility of wearable brain monitoring (electroencephalography, EEG) applications in the real world in the context of fully wearable 24/7 solutions. Based on advanced sensing technology and signal processing algorithms, recent consumer wearable devices can monitor physiological responses, especially cardiac activities, in out-of-the-laboratory environment. Such systems also give feedback to the users based on the recorded signal (responses). As a different modality of physiological responses, the EEG has been traditionally investigated in medicine as a signature of cognitive processes; recently the engineering research community has explored the use of EEG signal as a communication channel between humans and machines. In this research area, this thesis has suggested wearable solutions to facilitate a user's life, based on in-ear EEG, especially in the area of sleep and person authentication. In terms of sleep, an automatic sleep staging algorithm, which is capable of capturing the characteristics of human sleep using conventional on-scalp EEG, have been first introduced . The feasibility of our own viscoelastic in-ear sensor for sleep monitoring in out-of-clinic scenarios have been then demonstrated. Combining these two ideas, a viable wearable solution for monitoring sleep in the real world have been introduced, as an example of the *transparent EEG*. Regarding the biometrics, even though it is still at a proof-of-concept stage, the proposed in-ear EEG based person authentication system can resolve the cumbersome setup issues associated with conventional on-scalp EEG based biometrics.

This chapter presents a brief summary of the contributions of this thesis and outlines future work.

- **Feature extraction for automatic sleep staging.** The combination of two features, 1) structural complexity features (multi-scale entropy) and 2) non-linear frequency domain features (spectral edge frequency), are shown to be effective for classifying sleep stages from a limited number of EEG channels. The proposed methods are evaluated using benchmarking publicly available sleep recordings, PhysioNet Sleep-EDF and NSRR CCSHS datasets.

- **Proof-of-concept sleep studies with in-ear sensors.** The on-scalp and in-ear EEG data were simultaneously recorded in the research laboratory; the so-recorded data were scored blindly (from on-scalp and from in-ear) by a clinician. The EEG data recorded from in-ear sensor represent an individual's sleep profile when compared to conventional on-scalp EEG. The agreement between 1) automatic sleep stage prediction based on in-ear EEG, and 2) manually scored hypnogram based on on-scalp EEG was 76.8 % with a corresponding kappa value of 0.65.
- **Readily collectable person authentication based on in-ear EEG.** The in-ear EEG data were recorded over two days from fifteen participants, and the data were split into training and testing in a rigorous way, which minimised the performance bias. With two authentication scenarios, verification and identification, the feasibility of the proposed in-ear EEG biometrics were validated.
- **Overnight automatic sleep staging using in-ear EEG in out-of-laboratory scenario.** Simultaneous polysomnography (PSG) and in-ear EEG recordings were conducted in 22 participants' homes. The manual sleep scoring was performed based on the so-obtained PSG data by a clinician, whereas the recorded in-ear EEG data were used for automatic sleep scoring. From the extensive analysis, the in-ear sensor were proven to 1) be feasible for monitoring overnight sleep outside a controlled environment and 2) mitigate technical difficulties associated with the current PSG recordings. Consequently, it represents a wearable alternative to conventional, cumbersome and expensive sleep monitoring in an out-of-clinic scenario.

## 9.2 Future work

In order to avoid any confusion or ambiguity, it is important to emphasise that 'good' quality EEG can only be recorded when subjects remain still. Since the voltage of EEG is of the order of 10 $\mu$V, when the subjects move their body (especially eyes, jaw, neck), the electrical potential related to the brain activities is immediately smeared by these additional potentials. Therefore, for any EEG-based applications with a short recording window size, the participants must minimise their movements and focus on the task. For example, the latest visual speller proposed in the BCI community, which is based on steady-state responses, only requires 0.5 s of data to decode the users' intention via on-scalp EEG responses [87]. However, the BCI application requires that users sit still, minimising their movements, and focusing on the visual speller.

In this thesis, a key feature was the use of a longer window size. For example, the standard epoch size of sleep monitoring is 30 s according to the current scoring guideline. The window size of in-ear EEG biometrics was 20 s. If the EEG is recorded with a longer window size, it is feasible to detect target EEG responses through averaging; therefore, even though the area covered by our in-ear sensor is limited and the subjects move slightly

during the night, physiological responses from the brain is monitored.

Although the application of in-ear EEG was proposed and demonstrated, many modifications and extensions still remain to be explored. Here some ideas to be addressed, which enhance the robustness of in-ear EEG applications in the real world, are summarised. Some of these ideas are in progress in Professor Mandic's research group.

- **Sensor development.** Throughout the series of experiments in this thesis, the size of viscoelastic in-ear sensors was kept the same for all subjects, approximately 25 mm in length and 12 mm in diameter [41]. Some subjects have mentioned the discomfort of the in-ear sensor due to the size. Preparing multiple sizes of earpieces, or possibly different viscoelastic materials would help to alleviate this fitting problem. In addition, the fitting problem affects the quality of the EEG signal over a long period of time. The current design of the in-ear sensor uses two electrodes and two microphones (for a mechanical signal) [10], however, the mechanical signal recording from the in-ear [42] has not been examined yet in real world setups. Robust microphone recordings from in-ear enrichs the data modality; for example, cardiac activities from a single in-ear sensor, would allow us to monitor a users physiological responses from an additional perspective/modality.

- **Sensor development (as a group project).** In this thesis, all recordings were undertaken with sensors which were connected to the amplifier via wires. The all-in-one in-ear device, which uses embedded electrodes, microphones for the mechanical signal, battery, amplifier, and data logger or data transmitter, will facilitate the use of this in-ear EEG application. Figure 9.1 (Left) illustrates a prototype integrated wireless in-ear sensor.

- **Automatic sleep staging based on different data cohorts.** Until recently, the number of publicly available PSG (or EEG) datasets, such as the Sleep-EDF dataset [93], has been limited. In 2016, the National Sleep Research Resource (NSRR) provided more than 10,000 PSG recordings from multiple participants with different age, sex, and health condition [94]. Note that one of the ultimate goals for automatic sleep staging is to provide a viable proxy for human scoring. Therefore, such algorithms should provide a detailed sleep profile based on PSG recordings automatically, and ideally, diagnose sleep disorders. To this end, the first step is to validate the automatic sleep staging with multiple datasets; for example the classification model would be trained by Dataset A, which contains PSG recordings from young healthy participants, whereas the validation data would be selected from the Dataset B, which contains clinical PSG recordings from patients. The study in [138] proposed a 'deep learning' approach using four different datasets; the prediction model was trained using two datasets, which both contained recordings from healthy subjects, whereas the test data were from two cohorts of clinical data, recorded from

patients with sleep disorders. This validation approach would be the next standard for automatic sleep staging in the future.

- **In-ear sleep recordings from different groups (e.g. age, health condition).** The recorded in-ear EEG for the series of sleep studies were from healthy subjects, aged between 20 to 40. The ultimate goals of this in-ear EEG application is to provide a wearable sleep monitoring solution in the community. Therefore, investigating in-ear EEG responses using a heterogeneous dataset, such as recordings from different age groups and from patients of specific sleep disorders, is the subject of future work.

- **Prediction of in-ear EEG data based on on-scalp EEG data.** Sleep monitoring with the in-ear sensor is shown to be a viable proxy for providing sleep profiles in an out-of-clinic scenario. However, the current method requires the use of a scored hypnogram based on in-ear EEG sleep data (or based on simultaneously recorded PSG data). Since there are publicly available overnight sleep PSG (or EEG) datasets including Sleep-EDF and NSRR [93, 94], the ideal automatic sleep classification with in-ear EEG will 1) train the classification model based on on-scalp EEG from publicly available datasets and 2) provide the sleep profiles automatically based on new in-ear EEG recordings.

- **Permanence of in-ear EEG biometrics.** In Chapter 7, in-ear EEG biometrics are discussed through a rigorous validation scenario – data recorded over multiple recording days, and the training and test segments were not taken from the same recording days. However, this setup did not fully evaluate the feasibility, the permanence of biometrics in terms of the number of individuals, and the number of data acquisition days. For example, recently Maiorana *et al.* [178] have conducted recordings over six days from the same individuals across 30 participants, and rigorously assessed their on-scalp EEG biometrics.

- **In-ear EEG for evaluating driver fatigue (as a group project).** This is a current group project, which aims to explore the in-ear responses of participants during driving simulations, shown in Figure 9.1 (Right). Driver fatigue is a subjective symptom, which can be described as a feeling of tiredness, lack of attention and alertness (which is linked with sleepiness), and a difficulty in controlling the vehicle. This fatigue symptom is reflected in increased reaction times, the frequency of steering, and causes impairment to spatial recognition, such as the inability to identify the distance from the car in front. Previously, on-scalp EEG has been used to investigate the psychophysiological links to driver fatigue [156, 186, 187]. Due to the nature of our in-ear sensor, the driver can readily wear the earpiece, which monitors their physiological responses on the road. The in-ear responses can be analysed in real-time to give feedback to or alert the driver when physiological responses indicate

fatigue using signal processing and machine learning algorithms.

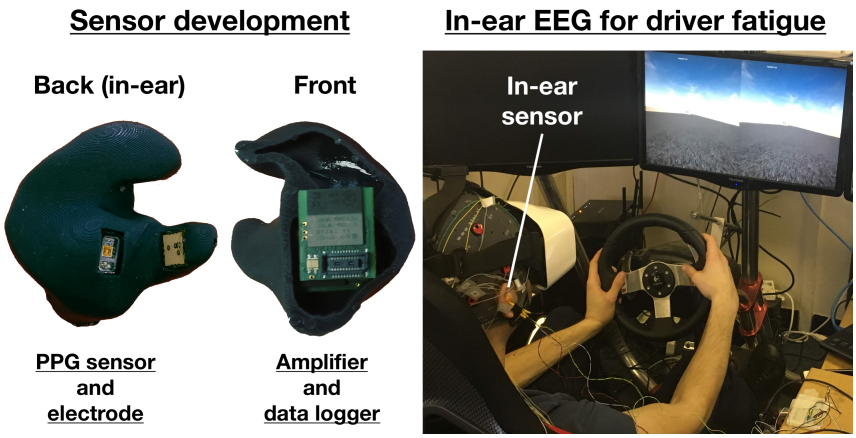



**Figure 9.1:** Future work. (Left) Sensor development; (Right) in-ear sensing for driver fatigue.

# References


[1] U. of Minnesota, "UMN research shows people can control robotic arm with their minds." https://twin-cities.umn.edu/news-events/umn-research-shows-people-can-control-robotic-arm-their-minds. [Online. Last Accessed: 26-Jun-2019].

[2] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Scientific Reports*, vol. 6, no. 1, pp. 1–15, Article no. 38565, 2016.

[3] A. Nijholt, D. Tan, G. Pfurtscheller, C. Brunner, J. D. R. Millón, B. Allison, B. Graimann, F. Popescu, B. Blankertz, and K. R. Müller, "Brain-computer interfacing for intelligent systems," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 72–79, 2008.

[4] K. Gramann, D. P. Ferris, J. Gwin, and S. Makeig, "Imaging natural cognition in action," *International Journal of Psychophysiology*, vol. 91, no. 1, pp. 22–29, 2014.

[5] K. Gramann, J. T. Gwin, D. P. Ferris, K. Oie, T.-P. Jung, C.-T. Lin, L.-D. Liao, and S. Makeig, "Cognition in action: Imaging brain/body dynamics in mobile humans," *Reviews in the Neurosciences*, vol. 22, no. 6, pp. 593–608, 2011.

[6] Y.-P. Lin, Y. Wang, and T.-P. Jung, "Assessing the feasibility of online SSVEP decoding in human walking using a consumer EEG headset," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, pp. 1–8, Article no. 119, 2014.

[7] Emotiv, "EPOC+." https://www.emotiv.com/epoc/. [Online. Last Accessed: 26-Jun-2019].

[8] Neurosky, "Brainlink." http://neurosky.com. [Online. Last Accessed: 26-Jun-2019].

[9] M. G. Bleichner and S. Debener, "Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG," *Frontiers in Human Neuroscience*, vol. 11, no. 4, pp. 1–14, Article no. 163, 2017.

[10] V. Goverdovsky, W. von Rosenberg, T. Nakamura, D. Looney, D. J. Sharp, C. Papavassiliou, M. J. Morrell, and D. P. Mandic, "Hearables: Multimodal physiological in-ear sensing," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[11] D. Looney, C. Park, P. Kidmose, M. L. Rank, M. Ungstrup, K. Rosenkranz, and D. P. Mandic, "An in-the-ear platform for recording electroencephalogram," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6882–6885, 2011.

[12] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear," *Scientific Reports*, vol. 5, no. 1, pp. 1–11, Article no. 16743, 2015.

[13] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG," *Entropy*, vol. 14, no. 7, pp. 1186–1202, 2012.

[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.

[15] D. Looney, V. Goverdovsky, I. Rosenzweig, M. J. Morrell, and D. P. Mandic, "A wearable in-ear encephalography sensor for monitoring sleep: Preliminary observations from nap studies," *Annals of the American Thoracic Society*, vol. 13, no. 12, pp. 2229–2233, 2016.

[16] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel, M. R. Pressman, and C. Iber, "The visual scoring of sleep in adults," *Journal of Clinical Sleep Medicine*, vol. 3, no. 2, pp. 121–131, 2007.

[17] I. D. Corporation, "IDC Forecasts Shipments of Wearable Devices to Nearly Double by 2021 as Smart Watches and New Product Categories Gain Traction." https://www.idc.com/getdoc.jsp?containerId=prUS43408517, 2017. [Online. Last Accessed: 26-Jun-2019].

[18] Forbes, "Apple Watch Series 4: When Will The Brilliant ECG Feature Arrive In The States And Worldwide?." https://www.forbes.com/sites/davidphelan/2018/09/27/apple-watch-series-4-when-will-the-ecg-feature-arrive-in-the-states-and-worldwide/, 2018. [Online. Last Accessed: 26-Jun-2019].

[19] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Wiley Online Library, 2007.

[20] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring The Brain*. Wolters Kluwer, 2015.

[21] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 4, no. 113, pp. 767–791, 2002.

[22] C. C. Chernecky and B. J. Berger, *Laboratory Tests and Diagnostic Procedures*. Elsevier, 2012.

[23] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. OUP USA, 2012.

[24] M. Wilson, "Six views of embodied cognition," *Psychonomic Bulletin & Review*, vol. 9, no. 4, pp. 625–636, 2002.

[25] T. R. Mullen, C. A. Kothe, Y. M. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2553–2567, 2015.

[26] C. E. Schroeder, D. A. Wilson, T. Radman, H. Scharfman, and P. Lakatos, "Dynamics of active sensing and perceptual selection," *Current Opinion in Neurobiology*, vol. 20, no. 2, pp. 172–176, 2010.

[27] K. Gramann, J. T. Gwin, N. Bigdely-Shamlo, D. P. Ferris, and S. Makeig, "Visual evoked responses during standing and walking," *Frontiers in Human Neuroscience*, vol. 4, no. 10, pp. 1–12, Article no. 202, 2010.

[28] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, no. 8, pp. 145–151, 1996.

[29] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Removal of movement artifact from high-density EEG recorded during walking and running," *Journal of Neurophysiology*, vol. 103, no. 6, pp. 3526–3534, 2010.

[30] M. De Vos, K. Gandras, and S. Debener, "Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away," *International Journal of Psychophysiology*, vol. 91, no. 1, pp. 46–53, 2014.

[31] R. Zink, B. Hunyadi, S. V. Huffel, and M. De Vos, "Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks," *Journal of Neural Engineering*, vol. 13, no. 4, pp. 1–11, Article no. 46017, 2016.

[32] L. D. Liao, I. J. Wang, S. F. Chen, J. Y. Chang, and C. T. Lin, "Design, fabrication and experimental validation of a novel dry-contact sensor for measuring electroencephalography signals without skin preparation," *Sensors*, vol. 11, no. 6, pp. 5819–5834, 2011.

[33] S. Ladouce, D. I. Donaldson, P. A. Dudchenko, and M. Ietswaart, "Understanding minds in real-world environments: Toward a mobile cognition approach," *Frontiers in Human Neuroscience*, vol. 10, no. 1, pp. 1–14, Article no. 694, 2017.

[34] M. Weiser, "The computer for the 21st century," *Scientific American*, vol. 265, no. 3, pp. 94–104, 1991.

[35] J. Plazak and M. Kersten-Oertel, "A survey on the affordances of Hearables," *Inventions*, vol. 3, no. 3, pp. 1–16, Article no. 48, 2018.

[36] D. Looney, P. Kidmose, C. Park, M. Ungstrup, M. Rank, K. Rosenkranz, and D. P. Mandic, "The in-the-ear recording concept: User-centered and wearable brain monitoring," *IEEE Pulse*, vol. 3, no. 6, pp. 32–42, 2012.

[37] P. Kidmose, D. Looney, L. Jochumsen, and D. P. Mandic, "Ear-EEG from generic earpieces: A feasibility study," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 543–546, 2013.

[38] P. Kidmose, D. Looney, and D. P. Mandic, "Auditory evoked responses from Ear-EEG recordings," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 586–589, 2012.

[39] P. Kidmose, D. Looney, M. Ungstrup, M. L. Rank, and D. P. Mandic, "A study of evoked potentials from ear-EEG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2824–2830, 2013.

[40] J. Hoon Lee, S. Min Lee, H. Jin Byeon, J. Sook Hong, K. Suk Park, and S.-H. Lee, "CNT/PDMS-based canal-typed ear electrodes for inconspicuous EEG recording," *Journal of Neural Engineering*, vol. 11, no. 4, pp. 1–11, Article no. 46014, 2014.

[41] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, "In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 271–277, 2016.

[42] V. Goverdovsky, D. Looney, P. Kidmose, C. Papavassiliou, and D. P. Mandic, "Co-located multimodal sensing: A next generation solution for wearable health," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 138–145, 2015.

[43] W. von Rosenberg, T. Chanwimalueang, V. Goverdovsky, N. S. Peters, C. Papavassiliou, and D. P. Mandic, "Hearables: Feasibility of recording cardiac rhythms from head and in-ear locations," *Royal Society Open Science*, vol. 4, pp. 1–13, Article no. 171214, 2017.

[44] J. J. S. Norton, D. S. Lee, J. W. Lee, W. Lee, O. Kwon, P. Won, S.-Y. Jung, H. Cheng, J.-W. Jeong, A. Akce, S. Umunna, I. Na, Y. H. Kwon, X.-Q. Wang, Z. Liu, U. Paik, Y. Huang, T. Bretl, W.-H. Yeo, and J. A. Rogers, "Soft, curved electrode systems capable of integration on the auricle as a persistent braincomputer interface," *Proceedings of the National Academy of Sciences*, vol. 112, no. 13, pp. 3920–3925, 2015.

[45] K. B. Mikkelsen, S. L. Kappel, D. P. Mandic, and P. Kidmose, "EEG recorded from the ear: Characterizing the ear-EEG Method," *Frontiers in Neuroscience*, vol. 9, no. 11, pp. 1–8, Article no. 438, 2015.

[46] K. B. Mikkelsen, P. Kidmose, and L. K. Hansen, "On the keyhole hypothesis: High mutual information between ear and scalp EEG," *Frontiers in Human Neuroscience*, vol. 11, no. 6, pp. 1–9, Article no. 341, 2017.

[47] S. L. Kappel, C. B. Christensen, K. B. Mikkelsen, and P. Kidmose, "Reference configurations for ear-EEG steady-state responses," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5689–5692, 2016.

[48] B. G. Do Valle, S. S. Cash, and C. G. Sodini, "Wireless behind-the-ear EEG recording device with wireless interface to a mobile device (iPhone/iPod touch)," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5952–5955, 2014.

[49] S. L. Kappel and P. Kidmose, "High-density ear-EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2394–2397, 2017.

[50] S. L. Kappel, M. L. Rank, H. O. Toft, M. Andersen, and P. Kidmose, "Dry-contact electrode ear-EEG," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 150–158, 2019.

[51] M. G. Bleichner, M. Lundbeck, M. Selisky, F. Minow, M. Jäger, R. Emkes, S. Debener, and M. De Vos, "Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see?," *Physiological Reports*, vol. 3, no. 4, pp. 1–9, Article no. e12362, 2015.

[52] H. Dong, P. M. Matthews, and Y. Guo, "A new soft material based in-the-ear EEG recording technique," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5709–5712, 2016.

[53] A. Favre-Felix, C. Graversen, T. Dau, and T. Lunner, "Real-time estimation of eye gaze by in-ear electrodes," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4086–4089, 2017.

[54] A. Martin and J. Voix, "In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1256–1263, 2017.

[55] S. L. Kappel, D. Looney, D. P. Mandic, and P. Kidmose, "Physiological artifacts in scalp EEG and ear-EEG," *BioMedical Engineering OnLine*, vol. 16, no. 1, pp. 1–16, Article no. 103, 2017.

[56] I. C. Zibrandtsen, P. Kidmose, C. B. Christensen, and T. W. Kjaer, "Ear-EEG detects ictal and interictal abnormalities in focal and generalized epilepsy - A comparison with scalp EEG monitoring," *Clinical Neurophysiology*, vol. 128, no. 12, pp. 2454–2461, 2017.

[57] I. C. Zibrandtsen, P. Kidmose, and T. W. Kjaer, "Detection of generalized tonic-clonic seizures from ear-EEG based on EMG analysis," *Seizure*, vol. 59, pp. 54–59, 2018.

[58] Y. Gu, E. Cleeren, J. Dan, K. Claes, W. Van Paesschen, S. Van Huffel, and B. Hunyadi, "Comparison between scalp EEG and behind-the-ear EEG for development of a wearable seizure detection system for patients with focal epilepsy," *Sensors*, vol. 18, no. 1, p. 29, 2017.

[59] Y. D. Alqurashi, T. Nakamura, V. Goverdovsky, J. Moss, M. I. Polkey, D. P. Mandic, and M. J. Morrell, "A novel in-ear sensor to determine sleep latency during the Multiple Sleep Latency Test (MSLT) in healthy adults with and without sleep restriction," *Nature and Science of Sleep*, vol. 10, pp. 385–396, 2018.

[60] T. Nakamura, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, "Automatic sleep monitoring using ear-EEG," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, no. 1, pp. 1–8, Article no. 2800108, 2017.

[61] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Automatic detection of drowsiness using in-ear EEG," in *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 5569–5574, 2018.

[62] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Hearables: Automatic overnight sleep monitoring with standardised in-ear sensor," *(in press)*, 2019.

[63] I. Zibrandtsen, P. Kidmose, M. Otto, J. Ibsen, and T. W. Kjaer, "Case comparison of sleep features from ear-EEG and scalp-EEG," *Sleep Science*, vol. 9, no. 2, pp. 69–72, 2016.

[64] A. Stochholm, K. Mikkelsen, and P. Kidmose, "Automatic sleep stage classification using ear-EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4751–4754, 2016.

[65] K. B. Mikkelsen, D. B. Villadsen, M. Otto, and P. Kidmose, "Automatic sleep staging using ear-EEG," *BioMedical Engineering OnLine*, vol. 16, no. 1, pp. 1–15, Article no. 111, 2017.

[66] A. Sterr, J. K. Ebajemito, K. B. Mikkelsen, M. A. Bonmati-Carrion, N. Santhi, C. della Monica, L. Grainger, G. Atzori, V. Revell, S. Debener, D.-J. Dijk, and M. De Vos, "Sleep EEG derived from behind-the-ear electrodes (cEEGrid) compared to standard polysomnography: A proof of concept study," *Frontiers in Human Neuroscience*, vol. 12, no. 11, pp. 1–9, Article no. 452, 2018.

[67] K. B. Mikkelsen, J. K. Ebajemito, M. A. Bonmati-Carrion, N. Santhi, V. L. Revell, G. Atzori, C. della Monica, S. Debener, D.-J. Dijk, A. Sterr, and M. De Vos, "Machine-learning-derived sleepwake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy," *Journal of Sleep Research*, pp. 1–13, Article no. e12786, 2018.

[68] A. Nguyen, R. Alqurashi, Z. Raghebi, F. Banaei-Kashani, A. C. Halbower, T. Dinh, and T. Vu, "In-ear biosignal recording system: A wearable for automatic whole-night sleep staging," in *Proceedings of the 2016 Workshop on Wearable Systems and Applications (WearSys)*, pp. 19–24, 2016.

[69] A. Nguyen, R. Alqurashi, Z. Raghebi, F. Banaei-kashani, A. C. Halbower, and T. Vu, "A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems*, pp. 230–244, 2016.

[70] A. Nguyen, R. Alqurashi, Z. Raghebi, F. Banaei-Kashani, A. C. Halbower, and T. Vu, "LIBS: A bio-electrical sensing system from human ears for staging whole-night sleep study," *Communications of the ACM*, vol. 61, no. 11, pp. 157–165, 2018.

[71] T. Hwang, M. Kim, S. Hong, and K. S. Park, "Driver drowsiness detection using the in-ear EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4646–4649, 2016.

[72] J. M. Morales, C. Díaz-Piedra, H. Rieiro, J. Roca-González, S. Romero, A. Catena, L. J. Fuentes, and L. L. Di Stasi, "Monitoring driver fatigue using a single-channel electroencephalographic device: A validation study by gaze-based, driving performance, and subjective data," *Accident Analysis and Prevention*, vol. 109, pp. 62–69, 2017.

[73] C. B. Christensen, J. M. Harte, T. Lunner, and P. Kidmose, "Ear-EEG-based objective hearing threshold estimation evaluated on normal hearing subjects," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1026–1034, 2018.

[74] C. B. Christensen, R. K. Hietkamp, J. M. Harte, T. Lunner, and P. Kidmose, "Toward EEG-assisted hearing aids: objective threshold estimation based on ear-EEG in subjects with sensorineural hearing loss," *Trends in Hearing*, vol. 22, pp. 1–13, 2018.

[75] M. G. Bleichner, B. Mirkovic, and S. Debener, "Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison," *Journal of Neural Engineering*, vol. 13, no. 6, pp. 1–13, Article no. 66004, 2016.

[76] B. Mirkovic, M. G. Bleichner, M. De Vos, and S. Debener, "Target speaker detection with concealed EEG around the ear," *Frontiers in Neuroscience*, vol. 10, no. 7, pp. 1–11, Article no. 349, 2016.

[77] F. Denk, M. Grzybowski, S. M. A. Ernst, B. Kollmeier, S. Debener, and M. G. Bleichner, "Event-related potentials measured from in and around the ear electrodes integrated in a live hearing device for monitoring sound perception," *Trends in Hearing*, vol. 22, pp. 1–14, 2018.

[78] L. Fiedler, J. Obleser, T. Lunner, and C. Graversen, "Ear-EEG allows extraction of neural responses in challenging listening scenarios - A future technology for hearing aids?," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5697–5700, 2016.

[79] L. Fiedler, M. Wöstmann, C. Graversen, A. Brandmeyer, T. Lunner, and J. Obleser, "Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech," *Journal of Neural Engineering*, vol. 14, no. 3, pp. 1–11, Article no. 36020, 2017.

[80] T. Nakamura, V. Goverdovsky, and D. P. Mandic, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648–661, 2018.

[81] M. T. Curran, J. K. Yang, N. Merrill, and J. Chuang, "Passthoughts authentication with low cost earEEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1979–1982, 2016.

[82] F. Farooq, D. Looney, D. P. Mandic, and P. Kidmose, "EarEEG based visual P300 brain-computer interface," in *Proceedings of 7th Annual International IEEE EMBS Conference on Neural Engineering (NER)*, pp. 98–101, 2015.

[83] M. Pacharra, S. Debener, and E. Wascher, "Concealed around-the-ear EEG captures cognitive processing in a visual Simon task," *Frontiers in Human Neuroscience*, vol. 11, no. 6, pp. 1–11, Article no. 290, 2017.

[84] Y.-T. Wang, M. Nakanishi, S. L. Kappel, P. Kidmose, D. P. Mandic, Y. Wang, C.-K. Cheng, and T.-P. Jung, "Developing an online steady-state visual evoked potential-based brain-computer interface system using earEEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2271–2274, 2015.

[85] S. I. Choi, C. H. Han, G. Y. Choi, J. Shin, K. S. Song, C. H. Im, and H. J. Hwang, "On the feasibility of using an ear-EEG to develop an endogenous brain-computer interface," *Sensors*, vol. 18, no. 9, pp. 1–14, Article no. 2856, 2018.

[86] E. D. Schroeder, N. Walker, and A. S. Danko, "Wearable ear EEG for brain interfacing," in *Proceedings of SPIE 10051, Neural Imaging and Sensing*, pp. 1–6, Article no. 15, 2017.

[87] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a non-invasive brain-computer interface," *Proceedings of the National Academy of Sciences*, vol. 112, no. 44, pp. 6058–6067, 2015.

[88] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 12, pp. 1412–1423, 2001.

[89] R. Dick, T. Penzel, I. Fietze, M. Partinen, H. Hein, and J. Schulz, "AASM standards of practice compliant validation of actigraphic sleep analysis from SOMNOwatch versus polysomnographic sleep diagnostics shows high conformity also among subjects with sleep disordered breathing," *Physiological Measurement*, vol. 31, no. 12, pp. 1623–1633, 2010.

[90] A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Communications of the ACM*, vol. 43, no. 2, pp. 90–98, 2000.

[91] L. C. Johnson and G. A. Ulett, "Quantitative study of pattern and stability of resting electroencephalographic activity in a young adult group," *Electroencephalography and Clinical Neurophysiology*, vol. 11, no. 2, pp. 233–249, 1959.

[92] J. Berkhout and D. O. Walter, "Temporal stability and individual differences in the human EEG: An analysis of variance of spectral values," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 165–168, 1968.

[93] PhysioNet, "The Sleep-EDF Database [Expanded]." https://physionet.org/physiobank/database/sleep-edfx/. [Online. Last Accessed: 26-Jun-2019].

[94] D. A. Dean, A. L. Goldberger, R. Mueller, M. Kim, M. Rueschman, D. Mobley, S. S. Sahoo, C. P. Jayapandian, L. Cui, M. G. Morrical, S. Surovec, G.-Q. Zhang, and S. Redline, "Scaling up scientific discovery in sleep medicine: The National Sleep Research Resource," *Sleep*, vol. 39, no. 5, pp. 1151–1164, 2016.

[95] NSRR, "National Sleep Research Resource: Cleveland Children's Sleep and Health Study." https://sleepdata.org/datasets/ccshs/files/datasets. [Online. Last Accessed: 26-Jun-2019].

[96] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, pp. 1–4, Article no. 68102, 2002.

[97] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-H. Pan, and Y.-H. Wang, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1649–1657, 2012.

[98] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266–272, 2007.

[99] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, pp. 1–4, Article no. 174102, 2002.

[100] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[101] T. Adjei, W. von Rosenberg, V. Goverdovsky, K. Powezka, U. Jaffer, and D. P. Mandic, "Pain prediction from ECG in vascular surgery," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, no. 1, pp. 1–10, Article no. 2734647, 2017.

[102] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, no. 6, pp. 1–10, Article no. 61918, 2011.

[103] S. A. Imtiaz and E. Rodriguez-Villegas, "A low computational cost algorithm for REM sleep detection using single channel EEG," *Annals of Biomedical Engineering*, vol. 42, no. 11, pp. 2344–2359, 2014.

[104] S. A. Imtiaz and E. Rodriguez-Villegas, "Automatic sleep staging using state machine-controlled decision trees," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 378–381, 2015.

[105] C. Iber, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine, 2007.

[106] H. Sun, J. Jia, B. Goparaju, G.-B. Huang, O. Sourina, M. T. Bianchi, and M. B. Westover, "Large-scale automated sleep staging," *Sleep*, vol. 40, no. 10, 2017.

[107] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–39, 2011.

[108] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.

[109] J. H. Friedman, R. Tibshirani, and T. Hastie, *The Element of Statistical Learning*. Springer, 2001.

[110] J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[111] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[112] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN)*, vol. 2, pp. 985–990, 2004.

[113] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[114] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.

[115] A. Akusok, K. M. Bjork, Y. Miche, and A. Lendasse, "High-Performance extreme learning machines: A complete toolbox for big data applications," *IEEE Access*, vol. 3, pp. 1011–1025, 2015.

[116] Z. Ghahramani, "An introduction to hidden Markov models and Bayesian networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 1, pp. 9–42, 2001.

[117] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[118] M. Kryger, T. Roth, and W. Dement, *Principles and Practice of Sleep Medicine*. Elsevier, 6th ed., 2017.

[119] P. Maquet, "The role of sleep in learning and memory," *Science*, vol. 294, no. 5544, pp. 1048–1052, 2001.

[120] A. Sehgal and E. Mignot, "Genetics of sleep and sleep disorders," *Cell*, vol. 146, no. 2, pp. 194–207, 2011.

[121] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. 1968.

[122] M. Ronzhina, O. Janoušek, J. Koláová, M. Nováková, P. Honzík, and I. Provazník, "Sleep scoring using artificial neural networks," *Sleep Medicine Reviews*, vol. 16, no. 3, pp. 251–263, 2012.

[123] G. Zhu, Y. Li, and P. P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1813–1821, 2014.

[124] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating," *Biomedical Signal Processing and Control*, vol. 24, pp. 1–10, 2016.

[125] T. L. T. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain," *Medical & Biological Engineering & Computing*, pp. 1–10, 2016.

[126] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Computers in Biology and Medicine*, vol. 42, no. 12, pp. 1186–1195, 2012.

[127] PhysioNet, "The Sleep-EDF Database." https://physionet.org/physiobank/database/sleep-edf/. [Online. Last Accessed: 26-Jun-2019].

[128] A. R. Hassan and M. I. H. Bhuiyan, "A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features," *Journal of Neuroscience Methods*, vol. 271, pp. 107–118, 2016.

[129] S. A. Imtiaz and E. Rodriguez-villegas, "An open-source toolbox for standardized use of PhysioNet Sleep EDF Expanded database," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6014–6017, 2015.

[130] G. Bremer, J. R. Smith, and I. Karacan, "Automatic detection of the K-complex in sleep electroencephalograms," *IEEE Transactions on Biomedical Engineering*, vol. BME-17, no. 4, pp. 314–323, 1970.

[131] D. Sommer, M. Chen, M. Golz, U. Trutschel, and D. Mandic, "Fusion of state space and frequency-domain features for improved microsleep detection," in *Proceedings of 15th International Conference on Artificial Neural Networks*, pp. 753–759, 2005.

[132] M. Chen, D. Sommer, S. L. Goh, T. Gautama, D. Obradovic, M. Golz, M. Morrell, H. Wang, and D. Mandic, "A novel tool for sequential fusion of nonlinear features: A sleep psychology application," in *Proceedings of 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 474–478, 2006.

[133] S. Devuyst, "The DREAMS Subjects Database." http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSubjects/. [Online. Last Accessed: 26-Jun-2019].

[134] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research," *Journal of Sleep Research*, vol. 23, no. 6, pp. 628–635, 2014.

[135] PhysioNet, "St. Vincent's University Hospital/University College Dublin Sleep Apnea Database." https://physionet.org/physiobank/database/ucddb/. [Online. Last Accessed: 26-Jun-2019].

[136] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.

[137] S. Biswal, J. Kulas, H. Sun, B. Goparaju, M. B. Westover, M. T. Bianchi, and J. Sun, "SLEEPNET: Automated sleep staging system via deep learning," *arXiv:1707.08262*, 2017.

[138] A. Patanaik, J. L. Ong, J. J. Gooley, S. Ancoli-Israel, and M. W. Chee, "An end-to-end framework for real-time automatic sleep stage classification," *Sleep*, vol. 41, no. 5, pp. 1–11, 2018.

[139] T. Nakamura, T. Adjei, Y. Alqurashi, D. Looney, M. J. Morrell, and D. P. Mandic, "Complexity science for sleep stage classification from EEG," in *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 4387–4394, 2017.

[140] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," in *Proceedings of 2017 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2017.

[141] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," *arXiv:1610.01683*, 2016.

[142] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[143] J. Zhang and Y. Wu, "A new method for automatic sleep stage classification," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 5, pp. 1097–1110, 2017.

[144] C. Liu, T. Weng, and Z. Liu, "ResSleepNet: Automatic sleep stage classification on raw single-channel," *Materials Science and Engineering*, vol. 466, no. 1, pp. 1–7, Article no. 12101, 2018.

[145] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, 2018.

[146] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, "Mixed neural network approach for temporal sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 324–333, 2018.

[147] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering (in press)*, 2019.

[148] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J. F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.

[149] X. Li, L. Cui, S. Tao, J. Chen, X. Zhang, and G. Q. Zhang, "HyCLASSS: A hybrid classifier for automatic sleep stage scoring," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 375–385, 2018.

[150] T. Nakamura, H. J. Davies, and D. P. Mandic, "Scalable automatic sleep staging in the era of Big Data," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC to appear)*, 2019.

[151] P. R. T. Ko, J. A. Kientz, E. K. Choe, M. Kay, C. A. Landis, and N. F. Watson, "Consumer sleep technologies: A review of the landscape," *Journal of Clinical Sleep Medicine*, vol. 11, no. 12, pp. 1455–1461, 2015.

[152] I. Rosenzweig, S. C. R. Williams, and M. J. Morrell, "The impact of sleep and hypoxia on the brain: Potential mechanisms for the effects of obstructive sleep apnea," *Current Opinion in Pulmonary Medicine*, vol. 20, no. 6, pp. 565–571, 2014.

[153] T. Q. Le, C. Cheng, A. Sangasoongsong, W. Wongdhamma, and S. T. S. Bukkapatnam, "Wireless wearable multisensory suite and real-time prediction of obstructive sleep apnea episodes," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 1, pp. 1–9, Article no. 2700109, 2013.

[154] T. Åkerstedt, L. Torsvall, and M. Gillberg, "Sleepiness in shiftwork. A review with emphasis on continuous monitoring of EEG and EOG," *Chronobiology International*, vol. 4, no. 2, pp. 129–140, 1987.

[155] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *International Journal of Neuroscience*, vol. 52, pp. 29–37, 1990.

[156] S. K. L. Lal and A. Craig, "Driver fatigue: Electroencephalography and psychological assessment," *Psychophysiology*, vol. 39, pp. 313–321, 2002.

[157] A. Vakulin, P. G. Catcheside, S. D. Baulk, N. A. Antic, S. Banks, J. Dorrian, and R. D. McEvoy, "Individual variability and predictors of driving simulator impairment in patients with obstructive sleep apnea," *Journal of Clinical Sleep Medicine*, vol. 10, no. 6, pp. 647–655, 2014.

[158] K. C. A. Patrick, S. A. Imtiaz, S. Bowyer, and E. Rodriguez-Villegas, "An algorithm for automatic detection of drowsiness for use in wearable EEG systems," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3523–3526, 2016.

[159] M. A. Carskadon and W. C. Dement, "The multiple sleep latency test: What does it measure?," *Sleep*, vol. 5, pp. 67–72, 1982.

[160] Y. Sato, F. Akazawa, D. Muramatsu, T. Matsumoto, A. Nakamura, and T. Sota, "An authentication method by high spectral resolution palm datacube," in *Proceedings of 2013 the International Conference on Biometrics and Kansei Engineering*, pp. 239–244, 2013.

[161] C. D. Holland and O. V. Komogortsev, "Complex eye movement pattern biometrics: The effects of environment and stimulus," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 2115–2126, 2013.

[162] I. Odinaka, P. H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, and J. W. Rohrbaugh, "ECG biometrics: A robust short-time frequency analysis," in *Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, 2010.

[163] Y. Liu and D. Hatzinakos, "Earprint: Transient evoked otoacoustic emission for biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2291–2301, 2014.

[164] S. Prabhakar, S. Pankanti, and A. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security & Privacy Magazine*, vol. 1, no. 2, pp. 33–42, 2003.

[165] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 738–742, 2007.

[166] P. Campisi and D. L. Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, 2014.

[167] D. La Rocca, P. Campisi, and G. Scarano, "Stable EEG features for biometric recognition in resting state conditions," *Biomedical Engineering Systems and Technologies*, vol. 452, pp. 313–330, 2014.

[168] M. K. Abdullah, K. S. Subari, J. L. C. Loong, and N. N. Ahmad, "Analysis of effective channel placement for an EEG-based biometric system," in *Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 303–306, 2010.

[169] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini, "Unobtrusive biometric system based on electroencephalogram analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–8, 2008.

[170] F. Su, L. Xia, A. Cai, Y. Wu, and J. Ma, "EEG-based personal identification: From proof-of-concept to a practical system," in *Proceedings of International Conference on Pattern Recognition*, pp. 3728–3731, 2010.

[171] S. Marcel and J. d. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 743–748, 2007.

[172] H. J. Lee, H. S. Kim, and K. S. Park, "A study on the reproducibility of biometric authentication based on electroencephalogram (EEG)," in *Proceedings of 6th Annual International IEEE EMBS Conference on Neural Engineering (NER)*, pp. 13–16, 2013.

[173] D. La Rocca, P. Campisi, and G. Scarano, "On the repeatability of EEG features in a biometric recognition framework using a resting state protocol," in *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS)*, pp. 419–428, 2013.

[174] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin, and S. Laszlo, "Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics," *Neurocomputing*, vol. 166, pp. 59–67, 2015.

[175] E. Maiorana, D. La Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 163–175, 2016.

[176] Y. Chen, A. D. Atnafu, I. Schlattner, W. T. Weldtsadik, M. C. Roh, H. J. Kim, S. W. Lee, B. Blankertz, and S. Fazli, "A high-security EEG-based login system with RSVP stimuli and dry electrodes," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2635–2647, 2016.

[177] C. Neuper, R. H. Grabner, A. Fink, and A. C. Neubauer, "Long-term stability and consistency of EEG event-related (de-)synchronization across different cognitive tasks," *Clinical Neurophysiology*, vol. 116, no. 7, pp. 1681–1694, 2005.

[178] E. Maiorana and P. Campisi, "Longitudinal evaluation of EEG-based biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1123–1138, 2018.

[179] I. Rosenzweig, M. Glasser, D. Polsek, G. D. Leschziner, S. C. Williams, and M. J. Morrell, "Sleep apnoea and the brain: A complex relationship," *The Lancet Respiratory Medicine*, vol. 3, no. 5, pp. 404–414, 2015.

[180] D. W. Beebe, G. Fallone, N. Godiwala, M. Flanigan, D. Martin, L. Schaffner, and R. Amin, "Feasibility and behavioral effects of an at-home multi-night sleep restriction protocol for adolescents," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 49, no. 9, pp. 915–923, 2008.

[181] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell, "Unobtrusive sleep monitoring using smartphones," in *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pp. 145–152, 2013.

[182] T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, T. Brown, A. Chesson, J. Coleman, T. Lee-Chiong, J. Pancer, and T. J. Swick, "Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An update for 2007," *Sleep*, vol. 30, no. 4, pp. 519–529, 2007.

[183] M. Willetts, S. Hollowell, L. Aslett, C. Holmes, and A. Doherty, "Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, Article no. 7961, 2018.

[184] L. D. Souza, A. A. Benedito-Silva, M. L. N. Pires, D. Poyares, S. Tufik, and H. M. Calil, "Further validation of actigraphy for sleep studies," *Sleep*, vol. 26, no. 1, pp. 81–85, 2003.

[185] M. T. Smith, C. S. McCrae, J. Cheung, J. L. Martin, C. G. Harrod, J. L. Heald, and K. A. Carden, "Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment," *Journal of Clinical Sleep Medicine*, vol. 14, no. 7, pp. 1231–1237, 2018.

[186] C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accident Analysis and Prevention*, vol. 45, pp. 83–90, 2012.

[187] S. K. L. Lal, A. Craig, P. Boord, L. Kirkup, and H. Nguyen, "Development of an algorithm for an EEG-based driver fatigue countermeasure," *Journal of Safety Research*, vol. 34, pp. 321–328, 2003.