Convolutional Neural Network Based Sleep Stage Classification with Class Imbalance

Authors:

Qi Xu[a,1,*], **Dongdong Zhou[b,1]**, Jian Wang[b], Jiangrong Shen[c], Lauri Kettunen[d], Fengyu Cong [a,b,d]

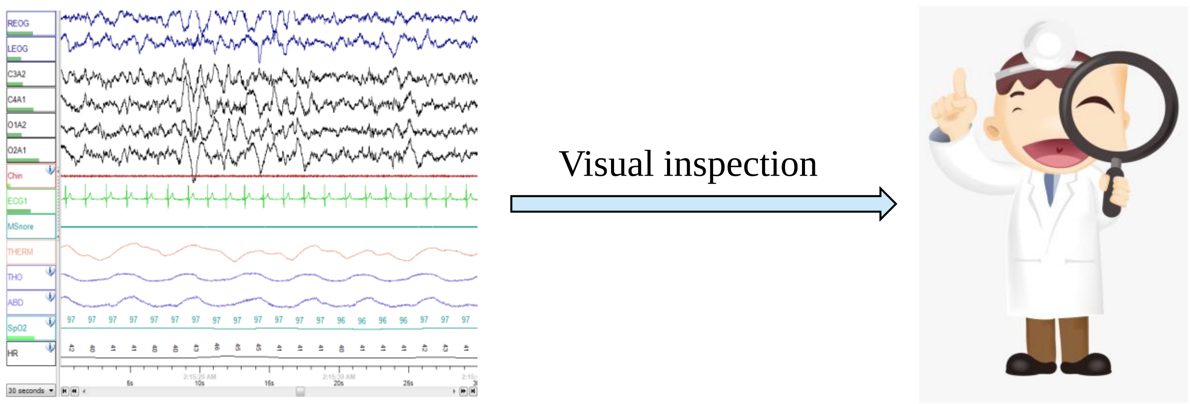
a School of Artificial Intelligence, Dalian University of Technology
b School of Biomedical Engineering, Dalian University of Technology
c College of Computer Science and Technology, Zhejiang University
d Faculty of Information Technology, University of Jyväskylä

Contents

Figure 1. Sleep
(https://wblog.wiki/uk/Sleeping)

1. Introduction

1.1 Manual labeling



Polysomnography (PSG)

Figure 2. The process of manual labeling of PSG recordings

Drawbacks: Time-consuming, Subjective error

1. Introduction

1.2 Two gold rules

- R&K manual (Rechtschaffen, 1968): Wake, N-REM (N1, N2, N3, N4), and REM.

- AASM rule (C. Iber et al., 2007) : Wake, N-REM (N1 ,N2, N3) and REM.

REM: Rapid eye movement;   N-REM: Non-REM.

The biggest change in AASM standard is that stages N3 and N4 are merged into N3 with the abolition of stage “movement time”. (C. Iber et al., 2007)

1. Introduction

1.3 Automatic sleep scoring methods

- Conventional classifier: SVM, KNN, Random forest, etc.
- Artificial neural networks: CNN, RNN, Transformer, etc.

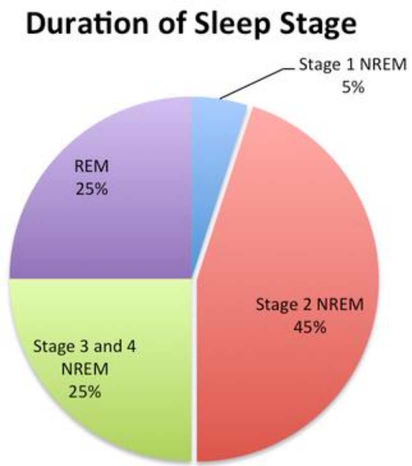


Figure 3. The percentage of each sleep stage in a typical overnight sleep

Class imbalance problem (CIP):

The number of each category is severely unequal. N1 is a representative of minority group.

1. Introduction

1.4 Main contributions

- We systematically analyze the class imbalance problem in PSG datasets and propose the class imbalance factor to quantify it.
- We introduce two balancing methods with the Gaussian white noise, termed signal-driven and image-driven, to balance the dataset samples.
- We propose a novel CNN-based model, implementing the multi-convolution (MC) block, for learning multi-scale feature presentations from the time-frequency image.

2. Materials and methodologies

2.1 Experimental datasets

- Sleep-EDF Database (Sleep-EDF-V1, 2013): 20 subjects (25-34 years), EEG Fpz-Cz
- Sleep-EDF Database Expanded (Sleep-EDF, 2018): 78 subjects (25-101 years), EEG Fpz-Cz
- Cleveland Children's Sleep and Health Study (CCSHS): (16-19 years), 515 subjects, EEG C4/A1

Table I. The data distribution of the experimental datasets

Stage	Sleep-EDF-V1	Sleep-EDF	CCSHS
W	10197 (23.1%)	69518(34.9%)	211030 (30.6%)
N1	**2804 (6.3%)**	21522 (10.8%)	**19211 (2.8%)**
N2	17799 (40.3%)	69132 (34.7%)	249681 (36.2%)
N3	5703 (13.0%)	**13039 (6.6%)**	110188 (16.0%)
REM	7717 (17.5%)	25835 (13.0%)	100252 (14.5%)

2. Materials and methodologies

2.2 Proposed class imbalance factor (*CIF*)

$$CIF = \frac{N}{2 \cdot c \cdot \min\{N_i\}} \quad i \in \{1, 2, \ldots, c\} \tag{1}$$

N： total samples. c： number of stages. N_i： the number of stage i.

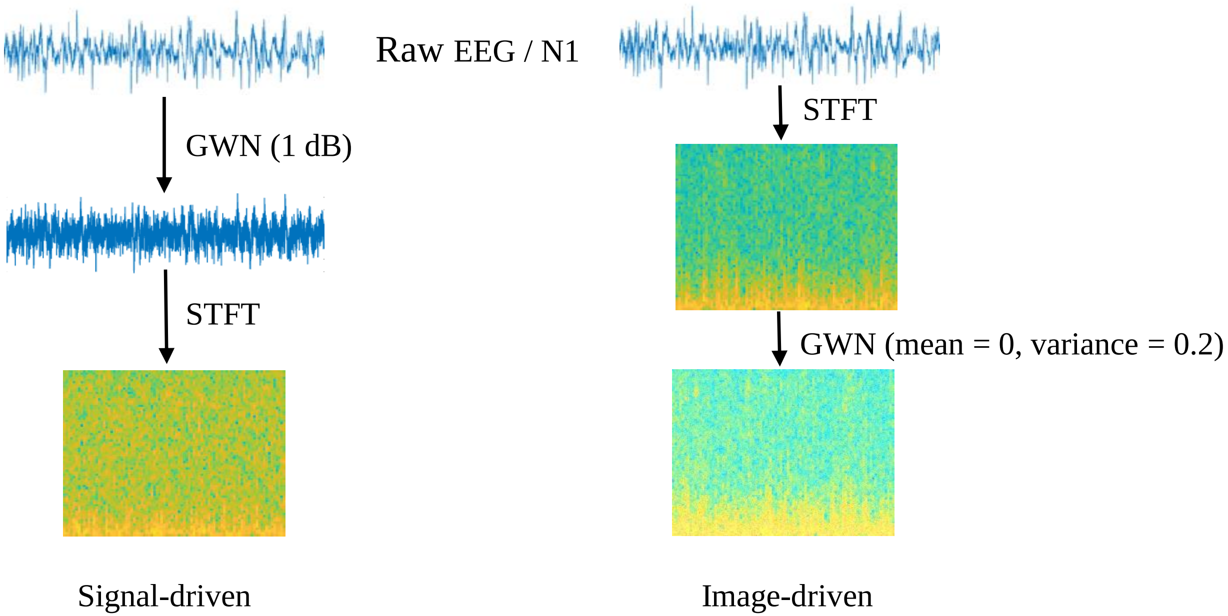
CIF

Sleep-EDF-V1: 3.6 Sleep-EDF: 1.5 CCSHS: 1.6

2. Materials and methodologies

2.3 Balancing Method

Subjective: Increasing the number of minority class (i.e., N1)



GWN: Gaussian white noise

2. Materials and methodologies

2.3 Balancing Method

Table II. The scheme of signal-driven and image-driven approaches

Intensity	signal-driven	image-driven
low	10 dB	mean = 0, variance = 0.05
moderate	5 dB	mean = 0, variance = 0.1
high	1 dB	mean = 0, variance = 0.2

Different intensities: low→moderate→high

2. Materials and methodologies

2.4 Evaluation model

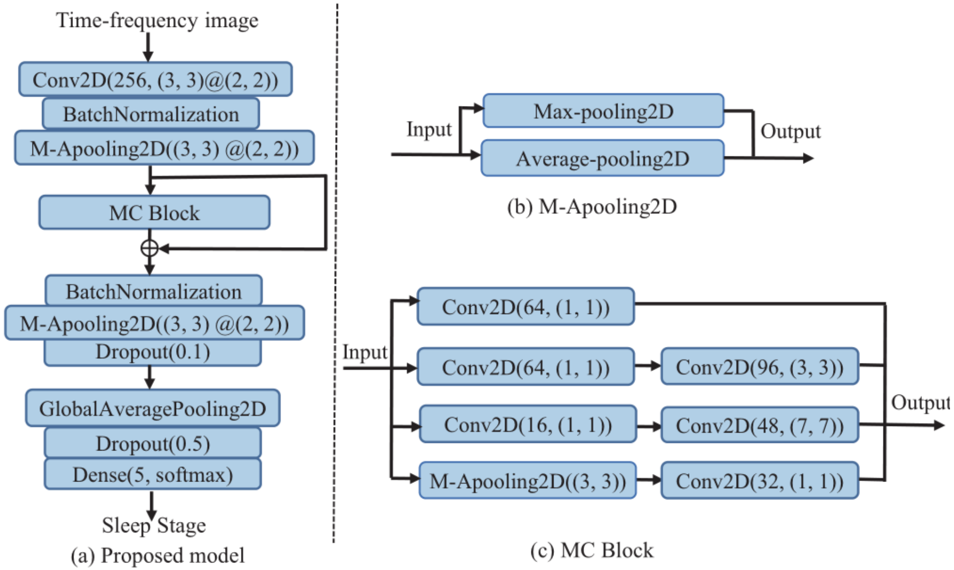


Figure 4. The overall construct of the evaluation model.

MC Block:

Large kernel: (7, 7), Small kernel: (3, 3)

3. Experimental results

3.1 Evaluation metrics

$$RE = \frac{TP}{TP + FN}. \tag{2}$$

$$ACC = \frac{\sum_{i=1}^{n} x_{ii}}{N} \tag{3}$$

$$K = \frac{\frac{\sum_{i=1}^{n} x_{ii}}{N} - \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}{1 - \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}. \tag{4}$$

RE: recall

ACC: accuracy

K: Cohen's kappa coefficient

TP: true positive

FN: false negative

N: total samples

n: number of stages

x_{ii}: diagonal value of the confusion matrix

3. Experimental results

3.2 Results

Table III. Performance comparison of different intensities of the Gaussian noise addition (signal-driven and image-driven ways) in this work

	Sleep-EDF-V1			Sleep-EDF			CCSHS		
	ACC (%)	K (%)	RE_N1 (%)	ACC (%)	K (%)	RE_N1 (%)	ACC (%)	K (%)	RE_N1 (%)
Without DA	86.3	81.1	38.9	84.5	79.0	24.6	87.0	82.0	22.9
DA (signal-driven, 10 dB)	85.4	80.0	35.2	84.5	79.0	26.1	87.3	82.4	25.7
DA (signal-driven, 5 dB)	86.8	81.1	42.7	84.3	78.6	18.7	87.2	82.3	24.1
DA (signal-driven, 1 dB)	87.1	82.3	34.8	84.7	79.3	24.0	87.5	82.5	27.3
DA (image-driven, V = 0.05)	87.0	82.1	30.8	84.6	79.0	15.6	87.1	82.1	21.9
DA (image-driven, V = 0.1)	87.0	82.2	34.5	84.6	79.1	21.1	87.3	82.3	22.2
DA (image-driven, V = 0.2)	86.1	80.7	30.3	84.6	79.1	25.9	87.3	82.3	23.0
DA (Combination, 1 dB & V = 0.1)	87.2	82.4	28.5	84.9	79.4	19.1	87.9	82.9	20.7

3. Experimental results

3.2 Results

Table IV. Performance comparison between the proposed methods and previous methods on the CCSHS dataset

Study	Method	Input channel	Input type	Subjects	$ACC(\%)$	$K(\%)$	RE_N1
Nakamura *et al.* [22]	HMM	C4/A1 + C3/A2	Spectrogram	515	-	73.0	-
Li *et al.* [23]	Random Forest	C4/A1	Features	116	86.0	80.5	7.3
Baseline	**CNN**	**C4/A1**	**Time-frequency image**	**515**	**87.0**	**82.0**	**22.9**
DA (signal-driven, 1 dB)	**CNN**	**C4/A1**	**Time-frequency image**	**515**	**87.5**	**82.5**	**27.3**
DA (image-driven, V = 0.1)	**CNN**	**C4/A1**	**Time-frequency image**	**515**	**87.3**	**82.3**	**23.0**

3. Experimental results

3.2 Results

Table V. Performance comparison between the proposed methods and previous methods on the Sleep-EDF-V1 and Sleep-EDF datasets

Study	Database	Method	Input type	Subjects	$ACC(\%)$	$K(\%)$	RE_N1
Fan *et al.* [14]	Sleep-EDF-V1	Deep CNN	Time series	20	74.8	66.0	-
Phan *et al.* [21]	Sleep-EDF-V1	1-max CNN	Time-frequency image	20	82.6	76	29.9
Zhou *et al.* [15]	Sleep-EDF-V1	CNN	Spectrogram	20	86.1	81.0	-
Baseline	**Sleep-EDF-V1**	**CNN**	**Time-frequency image**	**20**	**86.3**	**81.1**	**38.9**
DA (signal-driven, 5 dB)	**Sleep-EDF-V1**	**CNN**	**Time-frequency image**	**20**	**86.8**	**81.1**	**42.7**
DA (image-driven, V = 0.1)	**Sleep-EDF-V1**	**CNN**	**Time-frequency image**	**20**	**87.0**	**82.2**	**34.5**
Mousavi *et al.* [26]	Sleep-EDF	CNN + LSTM	Time series	78	80.0	73	-
Supratak *et al.* [27]	Sleep-EDF	CNN + LSTM	Time series	78	83.1	77	-
Phan *et al.* [11]	Sleep-EDF	RNN	Time series	78	84.0	77.8	-
Baseline	**Sleep-EDF**	**CNN**	**Time-frequency image**	**78**	**84.5**	**79.0**	**24.6**
DA (signal-driven, 10 dB)	**Sleep-EDF**	**CNN**	**Time-frequency image**	**78**	**84.5**	**79.0**	**26.1**
DA (image-driven, V = 0.2)	**Sleep-EDF**	**CNN**	**Time-frequency image**	**78**	**84.6**	**79.1**	**25.9**

Further information: dongdongzhou1017@foxmail.com
Corresponding author: xuqi@dlut.edu.cn