

Application of Deep Neural Network to Study the Sleep Stage Scoring on the Polysomnography

Shuwei Wang¹, Fuxian Xu¹, Xiangyu Qian¹, Huan Hu¹, Qingzu He¹, Hai Lin², Jianwei Shuai^{1,3*}

¹Department of Physics, College of Physical Science and Technology, Xiamen University, Xiamen Fujian

²Xiamen Zhonglingyi Youchuang Technology Co., Ltd., Xiamen Fujian

³National Institute for Big Data in Healthcare at Xiamen University, Xiamen Fujian

Email: *jianweishuai@xmu.edu.cn

Received: May 16th, 2019; accepted: May 30th, 2019; published: Jun. 6th, 2019

Abstract

In the field of medical informatics, the automatic sleep staging is a challenging and time-consuming task, and most existing automatic sleep staging methods are based on single channel electroencephalography (EEG) data. However, these methods ignore the physician's overall observation of multiple channel EEG and EOG signals for the sleep stage scoring. To resolve this problem, we propose an automatic sleep scoring method based on multi-channel EEG, including three-channel EEG and two-channel Electrooculogram (EOG) data. We introduce the use of a deep convolutional neural network (CNN) on raw EEG samples for supervised learning of sleep stage prediction, which does not require any signal preprocessing or feature extraction. We use the EEG and EOG of polysomnography (PSG) data which have been assessed by medical expert from a Hospital of Fujian Province to train and evaluate our system. Comparing with the staging result with single-channel EEG data, we indicate that the EOG data should not be ignored for a better sleep staging. It shows that the performance of our system is comparable to that of mid-level experts.

Keywords

Sleep Stage Classification, Multichannel, Convolutional Neural Network, EEG, EOG

应用深度神经网络对多导睡眠图的睡眠分期研究

王抒伟¹, 徐富献¹, 钱镶钰¹, 胡桓¹, 何情祖¹, 林海², 帅建伟^{1,3*}

¹厦门大学物理科学与技术学院物理系, 福建 厦门

²厦门中翊易优创科技有限公司, 福建 厦门

³厦门大学健康医疗大数据国家研究院, 福建 厦门

*通讯作者。

摘要

传统上,自动睡眠分期是一项非常具有挑战性且费时费力的任务。大多数现有的自动睡眠分期方法都基于单通道的脑电(electroencephalography, EEG)数据,然而,这些方法忽略了医师从整体上观测多个通道EEG信号进行睡眠阶段评分的过程。为了解决这一问题,我们优化了数据结构,对医师的评分过程进行了详细的学习与建模,提出了一种基于多通道脑电图的自动睡眠评分方法。我们介绍了在原始EEG与EOG样本上使用深度卷积神经网络(convolutional neural network, CNN)进行睡眠阶段评分的监督学习。该网络具有11层,每30 s的睡眠数据作为一个分期,并且不需要任何信号预处理或特征提取。本文使用来自福建省某医院的EEG与EOG及专家评估的多导睡眠图(polysomnography, PSG)数据对系统进行训练和评估。实验结果表明,在自动睡眠分期的研究中不应该忽略EOG数据。我们的系统性能与中级睡眠分期专家的结果相当。

关键词

睡眠分期, 多通道, 卷积神经网络, 脑电图, 眼电图

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人类大约三分之一的时间都被用来睡觉。睡眠对人类的健康来说是极其重要的。与睡眠有关的疾病,如睡眠呼吸暂停、失眠、阵发性发作性睡病,严重影响人类的生活质量。最近的研究表明,睡眠/昼夜节律紊乱可能是阿尔茨海默病和帕金森病等神经退行性疾病的重要标志,睡眠病理的治疗可以改善患者的生活质量[1]。

根据美国睡眠医学会(American Academy of Sleep Medicine, AASM)的睡眠分期标准[2],人类睡眠是一个动态过程,可分为清醒期(Awake, W),快速眼动期(rapid eye movement, REM)和非快速眼动期(non-rapid eye movement, NREM)三个主要期。通常,睡眠专家使用多导睡眠图(Polysomnographic, PSG)进行睡眠阶段分期的临床诊断。PSG 包含一组信号,例如脑电图(Electroencephalogram, EEG),肌电图(Electromyogram, EMG),心电图(Electrocardiogram, ECG)和眼电图(Electrooculogram, EOG)等。这些信号是通过连接到身体不同部位的传感器记录下来的。PSG 录音一般分为 20 秒或 30 秒的样本,睡眠专家或医生将根据美国睡眠医学会睡眠分期标准,把这些样本分为不同的睡眠期,该过程被称为睡眠阶段评分或睡眠分期。

睡眠多导图是睡眠质量的简单表示,可用于诊断睡眠障碍。此外,睡眠分期的质量取决于医师的经验和疲劳程度,睡眠专家分期的准确度通常不到 90% [3] [4],且人工标记 PSG 数据需要耗费大量的时间,因此迫切需要自动睡眠分期算法。

在传统统计学习方法中,有许多研究者一直试图开发一种基于 EEG、EOG 和 EMG 等多种信号[5] [6] [7],或单通道脑电图[8] [9]的自动化睡眠阶段分期的方法。这些方法首先从每个记录样本时期提取时域、频域和时频域特征。在多个信号的情况下,把来自一个时期的所有特征连接成一个特征向量。然后,这些特征用于训练分类器被用于识别单个样本的睡眠阶段分期。由于受试者和记录硬件之间的非均衡异质性,这些网络是根据研究所使用数据集的特征而进行手工设计的,我们认为这些方法并不具有广泛的应用价值。

近年来,深度学习已经被用于睡眠阶段分期,深度学习利用多层线性和非线性处理单元学习来自输入数据的分层表示或特征。其中,卷积神经网络被用于短脑电时间序列数据,例如脑计算机接口[10] [11],癫痫发作检测[12],驾驶员的认知表现[13]和眼动追踪[14]以及睡眠分期[15] [16]。CNN 已经在其他领域用于原始连续信号,从图像识别开始[17] [18],到许多其他领域,如自然语言处理[19],推荐系统[20]和其他监督模式识别任务。

一般来说,神经网络模型的发展源于双重目标:第一,更好地理解神经系统;第二,尝试构建生物功能启发的信息处理系统。虽然在某些领域计算机能够比人脑执行处理的更有效,但计算机无法与大脑相提并论。现阶段大多数网络结构并没有提供任何可以被用于建模系统的信息[21],它没有考虑原始系统的物理组织。神经网络的一个优点是它表现为非线性黑箱,几乎可以建模和描述任何非线性动态。就传统统计学而言,神经网络常常被认为是不可识别的模型,在某种意义上可以获得具有不同拓扑和参数的各种网络。与传统的数据分析方法相比,它们在解决现实问题方面具有很强的竞争力。

作为深度学习的一种[17],CNN 是一种可训练的多层非线性系统,旨在从图像中提取和分类高维模式。CNN 本质上是具有特殊拓扑结构的多层感知器(Multi-Layer Perceptron, MLP),它包含多个隐藏层[10]。CNN 已经成功应用于原始 EEG 数据[22],对象识别[23]和手写字符识别[17],如分析视觉图像[24],语音识别和预测人流量等。CNN 具有共享权重架构和转换不变性的特征。CNN 由输入和输出层以及多个隐藏层组成。隐藏层可以是卷积层、池化层或全连接层。与其他传统算法相比,CNN 使用相对较少的预处理。这意味着 CNN 网络可以学习传统算法中人工设计的滤波器。其主要优点便是可以代替先验知识和人工设计特征工程。

循环神经网络(Recurrent Neural Networks, RNN)让神经网络有了记忆,对于序列化的数据,循环神经网络能达到更好的效果。由于 RNN 容易出现梯度爆炸,不易训练的缺点,RNN 发展出了许多变种,其中长短记忆网络(Long Short-Term Memory, LSTM)网络是一种循环神经网络特殊的类型,可以学习长期依赖信息。LSTM 由 Hochreiter & Schmidhuber (1997)提出,并在近期被 Alex Graves 进行了改良和推广。

在[25]中, Längkvist 等人提出了第一个基于深度学习的睡眠分期系统;在[26]中, Ronan Collobert 提出了一种端到端深度学习方法,它采用多变量睡眠信号(即 EEG, EOG 和 EMG)进行时间睡眠阶段分期。随着可穿戴设备的快速发展和 EEG 信号采集设备的发展,例如, Supratak 等[16]提出 DeepSleepNet,该模型基于原始睡眠 EEG 信号,包含两个不同的 CNN 来提取时不变特征,一个双向长短记忆(Long Short-Term Memory, LSTM)用于序列残差学习[27]。然而这些工作都忽略了睡眠专家的分期过程,睡眠专家是通过经验和 AASM 规则从整体上观测多个通道 EEG 信号进行睡眠阶段分期。

在本文中,我们介绍了一种使用深度监督卷积神经网络(CNN)对原始信号样本进行的睡眠分期的方法。实验目标是利用深度学习的特征提取功能代替手动特征提取工作。这项工作的主要贡献如下:1) 构建了一个基于卷积神经网络的睡眠评分系统;2) 网络端到端训练,并在原始脑电图上学习特征检测;3) 系统在大型数据集上进行评估,保证了训练的模型能够得到良好的推广应用。

2. 数据与方法

此次实验所用的数据是通过 Alice Sleepware G3 软件添加 Alice5 设备进行采集的。Alice Sleepware G3 软件是飞利浦公司专门为 Alice 系列多导睡眠检测仪设计的软件。Alice Sleepware G3 软件可以直接对采

集的各种指标数据进行展示,之后对采集到的数据每30 s应用一次的睡眠自动分期。然后专家依据 Alice Sleepware G3 编辑界面观察各个通道的生理信号,对错误的分期进行排查,并对分错的进行手动标记矫正,分期结果以专家手工校正为准,最终得到睡眠结构图。

2.1. 数据及预处理

本文采用的数据集来自于福建省某睡眠中心医院以及公开的 Physionet Sleep-EDFx 数据库[28],睡眠中心医院使用飞利浦 Alice 5 多导睡眠监测仪来进行实验。实验前对来睡眠监测中心的患者进行简单编号并要求每一名患者填写问卷调查。问卷主要内容包括患者身高、体重、颈围、血压和心率以及医嘱现病史、既往病史、家族史、用药情况等。数据是通过 Alice Sleepware G3 软件添加 Alice5 设备进行采集。随机挑选 114 名患者的数据,这些患者没有服用任何睡眠相关的药物。我们对这些患者使用 Alice5 设备以 200 Hz 采样率进行一整夜的睡眠采样。经统计患者的睡眠时间为 8~10 小时。需要说明的是,福建省某睡眠中心医院所有的数据都采集自有睡眠障碍的患者。

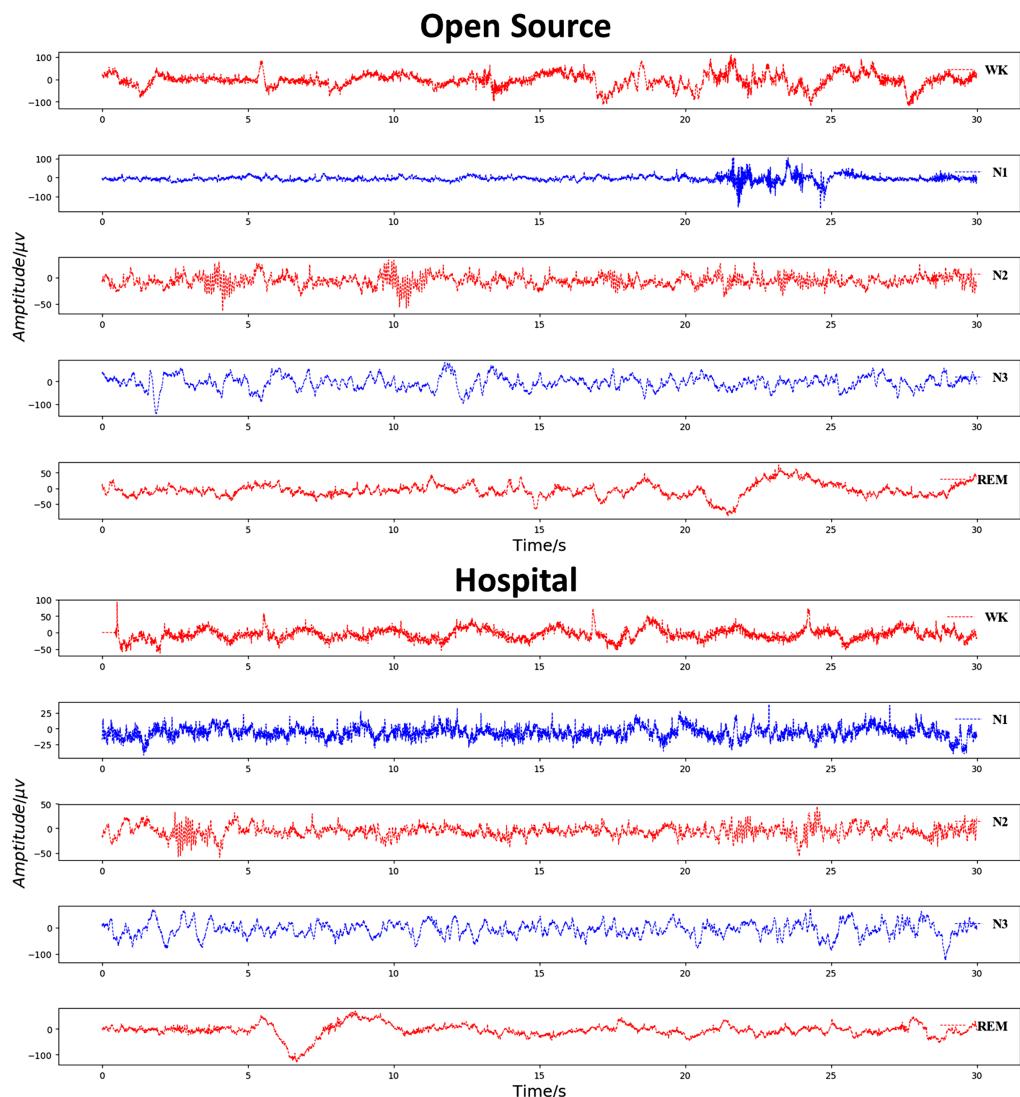


Figure 1. The characteristics of EEG waveform example
图 1. EEG 波形特征示例

图 1 展示了本次实验使用的两种不同期 EEG 波形示例，分别为公开数据以及睡眠中心医院数据集。根据 Amaud Sors 的研究表明，“可用的 EEG 通道是对称的，因此它们具有相当的性能”[29]。然而这个工作忽略了睡眠专家的分期过程，睡眠专家通过经验和 AASM 规则，从整体上观测多个通道 EEG 信号进行睡眠阶段分期。因此，在下文中也采用 EEG-F3-A2, EEG-C3-A2, EEG-O1-A2 通道，以及眼电信号采用 EOG-left 与 EOG-Right 等多个通道来执行自动睡眠阶段分期任务。每个受试者的整个 EEG 信号被分成 30 秒的样本数据(每个样本包含 6000 个数据点)。每个时期被标记为 W, N1, N2, N3, REM 和未分期(未记录的时期在每次记录的开始或结束时)，由睡眠专家进行分期。在本研究中，占比较少的 REM 和未分期的时期已经被删除。

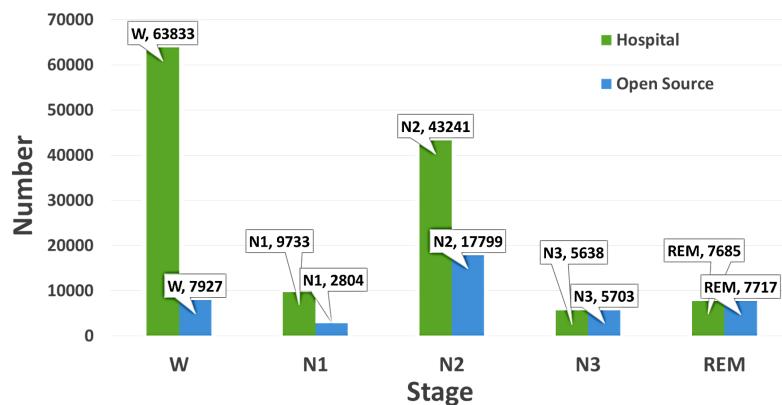


Figure 2. The number of sleep epochs

图 2. 睡眠时期的数量

某些特殊的睡眠样本例如含有 12 个小时的清醒期样本以及睡眠时长达到 25 小时的样本等，没有被包含到实验数据中，同时，通过专业睡眠专家的筛选，极少数患者被排除在外，因为它们可能是异常值。从图 2 可以明显看出不同期的数据分布不平衡，因此，我们对 N1 期的数据做了数据增强，提取全部 N1 期数据拼接成总长度为 81 h 的数据，然后去除数据前 10 s，再重新以 30 s 逐一切割，从而减少数据不平衡带来的误差。因为这样错位 10 s 的数据，既可以实现数据增强，又可以实现数据的多样性。从以上工作来看没有对 EEG 信号本身进行额外的预处理。

2.2. 网络结构

对于 CNN 架构的选择尝试了两种结构，一种是现有的主流 EEG 分期模型，这类模型是使用单通道 EEG 数据进行分期的结构；另外一种是本文提出的新分期模型，使用多通道 EEG 与 EOG 进行分期，下文将主要以此模型的结果进行分析和讨论。

2.2.1. 基于单通道 EEG 的模型 A

首先介绍基于单通道 EEG 的模型结构：这一部分，我们构建了基于卷积神经网络与循环神经网络的自动睡眠阶段分期模型 A。

为了更好的模拟睡眠专家的分期过程，CNN 的输入包括分类的时期未处理的 EEG 信号与 EOG 信号，以一个数字矩阵作为输入。模型训练好之后，如果当前要预测的样本信号为不确定的情况下，模型中表现为该期的分期概率值小于等于 0.25，偶尔会引用下一个和前一个样本的分期结果。

图 3 给出了该网络结构的基本视图，展示了基于 EEG 信号的卷积神经网络的自动睡眠阶段分期模型。该模型的输入是一个 30 s 的采样率为 200 Hz 的 EEG 信号，每个信号包含 30×200 个数据，表示为 1×6000

的矩阵。其中 1 代表采用 1 个通道的数据，6000 代表一个通道的数据量。实验采用 F3-A2 通道的 EEG，每个 EEG 样本信号按顺序由输入层，Cov1, dropout, Cov2, Pool1, Cov3, dropout, Pool2, FC1, dropout, FC2, dropout 和输出也即属于每个睡眠期的概率组成。需要说明的是，为了加快网络的训练，测试阶段采用 8 个人的样本数据。当网络准确率达到 75% 时，找到了合适的参数区间后采用 39 人数据样本，总的期数达到 41,000。

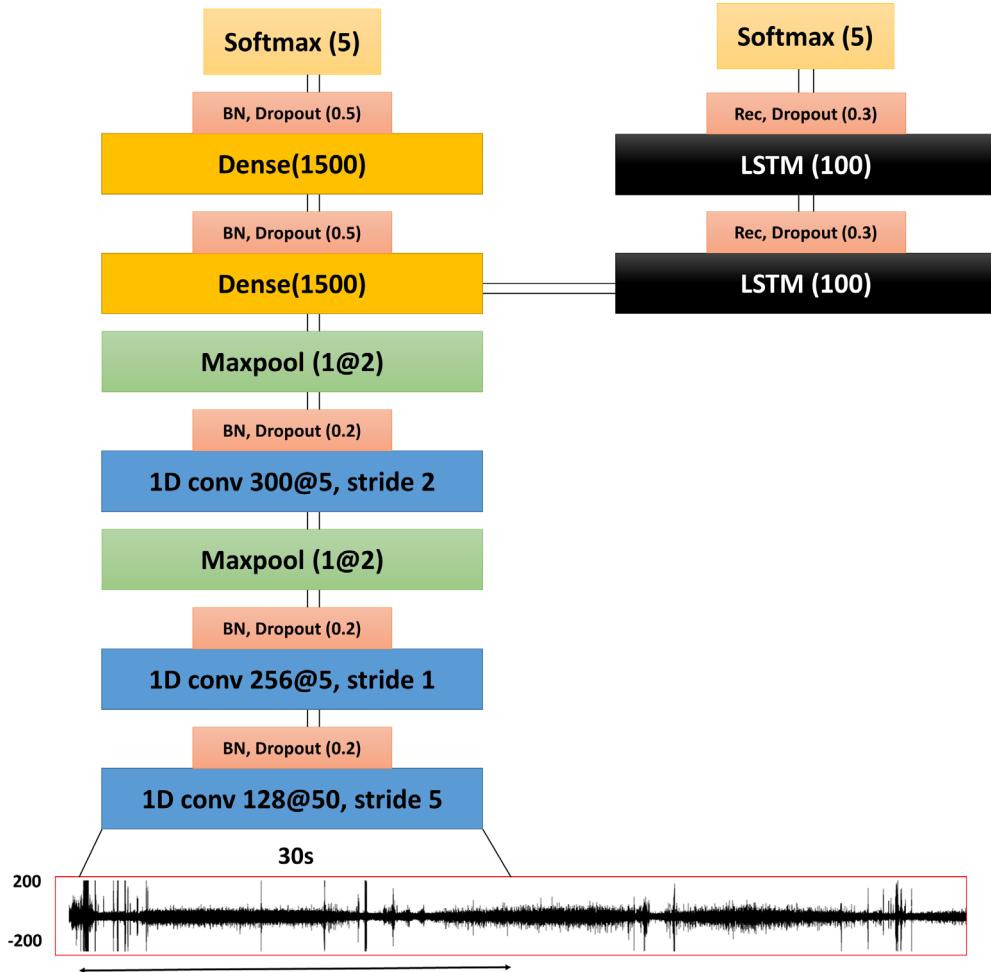


Figure 3. Architecture of Model-B based on single channel dataset

图 3. 基于单通道 EEG 模型 B 的网络结构

2.2.2. 基于多通道 EEG 与 EOG 的模型 B

以下是基于多通道 EEG 与 EOG 的模型结构：在这一部分中，我们构建了基于卷积神经网络的自动睡眠阶段分期模型 B。

模型 B 采用 4 个二维卷积层，激活函数采用线性整流函数(Rectified Linear Unit, ReLU)，每个二维卷积层紧接着采用了 dropout 避免过拟合的方法，都以 20% 的几率使得神经元会被关闭或丢弃。每个卷积层后都接一个最大池化层。紧接着是两个全连接层，第一个全连接层有 4096 个单元，另一个全连接层有 1500 个单元，这两个全连接层也采用了 dropout 避免过拟合的方法，都以 50% 的几率使得神经元会被关闭/丢弃。最后是一个大小为 5 的全连接层，激活函数采用 Relu，卷积层的激活函数都用的是 Relu，负斜率为 0。

图 4 展示了基于 EEG 信号的卷积神经网络的自动睡眠阶段分期模型，网络的输入为一个 30 s 的采样率为 200 Hz 的 EEG 信号，它以 5×6000 的矩阵形式表示。其中 5 代表采用了 5 个通道的数据，模型 B 采用 F3-A2, C3-A2, O1-A2 通道的 EEG 以及 Left EOG, Right EOG，每个 EEG 样本信号按顺序由输入层，Cov, dropout, Pool 交替搭建，直到 FC1 与 FC2 和输出。测试阶段采用 8 个人的样本数据，当网络准确率达到 75% 时，找到了合适的参数区间采用 114 人数据样本，图 2 显示了各个期的数量分布。

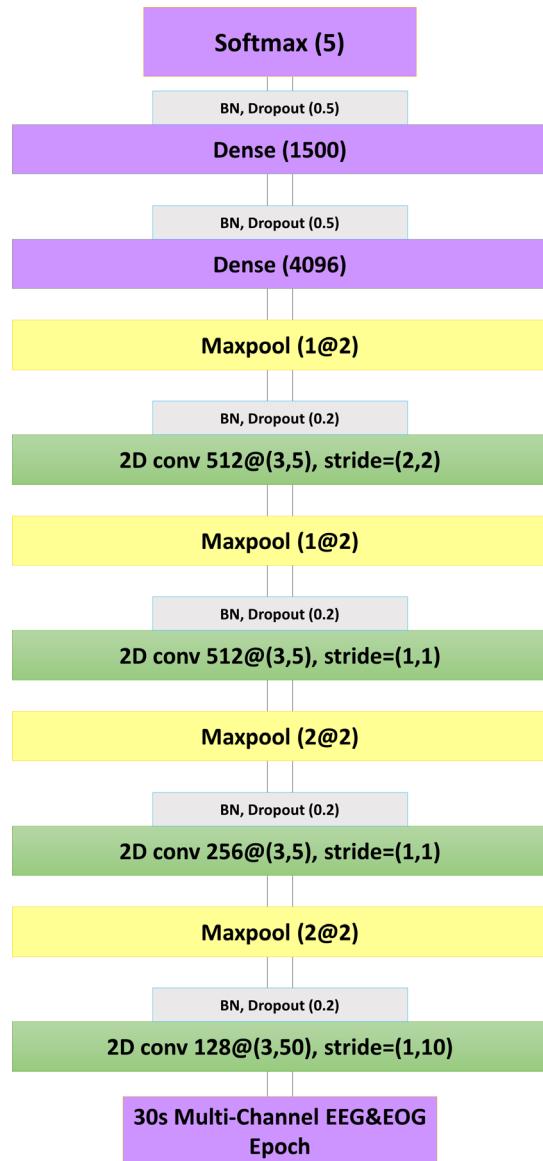


Figure 4. Architecture of Model-B based on multi-channel dataset
 图 4. 基于多通道数据模型 B 的体系结构

2.3. 模型评估与优化

本文的前半部分，详细介绍了本研究中使用的数据结构和模型设计。表 1 显示了本研究中使用的样本数。为了评估模型的性能，实验使用了 K-fold 交叉验证方法[30]，K 设置为 5，也即 5 折交叉验证。我们将所有类型的睡眠期 EEG 信号以及 EOG 数据划分为测试集和训练集。在本文中，K 被设置为 5，具

体来说，随机选择所有数据集的 20%作为测试集，其余作为每个训练期间的训练集。一般使用反向传播算法就可以直接得到梯度，但是对于模型 B，此优化方法结果并不是很好。为了优化，用了 Adam [31] 优化方法。模型 B 还采用了随机梯度下降 Stochastic Gradient Descent (SGD)，SGD 因为更新比较频繁，会造成损失函数有剧烈的震荡，最终停留在局部最小或者 Saddle Point，也即鞍点处。所以模型 B 采用 Adamwith Momentum 梯度更新方法，引入 moment 可以避免落入局部最优解最终可以加快收敛的同时减小震荡。

在固定长度的时间序列上使用 CNN 时，卷积部分的输出大小直接与输入大小、卷积层数和它们的步幅有关。如果最后一个卷积层的输出太大，则绝大多数权重将位于全连接层中。我们试验了 6~12 层，步长为 2~10，还尝试了 7 号，5 号和 3 号大小的卷积核，最终选择了 3 号和 5 号卷积核，尽管 5 到 7 之间的性能差异很小。经过各种功能配置的测试，模型 B 保留了前四层的 128 个特征图和 256 个特征图，以及最后两层卷积的 512 个特征图。最后，在网络架构中，使用 3 号和 5 号卷积核组合是一个很好的折衷。除此之外，还采用批训练的方法来加快训练过程。实验过程中尝试了 16、64、128、256 批次的训练。结果显示，当批次为 128 时效果最好。

最后要说明的是，本实验的代码使用 python (Python 3.6.6)语言和 Pytorch (Pytorch 0.4.1)深度学习框架。为了加速模型的训练，模型在 Pytorch 中实施，并在 Nvidia GTX1060 的 GPU 上进行模拟训练。

Table 1. The specific params of our CNN model based on multi-channel data
表 1. 基于多通道数据的卷积神经网络模型参数

层级	层类	卷积核数量	激活函数	大小	Dropout 率	步长	边缘填充
Input	EEG & EOG matrix			(1, 5, 6000)			
Cov1	Convolutional	128	Relu	(3, 50)	Dropout-0.2	(1, 10)	(1, 1)
Pool1	Max-Pooling			(2, 2)		(2, 2)	(0, 0)
Cov2	Convolutional	256	Relu	(3, 5)	Dropout-0.2	(1, 1)	(1, 1)
Pool2	Max-Pooling			(2, 2)		(2, 2)	(0, 0)
Cov3	Convolutional	512	Relu	(3, 5)	Dropout-0.2	(1, 1)	(1, 1)
Pool3	Max-Pooling			(1, 2)		(1, 2)	(0, 0)
Cov4	Convolutional	512	Relu	(3, 5)	Dropout-0.2	(2, 2)	(1, 1)
Pool4	Max-Pooling			(1, 2)		(1, 2)	(0, 0)
Fc1	Fully-Connected		Relu	(4096)	Dropout-0.5		
Fc2	Fully-Connected		Relu	(1500)	Dropout-0.5		
Output	Sleep Stage						

3. 结果

我们使用召回(Recall, REC)，准确率(Precision, PRE)和 F1 值(F1-Score, F1)以及总体准确度(accuracy, ACC)来评估模型 A 与模型 B 的分类表现。其中准确率 PRE 表示被模型预测为正例的样本中实际也是正例的比例；召回率 REC，又称查全率，是一种覆盖面的衡量标准，表示样本中的正例有多少被预测正确。其中指数 MF1 指的是每个睡眠期的所有 F1 的平均值。ACC 表示正确数量的睡眠阶段时期分类与所有睡眠阶段时期的比例。REC, PRE, F1 值, ACC 与 MF1 的计算方法分别如下：

$$PRE = \frac{TP}{TP + FP}, REC = \frac{TP}{TP + FN} \quad (1)$$

$$F1 = 2 \frac{PRE * REC}{PRE + REC} \quad (2)$$

$$ACC = \frac{\sum_{c=1}^C TP_c}{N} \quad (3)$$

$$MF_1 = \frac{\sum_{c=1}^C F1_c}{C} \quad (4)$$

其中方程 3.3 中的 TP_c 是每个类中被预测为正确的数量, $F1_c$ 是某一类的 $F1$ 分数, C 是睡眠分期的数量, N 是所有的 30 秒样本数量。 TP 是真阳性, FN 是假阴性和 FP 是假阳性。

3.1. 基于单通道 EEG 的模型结果

在单通道模型中, 时长为 30 秒的 EEG 信号样本总数是 41,000, 实验使用 5 折交叉验证方法来评估模型的性能。因此, 取出总样本的 20% 作为测试集, 这些样本没有参与模型的训练, 只作为测试集。在下面的内容中, 表 2 显示模型在每个实验中的测试集的平均性能。

表 2 显示了使用模型 A 从睡眠期的单通道 EEG 信号的实验获得的混淆矩阵。该实验的总体准确度 ACC 为 79.92%。在表 2 中, 每行和每列分别表示由睡眠专家判断并通过模型 A 预测的 30 秒 EEG 信号和 EOG 信号的睡眠分期的数量。

Table 2. Confusion matrix obtained from experiment on single channel EEG signals using our model

表 2. 使用基于单通道 EEG 的模型获得的混淆矩阵

	W	N1	N2	N3	REM	PRE	REC	F1
W	6601	728	373	74	391	0.871	0.808	0.838
N1	292	1399	829	27	539	0.454	0.453	0.454
N2	311	498	13362	1353	676	0.853	0.825	0.839
N3	28	8	490	5362	202	0.786	0.880	0.831
REM	350	447	609	6	6045	0.770	0.811	0.790
ACC = 79.92%, MF1 = 74.99%								

为了评估模型 A 的性能, 模型结果与几种现有方法的结果进行比较。需要说明的是这些方法都是采用 Fpz-Cz 通道 EEG 数据。我们应用这些现有方法, 因为它们的结果也基于模型 A 在本研究中使用的数据集, 即福建省某睡眠中心医院数据集, 表 3 列出了比较结果。根据表 3, 对于相同数据集, 模型 A 达到了最佳性能参考的准确性。

Table 3. Accuracy comparison matrix with other method

表 3. 与其他方法的精度比较矩阵

模型	作者	方法	基于公开数据集 PhysioNet 准确率	基于福建省某医院睡 眠中心数据准确率
模型 1	Akara Supratak (2017) [16]	CNN and LSTM	81.50%	70.52%
模型 2	Liangjie We (2017) [32]	Time-Frequency CNN	82.86%	72.60%
模型 A	Wang and Xu et al.	Based the multi-channels EEG and CNN	79.23%	79.92%

模型 A 比其他的单通道模型具有更好的性能，尽管它们都具有良好的性能，图 5 显示出了由睡眠专家分期的其中一位受试者一整夜的多导睡眠图的比较，即睡眠专家分期结果与模型 A 的结果对比。可以看出模型 A 是最优的，相对来说，对于福建省某睡眠中心医院数据集，模型 2 的分期结果是最差的。

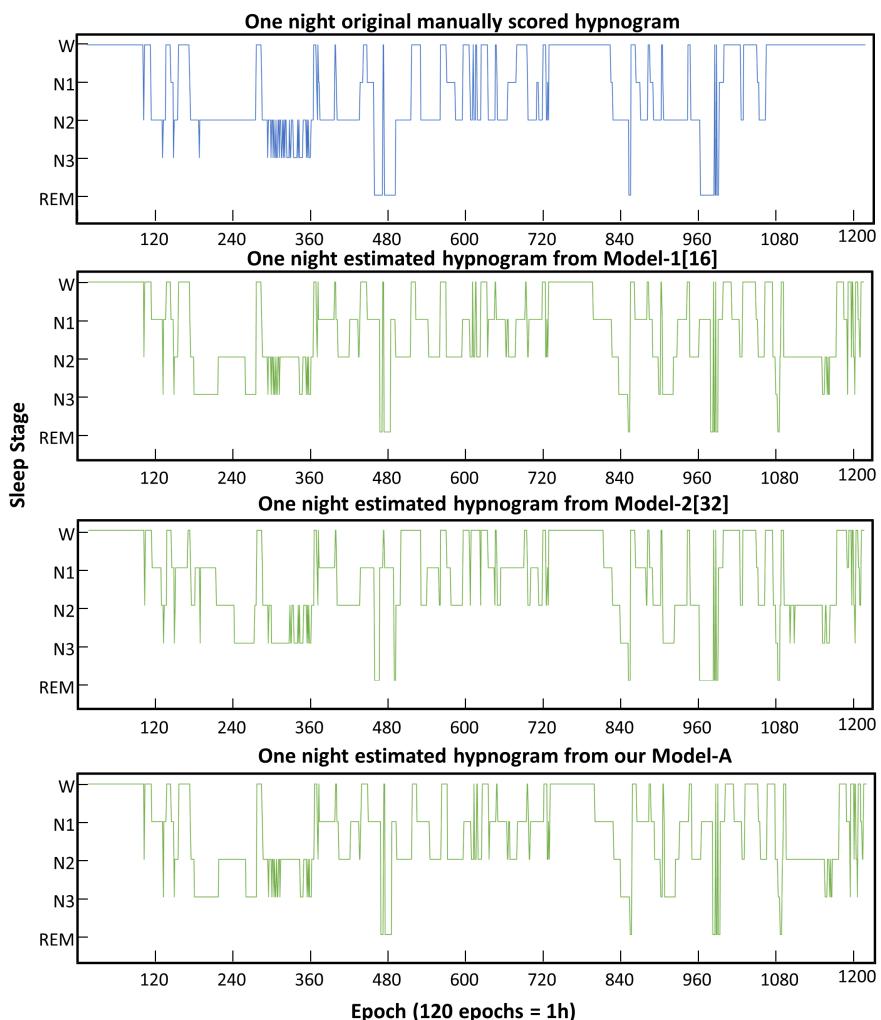


Figure 5. The comparison of hypnogram scored by sleep experts, and our Model-A, and other models
图 5. 睡眠专家分期的睡眠结构图和我们提出的模型 A 的结果比较

从表 3 来看，基于公开数据集模型 1 与模型 2 具有很好的性能，而当我们应用这两个模型测试于福建省某医院睡眠中心数据集时，发现模型 A 优于这两个模型，说明这些模型的泛化能力还是不足，依旧有很大的提升空间。反观模型 A 应用于福建省某医院睡眠中心数据集时却有更好的表现，相对来说有很强泛化能力。分析原因不乏有模型 1 与模型 2 本身是以健康人群的数据为训练数据，而模型 A 是以福建省某医院睡眠中心数据为训练数据，是采集自有睡眠障碍患者的数据集，具有很大的多样性，直观上可以从图 5 看出。接下来我们将主要对多通道模型 B 的结果进行分析与讨论。

3.2. 基于多通道 EEG 与 EOG 的模型结果

由图 2 可见，睡眠期 30 秒 EEG 信号样本的总数是 130130，实验使用 5 折交叉验证来评估模型的性能。因此，取出总样本的 20% 作为测试集，这些样本没有参与模型 B 的训练，只作为测试集。下面的内

容将显示模型 B 在每个实验中的测试集的平均性能。表 4 显示了使用模型 B 在多通道 EEG 和 EOG 信号的实验获得的混淆矩阵。总体上来看，该实验的总体准确度 ACC 为 81.93%。

Table 4. Confusion matrix obtained from experiment on multi-channel EEG signals using our proposed model
表 4. 使用基于多通道信号的 CNN 模型获得的混淆矩阵

	W	N1	N2	N3	REM	PRE	REC	F1
W	58875	2562	508	68	121	0.922	0.948	0.935
N1	2963	3962	3602	39	772	0.407	0.349	0.376
N2	935	1655	33600	1204	852	0.777	0.879	0.852
N3	95	51	2646	4299	50	0.763	0.602	0.673
REM	965	1503	2885	28	5863	0.766	0.521	0.620
ACC = 81.93%; MF1 = 68.60%								

正如表 4 所示，模型 B 的 ACC 达到了更高的值，表明模型 B 具有良好的性能。实验中最显著的结果是 W, N2 和 N3 的睡眠期具有优异的分类性能。例如，W 期分类的准确率接近 92%，N2 和 N3 期接近 77%。模型 B 在期 REM 上的表现略差，但其 REC 仍然超过 76%。可以看出，N1 期具有最差的分类性能，因为 N1 期主要被错误地分类为 W 期和 N2 期。从各个期的特征波来看，原因可能是背景 EEG 波与三个睡眠期非常相似。另一个原因可能是缺乏 N1 期的样本，平均每一个人的整夜睡眠中，仅有大约 70 个 30 秒的 EEG 信号属于 N1 期，五个期的分布不均可能导致使用 CNN 模型的 N1 期的分类结果差。因此，我们对 N1 期的数据做了数据增强，提取全部 N1 期数据拼接成总长度为 81 h 的数据，然后去除数据前 10 s，再重新以 30 s 逐一切割，从而减少数据不平衡带来的误差。因为这样错位 10 s 的数据，既可以实现数据增强，又可以实现数据的多样性。从以上工作来看，没有对 EEG 信号本身进行额外的预处理。

Table 5. Accuracy comparison matrix with other method

表 5. 与其他方法的精度比较矩阵

模型	作者	方法	基于公开数据集 PhysioNet 准确率	基于福建省某医院睡 眠中心数据准确率
模型-3	Fraiwan. <i>et al.</i> (2015) [33]	Based on time-frequency analysis of and random forest classifier	79.23%	70.45%
模型-4	Orestis Tsinalis, <i>et al.</i> (2016) [34]	Using Convolutional Neural Networks	74.00%	63.850%
模型-B	Wang and Xu <i>et al.</i>	Based the multi-channels EEG and CNN	80.05%	81.93%

为了评估模型 B 的性能，我们与几种现有方法进行了比较。因为它们的结果也基于模型 B 在本研究中使用的数据集，即福建省某医院睡眠中心的数据集。表 5 列出了比较结果，由表 5 可见，模型 B 达到了最佳性能，并且有着优于模型 3 与模型 4 的分期准确率。可以将这两个模型应用于更多可用数据集，并将结果与这些模型进行比较，以显示模型 B 的泛化能力。从结果中可以看出，我们的多通道 EEG 和 EOG 模型在福建省某医院睡眠中心数据集是优于现有的多通道模型方法。由于模型使用的数据是有睡眠障碍患者的数据，这说明模型 B 泛化能力显然高于公开数据集的模型。

虽然模型 3 与模型 4 都具有良好的性能，模型 B 比模型 3 与模型 4 具有更好的性能，图 6 显示出了由睡眠专家分期的代表性样本的一整夜的多导睡眠图 PSG 的比较，即睡眠专家分期结果与模型 B 的结果

对比。从不同模型的分期结果来看，模型 B 有着显著的优势。模型 B 与原标签基本一致，错误的部分大多数为清醒期与 N1 期，N3 期与 N4 期的错误分类其次。

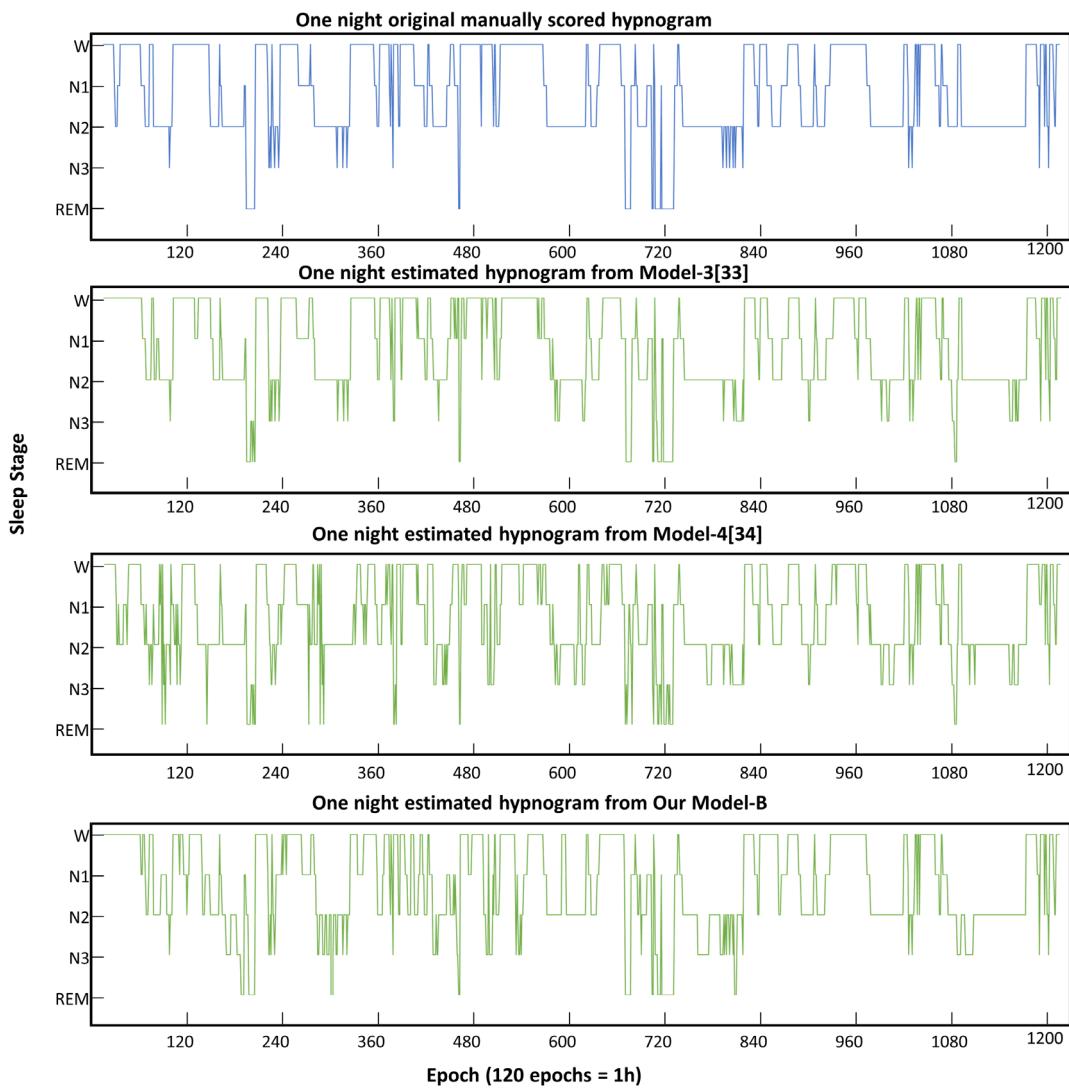


Figure 6. The comparison of hypnogram scored by sleep experts, and our Model-B and other models
图 6. 睡眠专家分期的睡眠结构图和我们提出的模型 B 以及文献方法结果比较

4. 结论

这项研究表明，使用多通道 EEG 和 EOG，使用卷积神经网络对睡眠阶段分期进行分类是可行的，在自动睡眠分期的研究中不应该忽略 EOG 数据，我们的系统性能与中级睡眠人类分期专家的性能结果相当。训练是端到端的，无需任何专业知识选择或任何信号预处理是一个优势。因为神经网络可以自主学习各个期的特征，最适合分类任务的功能。

从卷积网络的特征来看，卷积层已经能够学习出优秀的过滤器。另一个优点是该方法更易于适应另一种应用或者类似的医学序列数据。关于睡眠专家产生的误差，我们注意到误差主要对应于睡眠周期中连续的期。例如，N3 期最常与 N2 期混淆，几乎从不与 N1 混淆。类似的，N1 期虽然被称为人类医师分歧最小的期，但它们会被误分为 Wake，N2 或 REM，因为这些期都可以包含类似于 N1 期的特征。

作为睡眠阶段分期的数据集, 从图 2 可以看出我们的数据集的类别分布是不平衡的, 需要进一步研究以解决类别不平衡问题。集成学习[35]或人工平衡[36]是比较合适的。

表 3 与表 5 分别显示了近期单通道和多通道睡眠图睡眠分期研究的一些特征和性能指标。睡眠分期的研究是具有挑战性的, 因为它们并非都使用相同的数据库、样本数量和分期规则, 并且它们并非都以相同的方式平衡类别。例如, Physionet Sleep-EDFx 数据库[28]有比任何其他睡眠阶段更多的清醒期, 因为此数据库保留了夜晚前后数小时的清醒期的记录。

在本文中, 我们提出了一种基于深度卷积神经网络的多通道 EEG 信号的自动睡眠分期的方法。将 CNN 应用于脑电信号以及眼电信号的分类, 可以把 EEG 时间序列信号和 EOG 时间序列信号转换为有意义的数据矩阵, 类似于图的 CNN 可以处理的格式。睡眠分期结果表明, 我们提出的方法比其他现有方法获得更好的睡眠阶段分期性能, 并且对不同类型人群的 30 秒 EEG 信号和 EOG 信号睡眠分期数据具有更好的适应性, 避免了设计和提取分类器过程的不准确性和复杂性, 可用于自动睡眠阶段分期的 EEG 信号的特征。

随着睡眠数据的累积, 我们的模型性能将得到进一步提升。应该指出的是, 人类专家通常不会仅使用一个通道进行睡眠分期。例如, 在 AASM 指南中推荐至少三个通道, 并且通常是非 EEG 标记, 例如 EOG、EMG 或运动, 其帮助神经生理学家区分 N1 期或 REM 期。在现存的大多数的例子中, 虽然使用单个通道对轻便携式设备很有意义, 但它同时限制了模型性能[26]。未来将继续改进基于 CNN 的自动睡眠阶段分期方法, 并将进一步与 RNN 结合, 实现对同一个样本不同期之间的时序信息的学习。

此外, 我们的方法还为分析其他非线性和非平稳时间序列数据提供了一个思路方法。此模型原则上适用于使用 CNN 分类器的其他医学领域, 我们将探索使用其他医疗时序数据, 来处理相应的疾病检测和健康分析问题, 例如用于肺音分类[37]和心音分类[38]。

基金项目

国家自然科学基金资助项目(批准号: 11874310 和 11675134), 国家 111 项目(批准号: b16029)。

参考文献

- [1] Wulff, K., Gatti, S., Wettstein, J.G. and Foster, R.G. (2010) Sleep and Circadian Rhythm Disruption in Psychiatric and Neurodegenerative Disease. *Nature Reviews Neuroscience*, **11**, 589-599. <https://doi.org/10.1038/nrn2868>
- [2] Berry, R.B., Brooks, R., Gamaldo, C.E., Harding, S.M., Marcus, C.L. and Vaughn, B.V. (2012) The AASM Manual for the Scoring of Sleep and Associated Events, Rules, Terminology and Technical Specifications. American Academy of Sleep Medicine, Darien, IL, 176.
- [3] Stepnowsky, C., Levendowski, D., Popovic, D., Ayappa, I. and Rapoport, D.M. (2013) Scoring Accuracy of Automated Sleep Staging from a Bipolar Electroocular Recording Compared to Manual Scoring by Multiple Raters. *Sleep Medicine*, **14**, 1199-1207. <https://doi.org/10.1016/j.sleep.2013.04.022>
- [4] Wang, Y., Loparo, K.A., Kelly, M.R. and Kaplan, R.F. (2015) Evaluation of an Automated Single-Channel Sleep Staging Algorithm. *Nature and Science of Sleep*, **7**, 101-111. <https://doi.org/10.2147/NSS.S77888>
- [5] Huang, C.-S., Lin, C.-L., Ko, L.-W., Liu, S.-Y., Su, T.-P. and Lin, C.-T. (2014) Knowledge-Based Identification of Sleep Stages Based on Two Forehead Electroencephalogram Channels. *Frontiers in Neuroscience*, **8**, 263. <https://doi.org/10.3389/fnins.2014.00263>
- [6] Güneş, S., Polat, K. and Yosunkaya, Ş. (2010) Efficient Sleep Stage Recognition System Based on EEG Signal Using k -Means Clustering Based Feature Weighting. *Expert Systems with Applications*, **37**, 7922-7928. <https://doi.org/10.1016/j.eswa.2010.04.043>
- [7] Tsinalis, O., Matthews, P.M. and Guo, Y. (2016) Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders. *Annals of Biomedical Engineering*, **44**, 1587-1597. <https://doi.org/10.1007/s10439-015-1444-y>
- [8] Sharma, R., Pachori, R.B. and Upadhyay, A. (2017) Automatic Sleep Stages Classification Based on Iterative Filtering

- of Electroencephalogram Signals. *Neural Computing and Applications*, **28**, 2959-2978. <https://doi.org/10.1007/s00521-017-2919-6>
- [9] Hassan, A.R. and Subasi, A. (2017) A Decision Support System for Automated Identification of Sleep Stages from Single-Channel EEG Signals. *Knowledge-Based Systems*, **128**, 115-124. <https://doi.org/10.1016/j.knosys.2017.05.005>
- [10] Cecotti, H. and Graser, A. (2011) Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 433-445. <https://doi.org/10.1109/TPAMI.2010.125>
- [11] Manor, R. and Geva, A.B. (2015) Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI. *Frontiers in Computational Neuroscience*, **9**, 146. <https://doi.org/10.3389/fncom.2015.00146>
- [12] Page, A., Shea, C. and Mohsenin, T. (2016) Wearable Seizure Detection Using Convolutional Neural Networks with Transfer Learning. 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, 22-25 May 2016, 1086-1089. <https://doi.org/10.1109/ISCAS.2016.7527433>
- [13] Hajinoroozi, M., Mao, Z. and Huang, Y. (2015) Prediction of Driver's Drowsy and Alert States from EEG Signals with Deep Learning. 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Cancun, Mexico, 13-16 December 2015, 493-496. <https://doi.org/10.1109/CAMSAP.2015.7383844>
- [14] Drouin-Picard, A. and Falk, T.H. (2016) Using Deep Neural Networks for Natural Saccade Classification from Electroencephalograms. 2016 IEEE EMBS International Student Conference (ISC), Ottawa, ON, 29-31 May 2016, 1-4. <https://doi.org/10.1109/EMBSISC.2016.7508606>
- [15] Manzano, M., Guillén, A., Rojas, I. and Herrera, L.J. (2017) Combination of EEG Data Time and Frequency Representations in Deep Networks for Sleep Stage Classification. In: Huang, D.S., Jo, K.H., Figueroa-García, J., Eds., *Intelligent Computing Theories and Application (ICIC 2017)*, Lecture Notes in Computer Science, vol 10362, Springer, Cham, 219-229. https://doi.org/10.1007/978-3-319-63312-1_20
- [16] Supratak, A., Dong, H., Wu, C. and Guo, Y. (2017) DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **25**, 1998-2008. <https://doi.org/10.1109/TNSRE.2017.2721116>
- [17] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [18] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. In: Pereira, F., Burges, C.J.C., Bottou, L. and Weinberger, K.Q., Eds., *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Curran Associates Inc., New York, 1097-1105.
- [19] Collobert, R. and Weston, J. (2008) A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In: *Proceedings of the 25th International Conference on Machine Learning*, ACM, New York, 160-167. <https://doi.org/10.1145/1390156.1390177>
- [20] Van den Oord, A., Dieleman, S. and Schrauwen, B. (2013) Deep Content-Based Music Recommendation. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z. and Weinberger, K.Q., Eds., *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Curran Associates Inc., New York, 2643-2651.
- [21] Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Gorenne, P., Hjalmarsson, H. and Juditsky, A. (1995) Nonlinear Black-Box Modeling in System Identification: A Unified Overview. *Automatica*, **31**, 1691-1724. [https://doi.org/10.1016/0005-1098\(95\)00120-8](https://doi.org/10.1016/0005-1098(95)00120-8)
- [22] Mirowski, P.W., Madhavan, D. and LeCun, Y. (2007) Time-Delay Neural Networks and Independent Component Analysis for EEG-Based Prediction of Epileptic Seizures Propagation. *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*, Vancouver, 22-26 July 2007, 1892-1893.
- [23] LeCun, Y., Huang, F.J. and Bottou, L. (2004) Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington DC, 27 June-2 July 2004, 97-104.
- [24] LeCun, Y., Kavukcuoglu, K. and Farabet, C. (2010) Convolutional Networks and Applications in Vision. *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, Paris, 30 May-2 June 2010, 253-256. <https://doi.org/10.1109/ISCAS.2010.5537907>
- [25] Längkvist, M., Karlsson, L. and Loutfi, A. (2012) Sleep Stage Classification Using Unsupervised Feature Learning. *Advances in Artificial Neural Systems*, **2012**, Article ID: 107046. <https://doi.org/10.1155/2012/107046>
- [26] Chambon, S., Galtier, M.N., Arnal, P.J., Wainrib, G. and Gramfort, A. (2018) A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**, 758-769. <https://doi.org/10.1109/TNSRE.2018.2813138>
- [27] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE*

- Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 27-30 June 2016, 770-778.
<https://doi.org/10.1109/CVPR.2016.90>
- [28] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C. and Stanley, H.E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, **101**, e215-e220. <https://doi.org/10.1161/01.CIR.101.23.e215>
- [29] Sors, A., Bonnet, S., Mirek, S., Vercueil, L. and Payen, J.-F. (2018) A Convolutional Neural Network for Sleep Stage Scoring from Raw Single-Channel EEG. *Biomedical Signal Processing and Control*, **42**, 107-114. <https://doi.org/10.1016/j.bspc.2017.12.001>
- [30] Jung, Y. (2018) Multiple Predicting k -Fold Cross-Validation for Model Selection. *Journal of Nonparametric Statistics*, **30**, 197-215. <https://doi.org/10.1080/10485252.2017.1404598>
- [31] Kingma, D.P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, San Diego, CA, 7-9 May 2015, 1-13.
- [32] Wei, L., Lin, Y., Wang, J. and Ma, Y. (2017) Time-Frequency Convolutional Neural Network for Automatic Sleep Stage Classification Based on Single-Channel EEG. 2017 IEEE 29th International Conference on Tools with Artificial Intelligence, Boston, MA, 6-8 November 2017, 88-95. <https://doi.org/10.1109/ICTAI.2017.00025>
- [33] Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H. and Dickhaus, H. (2012) Automated Sleep Stage Identification System Based on Time-Frequency Analysis of a Single EEG Channel and Random Forest Classifier. *Computer Methods and Programs in Biomedicine*, **108**, 10-19. <https://doi.org/10.1016/j.cmpb.2011.11.005>
- [34] Tsinalis, O., Matthews, P.M., Guo, Y. and Zafeiriou, S. (2016) Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks. *Machine Learning*, arXiv: 1610.01683.
- [35] Dietterich, T.G. (2000) Ensemble Methods in Machine Learning. In: *Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 1-15. https://doi.org/10.1007/3-540-45014-9_1
- [36] Huang, C., Li, Y., Change Loy, C. and Tang, X. (2016) Learning Deep Representation for Imbalanced Classification. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 27-30 June 2016, 5375-5384. <https://doi.org/10.1109/CVPR.2016.580>
- [37] Cugell, D.W. (1985) Lung Sounds: Classification and Controversies. *Seminars in Respiratory and Critical Care Medicine*, **6**, 180-182. <https://doi.org/10.1055/s-2007-1011495>
- [38] Olmez, T. and Dokur, Z. (2003) Classification of Heart Sounds Using an Artificial Neural Network. *Pattern Recognition Letters*, **24**, 617-629. [https://doi.org/10.1016/S0167-8655\(02\)00281-7](https://doi.org/10.1016/S0167-8655(02)00281-7)



知网检索的两种方式：

1. 打开知网首页 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
 下拉列表框选择：[ISSN]，输入期刊 ISSN: 2330-1686，即可查询
2. 打开知网首页 <http://cnki.net/>
 左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>
 期刊邮箱：biphy@hanspub.org