

基于知识蒸馏的轻量化睡眠分期模型研究

王 丹*

河北地质大学信息工程学院, 河北 石家庄

收稿日期: 2026年8月2日; 录用日期: 2026年9月1日; 发布日期: 2026年9月9日

摘 要

面向大规模睡眠模型难以在资源受限环境部署以及轻量模型特征表示能力不足的问题, 本文提出一种基于知识蒸馏的轻量化睡眠分期方法, 实现大规模睡眠模型向轻量网络的知识迁移。该方法以SleepGPT作为教师模型, 将教师模型输出的软标签知识迁移至轻量学生网络SpectralContextSleepNet, 通过教师软目标指导学生模型训练, 在降低模型复杂度的同时提升轻量模型的睡眠阶段判别能力。本文在Sleep-EDF SC数据集上开展实验, 从睡眠分期性能和模型复杂度两个方面对所提出方法进行评估。实验结果表明, 基于知识蒸馏训练的轻量学生模型相比直接训练的轻量学生模型获得更优的睡眠分期性能, 并在保持有效分类能力的基础上显著降低模型计算和存储开销。其中, 学生模型包含1,048,725个参数, 相比教师模型参数量降低约130倍, 表明知识蒸馏在大规模睡眠模型轻量化迁移中具有应用潜力。本文方法通过利用大规模睡眠模型中的软标签知识增强轻量网络分类能力, 在模型性能与计算效率之间取得较好的平衡, 为构建高效、轻量化自动睡眠分期模型提供了一种可行方案。

关键词

睡眠分期, 知识蒸馏, 模型轻量化, 大规模睡眠模型, 深度学习

Research on a Lightweight Sleep Staging Model Based on Knowledge Distillation

Dan Wang*

School of Information Engineering, Hebei GEO University, Shijiazhuang Hebei

Received: August 2, 2026; accepted: September 1, 2026; published: September 9, 2026

Abstract

Large-scale sleep models face challenges in deployment under resource-constrained environments due to their high computational and storage costs, while lightweight models often suffer from limited feature representation capability. To address these issues, this paper proposes a knowledge

*通讯作者。

文章引用: 王丹. 基于知识蒸馏的轻量化睡眠分期模型研究[J]. 计算机科学与应用, 2026, 16(9): 74-84.
DOI: 10.12677/csa.2026.169290

distillation-based lightweight sleep staging method, which transfers knowledge from a large-scale sleep model to a lightweight network. Specifically, SleepGPT is employed as the teacher model, and the soft-label knowledge generated by the teacher model is transferred to the lightweight student network SpectralContextSleepNet. By optimizing the student model with teacher-guided soft targets, the proposed method improves the sleep-stage discrimination capability of the light-weight model while reducing model complexity. Experiments are conducted on the Sleep-EDF SC dataset to evaluate the proposed method from both sleep staging performance and model complexity perspectives. Experimental results demonstrate that the distilled lightweight student model achieves better sleep staging performance than the directly trained lightweight model while significantly reducing computational and storage costs. The student model contains 1,048,725 parameters, achieving approximately 130× parameter reduction compared with the teacher model, indicating the potential of knowledge distillation for lightweight migration of large-scale sleep models. The proposed method enhances the classification capability of lightweight networks by leveraging soft-label knowledge from large-scale sleep models, achieving a favorable balance between sleep staging performance and computational efficiency. It provides a feasible approach for developing efficient and lightweight automatic sleep staging models.

Keywords

Sleep Staging, Knowledge Distillation, Model Lightweighting, Large-Scale Sleep Model, Deep Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

睡眠分期是睡眠分析中的重要任务，其目标是根据多导睡眠监测(Polysomnography, PSG)信号[1]，将睡眠过程划分为清醒期(Wake)、非快速眼动睡眠阶段 N1、N2、N3 以及快速眼动睡眠阶段(Rapid Eye Movement, REM) [2]。传统人工分期通常由专业人员按照统一标准逐个判读 30 s 片段，不仅耗时，而且可能受到判读经验与个体差异影响。因此，利用人工智能方法实现自动睡眠分期成为生物信号分析领域的重要研究方向。

自动睡眠分期已由人工特征设计与传统分类器逐步发展为端到端深度学习模型。DeepSleepNet [3]联合卷积网络与循环网络提取局部形态和序列信息，SeqSleepNet [4]进一步加强连续睡眠序列建模能力。随后，AttnSleep [5]利用注意力机制刻画跨时段依赖，U-Sleep [6]进一步探索大规模训练与跨数据集应用。上述研究表明，单个 epoch 内的时频特征以及相邻 epoch 之间的阶段转移信息均与分期结果密切相关。然而，随着模型表示能力不断增强，网络结构复杂度和参数规模也持续增加，导致模型存储成本和计算开销不断提升，限制了其在资源受限环境中的应用[7][8]。随着大规模预训练技术的发展，睡眠预训练模型逐渐应用于脑电等生理信号分析任务[9]。相比传统任务驱动模型，睡眠预训练模型能够通过大规模数据学习更加丰富的信号表示。SleepGPT [9]等睡眠大模型进一步探索了大规模表示学习在睡眠分析中的应用，为下游任务提供了新的表示基础。然而，大规模睡眠预训练模型通常具有较大的参数规模和计算开销，在资源受限环境中的部署仍存在困难。

知识蒸馏(Knowledge Distillation, KD) [10]通过教师模型产生的软概率分布约束学生训练，使学生在真实标签之外获得类别间相对关系信息，为大模型向轻量化模型迁移提供了可行途径。针对睡眠分期任务中类别不平衡和部分阶段边界模糊的问题，利用教师模型知识辅助轻量模型学习，有望改善模型对困

难睡眠阶段的识别能力。

基于上述问题,本文围绕睡眠大规模预训练模型轻量化应用展开研究,构建基于 SleepGPT [9]教师模型和轻量学生模型的知识迁移框架,将教师模型中学习到的知识迁移至轻量睡眠分期网络 SpectralContextSleepNet,在降低模型复杂度的同时分析知识蒸馏对睡眠阶段识别能力的影响。本文主要贡献如下:

(1) 构建了一种基于教师-学生知识迁移的轻量化睡眠分期框架,实现 SleepGPT [9]知识向轻量网络的迁移。

(2) 设计基于 logit 软目标的蒸馏策略,分析不同蒸馏权重对轻量模型分类性能和类别平衡能力的影响。

(3) 从分类性能和模型复杂度两个角度验证知识蒸馏在睡眠分期模型轻量化迁移中的作用。

2. 相关工作

2.1. 自动睡眠分期

自动睡眠分期研究的主线,已经从依赖人工特征设计的传统分类器逐步发展为能够同时学习局部信号模式和阶段上下文的端到端模型[3]-[6]。卷积结构能够提取睡眠信号中的局部形态和频谱特征[3],循环网络[4]、注意力机制与 Transformer [5] [11]则进一步增强了对相邻睡眠阶段转移关系的建模能力。相关研究表明,多尺度时频表示以及有效的特征选择策略有助于提高生理信号的可分性[12]。

然而,不同睡眠阶段之间存在明显的识别难度差异。N1 处于清醒向稳定睡眠过渡阶段,容易与 Wake 和 N2 发生混淆;N3 和 REM 虽然具有较明确的生理特征,但仍会受到个体差异和阶段转换影响[1]。因此,睡眠分期模型不仅需要学习单个 epoch 中的信号特征,还需要利用连续睡眠过程中的上下文信息。同时,随着模型结构不断复杂化,网络参数规模和计算需求持续增加,使模型在存储效率和资源受限环境部署方面都面临新的挑战。

2.2. 睡眠预训练模型研究

近年来,随着自监督学习和大规模预训练技术的发展,预训练模型逐渐应用于脑电等生理信号分析任务[9] [13]。相比针对单一任务训练的深度学习模型,预训练模型能够利用大规模数据学习更加丰富的信号表示,并通过迁移方式适配不同下游任务,为复杂生理信号分析提供了新的研究思路。

针对睡眠信号分析任务, SleepGPT [9]等睡眠预训练模型探索了大规模表示学习在睡眠分析中的应用,通过预训练方式学习睡眠信号中的潜在模式,为睡眠分期等任务提供了新的表示基础。此类模型不仅能够完成下游分类任务,其输出信息中也包含不同睡眠阶段之间的关系特征,为进一步的知识迁移提供了潜在价值。然而,预训练模型通常依赖较大的参数规模和计算资源,在实际应用中面临存储成本高、推理效率低等问题。

2.3. 知识蒸馏与模型压缩

Hinton 等[10]提出利用温度调节后的软目标进行知识迁移,使学生模型不仅学习真实标签信息,还能够获得教师模型输出概率分布中包含的类别关系信息。相比传统监督学习方式,软目标能够提供更加丰富的监督信号,有助于提升轻量模型的表示能力[10]。

随着研究的发展,知识蒸馏逐渐扩展到特征蒸馏[14]、关系蒸馏[15]和自蒸馏[16]等不同形式。已有研究表明,教师-学生训练框架能够在保持较低模型复杂度的同时迁移大模型中的有效知识,并已应用于图像、文本以及时序任务等多个领域[17]。

对于睡眠分期任务而言，由于不同阶段存在类别不平衡和边界模糊问题，轻量模型容易出现困难阶段识别不足的情况[3]-[6][18]。因此，如何选择适合睡眠信号特点的蒸馏策略，在降低模型复杂度的同时保留有效判别信息，仍需要进一步探索。

3. 方法

3.1. 总体框架

本文构建了一种基于教师-学生学习模式的轻量化睡眠分期框架，整体结构如图 1 所示。该框架以 SleepGPT [9] 作为教师模型，以轻量网络 SpectralContextSleepNet 作为学生模型，通过知识迁移方式利用教师模型信息指导学生模型训练，实现大规模睡眠模型知识向轻量模型的迁移。

在训练阶段，教师模型和学生模型接收相同的睡眠脑电输入，并分别产生预测结果。其中，教师模型利用预训练获得的睡眠信号表示能力提供知识指导，学生模型结合自身网络结构学习睡眠特征，并完成轻量化睡眠阶段识别。通过教师-学生之间的信息传递，使学生模型在较低参数规模下获得有效的特征表示能力。在推理阶段，仅保留训练完成后的学生模型进行睡眠阶段预测，不再依赖大规模教师模型，从而降低模型存储和计算开销。

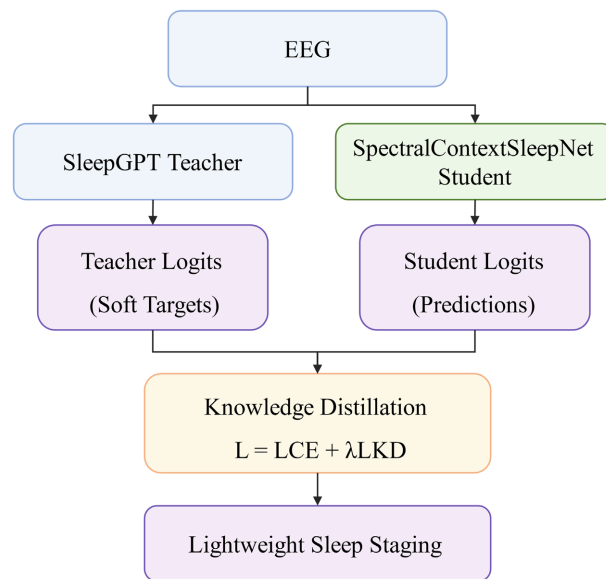


Figure 1. Overall framework for lightweight sleep staging with knowledge distillation

图 1. 知识蒸馏轻量睡眠分期总体框架

3.2. SleepGPT 教师模型

SleepGPT [9] 是一种面向睡眠信号分析的预训练模型，通过预训练过程学习睡眠信号中的潜在表示，为下游睡眠分析任务提供知识支持。相比针对单一任务训练的模型，预训练模型能够利用已有数据学习更加丰富的信号表示，因此适合作为知识迁移中的教师模型。

本文采用 SleepGPT [9] 作为教师模型，其网络结构保持原模型设计，不进行参数更新。在知识蒸馏过程中，教师模型接收睡眠脑电输入，并输出五分类睡眠阶段预测信息，为学生模型提供额外监督。本文仅利用教师模型输出层信息进行知识迁移，不引入中间层特征，以降低蒸馏过程复杂度。

由于 SleepGPT [9] 具有较大的参数规模，本文仅在训练阶段使用教师模型提供指导，在推理阶段不

再调用教师模型，而仅保留轻量学生模型完成睡眠阶段预测。

3.3. SpectralContext 学生模型

为降低大规模教师模型带来的计算和存储开销，本文采用轻量化网络 SpectralContextSleepNet 作为学生模型，用于承载教师模型知识并完成睡眠阶段预测。该模型针对轻量网络在有限参数条件下特征表达能力受限的问题，结合睡眠信号的频谱特征与连续阶段上下文信息，提高模型对不同睡眠阶段的判别能力。

SpectralContextSleepNet 主要由特征提取模块、上下文建模模块和分类模块组成。其中，特征提取模块用于学习睡眠脑电信号中的时频特征，上下文建模模块利用连续睡眠片段之间的时间依赖关系增强阶段识别能力。通过融合局部信号特征与阶段上下文信息，该模型能够在较低计算复杂度下完成睡眠分期任务。

在本文知识蒸馏框架中，SpectralContextSleepNet 作为最终部署模型，仅在训练阶段接受 SleepGPT [9]教师模型指导，训练完成后独立完成 Wake、N1、N2、N3 和 REM 五类睡眠阶段预测。该模型采用连续 20 个 epoch 作为上下文输入窗口，并通过分类头输出最终预测结果，具体网络结构配置如表 1 所示。

Table 1. Architecture of SpectralContextSleepNet

表 1. SpectralContextSleepNet 结构配置

模块	输入	配置	输出维度
时间域编码器	8×3000	Conv1d: $8 \rightarrow 64$ (k31/s4), $64 \rightarrow 128$ (k15/s3), $128 \rightarrow 256$ (k9/s3), $256 \rightarrow 256$ (k5/s2,g4); BN + GELU; AvgPool	256
频域分支	8×3000	rFFT 功率; 0.5~4、4~8、8~12、12~16、16~30 Hz; LayerNorm; Linear $40 \rightarrow 128$; GELU; Dropout 0.1	128
特征融合	$256 + 128$	特征拼接	384
上下文编码	20×384	1 层双向 GRU, hidden = 160	20×320
分类头	320	LayerNorm; Dropout 0.1; Linear $320 \rightarrow 5$	5

3.4. 知识蒸馏策略

知识蒸馏(Knowledge Distillation, KD) [10]通过利用教师模型输出信息指导学生模型训练，使轻量模型能够学习教师模型中的类别间预测关系。相比仅依赖真实标签的监督学习，教师模型输出的软目标能够提供不同类别之间的相似性信息，有助于学生模型获得更加平滑的分类边界。针对睡眠分期任务中不同阶段之间存在一定相似性和边界模糊的问题，本文采用基于输出层 logits 的知识蒸馏方法，将 SleepGPT [9]教师模型的预测信息迁移至轻量学生模型，如图 2 所示。

对于相同输入的睡眠脑电序列，教师模型和学生模型分别输出五类睡眠阶段对应的 logits:

$$z_T = [z_T^1, z_T^2, \dots, z_T^5] \quad (1)$$

$$z_S = [z_S^1, z_S^2, \dots, z_S^5] \quad (2)$$

其中， z_T 和 z_S 分别表示教师模型和学生模型的输出结果。为获得包含类别间关系的软目标，引入温度参数 τ 对预测分布进行平滑处理:

$$p_T(\tau) = \text{soft max}(z_T/\tau) \quad (3)$$

$$p_S(\tau) = \text{soft max}(z_S/\tau) \quad (4)$$

其中，当温度参数大于 1 时，预测概率分布更加平滑，可以保留非目标类别之间的相对关系，为学生模型提供额外监督信息。

在训练过程中，本文联合真实标签监督和教师知识约束优化学生模型。总体损失函数定义为：

$$L = \alpha L_{CE} + \beta L_{KD} \quad (5)$$

其中， L_{CE} 表示学生模型预测结果与真实睡眠标签之间的交叉熵损失， L_{KD} 表示教师模型与学生模型软预测分布之间的蒸馏损失，其定义为：

$$L_{KD} = \tau^2 KL(p_T(\tau) \| p_S(\tau)) \quad (6)$$

其中， $KL(\cdot)$ 表示 Kullback-Leibler 散度[19]， τ^2 用于补偿温度缩放对梯度幅度的影响。通过联合优化真实标签信息和教师预测信息，使学生模型在保持轻量结构的同时学习更加有效的睡眠阶段判别特征。

为分析不同蒸馏强度对轻量模型性能的影响，本文设置两种固定蒸馏配置。其中，KD005 采用较低蒸馏权重，设置 $\alpha = 0.95$ ， $\beta = 0.05$ ，以真实标签监督为主、教师知识为辅助，作为本文主要蒸馏方案；KD020 提高教师约束比例，设置 $\alpha = 0.80$ ， $\beta = 0.20$ ，用于分析增强蒸馏强度后的模型变化。

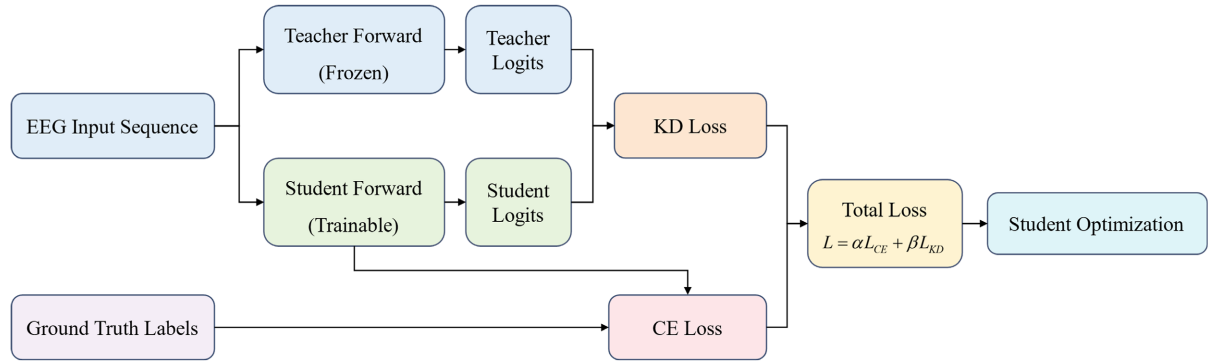


Figure 2. Teacher-student knowledge distillation pipeline

图 2. 教师 - 学生知识蒸馏流程

4. 实验与分析

4.1. 数据集与实验设置

本文采用公开睡眠数据集 Sleep-EDF Sleep Cassette (SC)[20][21]进行实验验证。该数据集包含 153 条整夜多导睡眠监测(Polysomnography, PSG)记录，是自动睡眠分期研究中常用的公开基准数据集。实验按照标准睡眠分期任务设置，将睡眠过程划分为清醒期(Wake)、非快速眼动睡眠阶段 N1、N2、N3 以及快速眼动睡眠阶段(Rapid Eye Movement, REM)五个阶段，并采用 30 s 作为基本分类单元。实验输入信号采样率为 100 Hz，经过统一预处理后用于模型训练与测试。

为保证不同训练策略之间的公平比较，本文固定学生模型 SpectralContextSleepNet 的网络结构和输入形式，仅改变模型训练过程中的监督方式。学生模型采用连续 20 个 epoch 组成上下文窗口作为输入，以利用相邻睡眠阶段之间的时间依赖关系，并输出五类睡眠阶段预测结果。实验中，SleepGPT[9]教师模型采用固定 checkpoint，仅用于生成教师预测信息，不参与参数更新。

本文设置三种学生模型训练方式进行比较：CE 表示仅采用真实标签监督训练的学生模型；KD005 表示蒸馏项权重 $\beta = 0.05$ 的蒸馏模型，以真实标签监督为主、教师预测信息为辅助；KD020 表示蒸馏项权重 $\beta = 0.20$ 的蒸馏模型，用于分析不同蒸馏强度对模型性能的影响。同时，SleepGPT[9]教师模型作为性

能参考，用于比较大规模模型与轻量学生模型之间的性能差异。

模型性能采用 Accuracy、Macro-F1 和 Cohen’s Kappa 进行整体评价，并进一步统计 Wake、N1、N2、N3 和 REM 五个睡眠阶段的分类 F1 值，以分析模型对不同睡眠阶段的识别能力。

4.2. 实验结果分析

为验证知识蒸馏对轻量睡眠分期模型性能的影响，本文比较了 SleepGPT [9]教师模型、SpectralContextSleepNet 采用交叉熵训练(CE)以及不同蒸馏权重下学生模型的性能。实验结果如表 2 所示。其中，SleepGPT [9]教师模型作为性能参考，CE、KD005 和 KD020 均基于相同学生网络结构，仅改变训练监督方式。

Table 2. Overall performance comparison
表 2. 总体性能比较

方法	参数量	ACC/%	Macro-F1/%	Kappa/%
SleepGPT Teacher	137,180,170	77.34	63.12	67.06
SpectralContext CE	1,048,725	69.23	44.25	54.51
SpectralContext KD005	1,048,725	68.54	55.12	57.18
SpectralContext KD020	1,048,725	69.21	55.02	57.83

由表 2 可知，SleepGPT [9]教师模型取得了较高的整体性能，体现了大规模预训练模型较强的睡眠信号表示能力。与仅采用真实标签监督训练的 CE 模型相比，KD005 在 Accuracy 基本保持稳定的情况下，Macro-F1 提高 10.87 个百分点，Kappa 由 54.51%提升至 57.18%。结果表明，教师模型输出的软预测信息能够补充类别间关系约束，使学生模型获得更加均衡的分类能力。对于睡眠分期任务中存在的类别不平衡和阶段边界模糊问题，知识蒸馏能够帮助轻量模型改善不同睡眠阶段之间的判别效果。

进一步比较不同蒸馏强度可以发现，提高蒸馏权重并未带来持续收益。KD020 模型的 Accuracy 为 69.21%，与 CE 模型基本一致，其中 Macro-F1 为 55.02%，略低于 KD005 模型的 55.12%。该结果说明，在当前轻量学生模型容量限制下，较强的教师约束未能进一步提升模型性能，而较低比例的蒸馏监督能够在真实标签信息与教师预测信息之间取得更好的平衡。

综上，实验结果表明，基于 logit 的知识蒸馏能够提升轻量睡眠分期模型的类别判别能力。虽然学生模型整体性能仍低于大规模教师模型，但通过引入教师预测信息，SpectralContextSleepNet 在较小参数规模下获得了更优的 Macro-F1 和 Kappa 表现，表明大规模睡眠模型知识向轻量网络迁移具有一定有效性。

4.3. 不同数据划分下的性能变化分析

表 3 给出了 CE 与 KD005 模型在不同 fold 下的性能比较结果，用于分析知识蒸馏效果在不同数据划分下的变化趋势。

Table 3. Cross-fold results of CE and KD005
表 3. CE 与 KD005 的跨折结果

Fold	CE ACC/%	CE Macro-F1/%	CE Kappa/%	KD005 ACC/%	KD005 Macro-F1/%	KD005 Kappa/%
Fold0	69.23	44.25	54.51	68.54	55.12	57.18
Fold1	63.41	53.54	51.84	65.49	57.20	54.65
平均	66.32	48.90	53.17	67.01	56.16	55.92

由表 3 可知, KD005 在两个 fold 中均表现出高于 CE 模型的 Macro-F1 结果。上述结果表明, 在不同数据划分下, 引入教师预测信息后, 轻量学生模型均表现出一定的类别判别能力改善趋势。

进一步分析不同指标变化可以发现, 知识蒸馏带来的主要收益体现在 Macro-F1 等类别均衡指标上, 而 Accuracy 变化相对有限。这说明蒸馏过程并非简单提高多数类别样本的预测比例, 而是利用教师输出的类别关系信息, 使学生模型学习更加平滑的类别表示, 从而改善不同睡眠阶段之间的识别平衡。

需要注意的是, 不同 fold 之间仍存在一定性能波动。例如, KD005 在 Fold0 中的 Macro-F1 提升幅度高于 Fold1, 说明数据划分差异仍会影响蒸馏收益。同时, 由于教师模型训练数据与部分测试受试者存在重叠, 该跨折结果主要用于描述不同数据划分下的性能变化趋势, 不能作为严格独立泛化性能验证。

总体而言, 跨折结果表明, KD005 在不同数据划分下均保持了相对一致的性能改善趋势, 说明低权重 logit 蒸馏对于轻量睡眠分期模型具有一定的迁移价值。

4.4. 类别性能与混淆矩阵分析

整体评价指标难以体现不同睡眠阶段之间的识别差异, 因此进一步分析各模型在五类睡眠阶段上的分类性能。表 4 给出了 CE、KD005 和 KD020 模型在不同睡眠阶段上的 F1-score 结果。

Table 4. Class-wise F1 scores on the test sets (%)

表 4. 各实验测试集类别 F1 (%)

Fold	方法	Wake	N1	N2	N3	REM
Fold0	Teacher	90.85	22.54	81.52	63.70	56.99
Fold0	CE	87.56	6.26	73.23	12.69	41.53
Fold0	KD005	90.48	21.56	64.26	32.45	66.87
Fold0	KD020	91.09	18.22	65.03	33.24	67.51
Fold1	CE	83.93	21.59	65.92	44.04	52.20
Fold1	KD005	91.66	28.10	56.27	36.99	72.99

由表 4 可知, 知识蒸馏带来的性能变化并非所有睡眠阶段同步提升, 而主要体现在部分阶段识别能力改善。说明知识蒸馏改变了不同阶段之间的预测分布, 使模型降低了对多数类别的预测偏向, 并提升了部分困难阶段的识别能力。

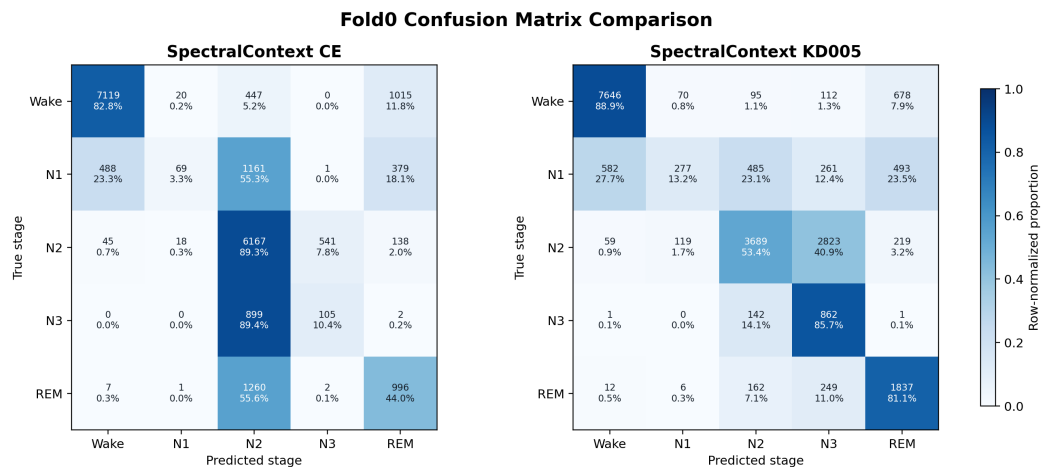


Figure 3. Normalized confusion matrices of CE and KD005 on Fold0

图 3. Fold0 中 CE 与 KD005 的归一化混淆矩阵

为进一步分析类别预测变化，图 3 展示了 CE 和 KD005 模型在 Fold0 测试集上的归一化混淆矩阵。可以观察到，CE 模型存在较明显的阶段混淆现象，其中部分 N1 和 REM 样本被错误预测为 N2，尤其 REM 阶段存在较多向 N2 的错误转移。加入知识蒸馏后，KD005 降低了部分类别向 N2 阶段的错误聚集，提高了 N1 和 REM 阶段的正确识别比例，说明教师预测信息有助于改善轻量模型的类别预测平衡。

进一步比较 KD005 与 KD020 可以发现，提高蒸馏权重并未带来所有阶段的持续提升。KD020 虽然在部分阶段略高于 KD005，但 N1 阶段识别能力下降，说明增加教师约束比例并不能保证所有类别获得收益。

总体而言，类别性能分析表明，知识蒸馏并非简单提升所有睡眠阶段分类性能，而是在保持整体性能基本稳定的情况下，改善部分困难阶段的识别表现，从而促进 Macro-F1 指标提升。

4.5. 典型错误与阶段转换边界分析

为进一步分析知识蒸馏对类别错误分布的影响，将相邻 epoch 真实标签发生变化时变化两侧的 epoch 定义为阶段转换边界，并与稳定区段进行比较，结果如表 5 所示。

Table 5. Error rates at stage-transition boundaries and stable segments (%)

表 5. 阶段转换边界与稳定区段错误率(%)

Fold	模型	转换边界错误率	稳定区段错误率
Fold0	CE	59.31	25.89
Fold0	KD005	58.18	26.93
Fold1	CE	56.89	32.94
Fold1	KD005	56.58	30.56

由表 5 可知，阶段转换边界的错误率明显高于稳定区段。KD005 在两个数据划分中均略微降低了边界错误率，但改善幅度有限。总体而言，知识蒸馏主要调整了轻量模型的类别预测平衡，改善了 N1 和 REM 等部分困难阶段的识别，但 N2 与 N3 之间的混淆以及阶段转换边界问题仍然存在。

4.6. 模型复杂度分析

模型规模和计算成本是影响睡眠分期模型实际应用的重要因素。为分析轻量学生模型的复杂度优势，本文比较了 SleepGPT [9]教师模型与 SpectralContext 学生模型的参数量、存储需求和计算量。

Table 6. Complexity of teacher and student models

表 6. 教师与学生模型复杂度

模型	参数量	FP32 存储/MiB	FLOPs/G
SleepGPT Teacher	137,180,170	523.30	323.89
SpectralContext Student	1,048,725	4.01	2.87

由表 6 可知，SpectralContext 学生模型参数量约为 1.05 M，相比 SleepGPT [9]教师模型降低约 130.81 倍。同时，模型 FP32 存储需求由 523.30 MiB 降低至 4.01 MiB，计算量由 323.89 GFLOPs 降低至 2.87 GFLOPs。

如图 4 所示，教师模型与学生模型在参数规模、计算量和存储需求方面存在明显差异。进一步结合前文结果可以发现，CE 与 KD005 采用相同学生模型结构，因此二者具有相同模型复杂度，而 KD005 通过引入教师预测信息获得更高 Macro-F1，说明性能改善并非来自模型规模增加。

总体而言，复杂度分析表明，SpectralContext 学生模型在保持较好睡眠分期性能的同时，实现了约 130 倍参数压缩和显著计算成本降低，说明知识蒸馏能够有效支持大规模睡眠模型向轻量模型迁移。

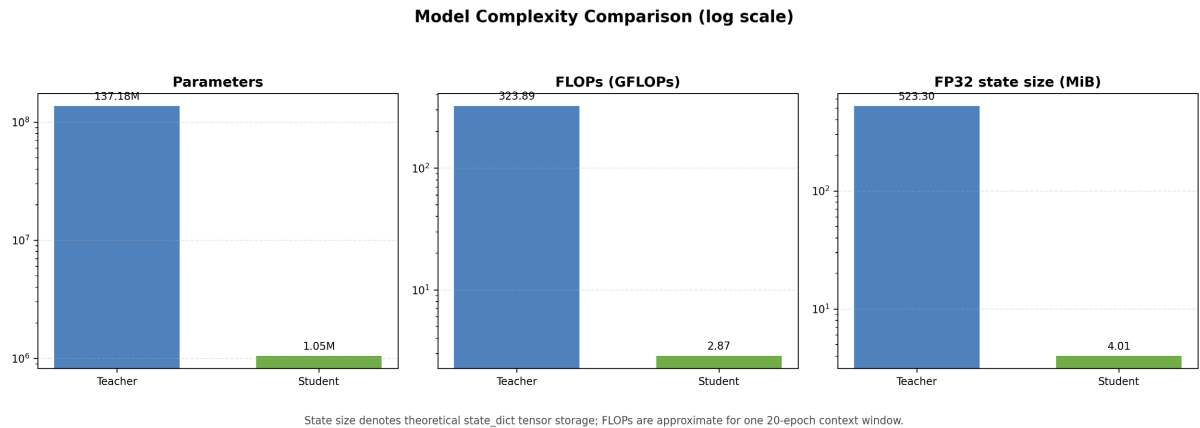


Figure 4. Complexity comparison between the SleepGPT teacher and SpectralContext student

图 4. SleepGPT 教师与 SpectralContext 学生复杂度对比

4.7. 讨论

本文实验结果表明，知识蒸馏能够改善轻量学生模型的睡眠分期性能，但其作用主要体现在类别判别平衡优化，而非所有睡眠阶段性能的同步提升。具体而言，低蒸馏权重设置下模型的 Macro-F1 和 Kappa 指标得到提升，部分困难阶段的分类性能有所改善。较高蒸馏权重并未带来进一步收益。这说明教师模型提供的软目标能够为轻量模型提供额外类别关系信息，但蒸馏强度需要合理控制，以避免影响原有分类能力。

从模型效率角度看，轻量学生网络本身具有较低参数规模和计算开销，知识蒸馏[10]能够在不增加学生模型推理复杂度的情况下改善其睡眠分期表现。由于推理阶段仅保留学生模型，不需要额外调用教师网络，因此该方法为资源受限环境下的轻量化睡眠分期部署提供了可能。进一步的错误分析表明，阶段转换边界的错误率仍明显高于稳定区段，说明相邻睡眠阶段之间的区分仍是当前模型的主要困难之一。

需要指出的是，本文实验仍存在一定限制。由于教师模型训练数据与部分测试受试者存在重叠，当前蒸馏结果主要用于分析知识迁移趋势，不能作为严格独立测试条件下的最终性能估计。此外，本文主要基于固定实验划开展验证，后续仍需在完全受试者隔离的数据协议和更多数据划分条件下进一步评估知识蒸馏方法的稳定性与泛化能力。

5. 结论

本文针对大规模睡眠模型参数规模较大、计算成本较高以及轻量模型分期能力受限的问题，提出了一种基于知识蒸馏的轻量化睡眠分期方法。该方法以 SleepGPT [9]作为教师模型，通过软标签知识迁移指导轻量学生网络 SpectralContextSleepNet 训练，实现大规模睡眠模型知识向紧凑网络的有效迁移，在降低模型复杂度的同时保持睡眠分期能力。实验结果表明，基于知识蒸馏训练的学生模型相比直接训练的轻量模型具有更好的睡眠分期性能，同时显著降低模型参数规模和计算开销。本文研究验证了知识蒸馏在睡眠分期模型轻量化迁移中的应用潜力，为构建高效、低计算开销的自动睡眠分期模型提供了参考。后续工作将进一步在严格受试者隔离的数据协议和更多数据集条件下验证模型的泛化能力，并探索更加有效的知识迁移策略。

参考文献

- [1] Iber, C., Ancoli-Israel, S., Chesson, A.L., *et al.* (2007) The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. American Academy of Sleep Medicine.
- [2] Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., *et al.* (2009) Sleep Classification According to AASM and Rechtschaffen & Kales: Effects on Sleep Scoring Parameters. *Sleep*, **32**, 139-149. <https://doi.org/10.1093/sleep/32.2.139>
- [3] 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>
- [4] Phan, H., Andreotti, F., Cooray, N., Chen, O.Y. and De Vos, M. (2019) SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **27**, 400-410. <https://doi.org/10.1109/tnsre.2019.2896659>
- [5] Eldele, E., Chen, Z., Liu, C., Wu, M., Kwok, C., Li, X., *et al.* (2021) An Attention-Based Deep Learning Approach for Sleep Stage Classification with Single-Channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **29**, 809-818. <https://doi.org/10.1109/tnsre.2021.3076234>
- [6] Perslev, M., Darkner, S., Kempfner, L., Nikolic, M., Jennum, P.J. and Igel, C. (2021) U-Sleep: Resilient High-Frequency Sleep Staging. *NPJ Digital Medicine*, **4**, Article No. 72. <https://doi.org/10.1038/s41746-021-00440-5>
- [7] Han, S., Mao, H. and Dally, W.J. (2016) Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. *International Conference on Learning Representations*, San Juan, 2-4 May 2016, 1-14.
- [8] Kaplan, J., Mccandlish, S., Henighan, T., *et al.* (2020) Scaling Laws for Neural Language Models. <https://arxiv.org/abs/2001.08361>
- [9] Huang, W., Wang, Y., Cheng, H., Xu, W., Li, T., Wu, X., *et al.* (2026) A Unified Time-Frequency Foundation Model for Sleep Decoding. *Nature Communications*, **17**, Article No. 1198. <https://doi.org/10.1038/s41467-025-67970-4>
- [10] Hinton, G., Vinyals, O. and Dean, J. (2015) Distilling the Knowledge in a Neural Network. <https://arxiv.org/abs/1503.02531>
- [11] Phan, H., Mikkelsen, K., Chen, O.Y., Koch, P., Mertins, A. and De Vos, M. (2022) SleepTransformer: Automatic Sleep Staging with Interpretability and Uncertainty Quantification. *IEEE Transactions on Biomedical Engineering*, **69**, 2456-2467. <https://doi.org/10.1109/tbme.2022.3147187>
- [12] Yang, B., Zhu, X., Liu, Y. and Liu, H. (2021) A Single-Channel EEG Based Automatic Sleep Stage Classification Method Leveraging Deep One-Dimensional Convolutional Neural Network and Hidden Markov Model. *Biomedical Signal Processing and Control*, **68**, Article ID: 102581. <https://doi.org/10.1016/j.bspc.2021.102581>
- [13] Banville, H., Chehab, O., Hyvärinen, A., Engemann, D. and Gramfort, A. (2021) Uncovering the Structure of Clinical EEG Signals with Self-Supervised Learning. *Journal of Neural Engineering*, **18**, Article ID: 046020. <https://doi.org/10.1088/1741-2552/abca18>
- [14] Romero, A., Ballas, N., Kahou, S.E., *et al.* (2015) FitNets: Hints for Thin Deep Nets. *International Conference on Learning Representations*, San Diego, 7-9 May 2015, 1-13.
- [15] Park, W., Kim, D., Lu, Y. and Cho, M. (2019) Relational Knowledge Distillation. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 11-15 June 2025, 3967-3976. <https://doi.org/10.1109/cvpr.2019.00409>
- [16] Zhang, L., Song, J., Gao, A., Chen, J., Bao, C. and Ma, K. (2019) Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27-28 October 2019, 3713-3722. <https://doi.org/10.1109/iccv.2019.00381>
- [17] Gou, J., Yu, B., Maybank, S.J. and Tao, D. (2021) Knowledge Distillation: A Survey. *International Journal of Computer Vision*, **129**, 1789-1819. <https://doi.org/10.1007/s11263-021-01453-z>
- [18] Gaiduk, M., Serrano Alarcón, Á., Seepold, R. and Martínez Madrid, N. (2023) Current Status and Prospects of Automatic Sleep Stages Scoring: Review. *Biomedical Engineering Letters*, **13**, 247-272. <https://doi.org/10.1007/s13534-023-00299-3>
- [19] Kullback, S. and Leibler, R.A. (1951) On Information and Sufficiency. *The Annals of Mathematical Statistics*, **22**, 79-86. <https://doi.org/10.1214/aoms/1177729694>
- [20] Kemp, B., Zwirnerman, A.H., Tuk, B., *et al.* (2000) Analysis of a Sleep-Dependent Neuronal Oscillation in the Human Sleep Electroencephalogram. *Journal of Neuroscience Methods*, **95**, 33-41.
- [21] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., *et al.* (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>