

Arquitetura Hierárquica CNN-BiMamba para Identificação de Estágios de Sono em Polissonografia com Aplicação à Detecção de REM sem Atonia

Fernando A. Soster¹, Leticia A. Soster², Mirela T. Cazzolato¹

¹ Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
Av. Trabalhador São-carlense, 400 - Centro, 13566-590 – São Carlos, SP – Brazil

² Unidade de Neurofisiologia Clínica, Instituto da Criança do Hospital das Clínicas da
Faculdade de Medicina da Universidade de São Paulo - HCFMUSP
Av. Dr. Enéas C. de Aguiar, 647 - Cerqueira Cesar, 05403-000 – São Paulo, SP – Brazil

fernando.soster@usp.br, leticia.soster@hc.fm.usp.br, mirela@icmc.usp.br

Abstract. *Accurate sleep staging is a prerequisite for quantifying REM Sleep Without Atonia (RSWA), the main biomarker of REM Sleep Behavior Disorder (RBD), an early indicator of several neurodegenerative diseases. Deep learning approaches typically operate on isolated 30-second epochs, a resolution insufficient for the short muscular events (0.5–5s) characteristic of RSWA. This work proposes a hierarchical CNN-BiMamba architecture for sleep staging from continuous sequences of 3-second mini-epochs. Evaluated on 151 PSG recordings via GroupKFold cross-validation, the model achieved macro-F1 of 0.736 ± 0.021 and Cohen's Kappa of 0.691 ± 0.027 , providing a viable first stage for future RSWA detection and quantification.*

Resumo. *O estagiamento automático do sono é etapa fundamental para a quantificação de REM Sem Atonia (RSWA), principal biomarcador do Transtorno Comportamental do Sono REM (RBD), marcador precoce de doenças neurodegenerativas. Métodos de aprendizado profundo geralmente operam em épocas de 30 segundos, resolução insuficiente para os eventos musculares breves (0,5–5s) característicos de RSWA. Este trabalho propõe uma arquitetura hierárquica CNN-BiMamba para o estagiamento do sono a partir de sequências contínuas de mini-épocas de 3 segundos. Avaliado em 151 exames de PSG via validação cruzada GroupKFold, o modelo alcançou F1-macro de $0,736 \pm 0,021$ e Kappa de Cohen de $0,691 \pm 0,027$, constituindo uma etapa inicial viável para futura detecção e quantificação de RSWA.*

1. Introdução

O sono é composto pelos estágios NREM (*Non-Rapid Eye Movement*) e REM (*Rapid Eye Movement*). O sono NREM é subdividido em N1, N2 e N3, enquanto o sono REM caracteriza-se por atividade cerebral semelhante à vigília, movimentos oculares rápidos e atonia muscular. [American Academy of Sleep Medicine 2023, Sobreira-Neto et al. 2023].

O Transtorno Comportamental do Sono REM (*REM Sleep Behavior Disorder* – RBD) é uma parassonia caracterizada por comportamentos motores e/ou vocais durante o sono REM. Seu diagnóstico requer a presença de REM Sem Atonia

(*REM Sleep Without Atonia* – RSWA) em Polissonografia (PSG) e evidências clínicas de atuação de sonhos. É um biomarcador precursor de doenças neurodegenerativas – como doença de Parkinson – precedendo o diagnóstico formal em 10 a 15 anos [Schenck et al. 1986, Postuma et al. 2019, American Academy of Sleep Medicine 2023].

Embora métodos atuais baseados em aprendizado profundo tenham alcançado elevado desempenho em estagiamento automático, eles perdem resolução e contexto temporal em um PSG completo operando em épocas de 30 segundos, inadequadas para eventos musculares de 0,5 – 5 segundos.

Neste contexto, este trabalho propõe uma arquitetura baseada em Redes Neurais Convolucionais (CNN) e Mamba Bidirecional (BiMamba) para o estagiamento automático do sono utilizando mini-épocas de 3 segundos. A arquitetura combina extração local de características e modelagem temporal contínua do PSG, constituindo a etapa inicial para futura detecção e quantificação de RSWA.

As contribuições deste trabalho são: (i) uma arquitetura CNN-BiMamba para estagiamento do sono em resolução de mini-época (3s), preservando contexto temporal do exame completo; e (ii) uma base metodológica para futura detecção e quantificação de RSWA.

2. Trabalhos Relacionados

Redes Neurais Convolucionais (CNN) constituem uma das principais abordagens para o estagiamento automático do sono. A evolução dessas arquiteturas inclui CNNs de canal único [Tsinalis et al. 2016], integração com LSTM para modelagem temporal [Supratak et al. 2017], processamento multimodal de Eletroencefalograma (EEG), Eletrooculograma (EOG) e Eletromiograma (EMG) [Chambon et al. 2018] e, mais recentemente, mecanismos de atenção e convoluções multi-escala [Goerttler et al. 2025, Chen et al. 2025].

Apesar dos ganhos com atenção e convoluções multi-escala, essas arquiteturas ainda enfrentam limitações na modelagem de dependências de sequências extensas, como PSG de noite completa. Os Modelos de Espaço de Estados (*State Space Models* – SSM), mais recentemente Mamba [Gu and Dao 2024], surgiram como alternativa, mantendo complexidade computacional linear, e tendo sido aplicado ao estagiamento automático do sono [Zhang et al. 2024, Zhou et al. 2024, Jiahao et al. 2025, Zhang et al. 2026], ainda que também restritos a épocas de 30 segundos.

Embora CNN e Mamba tenham avançado o estagiamento automático do sono, os trabalhos não operam em resolução inferior a 30 segundos. Este trabalho propõe uma arquitetura CNN-BiMamba para mini-épocas de 3 segundos.

3. Metodologia

3.1. Conjuntos de Dados

O modelo foi avaliado utilizando o conjunto público *PhysioNet HMC Sleep Staging Database* [Alvarez-Estevez and Rijsman 2022]. Composto por 151 PSGs sem critérios de exclusão (85 homens e 66 mulheres; $53,9 \pm 15,4$ anos), totalizando aproximadamente 1.144 horas de sinais fisiológicos e 1.372.430 mini-épocas de 3 segundos, distribuídas entre *Wake* (W) (17,2%), N1 (11,3%), N2 (36,5%), N3 (19,4%) e REM (15,5%).

3.2. Pré-processamento

Os sinais de PSG foram submetidos a filtros passa-banda (EEG 0,3–40 Hz, EOG 0,5–40 Hz e EMG 10–50 Hz) e rejeita-banda (50 Hz – atenuação da interferência da rede elétrica local). Estes filtros preservam as faixas espectrais relevantes para o estagiamento do sono e para a análise da atividade muscular [Chen et al. 2025, American Academy of Sleep Medicine 2023].

Para reduzir o custo computacional, todos os sinais foram reamostrados a 100 Hz utilizando um filtro *anti-aliasing*. Em seguida, foram segmentados em épocas de 30 segundos conforme as anotações de estagiamento e posteriormente subdivididos em dez mini-épocas consecutivas de 3 segundos – resolução temporal recomendada pela *American Academy of Sleep Medicine* (AASM) e SINBAR para avaliação de RSWA – cada uma herdando o rótulo de estágio da época de 30 segundos correspondente [American Academy of Sleep Medicine 2023, Frauscher et al. 2012].

3.3. Arquitetura

A arquitetura proposta é composta por dois estágios (Figura 1). Uma Análise Tradicional (AT) – extração local por CNN; e Análise Contínua (AC) – BiMamba para modelagem temporal do exame completo.

Na AT, cada mini-época é processada com seu contexto local com cinco janelas consecutivas ($t-2$ a $t+2$, definidas empiricamente considerando o custo computacional). CNNs específicas para EEG, EOG e EMG fornecem contexto temporal local sem perder a resolução de 3 segundos. Cada modalidade utiliza três kernels em diferentes escalas temporais, conforme a Tabela 1. Os resultados são ponderados por um bloco *Squeeze-and-Excitation* (SE) antes da fusão em um vetor por mini-época, que alimenta a AC.

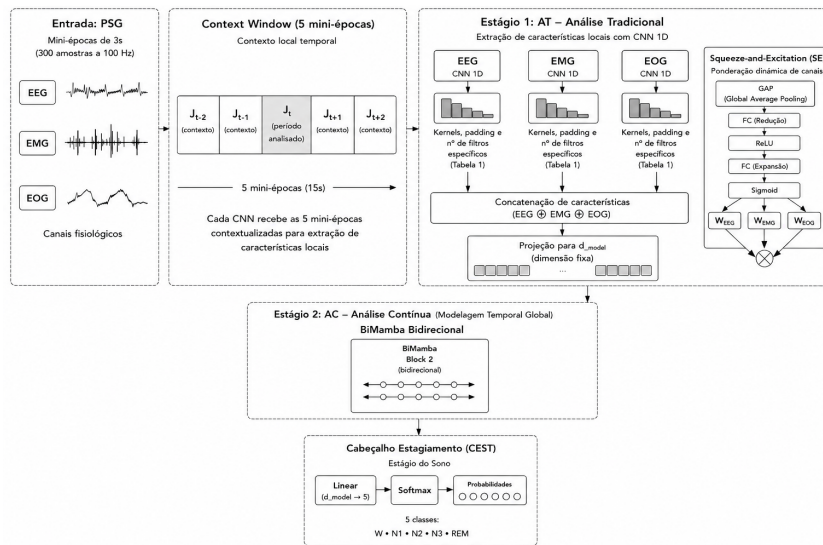


Figura 1. A arquitetura é composta por dois estágios: (i) extração local de características (AT) e (ii) modelagem temporal contínua (AC).

Na AC, a modelagem bidirecional (informações dos estados anteriores e posteriores do exame) favorece a identificação das transições entre estágios do sono. Foi utilizada uma única camada BiMamba, baseada em evidências da literatura, indicando desempenho semelhante ou superior a arquiteturas mais profundas [Zhang et al. 2024].

Tabela 1. Configuração das CNNs multi-escala para cada modalidade fisiológica.

Modalidade	Kernels	Janela (s)	Alvo
EEG	30 / 70 / 150	0.3 / 0.7 / 1.5	ondas β/α , θ e δ
EMG	10 / 30 / 70	0.1 / 0.3 / 0.7	Bursts musculares e atividade tônica
EOG	50 / 100 / 150	0.5 / 1.0 / 1.5	Movimentos oculares rápidos e lentos

A classificação final é realizada por uma camada linear seguida de função *softmax*, produzindo as probabilidades para os cinco estágios do sono.

3.4. Treinamento

O modelo foi implementado em PyTorch, utilizando MNE-Python para leitura e pré-processamento dos exames de PSG em formato EDF. A arquitetura possui aproximadamente 463.945 parâmetros treináveis e foi avaliada utilizando validação cruzada *GroupK-Fold* ($k = 5$) sobre os 151 exames do HMC Sleep Staging Database. O treinamento empregou o otimizador AdamW, *learning rate* de 1×10^{-4} , *dropout* de 0,35 e função de perda *Cross-Entropy* ponderada pelas frequências das classes (W: 1,01; N1: 1,13; N2: 0,69; N3: 0,99; REM: 1,03). O treinamento foi realizado por até 50 épocas, utilizando *early stopping* após 20 épocas sem melhora do F1 de validação. Os experimentos foram executados em uma GPU NVIDIA A100 de 80 GB, com tempo médio de treinamento de 110 minutos por *fold*.

4. Resultados

O modelo atingiu F1-macro de $0,736 \pm 0,021$, Kappa de $0,691 \pm 0,027$ e precisão balanceada de $0,750 \pm 0,021$, demonstrando desempenho consistente entre os cinco *folds* da validação cruzada.

**Figura 2. Matriz de confusão por classe (sensibilidade) agregada dos *folds*.**

A Figura 2 apresenta a matriz de confusão dos cinco *folds*. Os melhores índices foram REM (0,90), W (0,86) e N3 (0,85). O N2 apresentou bom desempenho (0,71), enquanto N1 obteve o menor índice de acerto (0,43), resultado compatível com a literatura devido à curta duração desse estágio e à elevada variabilidade entre transições [Anderer et al. 2023, Jiahao et al. 2025, Zhang et al. 2026].

A Tabela 2 apresenta uma comparação referencial com trabalhos da literatura. Embora o F1-macro obtido seja inferior ao de trabalhos que operam em épocas de 30

segundos, essa diferença é esperada: a granularidade de 3 segundos aumenta a exposição a transições e artefatos de curta duração, tornando a tarefa de estagiamento mais difícil e não diretamente comparável aos resultados em resolução padrão. Além disso, a avaliação foi restrita a uma única base de dados, sem validação externa nem estudos de ablação por componente, o que limita as conclusões quanto à generalização.

Tabela 2. Comparação com trabalhos representativos do estado da arte em estagiamento automático do sono.

Trabalho	Arquitetura	Dataset	Classes	Época	F1	Kappa	Obs.
<i>Modelos baseados em CNN</i>							
[Tsinalis et al. 2016]	CNN	Sleep-EDF-20	5	30s	0,840	—	—
[Supratak et al. 2017]	CNN+BiLSTM	Sleep-EDF	5	30s	0,769	0,760	—
[Chambon et al. 2018]	CNN multicanal	MASS	5	30s	0,817	—	—
[Goerttler et al. 2025]	MSA-CNN	ISRUC-S3	5	30s	0,777	0,784	Multivariada
<i>Modelos baseados em SSM/BiMamba</i>							
[Zhang et al. 2024]	MSSC-BiMamba	ISRUC+Sleep-EDF	5	30s	0,801	0,773	valid. cross-dataset
[Zhou et al. 2024]	BiT-MamSleep	Sleep-EDF-20	5	30s	0,785	0,796	—
[Jiahao et al. 2025]	Mamba-CAM-Sleep	DOD-H	5	30s	0,804	0,763	—
[Zhang et al. 2026]	BiMamba wearable	privado	3	30s	0,842	0,728	357 suj.
Soster et al.	CNN-BiMamba	HMC	5	3s	0,736	0,691	151 exames

O desempenho competitivo obtido, mesmo em resolução de mini-época, indica a viabilidade da abordagem como base para a futura detecção e quantificação de RSWA.

5. Conclusão

Este trabalho indica a viabilidade preliminar de estagiamento automático do sono na resolução de mini-épocas de 3 segundos. Com desempenho competitivo, os resultados sugerem uma base arquitetural promissora para a futura detecção e quantificação de RSWA em granularidade compatível com eventos musculares breves.

Como próximos passos, pretende-se estender a arquitetura à detecção de RSWA, incorporar exames clínicos do Instituto da Criança do HCFMUSP e realizar comparações estatísticas com métodos da literatura. Por ora, o código não é disponibilizado publicamente; sua publicação está prevista em trabalhos futuros, mediante adequação aos protocolos éticos do projeto.

Aprovação do Comitê de Ética. O presente estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas – FMUSP sob protocolo CAAE 84327424.7.0000.0068, em conformidade com a Resolução CNS 466/2012.

Agradecimentos. O trabalho recebeu apoio da CAPES – Cód. Fin. 001, FAPESP (#2024/13328-9, #2024/15430-5, #2026/11130-2), CNPq, e USP (PRPI 1032, #207).

Referências

- Alvarez-Estevez, D. and Rijsman, R. (2022). Haaglanden Medisch Centrum sleep staging database. *PhysioNet*. Version 1.1, DOI: 10.13026/t79q-fr32.
- American Academy of Sleep Medicine (2023). *The AASM Manual for the Scoring of Sleep and Associated Events: Terminology and Technical Specifications*. Version 3.
- Anderer, P., Ross, M., Cerny, A., Vasko, R., Shaw, E., and Fonseca, P. (2023). Overview of the hypnodeltensity approach to scoring sleep for polysomnography and home sleep testing. *Front. Sleep*, 2:1163477. DOI: 10.3389/frsle.2023.1163477.

- 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. DOI: 10.1109/TNSRE.2018.2813138.
- Chen, X., Zhang, Y., Tang, Z., Wu, H., Chen, C., and Chen, W. (2025). MtRBD: Advancing iRBD Analysis With Multi-Task Learning for Joint Sleep Staging and RSWA Detection. *IEEE J. Biomed. Health Inform.*, 29(12):8743–8750. DOI: 10.1109/JBHI.2025.3606944.
- Frauscher, B., Iranzo, A., et al. (2012). Normative emg values during rem sleep for the diagnosis of rem sleep behavior disorder. *Sleep*, 35(6):835–847.
- Goerttler, S., Wang, Y., Eldele, E., Wu, M., and He, F. (2025). MSA-CNN: A Lightweight Multi-Scale CNN with Attention for Sleep Stage Classification. DOI: 10.48550/arXiv.2501.02949.
- Gu, A. and Dao, T. (2024). Linear-Time Sequence Modeling with Selective State Spaces. DOI: 10.48550/arXiv.2312.00752.
- Jiahao, H., Ur Rahman, M. M., Al-Naffouri, T., and Laleg-Kirati, T.-M. (2025). Mamba-CAM-Sleep: A Mamba-based Channel Attention Model for Sleep Staging Classification. In *Annual Int. Conf. of the IEEE EMBC*, pages 1–6, Copenhagen, Denmark. IEEE. DOI: 10.1109/EMBC58623.2025.11251806.
- Postuma, R. B., Iranzo, A., et al. (2019). Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain*, 142(3):744–759. DOI: 10.1093/brain/awz030.
- Schenck, C. H., Bundlie, S. R., Ettinger, M. G., and Mahowald, M. W. (1986). Chronic Behavioral Disorders of Human REM Sleep: A New Category of Parasomnia. *Sleep*, 9(2):293–308. DOI: 10.1093/sleep/9.2.293.
- Sobreira-Neto, M. A., Stelzer, F. G., et al. (2023). REM sleep behavior disorder: update on diagnosis and management. *Arquivos de Neuro-Psiquiatria*, 81(12):1179–1194. DOI: 10.1055/s-0043-1777111.
- 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 TNSRE Journal*, 25(11):1998–2008. DOI: 10.1109/TNSRE.2017.2721116.
- 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. DOI: 10.1007/s10439-015-1444-y.
- Zhang, A. H., He-Mo, A., et al. (2026). Mamba-based deep learning approach for sleep staging on a wireless multimodal wearable system without electroencephalography. *SLEEP*, 49(4):zsag022. DOI: 10.1093/sleep/zsag022.
- Zhang, C., Cui, W., and Guo, J. (2024). MSSC-BiMamba: Multimodal Sleep Stage Classification and Early Diagnosis of Sleep Disorders with Bidirectional Mamba. DOI: 10.48550/arXiv.2405.20142.
- Zhou, X., Han, Y., et al. (2024). BiT-MamSleep: Bidirectional temporal Mamba for eeg sleep staging. URL: <http://arxiv.org/abs/2411.01589>.