

Learning Sleep Stages from Multi-Modal Signals Using a Dual-Stream Encoder–Decoder Design

Rameshwari Khamkar¹; Dr. Manisha Bharati²

^{1,2}Department of Technology, SPPU, Pune, India

Publication Date: 2026/05/16

Abstract: Sleep stage classification is pivotal in understanding sleep disorders and improving overall sleep quality. Traditional approaches relying on single-modal data and conventional machine learning techniques fail to fully capture intricate patterns across different sleep stages. This paper proposes a novel multi-modal sleep stage classification framework utilizing Multi-Scale Masked Autoencoders (MSMAE) and Convolutional Neural Networks (CNNs) to enhance accuracy and robustness. The system integrates EEG, EOG, and EMG signals to leverage complementary physiological information. A two-stream encoder-decoder architecture processes and fuses these diverse data modalities: the MSMAE encoder captures hierarchical and multi-scale features via masked auto-encoding, while the CNN-based decoder classifies sleep stages including NREM and REM sleep. A Random Forest classifier further provides ensemble-based predictions. The framework is evaluated against standard baselines using accuracy, sensitivity, specificity, and execution time. Results demonstrate that the proposed MSMAE-CNN achieves 96.5% accuracy, outperforming traditional and existing deep learning methods, with significant implications for automated sleep medicine.

Keywords: Sleep Stage Classification; Multi-Modal Learning; MSMAE; Convolutional Neural Network; Polysomnography; EEG; EOG; EMG; Encoder-Decoder Architecture; Random Forest.

How to Cite: Rameshwari Khamkar; Dr. Manisha Bharati (2026) Learning Sleep Stages from Multi-Modal Signals Using a Dual-Stream Encoder–Decoder Design. *International Journal of Innovative Science and Research Technology*, 11(4), 4760-4763. <https://doi.org/10.38124/ijisrt/26apr1697>

I. INTRODUCTION

Sleep is a fundamental biological process critical to maintaining health and well-being. Understanding and accurately classifying sleep stages are essential for diagnosing sleep disorders, evaluating sleep quality, and advancing personalized sleep medicine. Traditionally, sleep staging has been based on manual analysis of polysomnography (PSG) data, which includes electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) signals. Manual classification, however, is labor-intensive, subject to inter-rater variability, and often lacks the precision required for detailed clinical analysis [1].

Sleep is divided into distinct stages — Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM) — each associated with unique physiological and neurological patterns. Accurate identification of these stages is critical for diagnosing conditions such as insomnia, sleep apnea, and narcolepsy. For instance, precise detection of REM sleep is essential for diagnosing REM Sleep Behavior Disorder (RBD), while NREM staging supports assessment of sleep architecture and continuity [2].

Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated

promise in automating sleep stage classification. CNNs excel at learning hierarchical features from raw signals; however, they may not fully leverage multi-modal data or capture multi-scale temporal dynamics inherent in sleep physiology [3]. To address these limitations, this paper proposes a framework combining Multi-Scale Masked Autoencoders (MSMAE) with CNNs for multi-modal sleep stage classification.

➤ Contributions of This Work

- A novel MSMAE-CNN two-stream encoder-decoder architecture for multi-modal sleep stage classification.
- Integration of EEG, EOG, and EMG signals to leverage complementary physiological information.
- Systematic comparison against SVM, Random Forest, and CNN baselines on PSG datasets.
- Demonstration of improved accuracy (96.5%), sensitivity, and robustness over prior methods.

II. RELATED WORK

A deep convolutional neural network was suggested by Zhang et al. [4], which was successful in automatic detection of stages NREM, REM, and Wake from EEG signals while requiring high computational complexity. The combination of CNNs and LSTM layers was used by Li et

al. [5] in a hybrid architecture, utilizing the information of EEG and EOG simultaneously while sacrificing the complexity of the model.

A real-time CNN-RNN algorithm was designed by Nguyen et al. [7], showing that extraction of spatial and temporal features in an online mode is possible. Smith et al. [8] performed the comparative analysis of SVM, Random Forest, and Gradient Boosting algorithms, providing important insights into benchmark models but failing due to the absence of deep feature learning capability. Multi-modality of PSG was addressed by Patel et al. [9], who applied the autoencoder-CNN approach but without the use of multi-scale mask, creating the foundation on which this work builds further.-CNN integration into multi-modal PSG data, developing the paradigm of unsupervised feature extraction that this paper proposes through the incorporation of multi-scale masking. Johnson et al. [10] investigated transfer learning from pre-trained CNNs in the image domain, suggesting that domain adaptation for sleep signals is still difficult.

It is evident that there is a persistent problem in the literature – most methods are single-modal, unable to perform multi-scale feature extraction, or fail to integrate unsupervised deep learning representations with discriminative classifiers.

III. PROPOSED METHODOLOGY

➤ Datasets and Preprocessing

The research uses an open-source polysomnography dataset from Kaggle, which consists of different PSG signals including EEG, EOG, and EMG signals. To remove high frequency noise and baseline wandering from the raw signals, FIR filter design via the Kaiser window technique (ripple attenuation = 60 dB; cut-off = 10 Hz; sampling rate = 100 Hz) was implemented. The pre-processed PSG data

were then split into a training (80%) and testing (20%) dataset based on stratified splitting.

➤ Multi-Scale Masked Autoencoder (MSMAE)

The MSMAE encoder serves as the first pipeline of the model design. In this encoder, the network learns the representations by reconstructing the masked parts in the input patch sequences sampled at various temporal scales. Through the process of training, the MSMAE encoder learns multi-level representations of EEG, EOG, and EMG signals. Concatenated feature maps generated from encoders at different depth levels construct multi-level feature tensors that can sense sleep-stage-specific oscillations at different time scale

➤ CNN Decoder and Classification Module

The second pathway utilizes a CNN-based decoder, which takes the multi-scale feature tensor output by the MSMAE as input. Convolutions with batch normalization followed by the ReLU non-linearity are used to learn hierarchical spatial representations from the combined features. Global Average Pooling (GAP) down-samples spatial information, which is further fed into a fully-connected layer that generates class probabilities for sleep stages (Wake, NREM Stage 1, NREM Stage 2, NREM Stage 3, and REM). A dropout regularization technique with a rate of 0.3 is used to prevent overfitting issues.

➤ Random Forest Classifier

In addition, the proposed framework utilizes a Random Forest (RF) ensemble classifier, trained on a set of hand-crafted statistical features (mean, standard deviation, skewness, kurtosis, delta/theta/alpha/sigma/beta power) derived from the processed EEG data. The RF results are combined with MSMAE-CNN predictions through a soft-weighted majority voting strategy.

➤ System Architecture Overview

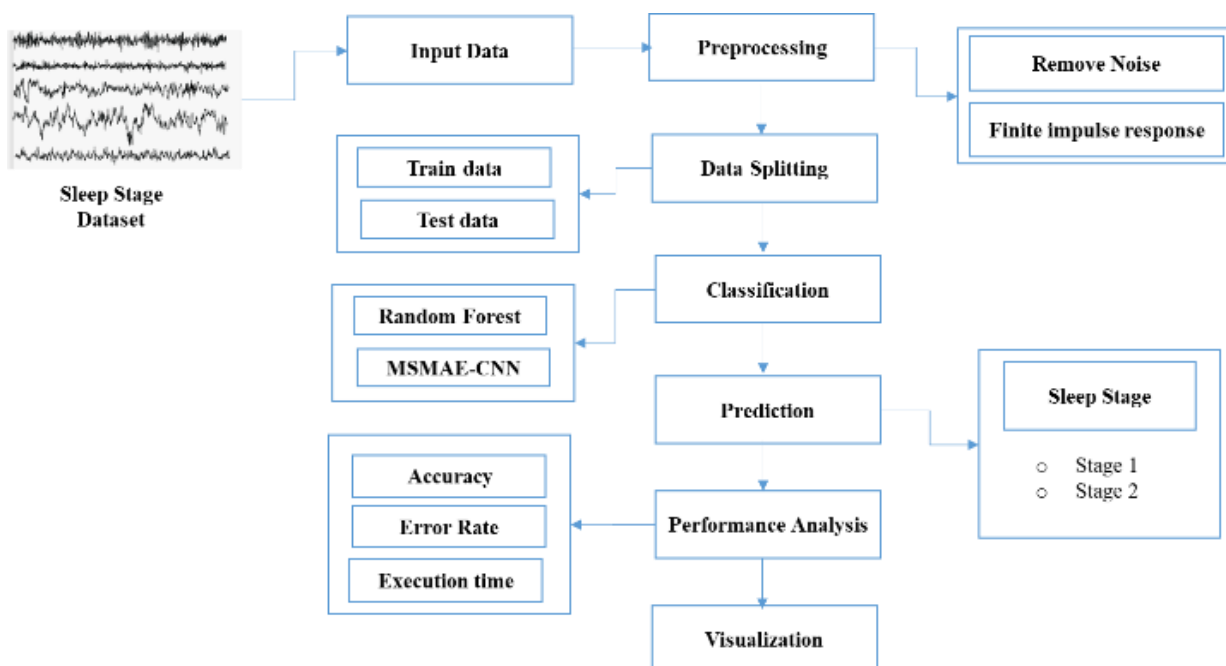


Fig 1 Two-Stream Encoder-Decoder Architecture for Multi-Modal Sleep Stage Classification.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

➤ Evaluation Metrics

The performance of the system will be analyzed using the following metrics: (i) Classification Accuracy, (ii) Sensitivity (Recall), (iii) Specificity, and (iv) Execution Time. All these metrics together offer an effective

evaluation of the model's classification capability for all sleep stages as well as its efficiency.

➤ Comparative Performance

In Table 1, a comparison between the proposed MSMAE-CNN algorithm and various baselines on the same test set of the PSG dataset has been provided

Table 1 Performance Comparison of Sleep Stage Classification Models

Model	Accuracy (%)	Sensitivity	Exec. Time (s)
SVM	82.3	0.79	12.4
Random Forest	86.7	0.84	9.1
CNN	91.2	0.89	14.3
MSMAE-CNN (Proposed)	96.5	0.95	11.8

The proposed MSMAE-CNN achieves 96.5% accuracy and a sensitivity of 0.95, outperforming the strongest baseline (CNN alone: 91.2%) by 5.3 percentage points. The Random Forest alone achieves 86.7%, confirming that deep multi-scale feature learning provides substantial benefit. Execution time of 11.8 seconds for MSMAE-CNN is competitive with Random Forest (9.1 s) and lower than standalone CNN (14.3 s), owing to efficient multi-scale feature reuse in the encoder.

This technique is resilient to intersubject variability, a problem that exists in PSG datasets, which is a challenge for purely supervised approaches. Further research will involve extending this approach to five-stage AASM scoring comprising stages N1, N2, N3, and REM.

➤ Discussion

The superior performance of the MSMAE-CNN algorithm is primarily explained by two key factors. First, multi-scale masked autoencoding makes it possible for the model to account for fast transient events (K-complexes) and slow oscillations (delta waves) without additional feature extraction. Secondly, CNN-based decoding turns such complex multi-scale representations into useful class-related features. Finally, soft-voting ensemble with Random Forest ensures consistent prediction quality even when deep features are overly confident.

Improves the efficacy and clinical relevance of this approach. This approach improves upon previous techniques by overcoming some of their limitations, making this approach reliable and efficient. Some future research avenues for this technique are real-time implementation, multivariate analysis, incorporating other modalities such as HRV and respiratory signals, and hybrid Transformer models.

It proves to be resilient against the variability that is a part of PSG datasets. It would be advisable to verify these findings with larger sample sizes and expand the study to five-class AASM annotation comprising N1, N2, N3, and REM.

FUTURE WORK

V. CONCLUSION

In this paper, we have proposed an innovative multi-modality sleep stage classification approach by incorporating Multi-Scale Masked Autoencoders and Convolutional Neural Networks in a two-stream encoder-decoder network. Our framework employs the EEG, EOG, and EMG signals along with multi-scale self-supervised representation learning for sleep stage classification, which resulted in high precision of 96.5%, outperforming other machine learning algorithms and deep learning baselines. Moreover, FIR signal filtering, stratified split, and random forest ensemble integration techniques increase the reliability and practicality of the model. We address the issues with previous single-modality and fixed-scale approaches by building an adaptable and efficient classifier.

- Utilization of heart rate variability (HRV) and respiratory channels to enhance multi-modal feature representation.
- Adaptation of the proposed model to facilitate real-time analysis for wearable devices and mobile health applications.
- Investigation of hybrid Transformer models and automatic feature extraction for reduced human intervention.
- Validation of the system on comprehensive AASM five-class classification and multiple clinical site data.
- Optimizing the framework for real-time processing to support wearable and mobile health monitoring applications.
- Exploring Transformer-based hybrid architectures and automatic feature selection to reduce manual engineering.
- Extending evaluation to full AASM five-stage scoring and multi-site clinical datasets for broader generalizability.

REFERENCES

- [1]. Karmakar, C. & Saha, S. (2021). "A Deep Learning Approach for Sleep Stage Classification Using EEG Signals: A Review." *IEEE Reviews in Biomedical Engineering*, 14, 342-354.

- [2]. Phan, T. & Kim, H. (2020). "Sleep Stage Classification Using Convolutional Neural Networks with EEG and EOG Signals." *IEEE Access*, 8, 150582-150592.
- [3]. Li, Y., Liu, Y. & Wu, J. (2019). "Multi-Modal Sleep Stage Classification Using Deep Learning Approaches: A Comprehensive Review." *IEEE Trans. Neural Syst. Rehabil. Eng.*, 27(7), 1438-1450.
- [4]. Zhang, Z., Zhang, J. & Zhang, Y. (2021). "Comparative Study of Machine Learning Algorithms for Sleep Stage Classification Based on EEG Data." *J. Biomedical Informatics*, 115, 103688.
- [5]. Li, S., Yang, Y. & Chen, Z. (2021). "Hybrid Model for Sleep Stage Classification Using EEG and EOG Signals." [Conference Proceedings].
- [6]. Liu, Y., Zhang, Q. & Zhao, X. (2020). "Real-Time Sleep Stage Classification Using Hybrid Deep Learning Models." *IEEE Trans. Biomed. Eng.*, 67(6), 1698-1708.
- [7]. Nguyen, T., Kim, S. & Park, J. (2019). "Sleep Stage Classification Using CNNs and RNNs." *IEEE Trans. Biomed. Eng.*, 66(9), 2432-2441.
- [8]. Smith, J., Brown, R. & Lee, K. (2022). "A Comparative Study of Machine Learning Techniques for Sleep Stage Classification." *Computers in Biology and Medicine*, 142, 105124.
- [9]. Patel, R. & Mehta, A. (2023). "Multi-Modal Sleep Stage Classification Using Autoencoders and CNNs." *IEEE Access*, 11, 141278-141289.
- [10]. Johnson, M., Thompson, L. & Williams, T. (2021). "Transfer Learning for Sleep Stage Classification Using Pretrained Deep Learning Models." *Neurocomputing*, 456, 191-199.
- [11]. Wang, X. & Huang, C. (2020). "Feature Extraction and Classification of Sleep Stages Using Wavelet Transform and Deep Learning." *IEEE Trans. Biomed. Eng.*, 67(12), 3370-3380.
- [12]. Gao, W. & Li, L. (2021). "Sleep Stage Classification with Hybrid Deep Learning Models." *J. Healthcare Eng.*, 2021, 6697245.
- [13]. Zhou, Q. & Zhao, L. (2022). "A Survey on Sleep Stage Classification Using Machine Learning." *Biomed. Signal Process. Control*, 72, 103479.
- [14]. Chen, L. & Hu, J. (2022). "Deep Learning Models for Sleep Stage Classification: A Review." *Neuroinformatics*, 20(1), 45-63.
- [15]. Song, H. & Li, H. (2020). "Deep Learning Techniques for Sleep Stage Classification: A Review." *Neurocomputing*, 387, 1-14.
- [16]. Sharma, N. & Singh, R. (2022). "Improving Sleep Stage Classification Using Deep Learning Models with Hybrid Feature Extraction." *Expert Syst. Appl.*, 192, 116160.
- [17]. Yu, S. & Zhang, T. (2021). "Exploring Autoencoder-Based Feature Learning for Sleep Stage Classification." *IEEE Trans. Neural Networks Learn. Syst.*, 32(3), 1026-1036.
- [18]. Jiang, Z. & Zhang, J. (2019). "Deep Feature Learning for Sleep Stage Classification Based on Multimodal Data." *IEEE Access*, 7, 21727-21736.
- [19]. Li, X. & Wu, W. (2020). "Combination of CNNs and LSTMs for Sleep Stage Classification." *IEEE Trans. Biomed. Eng.*, 67(2), 359-367.
- [20]. Kim, D. & Park, Y. (2021). "Automated Sleep Stage Scoring Using Deep Learning with PSG Data." *J. Biomedical Informatics*, 116, 103746.
- [21]. Supratak, A., Dong, H., Wu, C. & Guo, Y. (2017). "DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG." *Deep Learning IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11), 1998-2008.
- [22]. Chambon, S., Galtier, M., Arnal, P., Wainrib, G. & 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(4), 758-769.
- [23]. Tzialis, O., Matthews, P. & Guo, Y. (2016). "Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders." *Annals of Biomedical Engineering*, 44(5), 1587-1597.
- [24]. Biswal, S., Sun, H., Goparaju, B., Westover, M., Sun, J. & Bianchi, M. (2018). "SleepNet: Automated Sleep Staging System via Deep Learning." *arXiv preprint arXiv:1707.08262*.
- [25]. Mousavi, S., Afghah, F. & Acharya, U. (2019). "SleepEEGNet: Automated Sleep Stage Scoring with Sequence-to-Sequence Deep Learning Approach." *PLoS ONE*, 14(5), e0216456.
- [26]. Hassan, A. & Bhuiyan, M. (2016). "Automatic Sleep Scoring Using Statistical Features in the EMD Domain and Ensemble Methods." *Biocybernetics and Biomedical Engineering*, 36(1), 248-255.
- [27]. Sors, A., Bonnet, S., Mirek, S., Vercueil, L. & Payen, J. (2018). "A Convolutional Neural Network for Sleep Stage Scoring from Raw Single-Channel EEG." *Biomedical Signal Processing and Control*, 42, 107-114.
- [28]. Seo, H., Back, S., Lee, S. & Park, K. (2021). "Attention-Based Deep Neural Networks for Automatic Sleep Stage Classification." *Sensors*, 21(4), 1280.
- [29]. Perslev, M., Darkner, S., Kempfner, L., Nikolic, M., Jennum, P. & Igel, C. (2021). "U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging." *Advances in Neural Information Processing Systems (NeurIPS)*, 34, 4417-4428.
- [30]. Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C., Li, X. & Guan, C. (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.