Squeeze-Excite Fusion Based Multimodal Neural Network for Sleep Stage Classification with Flexible EEG/ECG Signal Acquisition Circuitry

Shauilin Tao[1,2], Jinhai Hu[1,2], Wang Ling Goh[1], Yuan Gao[2]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Institute of Microelectronics (IME), Agency for Science, Technology and Research (A*STAR), Singapore
Email: shuailin001@e.ntu.edu.sg, ewlgoh@ntu.edu.sg, gaoy@ime.a-star.edu.sg

***Abstract*—This paper presents a multimodal fusion strategy for sleep stage classification based on polysomnography (PSG) which consists of electroencephalogram (EEG) and Electrocardiogram (ECG) data. By implementing the Squeeze-Excite (SE) Fusion mechanism, a more effective contribution of EEG and ECG signals to the neural network classification is ensured. To address the challenges of imbalance in the dataset, a balanced sampler is introduced. The EEG signals are transposed to the frequency domain utilizing linear-frequency cepstrum coefficients (LFCC), enhancing the feature extraction process. A recurrent convolutional neural network (RCNN) is employed to reduce the number of model parameters and further optimize the architecture. The weight of the network is then quantized down to INT4 to ensure hardware compatibility, particularly for edge devices. By employing these methods on signals, this optimized approach has achieved a validation accuracy of 77.6% with only 23.5KB weight memory size on the MIT-BIH dataset across six classification categories.**

***Index Terms*—sleep stage classification, channel attention, multimodal neural network, flexible circuit and system**

I. Introduction

Sleep is an essential and crucial state of human body for individual's health and well-being. Sleep disorders or irregular patterns can lead to various health issues, including cognitive dysfunction, cardiovascular disease, and metabolic disorders [1]. The Polysomnogram (PSG) is a primary tool for analyzing sleep patterns and diagnosing related disorders. It records bio-physiological changes during sleep and offering detailed insights into one's sleep structure [2]. Electroencephalogram (EEG) and Electrocardiogram (ECG) are two primary components of PSG, contributing to a holistic understanding of an individual's physiological activities during sleep [3]. The EEG measures the brain's electrical activity, helping to identify different sleep stages, essential for assessing sleep quality and disorders. Each stage has unique EEG patterns representing sleep health [4]. The ECG, on the other hand, captures the heart's rhythms, which can suggest sleep quality and disturbances [5]. While traditional analysis relies on handcrafted features and standard machine learning methods, the emergence of deep learning has spurred interest in developing advanced models for interpreting these signals [6].

In the quest for enhanced classification accuracy, recent work reported using multimodal method to combine different types of data, specifically the amalgamation of ECG and EEG signals. This approach achieved remarkable improvement of the classification results [7]. Conventionally, feature maps from different signals are concatenated in the model without special processing, leading to a lack of weight information for different channels.

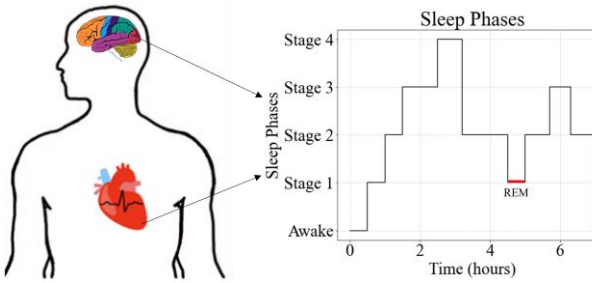

Fig. 1. Illustration of sleep stage recognition with EEG and ECG signal.

This paper presents a SE-Fusion RCNN model optimized for sleep stage categorization. This study employs a flexible circuitry board for signal acquisition, and transforming EEG signals into the frequency domain using Linear Frequency Cepstral Coefficients (LFCC). A refined mechanism, particularly channel attention inspired by SENet [8], is deployed. This mechanism is adeptly positioned post the convolutional layers and preceding the final classification Fully Connected (FC) layer. By applying channel attention mechanism to the concatenated feature map, this SE-Fusion method carefully assigns weights to each signal channel, emphasizing their importance in the subsequent classification process. The SE-Fusion channel attention model not only accentuates the crucial features intrinsic to each modality but also emboldens the model to glean insights from the inter-modal relationships, thereby substantially enhancing the accuracy and reliability of sleep stage classification. The integration of the SE- Fusion mechanism with the RCNN model enhances classifica- tion, and weight quantization to INT4 ensures low model size and hardware compatibility, broadening potential applications.

The remainder of the paper is organized as follows: Section II expounds the proposed framework, delineating the data processing, architecture and methodology. Section III delves into the flexible circuit system and data retrieval process, elucidating the mechanisms for signal acquisition. Then it presents a thorough examination of the classification outcomes and benchmarking against extant methodologies, followed by a comprehensive summary and discussion in Section IV.

This work was supported by the Agency for Science, Technology and Research (A*STAR), Singapore under the Nanosystems at the Edge Programme, grant no. A18A1b0055.

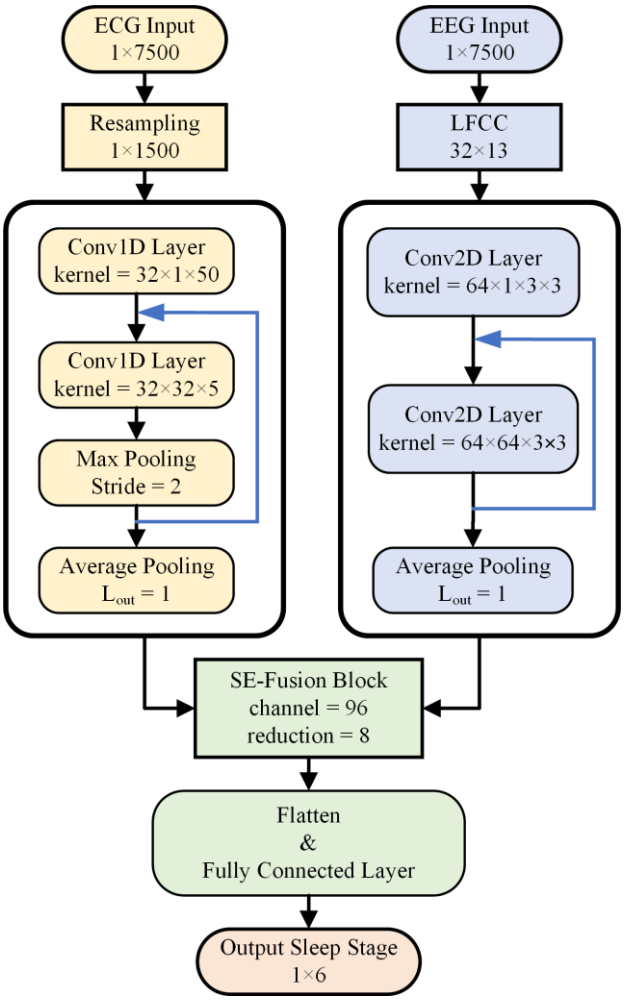

Fig. 2. Proposed PSG classification algorithm.

II. Proposed Framework

The proposed framework, as depicted in Fig. 2, elucidates the SE-Fusion RCNN algorithm designed to enhance PSG signal classification. The model architecture encompasses four pivotal layers: the input layer, RCNN layers, the SE-Fusion layer, and the final output layer. The neural network model takes both ECG and EEG inputs. The ECG data is resam- pled to 1x1500, followed by two 1-dimensional convolutions, where the second convolution layer is reused. Then it goes through the max pooling and average pooling, resulting in a 1-dimensional output. Simultaneously, the EEG data under- goes LFCC transformation to produce a 32x13 matrix, which then passes through two 2-dimensional convolutional layers followed by average pooling, also producing a 1-dimensional output. Similarly, the second convolutional layer is recurrent. These two outputs are combined with the SE-Fusion method, and then process through the final output layer. The subsequent subsections delineate the core components and operations involved in this framework.

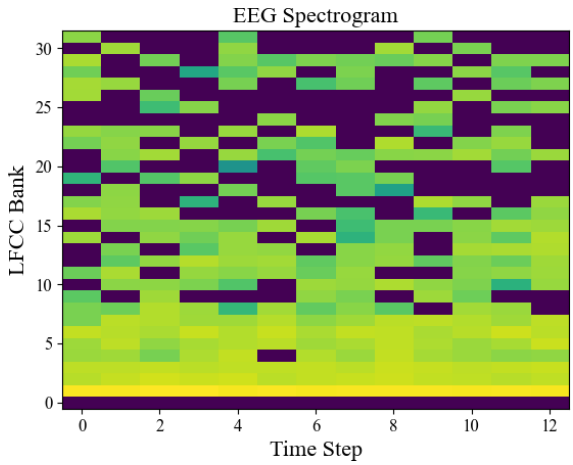

Fig. 3. LFCC spectrogram for EEG signal.

A. Data Pre-processing

The initial phase of the proposed framework necessitates a meticulous preprocessing of the raw EEG and ECG data, contingent on the unique attributes and diagnostic relevance of each signal type for sleep stage classification. A pivotal aspect of this preprocessing is the domain analysis, wherein the EEG signal is transitioned into the frequency domain while the ECG signal is retained in the time domain. The rationale behind this domain-specific preprocessing lies in the inherent characteristics and information content of the EEG and ECG signals. Specifically, the EEG signal is transformed into a 2D frequency domain feature map utilizing Linear Frequency Cepstral Coefficients (LFCC) as shown in Fig. 3. This transformation is instrumental in accentuating the powers of certain characteristic activities, thereby aiding in the delin- eation of various sleep stages [9] [10]. The frequency domain representation of EEG signals encapsulates critical information regarding the sleep state of the individual, thereby providing a fertile ground for further analysis and classification.

Conversely, the ECG signal is subjected to a resampling process, reducing the sampling rate from 250Hz to 50Hz, without undergoing a domain transformation. This approach is adopted to preserve the time domain characteristics of the ECG signal, which are imperative for extracting variability features [11]. These features are emblematic of the autonomic nervous system's activity during sleep, thereby contributing significantly to the ensuing classification process. The time domain analysis of ECG signals facilitates a more intuitive extraction of variability features, thus providing invaluable insights into cardiac behavior during different sleep stages. This domain-centric preprocessing approach underpins the subsequent analytical processes within the proposed frame- work, ensuring that the critical information intrinsic to both EEG and ECG signals is harnessed effectively for accurate sleep stage classification.

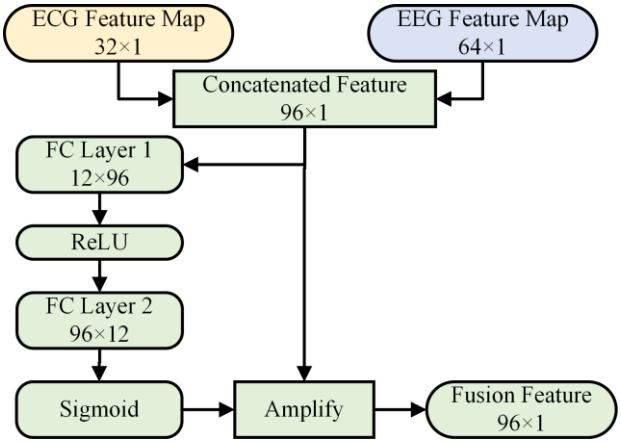

Fig. 4. SE-Fusion for ECG and EEG features.

B. RCNN Mechanism

Following preprocessing, the ECG and EEG signals are directed to their respective RCNN layers. The 1D CNN layers scrutinize the ECG signal's features through a time-series approach, translating them into a high-dimensional embedding vector. Simultaneously, 2D CNN layers explore the EEG's feature map to learn hierarchical structures and extract salient features, setting the stage for subsequent operations. Each block recurrents two times with the same weight matrices. This mechanism saves the parameter size while keeps the classification accuracy [12]. The CNN serves as token mixers, capturing local patterns and features efficiently.

C. SE-Fusion Mechanism

Post this phase, the signals converge at the SE-Fusion layer, as illustrated in Fig. 4. This layer employs a soft attention mechanism on individual channels to focus on channel-wise information between different features. It undergoes two operations: squeezing, which aggregates the feature maps across their spatial dimensions to produce a channel descriptor, and excitation, which utilizes this descriptor to modulate the original feature maps, enhancing informative features while suppressing less useful ones. This SE-Fusion mechanism improves the representation of ECG and EEG signals by allocating different weights to different channels, forming the basis for SE-Fusion method aimed at achieving an efficacious multi-modal fusion. Subsequent to the SE-Fusion operation, the re-calibrated feature maps are relayed to the output Fully Connected (FC) layer for final category classification, with provision for fine-tuning if necessitated, thereby completing the proposed framework.

D. Quantization

All weights in the framework are converted into signed INT4 format using a standard scaling factor of 0.0625, equiv- alent to 2^{-4}. This conversion limits the trainable parameters to a set of 31 values, spanning from -0.4375 to +0.4375. Such quantization bounds can help mitigate issues related to internal covariate shifts, specifically by preventing gradients from vanishing beyond the set weight limits. All the layers in this model framework are in the same scale, and trained via Quantization Aware Training (QAT), eliminating the need for batch normalization. The quantized model will be adaptable to use on portable devices.

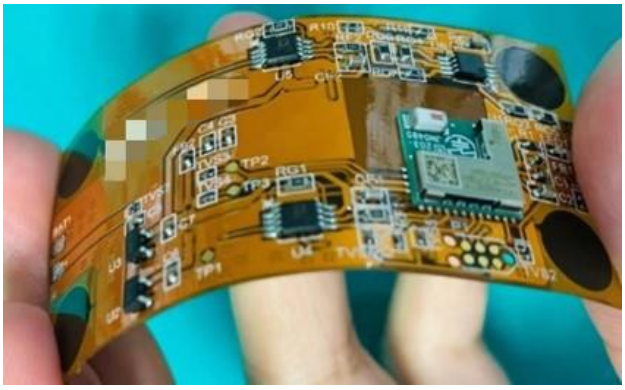

Fig. 5. The flexible circuit used.

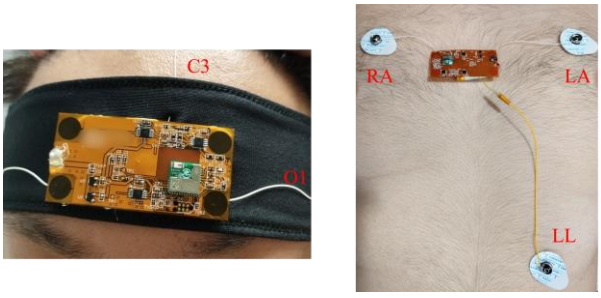

Fig. 6. ECG and EEG signal acquisation.

III. Experiment and Results

A. Bio-signal Acquisition

Two sets of bio-signal acquisition circuit modules are used to capture EEG and ECG signals, separately. Both acquisition circuit have the same architecture, which includes a low-noise analog front-end amplifier, an analog-to-digital converter (ADC) and a microcontroller with build-in Bluetooth wireless function for data transmission. The low-noise amplifier has a fully differential structure and variable gain [XX]. The amplified analog signal is passed to the SAR-ADC in the Bluetooth MCU for digitization and wireless transmission to mobile phone for signal processing. As shown in Fig. 6, a flexible circuit board is used so that the circuit module is conformal to the skin. For EEG signal acquisition, electrodes are placed at positions C3 and O1, with O1 serving as the reference electrode. Similarly, ECG data acquisition is via the electrodes placed at points LL, LA and RA is the reference electrode with a modified limb lead II (MLII) approach for data acquisition. The transmitted data will be saved by a data logger in the mobile phone for further signal analysis.

TABLE I
BENCHMARK TABLE

	ISCAS 2020 [13]	TAI 2020 [14]	EMBC 2021 [15]	ISCAS 2021 [16]	BioCAS 2022 [17]	BioCAS 2023 [18]	TBME 2023 [19]	JBHI 2023 [20]	**This Work**
Dataset	Sleep-EDF	MASS-SS3	Charite´	UCD	HMC	HMC	TBI	SleepEDF-20	**MIT-BIH**
Features	EEG	EEG, EMG, EOG	ECG, Respiration, Acceleration	ECG	ECG, EEG, EMG, EOG	EEG, EMG, EOG	ACC, PGG	EEG, EOG	**ECG, EEG**
Signal Channels	1	25	3	1	5	4	2	4	**2**
No. of Classes	5	5	4	4	5	5	4	5	**6**
Topology	Bi-LSTM	SleepPrintNet (CNN-based)	Random Forest	CNN	Transformer	LSTM-RNN	CNN	Transformer	**RCNN**
Feature Fusion	NA	Concatenation	NA	NA	NA	NA	NA	Encoder Fusion	**SE-Fusion**
No. of Layers	6	6	1	8	4	2	5	11	**5**
Weight Precision	FP32	FP32	FP32	Binary	FP32	FP32	FP32	FP32	**INT4**
No. of Parameters (K)	5309	8033	82.23	50.91	10138	8921	240.70	24025	**47.04**
Weight Memory Size (KB)	21236	32132	328.92	12.73	40548	35684	962.80	96101	**23.50**
Accuracy (%)	82.9	79.7	81.8*/69.0**	75.6	71.0	75.9	90.3*/71.0**	87.2	**91.3*/77.6****

*Train accuracy **Test accuracy

TABLE II
COMPARASION BETWEEN DIFFERENT CHANNELS AND FEATURE FUSION METHODS

	Single ECG Feature	Single EEG Feature	Concatenated Feature	SE-Fusioned Feature
Test Accuracy	59.0	69.3	71.0	77.6
Weighted Avg *Pre*	62.0	68.5	72.0	78.0
Weighted Avg *Rec*	60.0	70.0	71.0	78.0
Model Parameter	7.9 K	37.8 K	45.7 K	47.0 K

B. Result and Discussion

The effectiveness of the proposed algorithm was evaluated using the MIT-BIH dataset [23]. This dataset comprises 80 hours of sleep data obtained from ECG recordings and a single EEG channel. The physiological signals were sampled at 250 Hz, with sleep stages annotated every 7,500 samples. The dataset classifies sleep into six stages: W (wake), R (REM), and sleep stages 1 through 4, reflecting sleep depth. Table I summarizes the algorithm's performance. Despite having a significantly smaller network size than most, it not only had the most classification classes but also achieved higher accuracy than many leading methods.

To better illustrate the impact of SE-Fusion on feature concatenation, its performance is compared with single signals, basic feature concatenation, and SE-Fusion features as shown in II. When compared to the single ECG feature and single EEG feature, the fusion approach improved test accuracy by 18.6% and 8.3%, respectively. Additionally, the model size remained relatively compact at 47.04K, increased only 2.84% than the concatenated feature method but yield 6.6% higher test accuracy.

IV. CONCLUSION

This paper introduces an advanced methodology for sleep stage classification using PSG data. By incorporating the SE-Fusion mechanism, the study ensures a synergistic utilization of EEG and ECG signals. Addressing dataset imbalances with a balanced sampler and enhancing feature extraction through the use of LFCC are pivotal to this methodology. Additionally, the employment of a RCNN streamlines the architecture, while weight quantization to INT4 ensures broader hardware compatibility, notably with edge devices. This approach underscores the potential of combining various techniques and optimizations to achieve reliable sleep stage categorization. The backbone model will be considered to facilitate transfer learning on the collected data in future work.

REFERENCES

[1] M. Hirshkowitz, K. Whiton, S. M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E. S. Katz, L. Kheirandish-Gozal *et al.*, "National sleep foundation's sleep time duration recommendations: methodology and results summary," *Sleep health*, vol. 1, no. 1, pp. 40–43, 2015.

[2] N. R. Kramer and R. P. Millman, "Overview of polysomnography in adults," 2018.

[3] L. C. Markun and A. Sampat, "Clinician-focused overview and developments in polysomnography," *Current sleep medicine reports*, vol. 6, pp. 309–321, 2020.

[4] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, B. V. Vaughn *et al.*, "The aasm manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, vol. 176, p. 2012, 2012.

[5] I. Rechichi, M. Zibetti, L. Borz`ı, G. Olmo, and L. Lopiano, "Single-channel eeg classification of sleep stages based on rem microstructure," *Healthcare technology letters*, vol. 8, no. 3, pp. 58–65, 2021.

[6] D. Buongiorno, G. D. Cascarano, I. De Feudis, A. Brunetti, L. Carnimeo, G. Dimauro, and V. Bevilacqua, "Deep learning for processing electromyographic signals: A taxonomy-based survey," *Neurocomputing*, vol. 452, pp. 549–565, 2021.

[7] J. Li and Q. Wang, "Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation," *Information Fusion*, vol. 79, pp. 229–247, 2022.

[8] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[9] P. Kumar, A. K. Das, S. Halder *et al.*, "Time-domain hrv analysis of ecg signal under different body postures," *Procedia Computer Science*, vol. 167, pp. 1705–1710, 2020.

[10] A. Heima, S. Arun Karthick, and L. Suganthi, "Detection of sleep apnea based on hrv analysis of ecg signal," in *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*. Springer, 2019, pp. 1825–1834.

[11] Y. Zhang, B. Wang, J. Jing, J. Zhang, J. Zou, M. Nakamura *et al.*, "A comparison study on multidomain eeg features for sleep stage classification," *Computational Intelligence and Neuroscience*, vol. 2017, 2017.

[12] J. Hu, C. S. Leow, W. L. Goh, and Y. Gao, "Energy efficient software-hardware co-design of quantized recurrent convolutional neural network for continuous cardiac monitoring," in *2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, 2023, pp. 1–5.

[13] Y. Liao, M. Zhang, Z. Wang, and X. Xie, "Tri-featurenet: An adversarial learning-based invariant feature extraction for sleep staging using single-channel eeg," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2020, pp. 1–5.

[14] Z. Jia, X. Cai, G. Zheng, J. Wang, and Y. Lin, "Sleepprintnet: A multivariate multimodal neural network based on physiological time-series for automatic sleep staging," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 3, pp. 248–257, 2020.

[15] L. Weber, M. Gaiduk, R. Seepold, N. M. Madrid, M. Glos, and T. Penzel, "Enhancing current cardiorespiratory-based approaches of sleep stage classification by temporal feature stacking," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 5518–5522.

[16] A. John, B. Cardiff, and D. John, "A 1d-cnn based deep learning technique for sleep apnea detection in iot sensors," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5.

[17] I. Choi and W. Sung, "Performance assessment of automatic sleep stage classification using only partial psg sensors," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2022, pp. 670–674.

[18] ——, "Sleep model: A sequence model for predicting the next sleep stage," in *2023 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2023.

[19] M. Olsen, J. M. Zeitzer, R. N. Richardson, P. Davidenko, P. J. Jennum, H. B. D. Sørensen, and E. Mignot, "A flexible deep learning architecture for temporal sleep stage classification using accelerometry and photoplethysmography," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 1, pp. 228–237, 2023.

[20] Y. Dai, X. Li, S. Liang, L. Wang, Q. Duan, H. Yang, C. Zhang, X. Chen, L. Li, X. Li *et al.*, "Multichannelsleepnet: A transformer-based model for automatic sleep stage classification with psg," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[21] T. Moy, L. Huang, W. Rieutort-Louis, C. Wu, P. Cuff, S. Wagner, J. C. Sturm, and N. Verma, "An eeg acquisition and biomarker-extraction system using low-noise-amplifier and compressive-sensing circuits based on flexible, thin-film electronics," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 1, pp. 309–321, 2017.

[22] T.-T. Zhang, H. Son, J. Zhao, C.-H. Heng, and Y. Gao, "A 26.6–119.3-w 101.9-db snr direct digitization bio-impedance readout ic," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 9, pp. 2619–2631, 2023.

[23] Y. Ichimaru and G. Moody, "Development of the polysomnographic database on cd-rom," *Psychiatry and clinical neurosciences*, vol. 53, no. 2, pp. 175–177, 1999.