# Empowering Wearable Seizure Forecasting with Scheduled Sampling

Peikun Guo[1], Han Yu[1], Sruthi Gopinath Karicheri[2], Allen Kuncheria[2], Huiyuan Yang[1], Siena Blackwell[2], Zulfi Haneef[2], Akane Sano[1]

***Abstract*— The unpredictability of seizures imposes a significant burden on tens of millions of individuals with epilepsy worldwide. The ability to continuously monitor and forecast epileptic seizures would lead to a paradigm shift in epilepsy management. In this paper, we propose a novel progressive, personalized two-stage approach for seizure forecasting using 10-minute wearable time series data from wristbands worn by epilepsy patients. Our method effectively tackles the challenges posed by class imbalance and the complex nature of physiological signals. By measuring and ranking the reconstruction error and energy the normal samples present to a deep autoencoder and employing scheduled sampling, we demonstrate superior performance over existing deep learning models, anomaly detection methods, and class balancing during training. The proposed approach offers a promising solution for seizure forecasting and has potential applications in other medical problems characterized by imbalanced data and complex physiological signals.**

***Clinical relevance*— The study demonstrates the potential for seizure forecasting using wearable data and individualized treatment planning. Its findings also highlight the value of adaptive learning mechanisms in training deep learning models for imbalanced healthcare data.**

## I. INTRODUCTION

Epilepsy, one of the most common brain disorders, impacts over 70 million individuals globally and is characterized by unpredictable seizures [1]. This unpredictability is a major quality-of-life limiting factor for epilepsy patients [2], making seizure forecasting a critical research area in healthcare [3]. Approximately 80% of epilepsy sufferers live in low- and middle-income regions [4], thus, developing cost-effective seizure forecasting solutions can significantly narrow the epilepsy treatment gap .

The advent of wrist-worn wearable devices has spurred the development of algorithms to detect and forecast seizure events in real-world settings [5]–[8]. Despite advances, progress in improving detection accuracy and reducing false alarm rates has been impeded by class imbalance inherent to seizure datasets.

The emergence of deep learning methods has showcased their superiority over traditional feature engineering, especially in seizure detection tasks [9]–[15]. However, deep learning models bring challenges, such as capturing the temporal dependency in multi-sensor time series data, handling sparse annotations in large datasets, and the severe imbalance between positive and negative samples for tasks such as seizure events.

Anomaly detection techniques address the imbalanced label issue by identifying rare and unusual positive samples from normal samples [16]–[18]. For seizure detection, they have been adapted to use the loss of reconstructing encephalogram (EEG) signals to detect abnormal seizure events, showing significant improvements over supervised methods [19]–[22]. Yet, the application of anomaly detection in seizure forecasting remains challenging due to the heterogeneity of data across patients and the focus of anomaly detection on historical data.

Scheduled sampling for training, inspired by the concept of curriculum learning, has been used to gradually increase the complexity of the training data. This strategy of progressively introducing more complex tasks to the model is analogous to human learning [23]. It has seen application in the computer vision domain in tasks such as category discovery [24]. Scheduled sampling methods have been proposed for sequence-to-sequence predictions to improve Recurrent Neural Network (RNN) model performance [25]. In seizure detection and forecasting, the severe class imbalance makes it difficult for traditional deep learning models to learn discriminative features effectively. Employing scheduled sampling aims to improve the deep model's ability to provide more accurate and robust seizure forecasting.

In this study, we present a two-stage learning methodology, with a pretraining phase employing a deep autoencoder to obtain the representations of the normal samples and a fine-tuning stage employing scheduled sampling for classification. Initially, difficulty scores in anomaly detection are calculated as a combination of signal reconstruction loss from an autoencoder structure and distribution energy from a Gaussian mixture model (GMM)-based regularization module. Intuitively, the model prioritizes mastering easier tasks, such as identifying less ambiguous seizure patterns, before progressing to more complex challenges, such as recognizing subtle seizure indicators. This curriculum learning approach enables our model to efficiently learn from a diverse range of seizure patterns, ultimately reducing the false alarm rate for seizure events. We evaluate our proposed framework using two seizure datasets. Our experiments demonstrate the proposed method achieved substantial improvement when compared to supervised learning and anomaly detection baselines.

*This work was supported by Michael E. DeBakey VA Medical Center Bridge and Seed Award Program and National Science Foundation

[1]P. Guo, H. Yu, H. Yang, and A. Sano are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005, USA. Akane.Sano@rice.edu

[2]S. Gopinath Karicheri, A. Kuncheria, S. Blackwell, and Z. Haneef are with the Neurology Department, Baylor College of Medicine and the VA Medical Center, Houston, TX 77030, USA. Zulfi.Haneef@bcm.edu

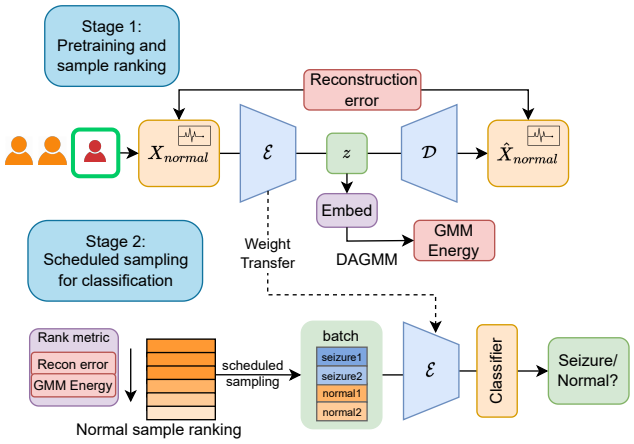


Fig. 1. The two-stage proposed method. Stage one focuses on representation learning, which reconstructs the normal samples from patient's wearable data, introducing a Gaussian Mixture Model (GMM) based regularization term to enhance diversity in latent representations. The second stage implements scheduled sampling during training, which leverages difficulty scores derived from the combination of reconstruction loss and the distribution likelihood of GMM regularization.

## II. METHOD

Figure 1 illustrates the two-stage framework of the proposed method. Initially, we employ an autoencoder to pre-train the network and impose a constraint on the latent space representations to unify the distribution of normal samples. In the supervised learning phase, we address the data imbalance by under-sampling negative samples based on energy scores obtained during the autoencoder phase. The seizure forecast is generated by a classifier built upon the pre-trained encoder.

### A. Signal reconstruction

We use a basic autoencoder structure for latent representation extraction from wearable signals. The autoencoder consists of an encoder $\mathscr{E}$ and a decoder $\mathscr{D}$. The encoder compresses the input signal $\mathbf{x}$ into a lower-dimensional latent space representation $\mathbf{z} = \mathscr{E}(\mathbf{x})$. The decoder reconstructs the original signal $\hat{\mathbf{x}} = \mathscr{D}(\mathbf{z})$. The goal is to minimize the reconstruction error, measured as the mean squared error between $\mathbf{x}$ and $\hat{\mathbf{x}}$.

### B. GMM-based Energy Constraint Module

To ensure the latent space representation of normal samples follows a unified distribution, we implement a GMM based energy constraint module, inspired by [17]. This module employs an embedding layer and a coefficient estimation network, designed to learn the mixture coefficients, means, and covariances of the GMM. We minimize the sample energy $E(\mathbf{z})$ during autoencoder training, encouraging the model to learn representations of normal samples under the same Gaussian mixture distribution.

### C. Encoder Fine-tuning with Scheduled Sampling during Training

After training the Deep Autoencoding Gaussian Mixture Model (DAGMM) on normal samples, we use these difficulty scores to fine-tune the encoder in the second stage. We implement scheduled sampling during training, defined by Algorithm 1, and initialize the encoder with the pre-trained encoder from the first stage. Our method pairs seizure-positive samples with normal samples drawn from the ranked pool of normal samples in sequential order. This strategy enables the model to learn progressively from easier to more challenging samples.

**Algorithm 1** Two-Stage Scheduled Sampling for Seizure Forecasting

**Input:** Data $\mathscr{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where $\mathbf{x}_i$ represents the multivariate time series and $y_i$ is the label (seizure or non-seizure)

**Output:** Trained model $M$

**Stage 1:** Train a DAGMM on normal samples, define the loss function as $\mathscr{L}_1(\mathbf{x}, \hat{\mathbf{x}}, \mathbf{z}) = \text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) + \text{GMM}(\mathbf{z})$.
**for** each normal sample $(\mathbf{x}_i, y_i) \in \mathscr{D}$ with $y_i = 0$ **do**
  Calculate reconstruction error $\varepsilon_i$, GMM energy $E_i$, and difficulty score $d_i = \varepsilon_i + E_i$
**end for**
**Stage 2:** Fine-tune the encoder using scheduled sampling during training, initialize the encoder $E$ with the pre-trained DAGMM encoder, and define the classification loss function as $\mathscr{L}(\mathbf{y}, \hat{\mathbf{y}}) = \text{CrossEntropy}(\mathbf{y}, \hat{\mathbf{y}})$.
**for** each iteration $t$ **do**
  Sample normal samples $(\mathbf{x}_i, y_i)$ based on difficulty scores $d_i$ and all seizure-positive samples $(\mathbf{x}_j, y_j)$ with $y_j = 1$.
  Update the model $M$ using the sampled data and the loss function $\mathscr{L}(\mathbf{y}, \hat{\mathbf{y}})$.
**end for**
**return** Trained model $M$

## III. DATA & PREPROCESSING

### A. My Seizure Gauge (MSG) Challenge dataset

This dataset, used in the My Seizure Gauge Project funded by the Epilepsy Foundation of America [26], provides physiological data recorded from epilepsy patients via Empatica E4 wristbands. The dataset contains 129829 10-minute segments from 6 patients, of which 1089 segments are labeled positive (pre-ictal). The device records 3D accelerometry (ACC), blood volume pulse (BVP), electrodermal activity (EDA), and temperature (TEMP) at various sampling rates, which were subsequently resampled to 4Hz for the sole purpose of reducing the computational complexity. The data span a minimum of six months for each patient, with seizures accurately logged via an implanted NeuroPace Responsive Neurostimulation (RNS) device.

Each ten-minute data segment in the public dataset is labeled 0 or 1, where a label of 1 denotes that more than half of the segment data falls within 75-15 minutes before a seizure onset, and a setback of 15 minutes is incorporated. Segments within four hours of a previous seizure are unlabeled due to potential influences from the preceding

seizure. For preprocessing, we resampled the signals to 4Hz and applied the Butterworth low-pass filter (cutoff frequency 64Hz) and min-max normalization.

### B. Baylor College of Medicine (BCM) Epilepsy dataset

A clinical study was conducted to collect data using an E4 sensor from 58 subjects diagnosed with epilepsy and admitted to the long-term video-EEG monitoring unit (IRB protocol #H-47804). The data include 24164 10-minute segments, of which 55 are labeled positive. The study accounted for all seizure types identified by board-certified epilepsy specialists. The labeling and preprocessing of the BCM data align with that of the MSG dataset.

We split each dataset by allocating 80% for training and 20% for validation, ensuring a consistent ratio of normal to seizure samples in both training and validation sets.

## IV. Experiments

This section offers both quantitative and qualitative evaluations of our two-stage scheduled learning approach to seizure forecasting from 10-minute wearable data segments. It was tested against anomaly detection techniques (DAGMM), end-to-end deep learning models (ResNet, LSTM) without scheduled sampling, with random balanced sampling, and with all patients data, and data balancing techniques (upsampling/downsampling). Both patient-wise models and general models were tested for the MSG dataset. However, for the BCM dataset, only the general model was tested due to limited seizure data. AUC-ROC, precision, recall, and F1 scores were used as performance measures.

### A. Implementation Details

A modified ResNet18 architecture, adapted for time series data, was used for the DAGMM encoder in the first stage of our method, focusing on normal non-seizure samples. The decoder was constructed with transposed convolution, mirroring the encoder's architecture. The mean square error (MSE) loss function was utilized to assess the decoder's output against the input. For the GMM layer, we followed [17], setting the number of GM components to 4.

In the second stage, we ranked normal samples based on the sum of GMM energy and reconstruction score, indicating their "difficulty" levels. Then, we fine-tuned the pre-trained encoder for supervised classification. We addressed the class imbalance issue by upsampling seizure data samples during each epoch of classification training.

Both stages were trained using the AdamW optimizer [27] with a learning rate of 0.001 and a batch size of 64. Early stopping was employed based on validation loss. We used 50 epochs for autoencoder training and 30 epochs for classifier training. To improve generalization, we applied random jittering and scaling as data augmentation techniques during training.

## V. Results & Discussion

*a) Impact of Scheduled Sampling during Training:* In our results shown in Tables I and II, the scheduled sampling process consistently enhances model performance metrics, confirming effective utilization of sample difficulties. This is especially evident when comparing our proposed two stage method (with scheduled sampling) with the model with random balanced sampling but no designed schedule (ResNet-18*), which presents lower AUC and recall, validating the benefits of our sampling strategy.

TABLE I

Performance comparison for MSG challenge dataset. ResNet-18* denotes ResNet-18 trained with random balanced sampling (RBS), as a comparison to scheduled sampling (SS). All models but ResNet-18 (all patients) are personalized. For all models, AUC-ROC is calculated patient-wise and then averaged.

| Method | AUC-ROC | Precision | Recall | F1 Score |
|---|---|---|---|---|
| DAGMM | 0.47 | 0.41 | 0.49 | 0.40 |
| ResNet-18 (w/o SS) | 0.56 | 0.50 | 0.51 | 0.34 |
| ResNet-18 (w/o SS, all-patients) | 0.55 | 0.50 | 0.55 | 0.26 |
| ResNet-18* (with RBS) | 0.59 | 0.51 | 0.54 | 0.30 |
| Ours | **0.64** | **0.51** | **0.66** | **0.48** |

TABLE II

Performance comparison for the BCM dataset (general models)

| Method | AUC-ROC | Precision | Recall | F1 Score |
|---|---|---|---|---|
| DAGMM | 0.52 | 0.49 | 0.51 | 0.32 |
| ResNet-18 (w/o SS) | 0.64 | 0.51 | 0.53 | 0.36 |
| ResNet-18* (with RBS) | 0.69 | 0.50 | 0.50 | 0.38 |
| Ours | **0.74** | **0.51** | **0.60** | **0.51** |

*b) Patient-wise Training and Evaluation:* We conduct our experiments patient-wise for two reasons: clinical relevancy and accuracy. Patient-wise modeling caters to individual differences in seizure patterns, physiological signals, and treatment responsiveness [28], an essential perspective for the highly heterogeneous nature of epilepsy [7], [29]. The results in Table I reveal a significant decrease in accuracy when a general model is trained on an all-patient basis. For evaluation, the patient-averaged AUC is more suitable than the overall AUC to reflect the true performance of the method on individual patients, as the models are trained individually.

*c) Two-stage Pipeline:* Autoencoder-based anomaly detection can reconstruct abnormal (seizure) samples with high fidelity when even trained solely on normal samples, limiting their ability to differentiate between seizure and normal data. Table I and II show that the inferior performance of DAGMM necessitates the use of a two-stage mechanism, incorporating a fine-tuning stage for the encoder. As illustrated in Figure 2, our scheduled sampling strategy implemented in the fine-tuning stage not only improves the validation AUC, but also facilitates faster convergence.

*d) Limitations and Future Work:* Our approach was evaluated on a small dataset of six epilepsy patients, limiting generalizability. Future research should aim to validate our method on larger, more diverse datasets. Moreover, it would be beneficial to investigate the potential of more advanced deep learning architectures and the applicability of our

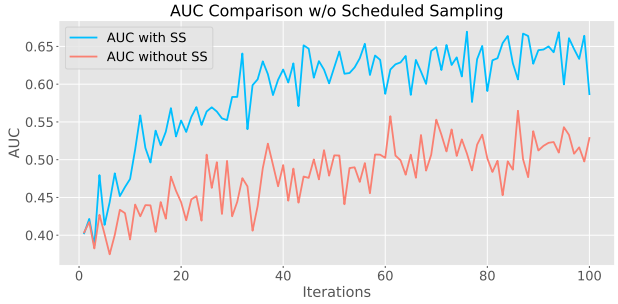


Fig. 2. Validation AUC by iterations, with/without scheduled sampling (SS).

method to other medical conditions characterized by highly imbalanced data and complex physiological signals.

## VI. CONCLUSIONS

Our two-stage scheduled learning approach provides a promising solution to the problem of seizure forecasting using wearable data segments. The method effectively addresses the issues of class imbalance and the complex nature of epilepsy-related physiological signals. By employing a scheduled sampling strategy inspired by the human learning process, our approach successfully outperforms existing deep learning models and anomaly detection methods, highlighting the importance of incorporating adaptive learning mechanisms in the training of deep neural networks for seizure forecasting.

## REFERENCES


[1] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in adults," *The Lancet*, vol. 393, no. 10172, pp. 689–701, 2019.

[2] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel Jr, L. Forsgren, J. A. French, M. Glynn, *et al.*, "Ilae official report: a practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.

[3] S. R. Haut, C. B. Hall, A. J. LeValley, and R. B. Lipton, "Can patients with epilepsy predict their seizures?," *Neurology*, vol. 68, no. 4, pp. 262–266, 2007.

[4] S. Saxena and S. Li, "Defeating epilepsy: a global public health commitment," *Epilepsia open*, vol. 2, no. 2, pp. 153–155, 2017.

[5] G. Regalia, F. Onorati, M. Lai, C. Caborni, and R. W. Picard, "Multimodal wrist-worn devices for seizure detection and advancing research: focus on the empatica wristbands," *Epilepsy research*, vol. 153, pp. 79–82, 2019.

[6] W. Li, D. Sheng, G. Wang, and G. Wang, "Generalized tonic-clonic seizure detection through the use of physiological signals," in *2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML)*, pp. 228–233, IEEE, 2022.

[7] M. Nasseri, T. Pal Attia, B. Joseph, N. M. Gregg, E. S. Nurse, P. F. Viana, G. Worrell, M. Dümpelmann, M. P. Richardson, D. R. Freestone, *et al.*, "Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning," *Scientific reports*, vol. 11, no. 1, p. 21935, 2021.

[8] F. Onorati, G. Regalia, C. Caborni, M. Migliorini, D. Bender, M.-Z. Poh, C. Frazier, E. Kovitch Thropp, E. D. Mynatt, J. Bidwell, *et al.*, "Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors," *Epilepsia*, vol. 58, no. 11, pp. 1870–1879, 2017.

[9] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for patient-independent epileptic seizure prediction using scalp eeg signals," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9377–9388, 2021.

[10] X. Wei, L. Zhou, Z. Chen, L. Zhang, and Y. Zhou, "Automatic seizure detection using three-dimensional cnn based on multi-channel eeg," *BMC medical informatics and decision making*, vol. 18, no. 5, pp. 71–80, 2018.

[11] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, 2019.

[12] Y. Sun, W. Jin, X. Si, X. Zhang, J. Cao, L. Wang, S. Yin, and D. Ming, "Continuous seizure detection based on transformer and long-term ieeg," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5418–5427, 2022.

[13] A. Bhattacharya, T. Baweja, and S. Karri, "Epileptic seizure prediction using deep transformer model," *International journal of neural systems*, vol. 32, no. 02, p. 2150058, 2022.

[14] R. Chen and K. K. Parhi, "Seizure prediction using convolutional neural networks and sequence transformer networks," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 6483–6486, IEEE, 2021.

[15] X. Zhang and H. Li, "Patient-specific seizure prediction from scalp eeg using vision transformer," in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, vol. 6, pp. 1663–1667, IEEE, 2022.

[16] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *International conference on machine learning*, pp. 4393–4402, PMLR, 2018.

[17] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[18] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 665–674, 2017.

[19] A. Farahat, D. Lu, S. Bauer, V. Neubert, L. S. Costard, F. Rosenow, and J. Triesch, "Diagnosing epileptogenesis with deep anomaly detection," 2022.

[20] S. You, B. H. Cho, Y.-M. Shon, D.-W. Seo, and I. Y. Kim, "Semi-supervised automatic seizure detection using personalized anomaly detecting variational autoencoder with behind-the-ear eeg," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106542, 2022.

[21] J. Xu, Y. Zheng, Y. Mao, R. Wang, and W.-S. Zheng, "Anomaly detection on electroencephalography with self-supervised learning," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 363–368, IEEE, 2020.

[22] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for medical anomaly detection–a survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 7, pp. 1–37, 2021.

[23] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

[24] Y. J. Lee and K. Grauman, "Learning the easy things first: Self-paced visual category discovery," in *CVPR 2011*, pp. 1721–1728, IEEE, 2011.

[25] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," *Advances in neural information processing systems*, vol. 28, 2015.

[26] EvalAI, "My seizure gauge seizure forecasting challenge." https://eval.ai/web/challenges/challenge-page/1693/overview, accessed 2023.

[27] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.

[28] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction—ready for a new era," *Nature Reviews Neurology*, vol. 14, no. 10, pp. 618–630, 2018.

[29] M. Nasseri, T. P. Attia, B. Joseph, N. M. Gregg, E. S. Nurse, P. F. Viana, A. Schulze-Bonhage, M. Dümpelmann, G. Worrell, D. R. Freestone, *et al.*, "Non-invasive wearable seizure detection using long–short-term memory networks with transfer learning," *Journal of neural engineering*, vol. 18, no. 5, p. 056017, 2021.