MelicientNet: Harnessing Mel-Spectrograms and EfficientNet Architectures for Predicting Neurological Recovery Post-Cardiac Arrest

Wenlong Wu, Ying Tan

Xiaomi Health Lab, Beijing, China

Abstract

Predicting neurological recovery in patients following cardiac arrest is a critical and intricate endeavor in medical research. This study by team MIWEAR introduces a groundbreaking methodology that leverages Electroencephalogram (EEG) signals transformed into mel-spectrograms as part of the 'Predicting Neurological Recovery from Coma After Cardiac Arrest: The George B. Moody PhysioNet Challenge 2023'. By adapting these complex sequences into two-dimensional image regression structures, we tailored the data for advanced convolutional neural networks. Our innovative "MelicientNet" model synergistically combines the mel-spectrogram approach with the EfficientNet family, particularly EfficientNet-B0 and EfficientNetV2-S. This union, bolstered by the architectures' compactness and precision, offers a formidable solution for our computer vision applications. Additionally, our strategy integrates comprehensive patient profiles, amalgamating demographic and recording data into our predictive framework. Employing data from seven diverse hospitals, our best performing algorithm achieved a Challenge score of 0.71, and a ranking of 4th place overall.

1. Introduction

Cardiac arrest is a pressing global health issue, with survival rates varying widely based on location. A significant post-resuscitation challenge is potential brain injury, the leading cause of death among survivors. Most survivors are comatose upon ICU admission, and early prognosis is crucial in guiding care decisions [1]. Current prognostic methods, however, have shown inconsistencies. Electroencephalography (EEG) presents a promising tool for objective brain monitoring post-cardiac arrest [2]. Yet, manual EEG interpretation is resource-intensive and requires specialized expertise.

To address this, the George B. Moody PhysioNet Challenge 2023 [3][4], backed by the International Cardiac Arrest REsearch (I-CARE) consortium, offers an extensive EEG dataset from over 1,000 post-cardiac arrest comatose patients [5]. Our study leverages this dataset, transforming EEG signals into mel-spectrograms, simplifying the data representation for machine learning applications. Our model, "MelicientNet", combines mel-spectrograms with EfficientNet architectures and integrates demographic and recording data for a comprehensive prognosis approach. The following sections detail our methodology, results, discussions, and conclusions.

2. Methodology

2.1 System Framework

Our proposed model, termed "MelicientNet," integrates the concept of mel-spectrograms with the architectural strengths of EfficientNet. As illustrated in Fig. 1, the MelicientNet framework begins with the preprocessing of EEG signals. These preprocessed signals are then converted into mel-spectrograms. Subsequently, we employ two variants of the EfficientNet architecture—EfficientNet-B0 and EfficientNetV2-S—to analyze the mel-spectrograms. The predictions derived from both EfficientNet versions are then amalgamated with demographic data and recording specifics through a random forest model to generate final predictions.

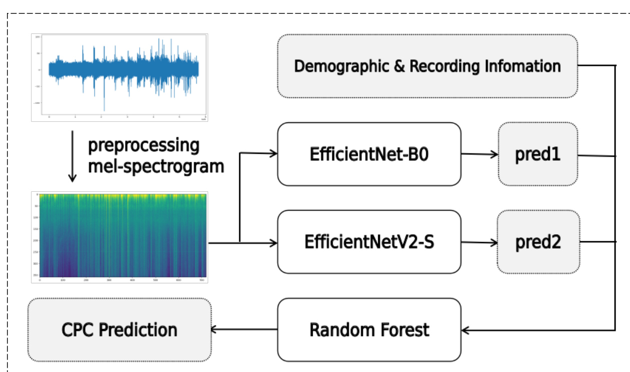

Fig. 1. Overview of the MelicientNet Framework

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.084

2.2 Evaluation Methodology

In our study, we utilized the Stratified K-fold validation method (K=5), grouped by resident, to ensure stratification according to the target variable: the Cerebral Performance Category (CPC) scale. This approach guarantees that each fold possesses a comparable distribution of CPC values. While the AUC serves as our primary evaluation metric, we also reference the challenge metric (TPR at a FPR of 0.05) for additional validation. Notably, considering the ordinal nature of the CPC value, we approached the CPC classification task as a regression problem, utilizing the MSE metric as our objective function.

2.3 Data Preprocessing

EEG recordings were consistently captured using 19 electrodes, adhering to the international 10-20 system. In the preprocessing phase, our initial step was to standardize the order of the 19 channels to the sequence. Subsequently, the data from these 19 channels was transformed into 18 bipolar channels, namely: 'Fp1-F7', 'F7-T3', 'T3-T5', 'T5-O1', 'Fp2-F8', 'F8-T4', 'T4-T6', 'T6-O2', 'Fp1-F3', 'F3-C3', 'C3-P3', 'P3-O1', 'Fp2-F4', 'F4-C4', 'C4-P4', 'P4-O2', 'Fz-Cz', and 'Cz-Pz'. We opted for bipolar referencing due to its prevalent utility in clinical settings, and the fact that many prior quantitative EEG analyses and models related to cardiac arrest have employed bipolar channels [2]. Additionally, each bipolar channel underwent processing with a 2nd order Butterworth band-pass filter, with cut-off frequencies established at 0.5 Hz and 50 Hz. This was followed by resampling to a uniform frequency of 128 Hz. During our analysis, we identified extended flat zero segments in many bipolar signals. To address this, we eliminated signal segments exhibiting flat zeros over a 10-second window.

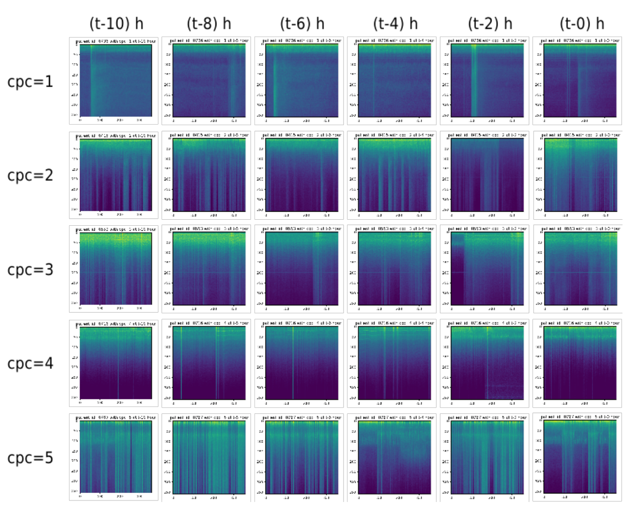

Fig. 2. Representative Mel-Spectrograms: Averages from 18 Channels Across Five Distinct CPC Levels Over Time

2.4 Spectrogram Creation

For our study, we converted one-hour segments from each bipolar channel into mel-spectrograms utilizing the 'librosa' library. We adopted a hop length of 10 seconds, ensuring that the duration of each hour is represented by 360 units. Our analysis concentrated on frequencies ranging from 0 to 45 Hz, resulting in mel-spectrograms with dimensions of 360x360. Fig. 2 provides illustrative examples, showcasing the average mel-spectrograms across 18 channels for five distinct CPC levels at varying times. For training purposes, we utilized the final 12 hours of data for each patient, translating to 12 mel-spectrogram images.

Before processing each mel-spectrogram image through our model, we enhanced its robustness by employing five specialized image augmentation techniques:

1. Adjustments to Brightness and Contrast, randomized within a 0.5 boundary (probability: 0.8).
2. Frequency Masking, capped at a 0.5 masking rate (probability: 0.8).
3. Application of Gaussian Blur, with a blur intensity set between 3 and 7 (probability: 0.6).
4. Time-based Shifting, limited to a 0.2 shift percentage (probability: 0.6).
5. The Cutout technique [6], which introduces up to 8 random holes, each with dimensions not exceeding 45x45 (probability: 0.6).

2.5 Model Architecture

EfficientNet-B0 [7] and EfficientNetV2-S [8] are compact, efficient CNN models optimized for image classification. EfficientNet-B0 achieved 77.1% top-1 accuracy on ImageNet with only 5.3M parameters. EfficientNetV2-S improved accuracy to 79.8% while reducing computations. Their efficiency and accuracy at processing complex patterns make them well-suited for analyzing EEG mel-spectrograms. We evaluate their effectiveness at predicting neurological outcomes within our proposed MelicientNet framework.

2.6 Ensemble Techniques

Alongside the EEG signals, our dataset also captures a comprehensive set of patient-specific attributes. This includes the patient's age, gender, the duration from cardiac arrest to the return of spontaneous circulation (ROSC), the location of the cardiac arrest (either out-of-hospital, denoted by OHCA, or in-hospital), the nature of the rhythm (whether it was shockable or not), and the targeted temperature management (TTM) applied. While this demographic and clinical data provides a rich context, our study further integrates these features with the predictions obtained from the mel-spectrograms. To optimize our predictive accuracy, we utilized a second-level random forest (RF) regressor model, amalgamating these diverse data sources for a more refined prediction.

2.7 Parameter Configurations

To optimize our model, we employed the Adam optimizer, initiating with a learning rate of 1e-4. This rate underwent a decay process via the CosineAnnealingLR scheduler, diminishing to a minimum value of 1e-6 throughout the training duration. Given the limitations of our hardware, we opted for a batch size of 16, ensuring efficient model training. The entire training spanned 10 epochs, a duration deemed adequate for model convergence as evidenced by the stabilization of the validation loss. The local training process was executed on an NVIDIA 2080Ti GPU. Our model, along with the optimizer, loss function, and associated configurations, was developed using the PyTorch framework. At the culmination of each epoch, checkpoints were established. The checkpoint exhibiting the most minimal validation loss was earmarked as the definitive model for evaluations.

3. Results

In our pursuit of predicting neurological outcomes, we meticulously analyzed the final 12 hours of the provided dataset. To validate the robustness of our results, we employed a 5-fold stratified cross-validation, benchmarking our outcomes using both the AUC and the challenge metric.

Table I: Comparison of different methods based on 5-fold cross validation results from training data.

Methodology	AUC	Challenge score
RF on Demographic and Recording Information	0.75	0.25
EfficientNet-B0	0.86	0.57
EfficientNetV2-S	0.85	0.59
Ensemble	0.88	0.62

Table I elucidates the AUC and challenge scores derived from diverse methodologies based on 5-fold cross validation results from training data. The ensemble model, which amalgamates predictions from EfficientNet-B0, EfficientNetV2-S, and the demographic and recording data, distinctly outperforms the other methods. Our ensemble algorithm provided a final test Challenge score of 0.71 (with an AUC of 0.92), achieving a final official ranking of 4th place.

Table II: Detailed examination of 5-fold cross validation results across training data folds.

Fold	AUC	Challenge score
0	0.89	0.61
1	0.89	0.75
2	0.92	0.68
3	0.90	0.70
4	0.82	0.60

Upon examining Table II, it's evident that the AUC value remains relatively stable across different validation folds, showcasing the model's consistent performance. In contrast, the challenge score exhibits more variability, underscoring the importance of using multiple metrics for a comprehensive evaluation.

Table III: Analytical breakdown by hospital based on training data

Hospital	AUC	Challenge score
A	0.90	0.65
B	0.69	0.36
D	0.86	0.64
E	0.94	0.95
F	0.85	0.69

The dataset spanned seven renowned hospitals, each bringing unique datasets to the table. Table III offers a granular breakdown of the AUC and challenge scores across these institutions. The variance in scores across hospitals highlights the heterogeneity in data and emphasizes the importance of a model that can generalize across diverse datasets. The amalgamation of datasets from two hospitals and the reservation of one for hidden testing further accentuates the complexity and challenges faced in this endeavor.

4. Discussions and Conclusions

4.1 Discussions

The ensemble model's standout performance, as evidenced by its final test AUC value of 0.92 and challenge score of 0.71, is a testament to the synergistic effect of combining diverse predictive models with demographic and recording data. This approach not only captures the intricate patterns within the EEG signals but also contextualizes them with patient-specific information, leading to more nuanced and accurate predictions. The ensemble model achieves a 4th place ranking on the leaderboard, demonstrating its robustness through external validation.

Diving deeper into the hospital-specific results presented in Table III, we observe a range of AUC values from 0.69 to 0.94 across training data folds. This variance underscores the inherent challenges in developing a universally applicable model across diverse datasets. Each hospital, with its unique patient demographics, equipment standards, and treatment protocols, contributes data with distinct characteristics. For instance, the relatively lower AUC value of 0.69 for Hospital B in the training data might indicate a more complex patient cohort or differences in data recording practices compared to other institutions.

Lastly, the stability of the AUC value across different validation folds, juxtaposed with the variability in the challenge score, offers a nuanced insight. While the model's ability to rank predictions remains consistent, its absolute predictive accuracy can fluctuate. This observation suggests that while the model has learned general patterns effectively, there might be specific scenarios or edge cases where its predictions are less reliable.

4.2 Conclusions

The challenge of predicting neurological recovery from coma after cardiac arrest is a complex and multifaceted one, with profound implications for patient care and clinical decision-making. In our endeavor to address this challenge, we have presented a novel approach that combines the power of mel-spectrograms with state-of-the-art convolutional neural network architectures, specifically the EfficientNet family.

Our methodological strengths lie in the innovative transformation of EEG signals into a two-dimensional format, allowing us to harness the prowess of image-based deep learning models. The incorporation of EfficientNet-B0 and EfficientNetV2-S, both of which strike a balance between computational efficiency and predictive accuracy, further bolsters our solution. Additionally, our ensemble strategy, which amalgamates predictions from multiple models with demographic and recording data, ensures a holistic and nuanced prediction.

However, as with any scientific endeavor, our approach is not without its limitations. The potential loss of temporal dynamics in EEG signals, variability across hospitals, and the inherent complexities of an ensemble model are challenges that we recognize. Moreover, the absence of trend analysis over consecutive hours in EEG signals might be a missed opportunity in capturing more granular insights.

In conclusion, our work represents a significant stride in the realm of neurological recovery prediction. By melding advanced deep learning techniques with clinical data, we hope to pave the way for more informed and patient-centric care decisions in the aftermath of cardiac arrests. As the field evolves, we remain committed to iterating on our approach, always with the aim of enhancing patient outcomes and aiding clinicians in their vital work.

References

[1] Rundgren M, Westhall E, Cronberg T, Rosen I, Friberg H. Continuous amplitude-integrated electroencephalogram predicts outcome in hypothermia-treated cardiac arrest patients. Critical Care Medicine. 2010;38(9):1838-44.

[2] Zheng WL, Amorim E, Jing J, Ge W, Hong S, Wu O, Ghassemi M, Lee JW, Sivaraju A, Pang T, Herman ST. Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks. Resuscitation. 2021;169:86-94.

[3] Reyna MA, Amorim E, Sameni R, Weigle J, Elola A, Rad AB, Seyedi S, Kwon H, Zheng WL, Ghassemi MM, van Putten MJ. Predicting Neurological Recovery from Coma After Cardiac Arrest: The George B. Moody PhysioNet Challenge 2023. Computing in Cardiology. 2023;50:1-4.

[4] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215-20.

[5] Amorim E, Zheng WL, Ghassemi MM, Aghaeeaval M, Kandhare P, Karukonda V, Lee JW, Herman ST, Adithya S, Gaspard N, Hofmeijer J, van Putten MJAM, Sameni R, Reyna MA, Clifford GD, Westover MB. The International Cardiac Arrest Research (I-CARE) Consortium Electroencephalography Database. Critical Care Medicine 2023 (in press); doi:10.1097/CCM.0000000000006074.

[6] DeVries T, Taylor GW. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552. 2017.

[7] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning 2019 (pp. 6105-6114). PMLR.

[8] Tan M, Le Q. Efficientnetv2: Smaller models and faster training. In International Conference on Machine Learning 2021 (pp. 10096-10106). PMLR.

Address for correspondence:

Wenlong Wu
33 Xi'erqi Middle Road, Haidian District, Beijing, China
wuwenlong3@xiaomi.com