2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR) | 979-8-3503-5142-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/MIPR62202.2024.00085

Harmful Brain Activity Classification of Spectrograms with Transfer Deep Learning

Shanker Ram
Lynbrook High School
San Jose, USA
shankernram@gmail.com

Sambhu Ganesan
Lynbrook High School
San Jose, USA
sambhu.ganesan150@gmail.com

Yajat Nagaraj Kiran
Lynbrook High School
San Jose, USA
yajat009@gmail.com

***Abstract*—Electroencephalograms are medical tests used by doctors to determine any irregular brain activity in patients. The drawn-out process of manually classifying each of these signals can be time-consuming, costly, and even subjective among neurologists. We make use of the HMS dataset to classify spectrograms of brain activity into one of six abnormal brain activities. By applying basic data preprocessing techniques, we show the potential and power of using signal processing techniques like the Short-Time Fourier Transform and the Continuous Wavelet Transform in creating an accurate prediction model. By training on images of spectrograms and EEG signal plots, we create a three-model ensemble that can accurately predict brain abnormalities, easing the load off physicians. Our research showcases the potential of transforms in signal processing tasks.**

***Index Terms*—ml, classification, health, data augmentation, deep learning**

I. Introduction

Electroencephalography (EEG) signals serve as vital indicators used by physicians in order to detect abnormal brain activity in patients and diagnose potential signs of brain damage such as epilepsy, sleep disorders, and brain tumors. Currently, EEG results are manually analyzed [17] by neurologists to detect abnormalities. Streamlining this process is crucial given that healthcare providers typically sift through approximately 100 pages of activity [1], which can be quite time-consuming. Furthermore, this process is prone to inter-observer variability, as the interpretation of EEG signals is subjective, which can lead to inconsistent assessments. Moreover, since this process requires multiple neurologists for review and lots of time, costs can add up quickly for patients.

Given that EEG stands as a basis for measuring electrical disturbances, there arises a critical demand for an automatic and efficient way to classify these signals, as such signals can be used in the early recognition of epileptic seizures. [2].

We propose the use of deep learning to classify spectrograms of brain activity into seizures, generalized periodic discharges, lateralized periodic discharges, lateralized rhythmic delta activity, generalized rhythmic delta activity, or other. Accurate classification of brain activity patterns can help clinicians better diagnose such conditions, as interpretations can be better standardized. We utilize the HMS dataset [3] from CCEMRC to accurately predict brain activity from spectrograms and EEG signals. This could help drastically reduce the cost and time used in manually diagnosing EEG signals.

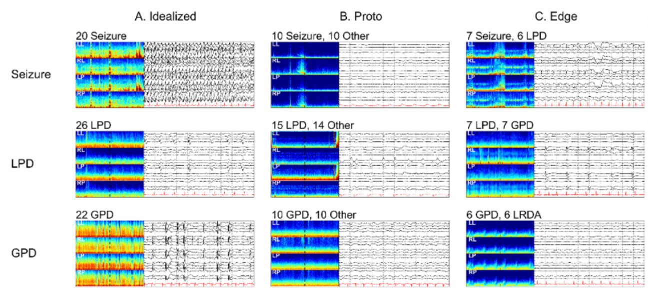

Fig. 1. Examples of Labeled Spectrograms

We divide our research into three major components–data preprocessing and augmentation, signal processing and transforms, and training of the model with checkpointing and learning rate tuning. Understanding that it is crucial to preprocess the data in order to remove irrelevant noise and also decrease the computational complexity [4], we normalized and standardized the spectrograms. After finetuning the dataset and model, we were able to achieve a peak Kullback-Leibler divergence score of 0.43 on the testing dataset.

II. Dataset

A. HMS Dataset

For this task, we made use of the Harmful Brain Activity dataset provided by CCEMRC. The dataset contains 50-second long EEG samples covering a 10-minute window as well as spectrograms matching each sample. The longer 10-minute window allows for the model to pick up on further insights regarding the nature of the brain activity. The EEG samples were labeled by expert annotators to determine which of the six features each spectrogram belonged to.

The dataset also provides a glimpse into the subjective nature of manually classifying EEG samples and spectrograms–many of the EEG segments did not have a consensus of opinion among labelers. CCEMRC has further stratified the data by labeling samples with consensus opinion as ”Idealized”, samples being split between one of the five conditions and ”other” as ”Proto”, and samples being split between two

2770-4319/24/$31.00 ©2024 IEEE
DOI 10.1109/MIPR62202.2024.00085

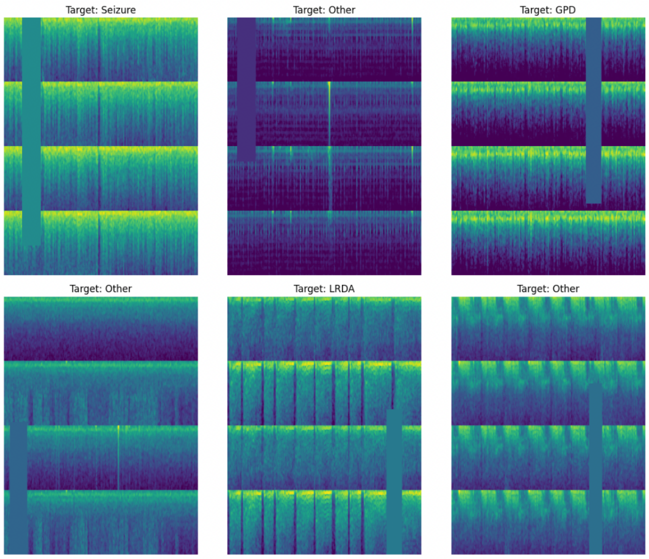

Fig. 2. Augmented Smoothed Spectrograms

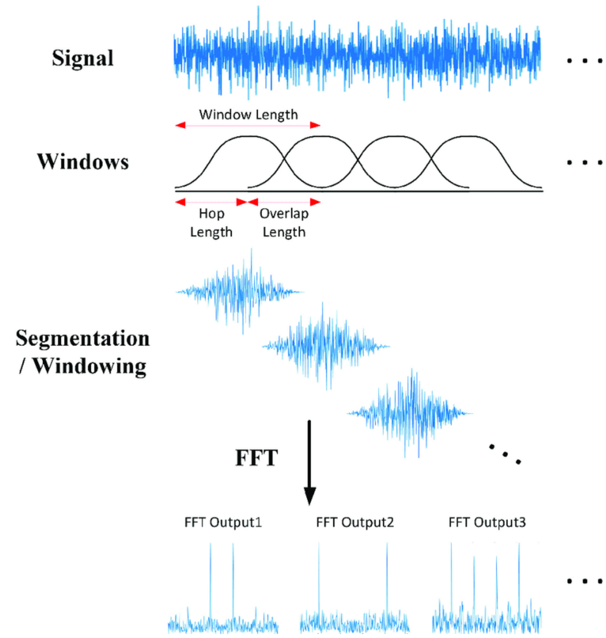

Fig. 3. Discrete Short-Time Fourier Transform

or more conditions as "Edge", as shown in Fig. 1. We make use of the difference in uncertainty levels in fine-tuning the dataset and training our model.

III. Data Augmentation and Preprocessing

A. Principle Preprocessing

To begin optimizing the form of the data, we normalize each spectrogram by ensuring that pixel values are bound in a range. We also apply a log transformation to each spectrogram so that the effect of the large number of outliers can be minimized. The data is centered around zero by subtracting the mean of the pixels from each pixel. Finally, we standardize the spectrograms to ensure an equal variance throughout images.

In order to perform data augmentation, we create an augmenter of two Keras layers: MixUp [10] and RandomCutout [12]. Research has shown that the MixUp layer, which smooths values, can stabilize the training of neural networks and improve the generalization of architectures. Similarly, the RandomCutout layer serves to randomly cut out rectangles from images (shown in Fig. 2), improving the ability of our model to generalize. The best KL divergence score achieved with this augmentation but without signal processing was 0.61.

IV. Signal Processing Techniques

We evaluated the use of different signal processing techniques to further preprocess the spectrograms, namely Short-Time Fourier Transforms (STFTs), Continuous Wavelet Transforms (CWTs), and Constant-Q transforms (CQTs). By using one or more of these transforms, the spectrograms can be manipulated and changed such that the neural network has a greater accuracy. We explore the advantages and disadvantages of each technique below.

A. Short-Time Fourier Transforms (STFT)

Let $X_m(\omega)$ be the Short-Time Fourier Transform of a window function $\omega(t)$ centered at time mR where R is the hop length. Then we have that

$$X_m(\omega) = \sum_{n=-\infty}^{\infty} x(n)\omega(n - mR)e^{-j\omega n}.$$

Let the signal function be $x(t)$ and the window function be $\omega(t)$. In our model, we used the Hanning Window [13] over the Gaussian Window as it helped minimize leakage. The Hop Length in our model was 0.5. Then we multiplied the signal function with the window function and with the resulting signal function we took the Fast Fourier Transform. Fig. 3 [8], shown above, shows the process of Discrete Short-Time Fourier Transform. Using only the STFT Processing we achieved a KL-Divergence score of 0.57 which was significantly better than the KL-Divergence score using no signal processing.

B. Continuous Wavelet Transforms (CWT)

The Continuous Wavelet Transform [15] [16] is a tool that allows to build an overcomplete model of the signal by letting the scale parameter $a \in \mathbb{R}^+$ and the translational value $b \in \mathbb{R}$ vary continuously. Mathematically,

$$X_\omega(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t)\overline{\psi}\left(\frac{t-b}{a}\right) dt$$

where $\psi = w(t)e^{it}$ is the mother wavelet and $\overline{\psi}$ is the conjugate of the mother wavelet. The main purpose of the mother wavelet is to create daughter wavelets which are just scaled and translated representations of the mother wavelet. Using the

CWT Processing we achieved a KL-Divergence score of 0.49 outperforming both the STFT and no signal processing models. Ultimately, we decided to go with the CWT in our final three-model ensemble because of its versatility and capability to detect more complex patterns within signals.

C. Constant-Q Transforms (CQT)

Although we opted not to test the viability of CQTs in our ensemble, the possibilities of using such a transform have many advantages and could be a viable alternative to the CWTs we ended up using in the end.

CQTs are very similar to STFTs, the main difference being that frequency bins are logarithmically spaced rather than being linearly spaced. CQTs are commonly used in audio signal processing tasks, but they are also advantageous in the processing of signals in biological systems, which is we propose the use of CQTs in future evaluations of brain signals.

However, we opted not to assess CQTs for this task because of how much more computationally intensive they are than STFTs and CWTs. Additionally, the output data structure they provide is more challenging to work with than that provided by STFTs and CWTs.

V. Model Architecture and Training

A. Dataset Splits

The full dataset was split into a training set, a testing set, and a validation set. The training and validation set made up 80% of the total data while the testing set was 20%. We make use of k-fold cross-validation [5] in order to get an accurate estimate of the model's predictive ability. The training data is divided into five folds and each fold is stratified based on labels, ensuring that each fold contains a uniform distribution of class labels. There was no overlap between the training and validation sets, ensuring no data leakage.

B. Loss Function

Because the output of the model is an array of predicted probabilities for each label class, we use the Kullback-Leibler divergence metric [11] to evaluate the model's performance. This metric allows for comparing two distributions, in our case, the predicted probabilities and the real label. Additionally, this metric can be used directly as a loss function training the model because its differentiability allows for not using an alternative function. The KL divergence metric is shown below.

$$D_{KL}(P||Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right).$$

C. Learning Rate Scheduler and Callbacks

In order to ensure smooth training for the model, we created a learning rate scheduler so that the neural network would be able to train efficiently. We took heavy inspiration from the proposed OneCycleLR. The learning rate quickly becomes high and slowly decreases, as shown in Fig. 4.

The smooth decrease of the learning rate is accomplished through cosine annealing, in which training can be done more efficiently and a model is converged upon in a more stable manner. We also used cosine annealing to act as a form of regularization, so the model did not converge too quickly. Finally, research has shown that cosine annealing results in better generalization, improving the model further.

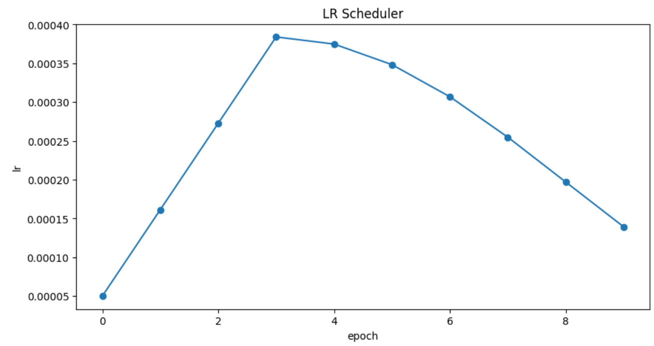

Fig. 4. Learning Rate Scheduling

Additionally, we implement callbacking and checkpointing to prevent against overfitting, as previous results showed that overfitting was a frequent problem with this type of pre-trained model.

D. Model Structure

We utilize transfer learning and use the pre-trained model EfficientNetV2 [7] from Keras for training the processed spectrograms and CWTs. The structure of the model is comprised of an input layer, an EfficientNetV2Backbone layer, a pooling layer, and a dense layer. The adam optimizer [6] is used for an adaptive learning rate. The neural network was trained for 10 epochs with a batch size of 64.

Additionally, we utilize two additional models to create a three-model ensemble. We use VIT image transformers for the two other models. These only use the 2d spectrograms and plots of EEGs as input data. Similar to the EfficientNetV2 model, they also output probability lists for each of the six features. The final probability list is created using a weighted average of the three models. The full ensemble's architecture is illustrated in Fig. 5. Although a more common approach for ensembling is fusing the output features and feeding the convolution block into a softmax function, we found that this method generally produced worse KL-Divergence scores than simply taking a weighted average.

After training for 10 epochs, the peak validation accuracy and best loss were achieved at epoch 7, the model for which we were able to obtain through the callbacks we added earlier. We achieved a peak Kullback-Leibler Divergence score of 0.43 with this three-model ensemble. All of the individual scores are shown below.

VI. Summary and Conclusion

A. Summary

Through our research, we show that the job of manually assessing potential brain abnormalities with EEG plots and

TABLE I
KULLBACK-LEIBLER DIVERGENCE METRIC FOR DIFFERENT MODELS

Model	KL Divergence
EfficientNetV2 without Signal Processing	0.61
EfficientNetV2 with STFT	0.57
EfficientNetV2 with CWT	0.49
Three-model Ensemble	0.43

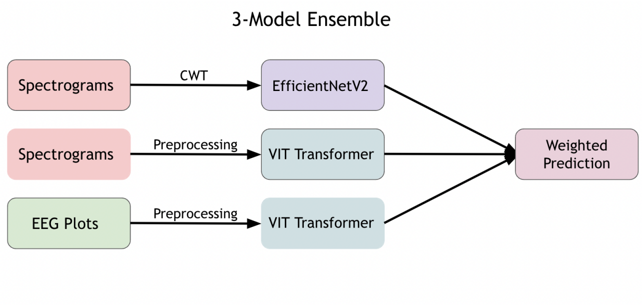

Fig. 5. 3-Model Ensemble Architecture

spectrograms can be accurately done with deep learning. We show that with the appropriate preprocessing of data, including the use of Continuous Wavelet Transforms, the neural network's KL divergence score can be increased significantly. Furthermore, we utilize ensemble learning and create an additional two VIT transformers [14] to provide predictions based on EEG plots and spectrograms. With a weighted average of the three models, we are able to achieve a peak score of 0.43.

B. Conclusion

Until now, most research dealing with signal processing focuses on other methods of altering the form of data, including basic spectrogram transformations and fundamental preprocessing regularization and standardization. However, we present a highly viable alternative of using Fourier transforms in order to achieve similar if not better results. The advantages of using such transforms are numerous, and in the future, we hope to look at the specifics of how exactly these changes alter the capability of trained models.

C. Future Work

In our research, we were not able to test the performance of the Constant-Q Transforms though we suspect that it would do a better job in classifying the models. Right now the best KL-Divergence score for classifying Harmful Brain Activity, is 0.23 and we feel that a better architecture would get us there. One proposition we have is to have a four-model ensemble with the first model running with no signal processing, the second using a Fourier Transform ensembled with another model running with no signal processing, the third using a CWT ensembled with another model running with no signal processing and finally the fourth using a CWT ensembled with a Fourier Transform. Then we ensemble each of these models for a better KL-Divergence score. At this point, this proposition is purely theoretical and has not been tested but we intend to try it in the future. Another avenue of improvement is to use other metrics to measure our model performance. For example, we could explore the Jensen-Shannon Divergence and use that metric on our ensembled model to validate the metrics in our results. Furthermore, while the HMS dataset is comprehensive and builds a strong foundation for future models that aim to detect harmful brain activity, its generalizability can be further assessed by introducing additional datasets, which includes CHB-MIT Scalp EEG Database, which has a larger range of patient demographics and neurological conditions.

REFERENCES

[1] Johns Hopkins Medicine, "Electroencephalogram (EEG)," John Hopkins Medicine, 2019. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg

[2] A. A. Ein Shoka, M. M. Dessouky, A. El-Sayed, and E. E.-D. Hemdan, "EEG seizure detection: concepts, techniques, challenges, and future trends," Multimedia Tools and Applications, Apr. 2023, doi: https://doi.org/10.1007/s11042-023-15052-2.

[3] Jin Jing, Zhen Lin, Chaoqi Yang, Ashley Chow, Sohier Dane, Jimeng Sun, M. Brandon Westover. (2024). HMS - Harmful Brain Activity Classification . Kaggle. https://kaggle.com/competitions/hms-harmful-brain-activity-classification

[4] C. Sun, C. Mou, "Survey on the research direction of EEG-based signal processing" in Frontiers in Neuroscience, Neural Technology, July 2023

[5] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, and S. Ridella, "The 'K' in K-fold Cross Validation," in European Symposium on Artificial Neural Networks, Computational Intelligence, and Machine Learning, ESANN, Apr. 2012.

[6] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014.

[7] Tan, M. and Le, Q., 2021, July. Efficientnetv2: Smaller models and faster training. In International conference on machine learning (pp. 10096-10106). PMLR.

[8] H. Jeon, Y. Jung, S. Lee, and Y. Jung, "Area-efficient short-time fourier transform processor for time–frequency analysis of non-stationary signals," MDPI, Oct. 2020.

[9] C. Schorkhuber and A. Klapuri, "Constant-Q transform toolbox for music processing," in 7th Sound and Music Computing Conference, researchgate, Jul. 2010.

[10] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond Empirical Risk Minimization," in International Conference on Learning Representations, arxiv, Oct. 2017.

[11] S. Ji, Z. Zhang, S. Ying, L. Wang, X. Zhao and Y. Gao, "Kullback–Leibler Divergence Metric Learning," in IEEE Transactions on Cybernetics, vol. 52, no. 4, pp. 2047-2058, April 2022, doi: 10.1109/TCYB.2020.3008248.

[12] K. Team, "Keras documentation: RandomCutout layer," keras.io. https://keras.io/api/keras_cv/layers/augmentation/random_cutout/ (accessed Apr. 16, 2024).

[13] N. Pielawski and C. Wählby, "Introducing Hann windows for reducing edge-effects in patch-based image segmentation," PLOS, Oct. 2019.

[14] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, arxiv, Oct. 2020.

[15] "THE CONTINUOUS WAVELET TRANSFORM: A TOOL FOR SIGNAL INVESTIGATION AND UNDERSTANDING." Accessed: Apr. 16, 2024. [Online]. Available: https://secwww.jhuapl.edu/techdigest/Content/techdigest/pdf/V15-N04/15-04-Sadowsky.pdf

[16] A. Prasad, J. Maan, and Sandeep Kumar Verma, "Wavelet transforms associated with the index Whittaker transform," Mathematical methods in the applied sciences, vol. 44, no. 13, pp. 10734–10752, Apr. 2021, doi: https://doi.org/10.1002/mma.7440.

[17] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the Degree of Human Attention Using EEG Signals from Mobile Sensors," Sensors, vol. 13, no. 8, pp. 10273–10286, Aug. 2013, doi: https://doi.org/10.3390/s130810273.