PROJECT IFT-7030 - DETECTING ALZHEIMER'S AND FRONTOTEMOPORAL DEMENTIA IN EEG READINGS

Jean-Francois Cadieux,
Martin Labrie

Laval University

ABSTRACT

As Alzheimer's and dementia cases are increasing in our society, unconventional methods must be developed to be able to detect and treat those medical conditions. Differentiating between the two conditions represents a complex clinical challenge, and electroencephalography (EEG) offers a non-invasive window into brain abnormalities. This article explores the comparative effectiveness of several machine learning models, including EEGVNet4, Dynamic Time Warping (DTW), Gaussian Mixture Models (GMM), Recurrent Neural Networks (RNN-LSTM), and Convolutional Neural Networks (CNN), in classifying patients into three categories: healthy EEG, EEG with Alzheimer's, and EEG with frontotemporal dementia.

Each model leverages the temporal and spectral characteristics of EEG differently, shedding light on the advantages and limitations of each approach. Therefore this project will be about exploring the field of EEG readings to be able to detect those conditions. Special attention to medical literature will be done to be able to supervise our models. More specifically, research specifying the abnormal signals to look out for in both medical conditions.

1. INTRODUCTION

Alzheimer's and frontotemporal dementia share similar cognitive and behavioral symptoms, making them difficult to distinguish using traditional methods such as neuropsychological assessments, standard cognitive tests, and conventional brain imaging such as MRI or CT scans. These methods, although effective in detecting general abnormalities, often lack specificity in differentiating the two pathologies, especially in the early stages. EEGs provide specific signatures of neurodegenerative diseases. Recent advances in deep learning and signal processing make it possible to use EEG to distinguish these pathologies with more precision.

The models analyzed in this study include EEGVNet4, DTW, GMM, RNN-LSTM, and CNN. These techniques are based on different strategies to analyze EEG data:

Laval University

- EEGVNet4 leverages advanced spectrogram-based approaches.
- DTW measures temporal similarities between signals.
- GMM models probabilistic distributions of EEG features.
- RNN-LSTM captures complex temporal dependencies.
- CNN automatically extracts spatial and temporal patterns.

Our methods will be iterative in some parts as we will first classify our EEG from normal to abnormal. Following this step, we will split the abnormal readings in two classes for Alzheimer's and dementia. After this step, we will try to predict the stage/phase of the medical condition. By doing that, our models will have to be able to learn the right characteristics in every step of the algorithm.

2. LITERATURE REVIEW

In our research for literature, we found a wide variety of models used for EEG readings but none seems to have applied it to the detection of Alzheimer's or dementia[4]. Also, only frequencies were used and not the transformed images. Therefore, our method seems to be a new approach for this field. Different models were tested like LightGBM, SVM , KNN and Random Forest. The maximum accuracy was about 77% for the Random Forest model.

Table 2. Classification Performance of Leave-One-Subject-Out validation for the AD-CN problem.

AD/CN	ACC	SENS	SPEC	F1
LightGBM	76.43%	76.01%	76.16%	76.12%
SVM	73.14%	71.89%	75.98%	73.74%
kNN	71.23%	69.67%	74.19%	72.81%
MLP	73.12%	73.00%	74.63%	74.82%
Random Forests	77.01%	78.32%	80.94%	75.31%

(a)

Table 3. Classification Performance of Leave-One-Subject-Out validation for the FTD-CN problem.

FTD/CN	ACC	SENS	SPEC	F1
LightGBM	72.43%	61.13%	80.74%	67.32%
SVM	70.14%	62.41%	75.98%	68.32%
kNN	67.34%	59.67%	76.13%	70.81%
MLP	73.12%	63.00%	78.63%	72.82%
Random Forests	72.01%	72.32%	80.94%	66.31%

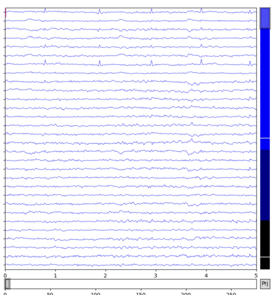
(b)

Fig. 1. Results from prior research [1]

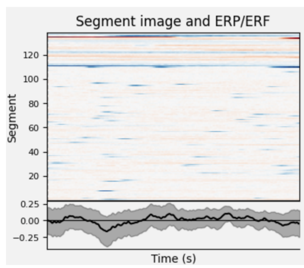
The models we are implementing don't seem to have been research previously, therefore this article is well placed within the literature.

3. INPUT-OUTPUT-DATASET

For the main input, we will use a public dataset from OpenNEURO [1] containing EEG readings from 88 patients. Of those 88 patients, 36 have Alzheimer's, 23 have dementia and 29 are normal subjects. For this project, we will test two input methods for our models. We will first use the 20 frequencies as inputs. Our focus will be on frequencies impacting each class the most by reducing the dimensions of the problem. ICA method will be applied to extract only the frequencies important to the problem by removing some signals like the blinking of the eyes or movements of the head.

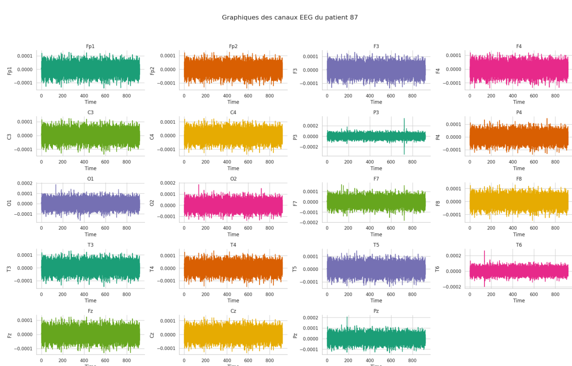


(c) EEG Frequencies



(d) EEG Frequencies

Fig. 2.

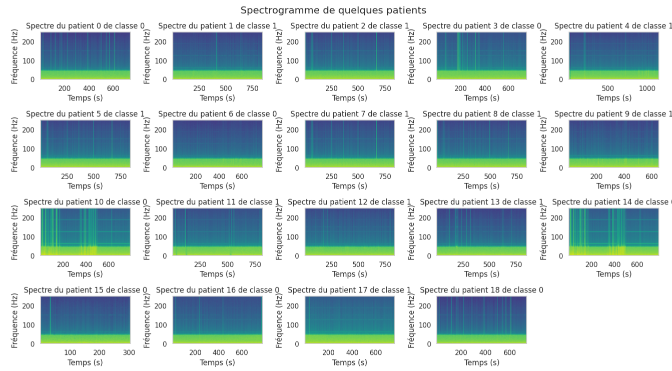


(e) Waveform of EEG

As for the output, our focus will be on the classification of each condition at each stage of the process. First, the classification of normal and abnormal followed by the different stages of the classification for Alzheimer's and frontotemporal dementia. From the EEG frequencies, we will extract a matrix which we will use to train our models and from that extract a predictive model to arrive at the classification.

4. PERFORMANCE METRICS

For the classification, we will analyze the confusion matrix which will be produced from our different models. The confusion matrix will be used to evaluate the performance of a

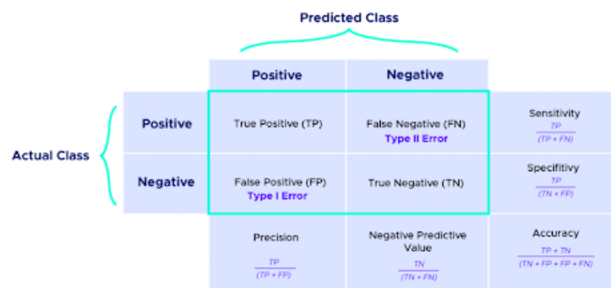


(f) Spectrogram of EEG

classifier on the objects of a class by making it possible to calculate the Accuracy (A), the Precision (P), the Recall (R) and the F-score (F) .

Accuracy will not be considered since our dataset is somewhat unbalanced as we have about 2/3 of the patients having an illness.

We will focus on precision, recall and the F1-score. The pre-



(c) Metrics used for performance evaluation[2]

cision happens to be the purity of the positive class. It will give us the real proportion of patients with a diagnosis. So it will be our focus at the first stage. The F1 score will ignore true negatives, it is the average of precision and recall that balances precision and recall. It will allow us to find the optimum, the best of both.

5. COMPUTATIONAL RESOURCES

The number of patients isn't high but the number of channels and the time of recordings implies that computational resource is a crucial factor in our project in order to have an acceptable response time.

Therefore, GPU's are going to be a big factor as we are going to work with PyTorch and tensors.

6. METHODOLOGY

6.1. Pretreatment of EEG [1] (Sources: https://www.kaggle.com/datasets/yosftag/open-nuro-dataset/data)

This dataset contains resting, eyes-closed EEG recordings from a total of 88 subjects. The EEGs of the 88 people are distributed as follows:

- Group AD: 36 have the diagnosis of Alzheimer's disease (41%)
- FTD group: 23 have a diagnosis of frontotemporal dementia (26%)
- CN group: 29 are healthy subjects (33%)

Each recording was made with participants in a seated position and eyes closed. 19 scalp electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1 and O2) according to the international 10-system 20 2 reference electrodes (A1 and A2) placed on the mastoids for impedance control, according to the manual of the device used for recording [3].

EEGs are transformed into spectrograms for CNN and EEGVNet4 approaches and they are segmented into time windows for DTW, GMM, and RNN-LSTM.

6.2. Models

6.2.1. EEGVNet4

EEGVNet4 is a hybrid architecture based on EEG spectrograms and convolutional layers followed by an attention mechanism. It captures the specific characteristics of pathologies by exploiting spatial and spectral relationships.

6.2.2. Dynamic Time Warping (DTW)

DTW is a classic method to evaluate the similarity between two time series by calculating the optimal alignment. The distances between EEG signals from patients with Alzheimer's and frontotemporal dementia are compared, providing insights into their temporal dynamics.

$$\mathrm{DTW}(X, Y) = \min_{P} \sum_{(i,j)\in P} d(x_i, y_j)$$

6.2.3. Gaussian Mixture Model

GMM models the statistical distributions of EEG features. GMMs segment patients into classes based on Bayesian probabilities.

6.2.4. Recurrent Neural Networks (RNN-LSTM)

Long Short-Term Memory (LSTM), a variant of RNNs, is well suited to capture complex temporal dependencies in EEG data, taking into account dynamic features over extended intervals. The loss function for this model is implemented from a mix of cross-entropy and DTW with a weighing parameter α and β for more flexibility.

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{CE}} + \beta \cdot \mathcal{L}_{\text{DTW}},$$

6.2.5. Convolutional Neural Networks

CNNs process EEGs transformed into spectrograms by automatically extracting discriminative patterns. Convolutional layers are followed by pooling and fully connected layers to perform the final classification.

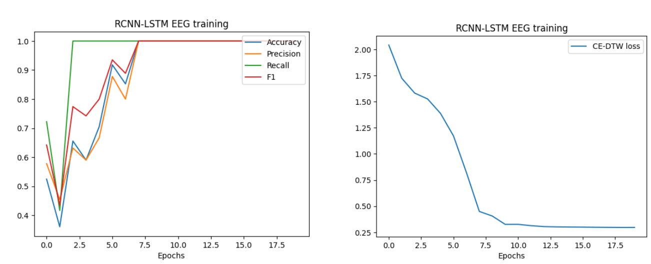
7. RESULTS AND DISCUSSION

7.1. Models performance

The models were evaluated using the following metrics: accuracy, precision, recall, F1-score (measures similarity in terms of combining precision and recall)

Model	Accuracy	Precision	Recall	F1-Score	Specificity
EEGVNet4	59%	60%	59%	59%	na
GMM	68%	68%	100%	81%	0%
RNN-LSTM-DTW	73%	88%	79%	83%	33%
CNN-DTW	86%	86%	100%	93%	0%

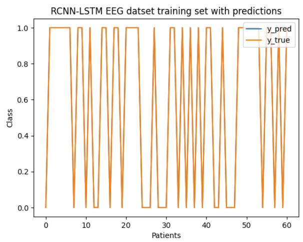
Table 1. Performance metrics for various EEG classification models.



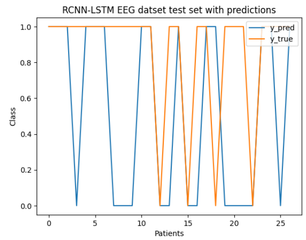
(d) (e)

Fig. 3. RCNN-LSTM CE-DTW metrics (d) and loss (e) during 20 epochs of training.

As shown in table 1, our models seems to perform somewhat well on the binary classification with the best performance for the CNN network with a DTW loss.



(f)



(g)

Fig. 4. RCNN-LSTM CE-DTW predictions on training set (f) and test set (g) after 20 epochs of training.

7.2. Analysis of Strengths and Limitations

- EEGVNet4: Better overall performance thanks to its ability to integrate spectral and temporal characteristics, but requires intensive calculation. We tested this model on 2 training sessions of 50 epochs. A stride of 64 was applied with a temporal size (ntime) of 128 per segment. The first training session was performed on raw data with binary classification between healthy and problematic EEGs. The second session was conducted on a spectrogram per patient (19 channels) for binary classification. The performance of our model, EEGVNet4, reveals weak metrics on both the training and test sets, whether using raw data or spectrograms for binary classification (0-1). The accuracy reached 59% on the training set but dropped to 48% on the test set, suggesting issues with overfitting and generalization. Other metrics, such as precision, recall, and F1-score indicate that the model makes too many errors, including a significant number of false positives and false negatives.

  These results highlight the model's difficulty in extracting relevant features from EEG data, likely due to their complexity.

  To improve performance, promising approaches include better preprocessing and data augmentation, exploring alternative pre-trained models, modifying the architecture, training for more epochs, analyzing individual channels, or adjusting the stride. These hypotheses could make the model more robust and enhance its generalization on test data.

- GMM: Effective probabilistic approach, but sensitive to the quality of the extracted features. A GMM was fitted to each class for training. After fitting the models, a maximum likelihood function was applied to classify each patients. The model only obtained 68% in accuracy and precision. It classifies the normal EEG's but has problems classifying the abnormal classes.

- RNN-LSTM: Very efficient at capturing temporal dependencies, but requires a large volume of data to avoid overfitting. The RNN classifier was implemented with an LSTM and a DTW loss. This network learned the classes in the training set perfectly but had trouble transferring it's learning to the test set. Obtaining 73% in accuracy and 88% in precision.

- CNN: Excellent ability to automatically extract relevant features, but requires data transformation (spectrograms). The CNN classifier was also implemented with a DTW loss. This network performed the best with 86% accuracy and 86% in precision. It had also a little bit of trouble transferring the learning to the test set. More training could help to identify a little bit better the classes.

7.3. Difficulty and solutions

- Limited data: Small EEG datasets can lead to overfitting. Data augmentation and transfer of learning are solutions being explored.

- Noise and artifacts: Raw EEG signals require rigorous preprocessing to ensure usable inputs for models.

- Class Imbalance: Approaches such as class weighting in loss functions or oversampling have been used to overcome this problem.

8. CONCLUSION

This article highlights the effectiveness of advanced models for classification of Alzheimer's versus frontotemporal dementia using EEG. We explored some machine learning techniques to classify EEG data into three categories: healthy, Alzheimer's disease (AD), and frontotemporal dementia (FTD). Among the tested models, CNN achieved the highest performance with 86% accuracy and 93% F1-score, effectively leveraging spectrogram features.

Other models, such as EEGVNet4 and GMM, showed potential but faced challenges like overfitting. Key difficulties included limited data, noise, and class imbalance. Solutions such as data augmentation, improved preprocessing, and class weighting were identified. This research highlights the potential of EEG and deep learning for non-invasive, early detection of neurodegenerative diseases.

References

[1] *A dataset of EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects*. URL: https://openneuro.org/datasets/ds004504/versions/1.0.6.

[2] *Matrice de confusion : qu'est-ce que c'est et comment l'utiliser ?* URL: https://datascientest.com/matrice-de-confusion.

[3] Andreas Miltiadous et al. "A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG". en. In: *Data* 8.6 (May 2023), p. 95. ISSN: 2306-5729. DOI: 10.3390/data8060095. URL: https://www.mdpi.com/2306-5729/8/6/95 (visited on 12/22/2024).

[4] Weining Weng et al. *Self-supervised Learning for Electroencephalogram: A Systematic Survey*. en. arXiv:2401.05446 [eess]. Jan. 2024. URL: http://arxiv.org/abs/2401.05446 (visited on 10/20/2024).