Seizure forecasting from EEG using signal processing engineering and Long Short Term Memory networks

Carlo Amodeo[1(✉)]

Department of Information, Electronics and Bioengineering, Politecnico di Milano, Milan, Italy
carlo.amodeo@mail.polimi.it

1 Summary

This report presents the solution I propose at Kaggle American Epilepsy Society Seizure Prediction Challenge. This document is part of the final submission for the course Advanced Signal and Data Processing in Medicine held by Professor Cerutti at Politecnico di Milano.

Epilepsy affects around 1% of the global population, leading to spontaneous seizures. While some patients use high-dose anticonvulsant medications to prevent seizures, these often cause side effects. A substantial portion (20-40%) of epilepsy patients find these medications ineffective, even after surgery. The fear of sudden seizures creates ongoing anxiety, despite their infrequent occurrence.
Seizure forecasting systems offer hope for a better quality of life for epilepsy patients. Effective EEG-based forecasting relies on precise algorithms to detect periods with an elevated risk of seizures. If we can pinpoint these brain states that allow seizures, it becomes feasible to create devices that alert patients. This can help them avoid potentially hazardous activities like driving or swimming, and minimize medication use, reducing side effects.
Recent research suggests that brain activity follows four states: Interictal (between seizures), Preictal (before seizures), Ictal (during seizures), and Post-ictal (after seizures). The key challenge in seizure forecasting is accurately distinguishing the preictal state from the interictal state.

The aim of this project is to detect epileptic seizures from EEG recordings; in a real world scenario, a model trained to predict seizures before they show would allow clinicians to warn patients in a timely manner, in order to take necessary precautions, improving people's life and open up possibilites for alternative treatments.

The strategy employed here to reach the proposed aim is double: on one hand, classic features engineer is applied to highlight meaningful predictors of the seizure, and used as input for a linear model classifier. On the other hand, the raw time series are reshaped and used as input to train a Long-Short-Term memory Recurrent Neural Network (LSTM RNN) which provides time series forecasting and seizure classification.

2 Dataset

Recordings of intracranial EEG in dogs with natural epilepsy were captured through a portable monitoring system. These EEG data were collected from 16 electrodes at 400 Hz, referenced to the group average. The recordings extended over several months to a year, and some dogs experienced up to a hundred seizures.
Additionally, the competition includes datasets from epilepsy patients undergoing intracranial EEG monitoring to pinpoint a resectable brain region to prevent future seizures. These datasets vary in the number of electrodes and are sampled at 5000 Hz, with voltage references outside the brain. The challenge involves distinguishing between ten-minute data clips occurring an hour before a seizure and interictal activity clips. For more details regarding the dataset, Kaggle has an insighful section about it **here**.

3 Methods

In this section, a very simplistic baseline method is presented to forecast seizures, together with two more robust approaches, a deterministic and a ML-based approach. The ML-based approach employs the recurrent neural networks architecture, while the deterministic ones is based on signal processing engineering and domain knowledge. In the first example, the relevant information are black-box extracted from the RRN; in the second example, domain knowledge is used to compute EEG-statistically relevant features in the forecasting of seizures.
For both the baseline method and the engineered features one, once the features are computed they are used to train a logistic regression model.

- An l2 penalization approach (*ridge*) is followed in order to avoid the risk of multicollinearity.
- A k-fold cross validation, with k=10, is performed in the training phase, using the AUC as score metrics. The cross entropy function is used as loss function.

3.1 Baseline Approach

The most simplistic approach implies some exploratory data analysis to check whether it is possible to find out any clue on ho the interictal and preictal clips differ. A plot is presented in fig 1.

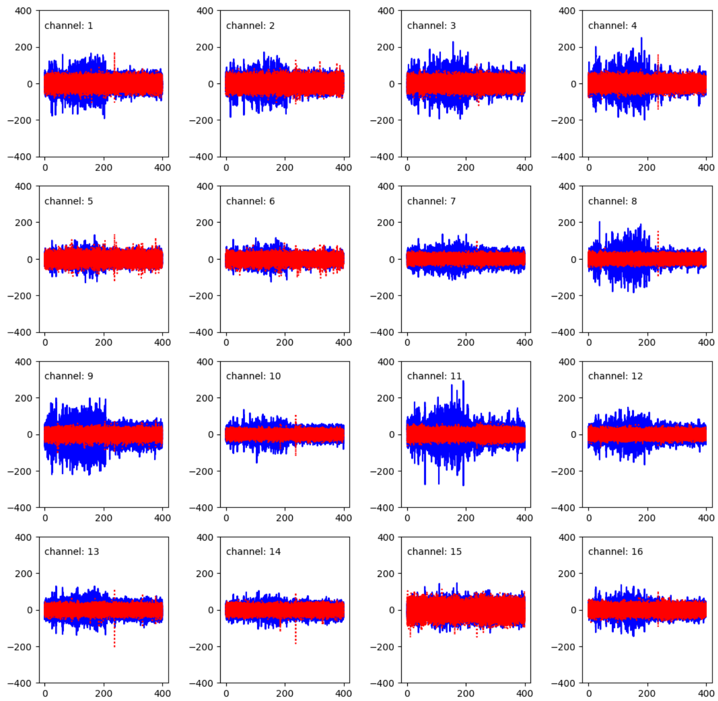

Fig. 1: Plot of the recording from all the channels of interictal (red) and preictal (blue) clips for one of the participant.

Considering the simplest features we can think of, namely the variance within each channel and the correlation between all the channels, we obtain n_{ch} variances + $n_{ch} * (n_{ch} - 1)/2$ correlations. Being $n_{ch} = 16$, we get 136 features to be used to forecast the seizures.

Figure 2 shows the two principal component analysis reduction of the features depicted before.

3.2 LSTM-RNN approach

Recurrent Neural Networks (RNNs) are well-suited to handle the challenging task of EEG-based seizure prediction. They excel in this context due to their inherent capability to capture temporal dependencies in sequential data. EEG data, being a time series, contains valuable information in its temporal patterns and trends, especially in the moments leading up to a seizure. RNNs can effectively model these dependencies, making them suitable for forecasting seizures.

However, it's important to acknowledge some challenges inherent to traditional

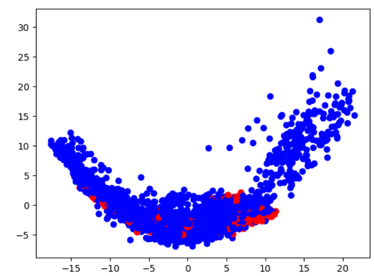

Fig. 2: First two principal components representation of the baseline features. Blue is preictal, red is interictal. Note how the two features overlap in the central portion of the space.

RNNs, such as vanishing gradient problems when dealing with long sequences. This issue can hinder the network's ability to capture long-range dependencies in the data. This is where Long Short-Term Memory (LSTM) networks come into play. LSTMs are a specialized type of RNN designed to address these problems. They incorporate a memory cell that can store and retrieve information over long sequences, allowing them to better handle temporal dependencies and alleviate vanishing gradient issues. In this context, LSTMs are a superior choice compared to traditional RNNs because they can effectively capture the intricate dynamics of EEG data leading to seizures, making them more robust and accurate for this specific application.

Importantly, LSTM can effectively handle feature extraction in an automatic way. The network's gating mechanisms, including input, forget, and output gates, regulate the flow of information in and out of these cells. Through these gates, LSTMs can selectively focus on relevant patterns and discard less important information, effectively identifying salient features within the EEG data. As a result, LSTMs autonomously learn to extract and weigh the most significant temporal features from the raw EEG signals, enabling them to adapt to the inherent complexity and dynamics of the data and improve the accuracy of seizure prediction.

Preprocessing and training strategy The major issue in using the raw data as input for the LSTM network lies in the length of the data itself, which is very long. In fact, considering a sampling frequency $f_s = 600Hz$ and the duration of the recording equal to 400 seconds, we obtain 24000 samples points in time. LSTM suffer for such a long input, and therefore a data reduction is necessary. Here we decided to emply a 1d convolution to reduce the number of points, as shown in figure 3

The drawback here is that thee convolution window size and the strides have to be treaten as hyperarameter to be tuned. The choice of these hyper-

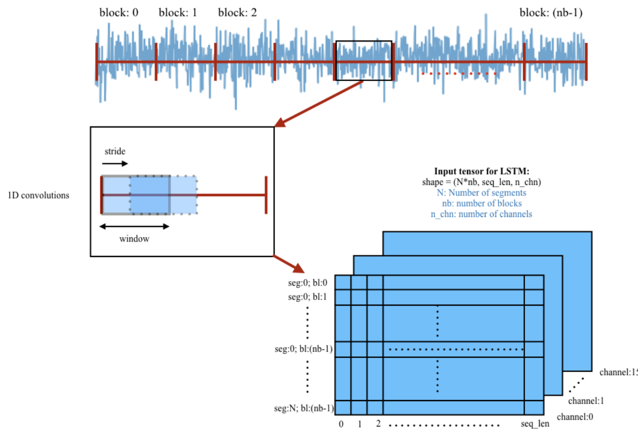

Fig. 3: Illustration of the data reduction strategy.

parameters was driven by the need to balance dimensionality reduction with the preservation of essential temporal information within the data; we selected windowing equal to 40 seeconds (16000 samples) and stride equal to 100. This configuration allowed us to effectively capture high-level patterns and reduce the sequence length while maintaining the integrity of critical features for our task, ultimately leading to favorable results Although further experiments were not conducted due to resource constraints, the selected configuration proved to be a successful strategy for our specific problem, demonstrating the feasibility of using 1D convolution for input preprocessing in long sequence scenarios.The resulting features' sequence length is now 200, handable by our LSTM. The utilities code to compute these 1d convolution features is avaliable in the github repo.

As for the LSTM itself, it has been implemented through the tensorflow framework. A number equal to 32 neurons has been placed in each of the LSTM layers, which are 2. A batch size equal to 200 has been select, and the sequence lenght is constrained to be equal to the input lenght, being hence 200. The model has been optimized on the sigmoid cross entropy loss function, with a learning rate equal to 1e-5. The selected number of epochs has been 500, and a k-fold cross validation strategy has been employed, with k=10.

3.3 Features Engineering

When substantial domain knowledge is available, it can often be advantageous to opt for standard feature engineering over a black-box ML approach. In the case of EEG data, where understanding of relevant features is well-established, feature engineering can provide clear interpretability and efficiency.

One variable that has not been taken in account so far but has been proven to be highly significant in EEG analysis is the frequency band of the signals. Analyzing EEG data within specific frequency ranges, such as delta, theta, alpha, beta, low-gamma, and high gamma, is valuable because different brain activities and states are associated with different frequency bands.

With this in mind, the signals have been divided in blocks 60 seconds long, and different spectral features have been engineered for each block and frequency band as described in Howbert et al. work. [1]. Figure 4 shows the division in blocks for each of the channels in one of the patient, while figure 5 shows a frequency decomposition of one of the blocks.

Here follows a list of the selected features, together with a brief explanation of the reasons behind the choice.Here follows a list of the selected features, together with a brief explanation of the reasons behind the choice.

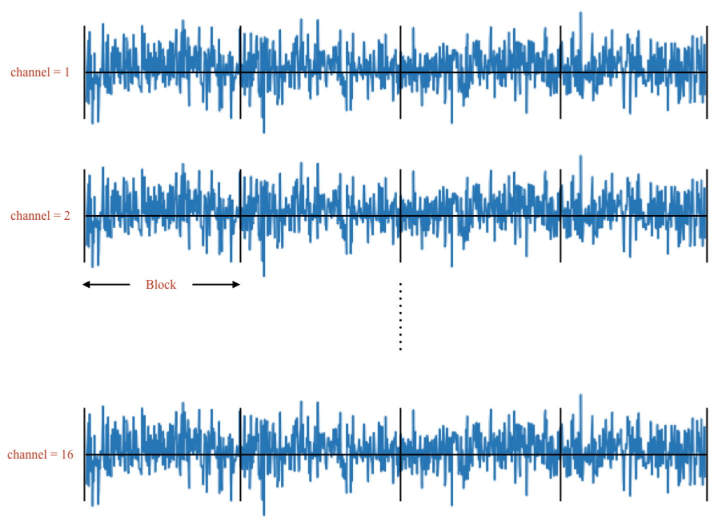

Fig. 4: Block decomposition of the signals. This increases the number of data points per segment of EEG recording, useful to reduce the model's variance in training phase.

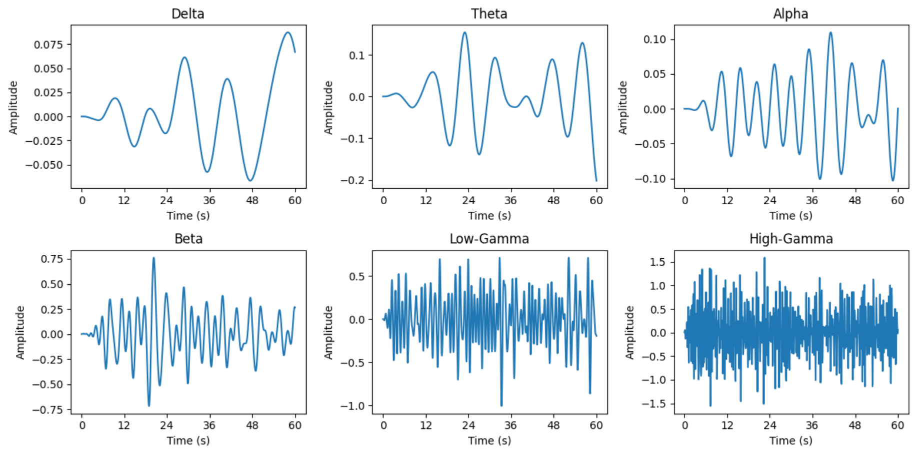

Fig. 5: Frequency decomposition of the time series. It has been shown how changes in different bandwidths can be used as seizures predictors.

- **Eigenvalues of correlation matrices** within each band and channel can capture the degree of linear relationship and synchronization between different frequency bands and channels. This is particularly relevant because changes in correlation patterns may indicate abnormal brain activity associated with seizures.
- **Shannon's entropy** for power within each band can measure the complexity and irregularity of EEG signal power. A significant increase in entropy may indicate a transition from normal brain activity to seizure onset.
- **Power at dyadic levels** provides a multiscale perspective, which can be crucial for capturing complex dynamics in EEG data. The eigenvalues of correlation between channels and Shannon's entropy for dyadic power levels offer insights into how brain activity is distributed across different frequency components and channels.
- **Hjorth parameters** provide information about signal mobility, complexity, and how localized or dispersed the signal is. These parameters can help in distinguishing between normal and abnormal EEG patterns.
- **Skewness** and **kurtosis** within each block can reveal the shape of the signal distribution and detect potential outliers or extreme values that may indicate abnormal activity.

For each 60s block we hence have a total of 166 features, 28 per frequency band. The utilities code to compute the engineered features is available in the github repo.

As show previously for the baseline method, figure 6 highlights the 2 principal component for the engineered features, which shows a clearer division in the plane between the preictal and the interictal clips.

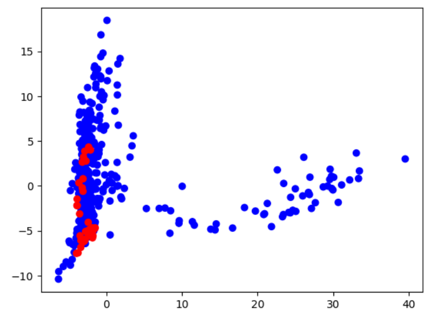

Fig. 6: First two principal components of the engineered features. Notice how the preictal clips correspond to PCs which span a large range, while the interictal clips align on a line.

4 Results

This section shows the numerical results obtained in the task of seizure forecasting using the three different strategies described in section 3.

Table 1: Seizure prediction results

	Baseline	**LSTM**	Features Engineering
Precision	0.69±0.03	**0.85±0.09**	0.73±0.03
Recall	0.61±0.04	**0.87±0.07**	0.71±0.04
F1-score	0.62±0.04	**0.78±0.08**	0.7±0.04
AUC Roc	0.64 ±0.03	**0.87 ±0.06**	0.73 ±0.03

The experiments have been run 10 times and the results show how the baseline features model is inferior to both the LSTM and the features engineering method, which is expected. Also, the LSTM method outperforms the latter; but it is worth noting how the variability of the classification results is higher in this case.

5 Conclusion

This work shows how it is possible to employ basic statistic and more complex ML-based or Features Engineering based methods to forecast seizures in human and dogs. The recurrent neural networks outperforms both the other two methods, at a cost of a significantly higher computational expense, as well as time necessary in the training phase. The classification results also shows higher variability between different experiment using the LSTM, which is expected being it a not-deterministic approach. Possible improvements can be made by combining the two approaches (i.e. using a neural network to predict the seizures using the engineered features, instead of using the logistic regression). On the other hand, the ML-based result can be furtherly improved via a fine tuning of its hyperparameters (number of LSTM layers, number of neurons...).

References

1. J Jeffry Howbert, Edward E Patterson, S Matt Stead, Ben Brinkmann, Vincent Vasoli, Daniel Crepeau, Charles H Vite, Beverly Sturges, Vanessa Ruedebusch, Jaideep Mavoori, et al. Forecasting seizures in dogs with naturally occurring epilepsy. *PloS one*, 9(1):e81920, 2014.