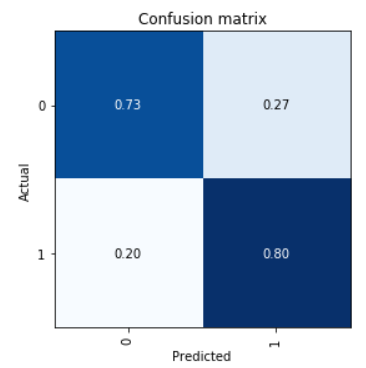
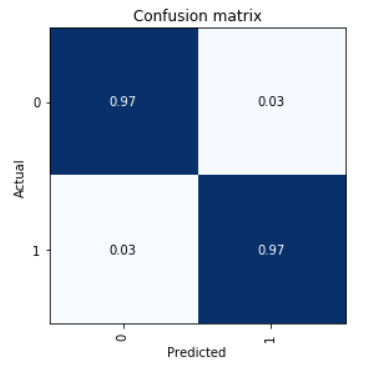
**Goal:** Analyze how a neural network can be useful in gamma burst prediction and adapt XAI techniques to explore the inner-workings of the trained network to extract additional insights.

**Methods**

We analyzed data from our previously published computational model of gamma oscillations in the basolateral amygdala (Feng et al. 2019). For variables to the model, we included (1) the calculated LFP based on the line-source approximation (details… citations). (2) The firing rate of interneurons and pyramidal cells, with 1 ms bins. We divided the neurons into five different groups based on their distance to the electrode (details to come). Thus, the ML models were given 16 variables; LFP, A-type PN firing rate in 5 groups, C-type PN firing rate in 5 groups, and INT firing rate in 5 groups.

**Results**

1. Forecasting raw LFP 10 ms into the future
   1. When trained on experimental data from BLA, our ensemble model did not perform better than autoregression (Model A in figure 1A). However, when we included all 64 electrodes as input to the model, we were able to improve the accuracy significantly (Model B in figure 1A).
   2. When trained on data from the biophysical model, we could greatly improve on the autoregression prediction accuracy by using an ensemble model of a CNN, LSTM, and Autoregression (Model A in figure 1B). Further, we showed that the forecast could be improved even more my including the firing rates of cells as input to the model (Model B in figure 1B).
2. Classifying LFP – Next we changed the task to one of classification. We trained a two layer CNN to classify whether the future 10 ms contained a peak in the gamma band above threshold, given 40 ms of past data. The results are shown in table 1a.
3. Next we explored the reason for misclassification. Including one additional variable, the non-causal filtered LFP, improved accuracy significantly, as expected (table 1b). This is because the non-causal filter includes information about the future of the signal. However, there are still misclassifications. Taking the mean of the peak of the misclassified cases revealed that the values for false negative and false positive were 1.9 and 2.1, respectively, very close to the threshold. Thus, samples that were misclassified were all very close to threshold.
4. We applied Guided GradCam, Occlusion, Named Neurons, Gradient descent on input, and failure analysis to the more realistic case i.e. without filtered LFP, and produced the following insights…
5. Hilbert Classification - Next we tried taking a window of LFP (with the old data) and trying to classify the next 10 ms as having either low, medium, or high average hilbert value (of the filtered LFP). I did not get good results on my model but Ben seemed to do much better, maybe worth looking at again with new data?
6. 4 Path Classification - The most recent task we tried was using a few different classification tasks. The tasks were: classifying if the next peak/trough will be above threshold, classifying how long until the next peak/trough, and classifying how many consecutive peaks/troughs above threshold there will be. The classification techniques about being above threshold did not work very well but the classification on time until peak worked pretty well. Then, we improved the classification of whether the next peak will be above threshold to about 65% above accuracy and 86% bellow accuracy using only samples where the peak is less than 3 ms away. We then implemented GradCam and Guided Backpropagation on that model.



B

A

**Table 1.** Confusion matrices for classification. **A:** Without non-causal filtered LFP. **B:** With non-causal filtered LFP.



**Figure 1.** AI models improve on linear forecasting methods. **A:** Log of the root mean squared error for an ensemble model trained on experimental data with one electrode (Model A) and 64 electrodes (Model B). **B:** Ensemble model trained on the LFP (Model A) and LFP combined with firing rates (Model B). \* p<0.001