Forecasting events in multidimensional electroencephalographic brain data: Application to epileptic seizure prediction.

Yueyang Liu
Dept. of Data Science and AI, Faculty of Information Technology
Monash University
Clayton, Australia
yueyang.liu@monash.edu

Artemio Soto-Breceda
Dept. of Biomedical Engineering
The University of Melbourne
Melboourne, Australia
artemio.soto@unimelb.edu.au

Mark J. Cook
Dept of Medicine, St Vincent's Hospital, The University of Melbourne, Fitzroy, Australia
markcook@unimelb.edu.au

Philippa Karoly
Dept. of Biomedical Engineering
The University of Melbourne
Melboourne, Australia
karolyp@unimelb.edu.au

David B. Grayden
Dept. of Biomedical Engineering
The University of Melbourne
Melboourne, Australia
grayden@unimelb.edu.au

Levin Kuhlmann
Dept. of Data Science and AI, Faculty of Information Technology
Monash University
Clayton, Australia
levin.kuhlmann@monash.edu

Dean R. Freestone
Seer Medical Pty
Melbourne, Australia
dean@seermedical.com

Daniel Schmidt
Dept. of Data Science and AI, Faculty of Information Technology
Monash University
Clayton, Australia
daniel.schmidt@monash.edu

***Abstract*—Forecasting events in multichannel electroencephalographic (EEG) brain recordings remains a formidable task given the noise and complexity in neural systems. Here we compare two dynamical systems motivated approaches to forecasting brain events. The first follows previous state-of-the-art (SOTA) research of time-series features of critical slowing down (autocorrelation, variance) as biomarkers of impending events. The second involves a novel long-term-short-term (LSTM) neural network-based filter to estimate the neurophysiological feature variables of mathematical neural population models of the EEG. Previous critical slowing research presented forecasting results for the best EEG channel, however, in practice the best channel cannot be known a priori. Therefore, here we also consider forecasting by combining the different features across the different EEG channels using logistic regression. One application area where forecasting brain events is important is epileptic seizure prediction. Epileptic seizures are debilitating events and up to 50 million people worldwide with drug-resistant epilepsy could benefit by receiving warnings of impending seizures. Here we apply the above methods to a long-term epileptic seizure prediction dataset from 15 patients. It was found that seizure forecasting with (1) logistic regression and critical slowing features, (2) logistic regression and neurophysiological features, and (3) the best channel using critical slowing features, respectively, achieved median sensitivities of 70, 54 and 67% and median time in low seizure risk of 84, 84, and 81%. This indicates that a multichannel model approach can perform as well as the best channel approach, removing the need to find the best channel. It also suggests neurophysiological features could be used to increase time in low risk. Future work exploring other features, machine learning models and their various combinations could yield further improvements.**

Keywords—brain event forecasting, dynamical systems, critical slowing down, neural filtering, epileptic seizure forecasting

I. INTRODUCTION

While forecasting the next samples in electrical signal time series recorded from the brain, such as the EEG, has been around for a while [1] and continues to evolve [2], this paper focuses on forecasting neurologically or psychologically significant events in the brain [3,4,5,6,7]. To understand the brain from a theoretical perspective it is often modelled as a dynamical system. As such one can potentially view significant events in the brain as a critical transition resulting from a parameter bifurcation or noise perturbations or a combination of both [8]. Critical slowing down represents one marker of predicting critical transitions in dynamical systems [9]. Time-series features such as zero-lag autocorrelation, autocorrelation function width (ACFW) and variance all increase in the lead up to critical transitions and as such can be used to forecast brain events by calculating these features using the EEG time-series. In particular, SOTA epileptic seizure forecasting results have been achieved when using critical slowing down features and hand-picking the best intracranial EEG channel [10].

Another way to gain insight into the dynamical changes in the brain is use neural mass models (NMMs) of the brain regions underlying the recorded EEG [8] and estimate the time-varying neurophysiological parameters of the model from EEG data. Estimated parameter changes could be indicative of critical transitions or changes in brain state and as such hold promise for forecasting brain events. Methods that estimate these time-varying neurophysiological parameters from data have primarily relied upon variations of Kalman filtering [11,12,13], however, we recently developed a novel LSTM [14] neural network-based filter for this purpose that does not require knowledge of the state and parameter initial conditions like Kalman filters do [15]. This overcomes the fact that it is difficult to obtain such initial conditions from living human brains. These

neurophysiological parameter estimation methods have yet to be applied in the context of brain event forecasting in long-term (> 6 months) data recordings.

Brain event forecasting methods have been applied in various contexts such as epileptic seizure prediction [3,4] or forecasting depression and anxiety [5,6,7]. Here we focus on the development of brain event forecasting methods within the context of epileptic seizure prediction. Around 30% of epilepsy patients are unresponsive to medication or surgery [4]. Reliable seizure prediction could provide these patients warning prior to seizures to avoid dangerous environments or the warnings could be used to activate seizure control interventions via electrical stimulation or drug delivery via a brain implant [3].

A plethora of seizure prediction methods based on a combination of feature extraction, machine learning or both have been developed [3,4]. However, most studies rely on the use of the most readily available data, short-term intracranial EEG data (<2 weeks per patient), and this limits their ability to be reliably assessed because seizures are rare events and sometimes only a handful can occur within a week. A handful of studies which have been able to access long-term (>6 months) recordings obtained with proprietary brain implants have provided the most reliable findings [10,16,17]. An example of such a brain implant and a short intracranial EEG segment recorded by a device. This device was used to provide the data in this study. SOTA seizure forecasting methods based on critical slowing down features [10] or sub-clinical epileptiform activity [17] both rely on the analysis of signal features on multiple long time-scales (>40 minutes). Analyzing the signals on longer time-scales seems to help to remove noise from the signals and leads to better forecasting performance.

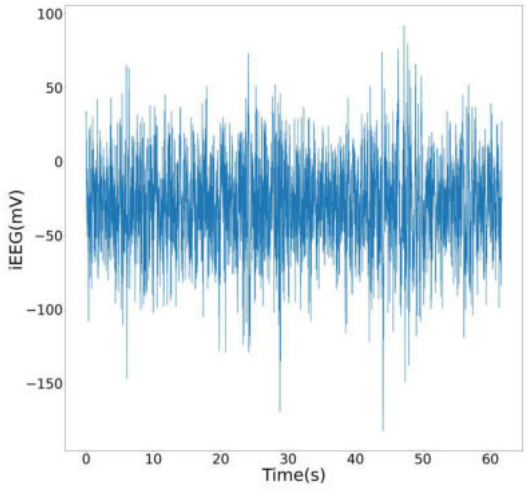

Fig. 1. Example of 1 minute of interseizure iEEG data from one of the 16 channels. Preseizure data has a similar visual appearance.

Here we seek to improve upon SOTA epileptic seizure forecasting achieved when using critical slowing down features and hand-picking the best intracranial EEG channel [10]. To do this we will use probabilistic regression methods to combine critical slowing down features across channels with the intention of alleviating the need to select the best channel. In addition, we will apply novel LSTM filter [15] based neurophysiological parameter estimation to epileptic seizure forecasting for the first time and compare it against the critical slowing down based methods.

In this paper, we will elaborate the methodology we have applied to improve SOTA seizure forecasting. The preprocessing and feature extraction techniques will be discussed first, including the Neural Mass Model implemented in this paper, and the LSTM filter we have developed. The two forecasting approaches with the comparison against the original approach will be covered with the detailed illustration and results. Lastly, we will discuss the limitation and future works.

II. Methods

A. Data

Data of 15 patients from a seizure advisory system clinical trial was analysed [16]. Use of this data in this paper is approved by Human Research Ethics Committee, St Vincent's Hospital, Melbourne - approval LRR145/13. Sixteen subdural electrodes were implanted in each patient on top of the cerebral hemisphere containing the presumed seizure focus. The electrode cables were tunnelled to a subclavicularly-position implanted telemetry unit. Data were sampled at 400 Hz with signed 16-bit resolution and wirelessly transmitted to an external personal advisory device. Figure 1 shows an example of a one-minute intracranial EEG (iEEG) data from a single channel. iEEG recorded from the 16 electrode contacts were referenced to the group average across all electrode channels. Recordings from the 15 subjects lasted a duration of between 6 months and 3 years per patient. Further details on the data can be found in [16]. As mentioned above recordings of this duration are needed to adequately capture enough seizures per participant to develop reliable seizure prediction methods. This is also because patient-specific seizure prediction models tend to work better than building a general model based on many patients [18]. The following work focuses on patient-specific brain event forecasting.

B. Preprocessing and Feature Exctraction

Following the analysis method of SOTA critical slowing down based seizure forecasting [10], features were calculated for every two minutes on each channel. The data of the first three seconds were obtained for each two-minute window, and were filtered using a finite impulse response filter with a cut-off of 170 Hz. This is to remove the 200 Hz artificial noise in the data when the device was charged. The first and the last

second were discarded after filtering. That is to say, for every two-minutes, we only keep a one-second filtered recording.

Based on the one-second filtered recording, we calculate the features, which include (1) the two critical slowing down features of ACFW and Variance [10] and (2) estimated NMM Parameters [15]. The ACFW was taken as the width at the half maximum of the autocorrelation function. On the other hand, the NMM parameters were estimated using a novel LSTM filter [15]. The NMM used to model a single channel of intracranial EEG is depicted in Fig. 2. The model consists of three neural populations: pyramidal, excitatory stellate and inhibitory. The populations approximate the neural populations present in cerebral cortex. The NMM is dependent on its state vector and a set of parameters that define external input levels, inhibitory and excitatory time constants, and the connection strengths between the neural populations. The intracranial EEG is assumed to be linearly dependent on the model state vector. This enables one to estimate the state vector and parameters of the model by feeding single channel intracranial EEG data into the LSTM filter. Full details of the NMM and the LSTM filter are given in [15]. To keep the number of features in this study low, only the time-varying average external input, inhibitory time constant and excitatory time constant parameters were estimated using each intracranial EEG channel recording. Other parameters were held constant in accordance with [15].

The NMM is a forward model which generates EEG like signals given a set of parameters. The post-synaptic potential of population m arising as a result of input from pre-synaptic population n is expressed as

$$v_{mn}(t) = \alpha_{mn} \int_{-\infty}^{t} h_{mn}(t-t')\phi\big(v_n(t')\big)dt', \qquad [\,1\,]$$

where α_{mn} is the population averaged synaptic connectivity strength.

The convolution in equation [1] can also be written as two coupled, first order, ordinary differential equations.

$$\frac{dv_{mn}}{dt} = z_{mn}, \qquad [\,2\,]$$

$$\frac{dz_{mn}}{dt} = \frac{\alpha_{mn}}{\tau_{mn}}\phi_{mn} - \frac{2}{\tau_{mn}}z_{mn} - \frac{1}{\tau_{mn}^2}v_{mn}, \qquad [\,3\,]$$

where τ_{mn} is a lumped time constant, and m and n indicate the pre-synaptic and post-synaptic neural population, respectively. α_{mn} is the parameter we are mainly estimating to represent the connectivity strength between neural populations.

Furthermore, it is helpful to express the NMM in matrix-vector notation so the operations can be expressed more compactly.

$$\dot{x}(t) = Ax(t) + B\vec{\phi}\big(Cx(t)\big), \qquad [\,4\,]$$

$$y(t) = Hx(t) + v(t), \qquad [\,5\,]$$

where H is the observation matrix, $v(t) \sim N(0, R)$ is the observation noise and $y(t)$ is the membrane potential of the pyramidal population, which is considered to be the contributor to the generation of the EEG signal in our model.

In order to show the set of parameters for estimation, we define a vector of parameters as $\theta = \left[u\ \alpha_{pe}\ \alpha_{pi}\ \alpha_{ip}\ \alpha_{ep}\right]^T$. This set corresponds to the input μ to the model and the population averaged connectivity strength parameters. Moreover, we assume the time constants of the model to be constant to simplify the estimation problem. The parameter vector is combined with the state vector X to from the augmented state vector,

$$\xi = [\mathrm{X}^{\mathrm{T}}\theta^{\mathrm{T}}]^T, \qquad [\,6\,]$$

The augmented state-space model is then

$$\xi_t = A_\theta \xi_{t-1} + B_\theta \phi(C_\theta \xi_{t-1}) + W_{t-1}, \qquad [\,7\,]$$

where W_t is Gaussian noise.

To link an LSTM neural network with the NMM, we customised the loss function to allow the LSTM model to follow the mathematical relationship between the observation and the state. Since the LSTM is able to produce the values of both the observation and the state for the current timestep t given the state of the last timestep $t-1$, it is possible to compare the state at timestep t with the state provided by the NMM, so we can know the difference between these two predictions. If we link the NMM to the loss function, the LSTM model would learn to minimise the difference.

The basic loss function for a regression problem at a single timestep is defined as

$$\text{Squared Error} = \left[\left(\xi_{\mathrm{t}} - \hat{\xi}_t\right)^2,\ \left(y_t - H\hat{\xi}_t\right)^2\right], \qquad [\,8\,]$$

where ξ is the augmented state, y is the observation of the training data and $\hat{\xi}$ and $H\hat{\xi}_t$ are the augmented state and observation predicted by the LSTM model, respectively.

The model error is then added, which is the error between the state of the LSTM model and the state generated by the NMM:

$$\text{Model Error} = \left[\left(\hat{\xi}_t - \left(A_\theta\hat{\xi}_{t-1} + B_\theta\phi\left(C_\theta\hat{\xi}_{t-1}\right)\right)\right)^2, 0\right] \qquad [\,9\,]$$

By minimising the model error, the LSTM model will learn to follow the mathematical expression of the NMM, since the training has to minimise the error between its own prediction and the state generated by the NMM.

Lastly, as we consider the connectivity strength parameters to be slowly changing parameters compared to the membrane potential and simulated EEG signals, we have to control the rate of change for these parameters. We can add the standard deviation of the parameters to the loss function to limit the changing rate. However, the parameters also have to be adjusted rapidly when they are too far away from the NMM. Thus, the standard deviation is combined with the model error as

$$Std = s(\alpha)\left[\left(\hat{\xi}_t - \left(A_\theta\hat{\xi}_{t-1} + B_\theta\phi\left(C_\theta\hat{\xi}_{t-1}\right)\right)\right)^2, 0\right] * k \qquad [\,10\,]$$

$$s(\alpha) = [0, \ldots 0, std(\alpha), 0]^T, \qquad [\,11\,]$$

where α is the vector of the four connectivity strengths parameters and k is an adjustable weight that is set to 0.1 as the default value.

The final loss function is then the summation of equation Squared Error = $\left(\xi_{\mathrm{t}} - \hat{\xi}_t\right)^2$, $\left(y_t - H\hat{\xi}_t\right)$2, [8], [9], and [10].

The LSTM filter was trained using a dataset comprising simulated EEG signals, where the true NMM parameters are systematically varied. This approach allows the LSTM network to associate specific signal patterns with corresponding parameter values. Training involved the minimization of a loss function involving the mean squared error between the network's predictions and the actual NMM parameters. However, in this case it was also customised incorporating the NMM, so the LSTM would follow the behaviour of the model [15]. Once trained, the LSTM model was employed to estimate NMM parameters from new EEG data. This was done for each one second of data taken from each 2 minute window. The average parameter estimates over that one second were taken to be the feature values for that period. Previously in controlled simulations the model's estimation accuracy has been evaluated for a large range of benchmark parameter values using metrics such as R-squared and root mean squared error (RMSE) to demonstrate the high estimation accuracy and robustness of the LSTM filter [15].

As mentioned above SOTA seizure forecasting methods [10,17] rely on the analysis of signal features on multiple long time-scales (>40 minutes). In accordance with prior studies [10], 'long/slow' and 'short/fast' cycle time-series were obtained from each feature time series in a causal way, to ensure the data would not be affected by later data. After the calculation of the feature (ACFW, variance or NMM parameters) time series, a trailing moving average was applied with a window of 2 days to identify long/slow cycle data. Separately, a trailing moving average with a window of 40 minutes was applied to identify the short/fast cycle. The Hilbert Transform [19] was implemented on both the fast and the slow cycles, from which signal phases were derived.

To summarise, for the best channel critical slowing down case this amounted to a total number of 4 features per channel (2 base features (ACFW, variance) x 2 cycles (slow and fast)).

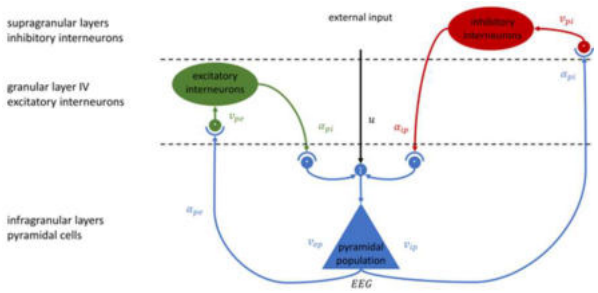

Fig. 2. Neural Mass Model (NMM) of the Intracranial EEG

For the multichannel critical slowing down case this amounted to 64 features (2 base features x 2 cycles x 16 channels). While for the multichannel neural parameter case this amounted to 80 features (3 base features (external input, inhibitory time constant, excitatory time constant) x 2 cycles x 16 channels).

C. *Forecasting Approach*

We aim to provide warning in the 2 to 4 minutes before a seizure, the event of interest in the current application. Since each data point represents the data in a two-minute window, the data point before the window where a seizure happens was labelled as preseizure stage. Training and testing was then based on these labels, meaning that we only consider providing warning during the preseizure stage as a success. In a perfect scenario, the system would be silent at all times except for the preseizure stage window.

Best channel critical slowing down case: For this we reimplemented the existing SOTA method [10], seizure forecasting was done by analysing the signal phases on both slow and fast cycles for the ACFW and variance features for each channel separately to find the best performing channel. The probability of the preseizure stage was calculated by evaluating the historical data. As illustrated in Fig. 3. the Hilbert transform phase, between $-\pi \leq \theta < \pi$, for a given feature was broken into 20 equally spaced bins. The probability of the preseizure stage occurring in one single bin is the number of historical preseizure stages falling into this bin S_θ, divided by the number of total phases in this bin N_θ:

$$P(S|\theta) = \frac{S_\theta}{N_\theta} \ . \qquad [12]$$

For each EEG channel four probabilities obtained for the four features were multiplied together to form the joint seizure probability used for forecasting based on an online method outlined in Fig. 3. To construct an online pseudo-prospective forecasting method we calculated the joint seizure probability distribution conditioned on the features based on the first 50 days of data. However, there were some patients who have very limited amount of seizures in the first fifty days. In this case, the data until the tenth seizure were included to train the initial model. This initial distribution model is then used to determine seizure probability in the next 2 to 4 minutes given the current features as a sliding 2 minute window scrolls through the data. Every time there is a new seizure the data leading up to that seizure is added to the seizure probability distribution to update it and deal with non-stationarity in the data. Seizure predictions are generated when the seizure probability crosses a threshold. The threshold is iteratively determined in an online manner based on the data up until the previous seizure and optimising by maximising the

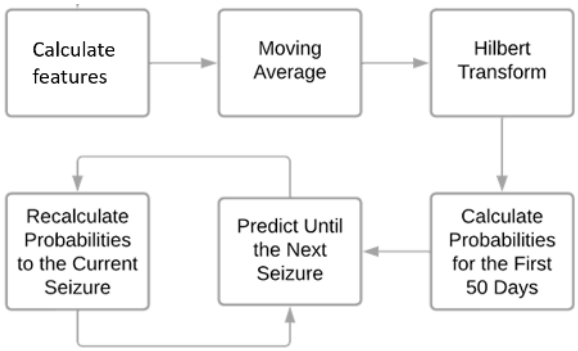

Fig. 3. Schematic of the online method

multiplication of the time spent in low risk and the sensitivity. Here sensitivity is defined as percent of seizures correctly predicted and time in low risk is the percent of time the forecaster is below threshold. Out-of-sample seizure forecasting performance is based on an online approach where we iteratively evaluate performance between each previous seizure and the next seizure, before the next seizure and the data preceding it is added to the probability distribution used to generate predictions. The best channel was found by evaluating out-of-sample sensitivity and time of low risk for each channel for each patient.

Multichannel critical slowing down case: In this case the phases of the slow and fast cycles of ACFW and variance for all channels were fed into a logistic regression model. The same online method used in the best channel case was used to iteratively train and evaluate the logistic regression model

(with fixed threshold of 0.5), where logistic regression takes the place of the seizure probability distribution. The nature of the data is extremely imbalanced, as preseizure stages are rare compared to interseizure stages. Therefore, the positive class was given a higher weight each time the logistic regression model was trained by matching the ratio of the two classes. Cross validation was performed to control the regularisation strength, while the performance was evaluated based on the AUC score. Cross validation was designed for the nature of the time-series data where preseizure stages were rare. Data were divided based on the time of the preseizure stages trying to make sure each fold would have the same amount of preseizure stages.

Multichannel neural parameter case: The only difference between this case and the multichannel critical slowing down case is that the NMM parameter estimate features are fed into logistic regression instead of the critical slowing down features.

III. Results

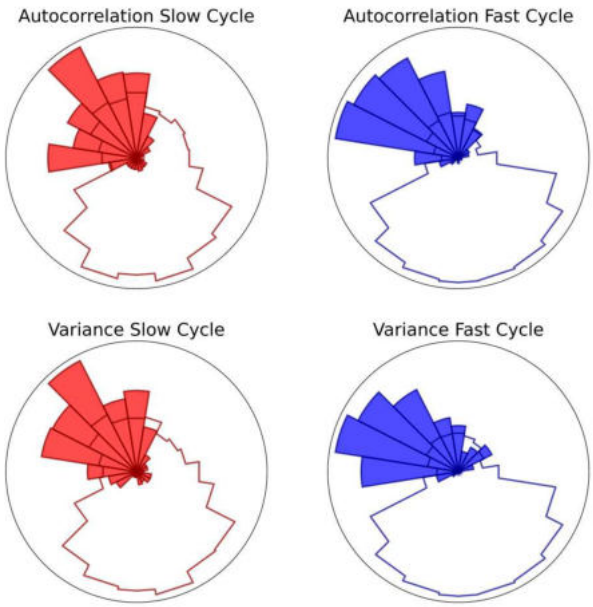

Fig. 4. Single patient example autocorrelation (top) and variance (bottom) normalized phase histograms for slow (red) and fast (blue) cycles. In each polar plot, the colored lines represent the normalized distribution of phases of the entire signal and the filled histogram represents the distribution of phases at the sample prior to the seizure times.

Examples of phase histograms of the critical slowing down features (ACFW and variance) and the NMM parameter features are shown for a single patient in Fig. 4 and 5, respectively. It can be seen that both feature groups show a tight tuning to particular phases of the slow and fast cycles. This suggests that phase of these features can be used for seizure forecasting.

For an example patient and EEG channel, Fig. 6 shows the raw (gray), fast (blue) cycle and slow (orange) cycle time series for ACFW (top) and the external input parameter of the NMM (middle). It can be seen that the signals tend to peak near the seizure times (red triangles). For the same patient the seizure risk probability output by logistic regression for the multichannel critical slowing down case is displayed in the bottom plot of Fig. 6. It can be seen that seizure risk also tends to increase near seizure times.

Out-of-sample seizure forecasting performance in terms of sensitivity and time in low risk is captured in Fig. 7 for each patient. It can be seen that the best channel critical slowing down (original/gray dots), multichannel critical slowing down (ACFW+Var/blue dots) and multichannel neural parameter (NMM/red dots) cases lead to a diverse range of performances depending on the patient. The methods tend to have comparable time in low risk and sensitivity, except for a few patients where the multichannel neural parameter method is less sensitive. This is reflected in the median performance values provided in Table 1. This indicates that a multichannel critical slowing down model approach can perform as well as the best channel approach, removing the need to find the best channel. It also suggests neurophysiological features could be used to increase time in low risk at a potential cost to sensitivity.

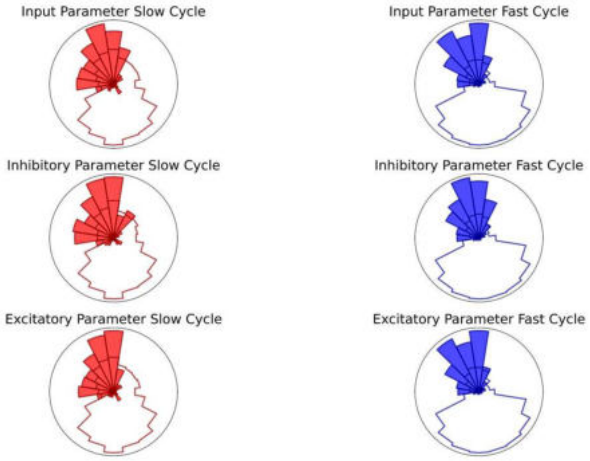

Fig. 5. Single patient example neural input (top), inhibitory (middle) and excitatory (bottom) parameter normalized phase histograms for slow (red) and fast (blue) cycles. The colored lines and filled histograms follow the same description as Fig. 4.

TABLE I. Out-of-sample seizure forecasting performance

Method	***Median Sensitivity (%)***	***Median Time in low risk (%)***
Best channel critical slowing down	67	81
Multichannel critical slowing down	70	84
Multichannel neural parameter	54	84

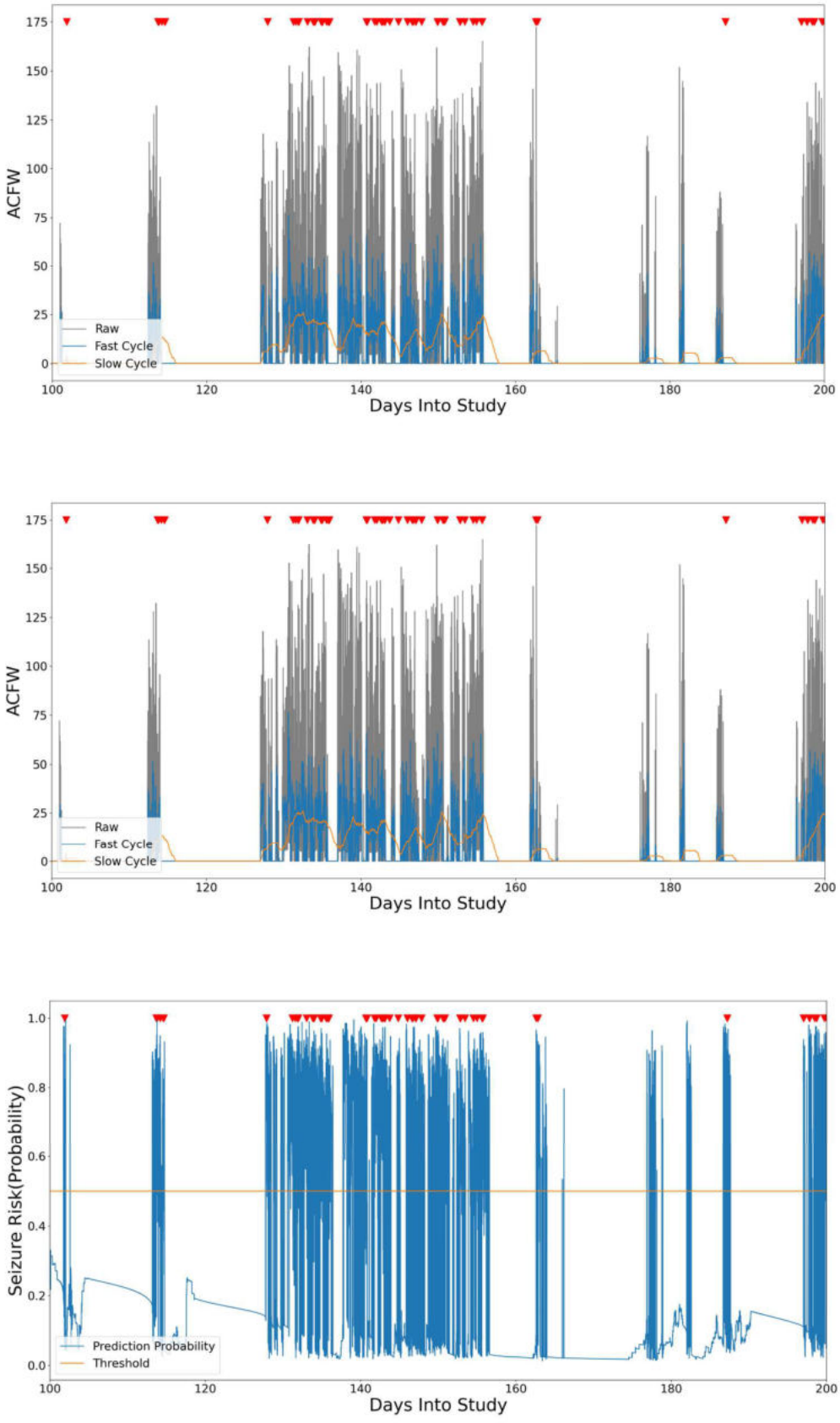

Fig. 6. Example ACFW (top), Model External Input Parameter Estimate (middle), and Seizure Risk (bottom; derived for the multichannel critical slowing down case) time series from a patient over 100 days. Seizure times are labelled with red triangles.

IV. Discussion

It is proven with the result in this paper that it is possible to find a multichannel critical slowing down feature based method which is comparable to the best channel method [10]. As the best channel must be chosen prior to applying the methodology, it is hard to determine which channel has the best performance, or is at least beneficial to be included in the prediction task. The proposed method does not require any channel to be included or excluded prior to implementation. This can also be further studied to find the features that are important and their channels, which lead to the brain areas that are more suitable for prediction.

A slightly higher median sensitivity is achieved by the proposed multichannel critical slowing down method, while also reaching a higher median time in low risk. Compared to the original method where fast and slow cycles of features are multiplied, the proposed method has treated them individually by feeding the phase directly into a logistic regression model. This provides more information to the machine learning model, as some feature vectors might show more separable patterns in higher dimensional space than simply by multiplying their probabilities together (which effectively assumes features have independent influences on seizure risk). We have also experimented on feeding the multiplication of probabilities into the logistic regression model, and found that to have a lower performance, which proves each feature can provide more information by its own.

The current result shows the performance is highly influenced by the imbalance of sensitivity and time in low risk, as some patients have shown very low sensitivity. It suggests the overall performance can be improved if we can fix this problem. The AUC score can be used for balancing the time spent in low risk and the sensitivity. Since we were applying the logistic regression model, the threshold of determining the positive or negative prediction can be changed. At the moment, the threshold is fixed at 0.5, but it can also be dynamic based on the historical data. For example, if the probability is high all the time, the threshold can be increased to balance the sensitivity and the time spent in low risk. If we aim to change the threshold based on the historical data each time a new seizure is predicted, the performance can be improved.

As the decision boundary of the logistic regression is linear, the performance of the model is also not the best. The linear nature of the machine learning model limits the performance. Non-linear machine learning algorithms like tree-based algorithms might achieve a higher performance, while keeping the imbalance under consideration.

This work is also the first time NMM parameter estimation has been applied to seizure forecasting on long-term data. While the multichannel neural parameter method achieved high time in low risk, it only achieved high sensitivity for some patients. This suggests it might still be useful for some patients, or this method could be used to keep a high time in low risk. Potentially the NMM parameter estimates could have complementary time series to the ACFW and variance values, and in such cases both feature cases could be combined into a single model to seek higher performance. Here the LSTM filter was constructed using a generic NMM used in many brain imaging studies [15]. It would be possible to reconstruct the LSTM filter using a NMM that might better model the epileptic brain. This in turn may lead to better performance. Future work exploring other features considered in the seizure prediction literature [3,4], machine learning models and their various combinations could also yield further improvements.

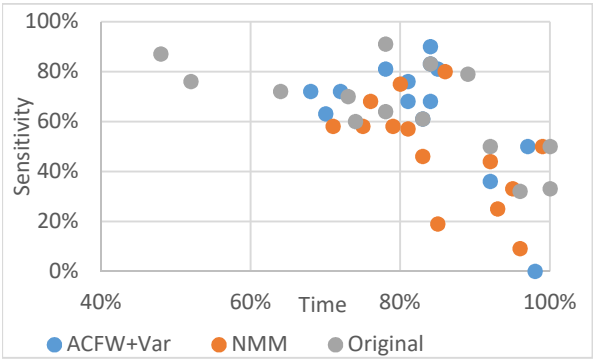

Fig. 7. Patient-wise Seizure forecasting performance for the best channel critical slowing down (original/gray dots), multichannel critical slowing down (ACFW+Var/blue dots) and multichannel neural parameter (NMM/red dots) cases

V. Conclusion

Here we have shown a framework for seizure forecasting based on two dynamical systems approaches (critical slowing down features and inferred neural model features). Moreover, the framework analyses these features on multiple long-time scales to extract signals that vary in synchrony with the occurrence of seizure events, and leverages machine learning models to combine this information to forecast seizures. It is hoped that this framework will have general utility to other applications of brain event forecasting [6,7].

Acknowledgment

This work is supported by grants from the Australian Research Council (DP210100045) and Monash University. We thank MASSIVE (https://www.massive.org.au) and Spartan (https://dashboard.hpc.unimelb.edu.au/) for computational resources.

References

[1] Blinowska, K. J., & Malinowski, M. (1991). Non-linear and linear forecasting of the EEG time series. Biological cybernetics, 66(2), 159-165.

[2] Wein, S., Schüller, A., Tomé, A. M., Malloni, W. M., Greenlee, M. W., & Lang, E. W. (2022). Forecasting brain activity based on models of spatiotemporal brain dynamics: A comparison of graph neural network architectures. Network Neuroscience, 6(3), 665-701.

[3] Mormann, F., Andrzejak, R. G., Elger, C. E., & Lehnertz, K. (2007). Seizure prediction: the long and winding road. Brain, 130(2), 314-333.

[4] Kuhlmann, L., Lehnertz, K., Richardson, M. P., Schelter, B., & Zaveri, H. P. (2018). Seizure prediction—ready for a new era. Nature Reviews Neurology, 14(10), 618-630.

[5] Moore, P. J., Little, M. A., McSharry, P. E., Geddes, J. R., & Goodwin, G. M. (2012). Forecasting depression in bipolar disorder. IEEE Transactions on Biomedical Engineering, 59(10), 2801-2807.

[6] Suhara, Y., Xu, Y., & Pentland, A. S. (2017, April). Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In Proceedings of the 26th International Conference on World Wide Web (pp. 715-724).

[7] Dang, V. (2023). Anxiety and Affective Forecasting (Doctoral dissertation).

[8] Breakspear, M. (2017). Dynamic models of large-scale brain activity. Nature neuroscience, 20(3), 340-352.

[9] Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., ... & Vandermeer, J. (2012). Anticipating critical transitions. science, 338(6105), 344-348.

[10] Maturana, M. I., Meisel, C., Dell, K., Karoly, P. J., D'Souza, W., Grayden, D. B., ... & Freestone, D. R. (2020). Critical slowing down as a biomarker for seizure susceptibility. Nature communications, 11(1), 2172.

[11] Schiff, S. J. (2011). Neural control engineering: the emerging intersection between control theory and neuroscience. MIT Press.

[12] Karoly, P. J., Kuhlmann, L., Soudry, D., Grayden, D. B., Cook, M. J., & Freestone, D. R. (2018). Seizure pathways: A model-based investigation. PLoS computational biology, 14(10), e1006403.

[13] Zhao, Y., Boley, M., Pelentritou, A., Karoly, P. J., Freestone, D. R., Liu, Y., ... & Kuhlmann, L. (2022). Space-time resolved inference-based neurophysiological process imaging: Application to resting-state alpha rhythm. NeuroImage, 263, 119592.

[14] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[15] Liu, Y., Soto-Breceda, A., Karoly, P., Grayden, D. B., Zhao, Y., Cook, M. J., ... & Kuhlmann, L. (2023). Brain model state space reconstruction using an LSTM neural network. Journal of Neural Engineering, 20(3), 036024.

[16] Cook, M. J., O'Brien, T. J., Berkovic, S. F., Murphy, M., Morokoff, A., Fabinyi, G., ... & Himes, D. (2013). Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. The Lancet Neurology, 12(6), 563-571.

[17] Proix, T., Truccolo, W., Leguia, M. G., Tcheng, T. K., King-Stephens, D., Rao, V. R., & Baud, M. O. (2021). Forecasting seizure risk in adults with focal epilepsy: a development and validation study. The Lancet Neurology, 20(2), 127-135.

[18] Massoud, Y. M., Abdelzaher, M., Kuhlmann, L., & Abd El Ghany, M. A. (2023). General and patient-specific seizure classification using deep neural networks. Analog Integrated Circuits and Signal Processing, 1-16.

[19] King, F. W. (2009). Hilbert Transforms: Volume 2 (Vol. 2). Cambridge University Press.