Multi-task multiple kernel learning reveals relevant frequency bands for critical areas localization in focal epilepsy

Vanessa D'Amario VANESSA.DAMARIO@DIBRIS.UNIGE.IT
Federico Tomasi FEDERICO.TOMASI@DIBRIS.UNIGE.IT
Veronica Tozzo VERONICA.TOZZO@DIBRIS.UNIGE.IT
Gabriele Arnulfo GABRIELE.ARNULFO@EDU.UNIGE.IT
Annalisa Barla ANNALISA.BARLA@UNIGE.IT
Department of Informatics, Bioengineering, Robotics and Systems Engineering
Università degli Studi di Genova
Genova, Italy

Lino Nobili LINO.NOBILI@OSPEDALENIGUARDA.IT
Ospedale Ca' Granda Niguarda
Milano, Italy

Abstract

The localization of epileptic zone in pharmacoresistant focal epileptic patients is a daunting task, typically performed by medical experts through visual inspection over neural recordings. For a finer localization of the epileptogenic areas and a deeper understanding of the pathology both the identification of pathogenical biomarkers and the automatic characterization of epileptic signals are desirable. In this work we present a data integration learning method based on multi-level representation of stereo-electroencephalography recordings and multiple kernel learning. To the best of our knowledge, this is the first attempt to tackle both aspects simultaneously, as our approach is devised to classify critical and non-critical recordings while detecting the most discriminative frequency bands. The learning pipeline is applied to a data set of 18 patients for a total of 2347 neural recordings analyzed by medical experts. Without any prior knowledge assumption, the data-driven method reveals the most discriminative frequency bands for the localization of epileptic areas in the high-frequency spectrum ($\geq$ 80 Hz) while showing high performance metric scores (mean balanced accuracy of 0.89 ± 0.03). The promising results may represent a starting point for the automatic search of clinical biomarkers of epileptogenicity.

1. Introduction

Epilepsy is a neurological disorder that affects more than 50 million people worldwide. This disease is characterized by abrupt loss of consciousness and convulsions which may cause severe impairments in daily life. The occurrence of epileptic symptoms can be localized (*focal seizure*) or general (*general seizure*). In the first case, the seizure onset zone is restricted to a portion of the brain which produces both hyper synchronization and hyper activity typical of the pathology (Jiruska et al., 2013). The onset zone can be further sub-categorized into *epiloptogenic* areas which generate the epileptic activity and *irritative* areas that actively

© 2018 V. D'Amario, F. Tomasi, V. Tozzo, G. Arnulfo, A. Barla & L. Nobili.

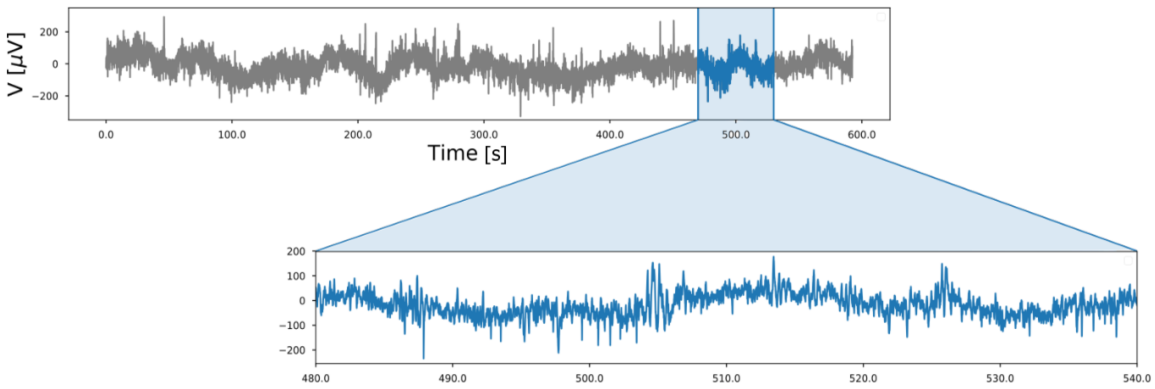


Figure 1: Example of epileptic signal corresponding to 10 minutes of acquisition. The SEEG recordings are characterized by high sampling frequencies (1 kHz). These signals are usually analyzed by clinical experts that look for biomarkers. This process is very subjective and prone to error.

contribute to the propagation. In the rest of the work we refer more generally to these areas as *critical* (or, equivalently, *pathological*).

About 30% of focal epileptic patients do not respond to pharmacological treatments and therefore may need surgical ablation of the pathological area. In such cases, the identification of the minimal amount of neural cortex to ablate for seizure-free outcomes, namely the *epileptic zone* (EZ), is an extremely delicate, precise and necessary task. To this aim, clinicians use non-invasive methodologies such as magnetic resonance imaging, computed tomography and scalp electroencephalography as first clinical tests, seeking for clear evidence of tumors or dysplasia which may cause the seizures. Nonetheless, in some cases, EZ borders may be difficult to localize. Medical experts often resort to the use of invasive investigation techniques such as *stereo-electroencephalography* (SEEG) to assess critical areas.

SEEG measures the electrical activity from intracranial areas through filiform electrodes implantation, where each sensor is endowed with dozen of acquisition channels. These recordings are characterized by high spatial and temporal correlation (see Figure 1). The former is due to the complexity of the brain structure, while the latter is intensified by the long period acquisitions at high sampling frequencies ($\geq$ 1 kHz). The characterization of SEEG signals is an extremely challenging and time-consuming task, usually based on visual inspection or signal processing tools, and it is intrinsically subjective, possibly leading to misclassification (Soriano et al., 2017; Yardi et al., 2016; Staba et al., 2014). Therefore, automatic classification of neural recordings is an emerging field. In this context, methods consider both temporal and spectral representation of the signal (Omerhodzic et al., 2013). The signal can in fact be described by its energy at different frequency bands, which shows highest discriminative power in seizure onset zone detection (γ frequency band, 20-70 Hz) (Vila-Vidal et al., 2017). Also, the quantification of energy concentration at different bands can be a measured with wavelet entropy (Rosso et al., 2001; Mooij et al., 2016).

Another typical approach for the classification of epileptogenic channels, is through the definition of the most informative biomarkers. Interictal spikes and spike-and-wave complexes are considered a well established evidence of the pathological condition (de Curtis

and Avanzini, 2001; Avoli et al., 2006). High-frequency oscillations (HFOs) are short events (2-5 ms) at frequencies in the range of (80-500 Hz), usually sub-categorized in *ripples* (80-200 Hz) and *fast ripples* (200-500 Hz). HFOs have been pointed out as a good predictors for the EZ localization (Fedele et al., 2017). The role of HFOs in seizure generation has been object of investigation, which attest reliable co-occurrence of HFOs in critical areas (Jacobs et al., 2012). Indeed, several works reveal the primary role played by HFOs as biomarkers for epilepsy, see Zijlmans et al. (2012) and Höller et al. (2015) for a comprehensive review.

Clinical Relevance. The main contribution of this work is a machine learning method which simultaneously tackles the problem of searching for informative frequency bands and localizing critical areas in focal epilepsy, through a multi-scale integration of SEEG recordings. The method leverages on continuous wavelet representation of the signal, and exploits its multi-level nature to obtain a redundant description that is integrated through pairwise similarity measures in the learning pipeline.

Technical Relevance. The number of channels and the type of signal acquired changes from patient to patient. Therefore, a direct comparison between patients is not feasible. Our procedure overcomes this issue by extending a *multiple kernel learning* (MKL) algorithm (Borgwardt, 2011; Wang et al., 2010). This learning method, namely *multi-task multiple kernel learning* (MT-MKL), performs a multi-task classification problem. It includes some additive constraints that guarantee robustness to noise and provides interpretable, stable results. The outcome of the method incorporate both a personalized description for each patient and the selection of the best descriptors of the pathology across the population.

Outline. The rest of the paper is organized as follows. Section 2 presents the methodological background needed for our analysis. Section 3 describes our data set and the proposed learning pipeline. The related results of the method and further discussion are detailed in Section 4. Finally, Section 5 concludes by giving possible future directions.

2. Methodological background

Here we recall state-of-the-art multi-scale representation methods with the aim of differentiate the SEEG recordings in several frequency bands. Then we introduce standard phase and amplitude correlation measures computed for each band to quantify the pairwise similarity between neural signals. Finally we present the background concepts of multiple kernel learning.

2.1. Data representation through multi-scale analysis

A multi-scale representation of a 1D signal allows to separately consider its behavior at different frequency bands. To such purpose, we used the wavelet transform that, due to its local nature, is able to detect transients through a time-frequency representation. In particular, the *continuous wavelet transform* (CWT) with a set of generators (*mother wavelet*) gives a rich and redundant decomposition of the signal (Mallat, 1999).

We choose as mother wavelet the complex Morlet transform:

$$\Psi_{\tau,s}(t) = \frac{1}{\sqrt{\pi s}} e^{2i\pi \frac{t-\tau}{s}} e^{-\frac{(t-\tau)^2}{s^2}}, \tag{1}$$

where τ, s denote respectively the temporal shift and the scaling parameter. For any one dimensional signal $x(t)$, its representation through wavelet transform Ψ at a fixed scale s is given by the coefficients $W_{\tau,s}(x)$. Therefore, CWT results in a two dimensional representation of the original signal.

2.2. Similarity measures

The decomposition of neural signals through CWT is necessary in the perspective of a multi-level pairwise comparison and a deeper insight in the role played by different frequency rhythms in the discrimination of EZ. In order to obtain a comprehensive description of the signal, it is necessary to consider measures involving both phase and amplitude correlation.

In the following, for any 1D signal $x(t)$ we denote as $|W_{\tau,s}(x)|$ the element-wise magnitude of the wavelet coefficients and its phase as $\Phi_{\tau,s}(x) = \arctan(\Im(W_{\tau,s}(x))/\Re(W_{\tau,s}(x)))$.

All the considered measures are bivariate and computed by taking into consideration two signals $x(t)$ and $y(t)$ at a specific scale s, resulting in positive semi-definitive symmetric matrices.

Phase locking value. A measure of phase synchrony between bivariate measures in neuroscience is known as *phase locking value* (PLV) computed as follows (Lachaux et al., 1999):

$$\mathrm{PLV}_s(x,y) = \frac{1}{T}\left|\sum_{\tau=1}^{T} \exp\Big(-i\big(\Phi_{\tau,s}(x) - \Phi_{\tau,s}(y)\big)\Big)\right|, \tag{2}$$

that is, the phase difference of the two signals at each time point. PLV ranges in $[0,1]$, where the maximum value corresponds to two signals perfectly synchronous (*e.g.*, when $x(t)$ is the same as $y(t)$ during all the acquisition period).

Normalized correlation. The amplitude similarity of two neural recordings is quantified through normalized correlation, as follows:

$$\mathrm{cov}_s(x,y) = \frac{\mathbb{E}_\tau\left[\left(|W_{\tau,s}(x)| - \mu_{|W_{\tau,s}(x)|}\right)^T\left(|W_{\tau,s}(y)| - \mu_{|W_{\tau,s}(y)|}\right)\right]}{\sigma_{|W_{\tau,s}(x)|}\sigma_{|W_{\tau,s}(y)|}}, \tag{3}$$

where $\sigma_{|W_{\tau,s}(x)|}$ is the standard deviation of the coefficients $|W_{\tau,s}(x)|$. Normalized correlation measures the amplitude similarity of two signals, but it is not invariant to relative temporal shift. For this reason, correlation is a reliable quantity when the lag between similar patterns is negligible compared to the pattern length.

Spectral measures. To overcome the problem of temporal lag between two time series, we resort to shift invariant spectral measures of similarities. We define the cross power spectral density as the Fourier transform of the convolutional product, denoted by $*$, of the absolute value of wavelet coefficients for the two signals, computed as follows:

$$(P_{x,y})_s = \mathcal{F}(|W(x)_{\tau,s}| * |W(y)_{\tau,s}|). \tag{4}$$

In fact, the module of cross power spectral density is invariant in the time and frequency domain, as stated in Parseval theorem (Tolstov, 2012). In order to get the similarity between spectra, Equation (4) is then normalized as follows:

$$\tilde{P}_s(x,y) = \frac{\|(P_{x,y})_s\|^2}{|(P_{x,x})_s| \cdot |(P_{y,y})_s|}. \tag{5}$$

2.3. Multiple kernel learning

Multiple kernel learning (MKL) integrates data by combining sets of kernel functions (Lanckriet et al., 2004; Borgwardt, 2011). *Kernels* are positive semi-definite matrices whose entries $K_{i,j} = k(x_i, x_j)$ encode pairwise similarities between data points (x_i, x_j). However, the choice of the most suitable kernel function for each problem at hand is tricky, and heavily depends on available data. Therefore, the idea behind MKL is to construct different measures of similarity on the same data set and then integrate them into a single kernel (Gönen and Alpaydın, 2011). For example, a straightforward MKL may be a linear combination of different kernels (Borgwardt, 2011). This is possible given the fact that kernels allow linear operations while preserving their mathematical properties, such as positive semi-definiteness and symmetricity (Friedman et al., 2001). Formally, consider k kernels $\{K_1, \ldots, K_k\}$ that represent different similarities measures among points of a data set. Such kernels can be combined linearly as a weighted sum $\sum_{i=1}^{k} w_i K_i$, where $\boldsymbol{w} = (w_1, \ldots, w_k)^\top \in \mathbb{R}_+^k$ is a list of (non-negative) coefficients, obtained through an optimization procedure. Such coefficients measure the importance of each kernel for the particular problem at hand.

3. Materials and methods

The data set was acquired at Ospedale Ca' Granda Niguarda, Milan (Italy). Patients provided written consent for further analysis with scientific research purpose. We registered local field potential with common reference in white matter, using multi-lead electrodes in platinum iridium. Each electrode was endowed with a number of channels that varies from 8 to 15. For further details on the acquisition system we refer to Arnulfo et al. (2015).

We analyzed signals recorded from 18 patients. For each patient we considered 590 seconds of spontaneous interictal activity with closed eyes at resting state, at 1 kHz sampling frequency. The total number of channels was 2347, 984 of which were tagged as critical by medical experts.

3.1. Multi-task Multiple Kernel Learning

The implantation settings are strongly dependent on the patient clinical condition and on preliminary medical evaluations, derived from previous non-invasive clinical tests. Such variability does not allow for a direct comparison of the neural activity across patients. In other words, given the fact that data from each patient was acquired differently, it is not possible to use a unified regression model for all patients. Therefore we need to look at part of the channels of each patient in order to classify the rest, for this reason classifying a completely new patient is not a straightforward task. In order to handle such problem we extended the MKL to multi-task multiple kernel learning (MT-MKL) to account for different patient conditions. Each kernel represents a particular similarity matrix among all the channels in a single patient at a specific scale. The innovation in our method consists in the capability of jointly analyzing the patients by taking into account their diversity. Our goal is two-fold: *(i)* to combine kernels to predict whether a channel is epileptic or not, and *(ii)* to identify the most informative kernels for prediction across patient. For each patient we have a matrix $\boldsymbol{X}^{(p)} \in \mathbb{R}^{c_p \times T}$ and a vector of labels $\boldsymbol{y}^{(p)}$ which denotes the pathological or physiological nature of each signal in $\boldsymbol{X}^{(p)}$. Note that the number of

channels c_p varies across patients, and the proportion of epileptic channels is not uniform across the population considered. We denote with $(K_1^{(p)}, \ldots, K_k^{(p)})$, where $K_j^{(p)} \in \mathbb{R}^{c_p \times c_p}$ the set of k kernels for the generic patient p. The decision function $f^{(p)}$ for a patient p and a channel x is defined as:

$$f^{(p)}(x) = \alpha_0^{(p)} + \sum_{i=1}^{c_p} \left[\alpha_i^{(p)} \sum_{j=1}^{k} w_j K_j^{(p)}(x_i, x) \right], \tag{6}$$

where $\alpha_i^{(p)}$ denotes the i-th component of the regression parameter $\boldsymbol{\alpha}^{(p)}$ specific for each patient p. Having separate parameters ($\boldsymbol{\alpha}^{(p)}$ and $\boldsymbol{w}$) is fundamental for the resolution of our problem. In fact, $\boldsymbol{\alpha}^{(p)}$ allows to better approximate the labels $\boldsymbol{y}^{(p)}$ by capturing the variance of each patient, while $\boldsymbol{w}$ combines the kernels by weighting them and, as it holds across patients, it provides relevant indication of the most discriminative kernels.

In order to obtain interpretable results and a more stable solution we also add an $\ell_1\ell_2$ penalty on $\boldsymbol{w}$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(n)})$. By considering all the patients, our goal translates into minimizing the following objective function:

$$\begin{aligned} \underset{\boldsymbol{\alpha}^{(1)},\ldots,\boldsymbol{\alpha}^{(n)},\boldsymbol{w}}{\text{minimize}} \Bigg\{ & \sum_{p=1}^{n} \left(\ell_{f^{(p)}} \left(\boldsymbol{X}^{(p)}, \boldsymbol{y}^{(p)} \right) + \lambda (r_\lambda \|\boldsymbol{\alpha}^{(p)}\|_1 + (1 - r_\lambda) \|\boldsymbol{\alpha}^{(p)}\|_2^2) \right) \\ & + n\beta \left(r_\beta \|\boldsymbol{w}\|_1 + (1 - r_\beta) \|\boldsymbol{w}\|_2^2 \right) \Bigg\} \\ & \text{s.t. } w_j \geq 0 \text{ for each } j = 1, \ldots, k \end{aligned} \tag{7}$$

where $\ell_{f^{(p)}} \left(\boldsymbol{X}^{(p)}, \boldsymbol{y}^{(p)} \right) = -\sum_{i=1}^{c_p} \log(1 + \exp(-y_i^{(p)} f^{(p)}(x_i)))$ is the negative log-likelihood of the logistic probability function and r_λ and r_β are the elastic-net penalty penalty ratios on $\boldsymbol{\alpha}$ and $\boldsymbol{w}$, respectively. The elastic-net penalty benefits indeed from the well-known stability property of the ℓ_2 regularization term (Zou and Hastie, 2005).

For the optimization of the functional in Equation (7) we rely on alternating minimization (Bolte et al., 2014). Note that the problem is not jointly convex in both $\boldsymbol{\alpha}$ and $\boldsymbol{w}$ and therefore there is no theoretical guarantee to converge to a global minimum. Further details on the minimization procedure are given in Appendix A.

3.2. Pipeline design

Our pipeline consists of three main steps: *(i)* signal preprocessing and multi-scale representation of the signals, *(ii)* computation of similarity measures, and *(iii)* learning the optimal combination of kernels and channels weights for signal classification (Figure 2).

In step *(i)*, given an input matrix $\boldsymbol{X}^{(p)}$ we re-refer the potential using the bipolar montage, which consists in the differential measures between two adjacent channels. Local reference is standard for phase measures, as it reduces volume conduction effects caused by white matter (Mercier et al., 2017). The output of this operation is then filtered, in order to remove power line effect (50 Hz and harmonics in Europe), using a FIR bandstop filter with 2 Hz bandwidth. Then, we compute the continuous wavelet transform over each SEEG recording. With respect to Equation (1) the shift parameter τ takes discrete values

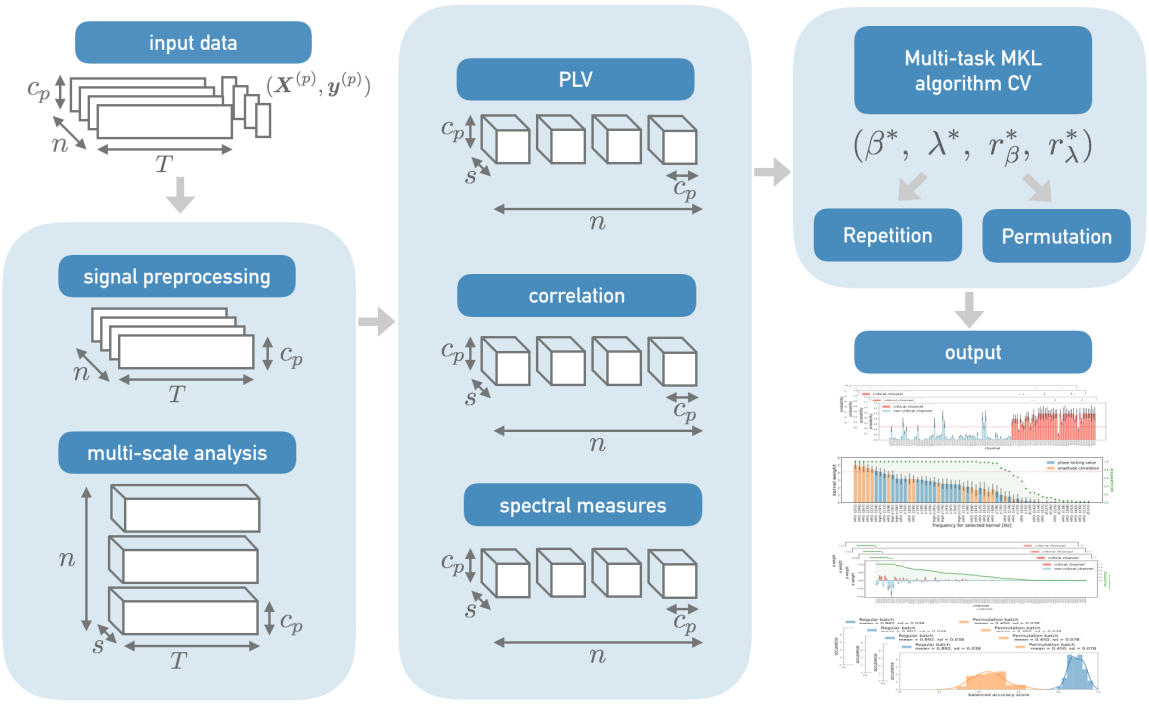


Figure 2: Schematic representation of the learning pipeline. From top left, SEEG recordings are first filtered. In the multi-scale step, we represent each time series through CWT obtaining a 2D representation for each signal. The central panel represent the similarity measure computation step, applied for each scale of the wavelet transform. The output of this operation consists in 300 kernels for each patient. In the last panel, through multi-task multiple kernel learning algorithm we learn, on the final data set the optimal hyperparameters. Then we repeat MT-MKL 50 times to obtain statistics on the parameters $(\boldsymbol{w},\ \boldsymbol{\alpha})$, the vector of classification probabilities and permutation test results.

in $[1, T]$, with T number of points of each time series. The scaling parameter s, instead, is chosen to be a list of 100 elements equally spaced in the logarithmic scale in the interval $[0.3, 3]$. Fixing s, the central frequency f_a of the mother wavelet corresponds to $f_a = \frac{1.}{st_s}$; with $t_s = 1$ ms, sampling period. Consequently, the values of f_a vary in the range between 0.5 Hz and the Nyquist frequency, corresponding to 500 Hz.

In step *(ii)* the multi-scale representation is given as input to different similarity measures, namely PLV, correlation and spectral measures. For each patient p, data are transformed into $k = 3 \times 100$ kernels, each of dimension $c_p \times c_p$. The computation of spectral measure (see Equation (5)) over all the time series is heavily intensive, therefore we approximate this quantity by averaging its estimation on smaller, non-overlapping windows of the signal (5.9 seconds length each).

Finally, in step *(iii)*, we apply MT-MKL on the resulting kernels. To assess the mean performances we bootstrap 50 times the procedure. In particular the dataset was split, for each patient, in half channels for the learning set and the other half for the validation set. The proportion between critical and non-critical channels in each set is kept fixed. The learning set is then used to select the optimal hyperparameters with a 3-fold grid search

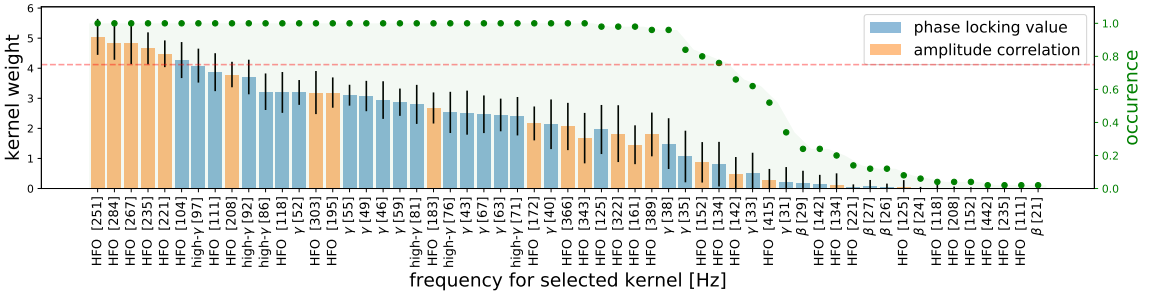


Figure 3: Kernels which mostly contribute in the characterization of the epileptogenic areas. These measures are reported on the x-axis. In square bracket we put the central frequency of the mother wavelet, and the typical event type related to each frequency. We assign blue color to phase measures and orange to amplitude. Each bar corresponds to weight mean value and standard deviation through the 50 repetitions of the experiment. The right y-axis denotes the occurrence value, the green dots correspond to the number of times each kernel was selected throughout the repetitions. The dashed line indicates the 0.75% of occurrence value.

cross validation and the score is computed on the validation set. The procedure selected as mean hyperparameters (across all the 50 repetitions) $\overline{\lambda} = 0.1$, $\overline{\beta} = 5 \cdot 10^{-3}$, $\overline{r_\lambda} = 0.95$, $\overline{r_\beta} = 0.5$, using balanced accuracy averaged on all patients. We also perform a permutation test to quantify the difference between the normal and the permutated distributions.

The outcomes of the pipeline are: a vector $\boldsymbol{w}$, whose components weight each similarity measure, shared across all patients; measures on single subject that include statistics on the set of coefficients $\boldsymbol{\alpha}$, through which we classify previously unseen channels; statistics on the probability of each channel of being critical; metrics scores of the classification task and permutation test value.

4. Results

For a more compact writing we present here results that are shared across patients, we refer to Appendix C for discussion and results on single patient. In particular, we show, for each patient, the most informative channels and the probability for each of them to belong to a certain class. The selection of relevant channels is enlighten for further clinical evaluations, since it constitutes an informative tool to detect areas which are likely characterized by abnormal and relevant activity for the EZ discrimination task.

Figure 3 shows the kernels which were selected at least once and ordered by their occurrence throughout the 50 repetitions of the experiment. In fact, through $\ell_1\ell_2$ penalty imposition in Equation (7) we obtain a sparse, stable and small-normed vector $\boldsymbol{w}$. The non-zero components of this vector can be analyzed to extract information about the importance of similarity measures at specific frequency bands for the prediction task. The most representative similarity measures ($\geq 75\%$ occurrence) and the related frequency bands for the localization of critical areas are summarized in Table 0(a) and Table 0(b), ordered by the mean value of their coefficient.

(a) event type	frequency[Hz]
HFO (fast ripples)	251
HFO (fast ripples)	284
HFO (fast ripples)	267
HFO (fast ripples)	235
HFO (fast ripples)	221
HFO (fast ripples)	208
HFO (fast ripples)	303
HFO (ripples)	195
HFO (ripples)	183
HFO (ripples)	172
HFO (fast ripples)	366
HFO (fast ripples)	343
HFO (fast ripples)	322
HFO (ripples)	161
HFO (fast ripples)	389
HFO (ripples)	152
HFO (ripples)	134

(b) event type	frequency[Hz]
HFO (ripples)	104
high-γ	97
HFO (ripples)	111
high-γ	92
high-γ	86
HFO (ripples)	118
γ	52
high-γ	81
high-γ	76
high-γ	43
high-γ	67
high-γ	63
high-γ	71
high-γ	40
HFO (ripples)	125
γ	38
γ	35

Table 1: Most selected ($\geq 75\%$) frequency bands divided in amplitude correlation (a) and PLV (b). (a) amplitude similarity measures: most relevant frequencies for the characterization of critic areas corresponding to amplitude correlation. The high frequency bands, typical of HFO events have a strong predictive power. (b) phase similarity measures: PLV measures influence prediction at lower frequencies than amplitude correlation. There is a strong prevalence of the γ and high-γ frequency bands On the right of both (a) and (b) we report the central frequency of the wavelet filter. On the left, the corresponding event type ordered by the mean value of its coefficient. We denote the high frequencies using the same terminology for sub-cathegorization of HFOs, in ripples ($80 - 200$ Hz) and fast ripples (≥ 200 Hz). Hence, it is possible to see how high frequency events mostly contribute to the signal classification.

We can notice that the greatest components of $\boldsymbol{w}$ correspond to amplitude correlation at high frequency and phase synchrony at γ and high-γ bands. Note that the learning pipeline never selected the spectral measures (averaged power spectral density) as features of the model throughout all the repetitions of the experiment. This can be due to how this similarity measure has been performed, since non-overlapping windows cause an underestimate of this quantity.

We retain the selection of relevant similarity measures at specific frequency bands to constitute the most statistical reliable payback of the entire procedure, since it is computed across patients. Without imposition of any prior, the method confirms the relevance of high frequencies and abrupt changes in the brain activity for the localization of pathological areas in focal epilepsy, in a data-driven fashion. The selection of ripples and fast ripples frequencies is in line with the recent result of Fedele et al. (2017), which shows that the co-occurrence of the two patterns allows a more precise localization of the EZ. Also, we confirmed that the co-occurrence of γ and high-γ bands play a relevant role in the definition of EZ, as observed by van Diessen et al. (2013).

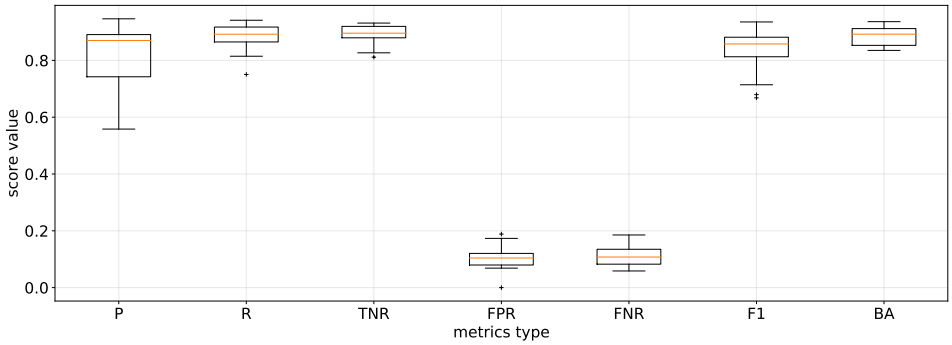


Figure 4: Average of performance scores across patients. Mean and standard deviation are computed across 50 repetitions for each patient over all channels. Several scores are used to evaluate the goodness of the method, given the unbalance of the data set.

The MT-MKL results in classification over the validation set, given the unbalance of the data set and the different proportion of critical areas across patients, are firstly evaluated on each patient. By assigning the label +1 to critical channels, and -1 to non-critical ones, we measured the performance of our model according to the following metrics scores: precision (P), recall (R), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), F1 score and balanced accuracy (BA); see Appendix B for more details. Figure 4 shows mean and standard deviation of metrics obtained across 50 repetitions of the experiment on all patients. Our results show excellent performance in the tag assignment, in particular for what concerns FPR and FNR, which are essential for a future definition of a reliable support in the clinical context. Given the double purpose of our method, we did not find a straightforward way to compare it with state-of-the-art classification methods for epileptic signal classification (Omerhodzic et al., 2013; Ghaffari and Ebrahimi Orimi, 2014).

5. Conclusions and future work

The proposed method allows the pairwise comparison between multi-scale representation of time series, in such a way to automatically select the most relevant similarity measures at specific frequency bands and to differentiate pathological activity from physiological in focal epilepsy. To the best of our knowledge, this represents a first attempt to integrate multi-scale kernel representation of neural signals for EZ localization. In order to achieve this purpose we devised an analysis pipeline which core is an extension of multiple kernel learning method, able to perform multi-task classification and kernels selection.

The methodology provides a promising starting point in the direction of data-driven definition of clinical biomarkers and of a general deeper understanding of focal epilepsy.

In the next future we plan to further inspect the achieved results in order to gain insights on the neural activity of epilepsy. In the light of our results the very first step will be the characterization of these signals at high frequencies ($\geq$ 80 Hz) in order to look for high frequency patterns which are informative for the critical/non-critical state discrimination. Given the automatic nature of the defined procedure, we have not removed interictal spiking activity and we do not exclude the presence of spikes contribution at high frequencies. In the

next future we aim to specifically characterize this activity, in order to define the role played by interictal spikes and high frequency oscillations in the discrimination of the epileptogenic areas.

Moreover we restricted our consideration to the interictal phase, but it would be interesting to consider also the preictal and ictal phases. Indeed, it is known that the synchronization level changes across these stages (Burns et al., 2014). This could result in both identification of different frequency bands and a better classification outcome. We retain that the analysis of these phases is mandatory to completely validate our work. Furthermore, we aim at looking at correlation between different brain areas in focal epilepsy through network inference methods that reduce spurious signal correlations (Tomasi et al., 2018).

All these steps represent further efforts in the direction of a personalized medicine approach to focal epileptic patients, with the double aim of explicitly understand the main features of the pathology and detect the EZ in the most efficient and precise way possible.

6. Implementation

All of the code we used in the analysis is freely available online, as an open-source Python framework, available under BSD-3-Clause at https://github.com/fdtomasi/multikernel. The implementation is largely based on high-performance libraries for numerical computations, scaling properly to an arbitrary number of patients and acquisition areas per patient.

References

Gabriele Arnulfo, Jonni Hirvonen, Lino Nobili, Satu Palva, and J Matias Palva. Phase and amplitude correlations in resting-state activity in human stereotactical eeg recordings. *Neuroimage*, 112:114–127, 2015.

Massimo Avoli, Giuseppe Biagini, and M De Curtis. Do interictal spikes sustain seizures and epileptogenesis? *Epilepsy currents*, 6(6):203–207, 2006.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146 (1):459–494, Aug 2014. ISSN 1436-4646. doi: 10.1007/s10107-013-0701-9. URL https://doi.org/10.1007/s10107-013-0701-9.

Karsten M Borgwardt. Kernel methods in bioinformatics. In *Handbook of statistical bioinformatics*, pages 317–334. Springer, 2011.

Samuel P Burns, Sabato Santaniello, Robert B Yaffe, Christophe C Jouny, Nathan E Crone, Gregory K Bergey, William S Anderson, and Sridevi V Sarma. Network dynamics of the brain and influence of the epileptic seizure onset zone. *Proceedings of the National Academy of Sciences*, 111(49):E5321–E5330, 2014.

Patrick L Combettes and Băng C Vũ. Variable metric forward–backward splitting with applications to monotone inclusions in duality. *Optimization*, 63(9):1289–1318, 2014.

Marco de Curtis and Giuliano Avanzini. Interictal spikes in focal epileptogenesis. *Progress in neurobiology*, 63(5):541–567, 2001.

Tommaso Fedele, Sergey Burnos, Ece Boran, Niklaus Krayenbühl, Peter Hilfiker, Thomas Grunwald, and Johannes Sarnthein. Resection of high frequency oscillations predicts seizure outcome in the individual patient. *Scientific Reports*, 7(1):13836, 2017.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

Ali Ghaffari and H Ebrahimi Orimi. Eeg signals classification of epileptic patients via feature selection and voting criteria in intelligent method. *Journal of medical engineering & technology*, 38(3):146–155, 2014.

Mehmet Gönen and Ethem Alpaydın. Multiple kernel learning algorithms. *Journal of machine learning research*, 12(Jul):2211–2268, 2011.

Yvonne Höller, Raoul Kutil, Lukas Klaffenböck, Aljoscha Thomschewski, Peter M Höller, Arne C Bathke, Julia Jacobs, Alexandra C Taylor, Raffaele Nardone, and Eugen Trinka. High-frequency oscillations in epilepsy and surgical outcome. a meta-analysis. *Frontiers in human neuroscience*, 9:574, 2015.

J Jacobs, R Staba, E Asano, H Otsubo, JY Wu, M Zijlmans, I Mohamed, P Kahane, F Dubeau, V Navarro, et al. High-frequency oscillations (hfos) in clinical epilepsy. *Progress in neurobiology*, 98(3):302–315, 2012.

Premysl Jiruska, Marco de Curtis, John GR Jefferys, Catherine A Schevon, Steven J Schiff, and Kaspar Schindler. Synchronization and desynchronization in epilepsy: controversies and hypotheses. *The Journal of physiology*, 591(4):787–797, 2013.

Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, and Francisco J Varela. Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208, 1999.

Gert RG Lanckriet, Tijl De Bie, Nello Cristianini, Michael I Jordan, and William Stafford Noble. A statistical framework for genomic data fusion. *Bioinformatics*, 20(16):2626–2635, 2004.

Stéphane Mallat. *A wavelet tour of signal processing*. Academic press, 1999.

Manuel R Mercier, Stephan Bickel, Pierre Megevand, David M Groppe, Charles E Schroeder, Ashesh D Mehta, and Fred A Lado. Evaluation of cortical local field potential diffusion in stereotactic electro-encephalography recordings: A glimpse on white matter signal. *Neuroimage*, 147:219–232, 2017.

Anne H Mooij, Birgit Frauscher, Mina Amiri, Willem M Otte, and Jean Gotman. Differentiating epileptic from non-epileptic high frequency intracerebral eeg signals with measures of wavelet entropy. *Clinical Neurophysiology*, 127(12):3529–3536, 2016.

Ibrahim Omerhodzic, Samir Avdakovic, Amir Nuhanovic, and Kemal Dizdarevic. Energy distribution of eeg signals: Eeg signal wavelet-neural network classifier. *arXiv preprint arXiv:1307.7897*, 2013.

Osvaldo A Rosso, Susana Blanco, Juliana Yordanova, Vasil Kolev, Alejandra Figliola, Martin Schürmann, and Erol Başar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods*, 105(1):65–75, 2001.

Miguel C Soriano, Guiomar Niso, Jillian Clements, Silvia Ortín, Sira Carrasco, María Gudín, Claudio R Mirasso, and Ernesto Pereda. Automated detection of epileptic biomarkers in resting-state interictal meg data. *Frontiers in neuroinformatics*, 11:43, 2017.

Richard J Staba, Matt Stead, and Gregory A Worrell. Electrophysiological biomarkers of epilepsy. *Neurotherapeutics*, 11(2):334–346, 2014.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

Georgi P Tolstov. *Fourier series*. Courier Corporation, 2012.

Federico Tomasi, Veronica Tozzo, Saverio Salzo, and Alessandro Verri. Latent variable time-varying network inference. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 2338–2346, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5552-0. doi: 10.1145/3219819.3220121. URL http://doi.acm.org/10.1145/3219819.3220121.

Eric van Diessen, Judith I Hanemaaijer, Willem M Otte, Rina Zelmann, Julia Jacobs, Floor E Jansen, François Dubeau, Cornelis J Stam, Jean Gotman, and Maeike Zijlmans. Are high frequency oscillations associated with altered network topology in partial epilepsy? *Neuroimage*, 82:564–573, 2013.

Manel Vila-Vidal, Alessandro Principe, Miguel Ley, Gustavo Deco, Adrià Tauste Campo, and Rodrigo Rocamora. Detection of recurrent activation patterns across focal seizures: Application to seizure onset zone identification. *Clinical Neurophysiology*, 128(6):977–985, 2017.

Shuhui Wang, Shuqiang Jiang, Qingming Huang, and Qi Tian. S3MKL: scalable semi-supervised multiple kernel learning for image data mining. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 163–172. ACM, 2010.

Ruta Yardi, Juan Bulacio, William Bingaman, Jorge Gonzalez-Martinez, Imad Najm, and Lara Jehi. Interictal spikes onintracranial eeg as a potential biomarker of epilepsy severity (p4. 071). *Neurology*, 86(16 Supplement):P4–071, 2016.

Maeike Zijlmans, Premysl Jiruska, Rina Zelmann, Frans SS Leijten, John GR Jefferys, and Jean Gotman. High-frequency oscillations as a new biomarker in epilepsy. *Annals of neurology*, 71(2):169–178, 2012.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Appendix A. Minimization method

Problem (7) is bi-convex — *i.e.*, it is convex in each variable keeping the other fixed. Its optimization is based on an alternating forward-backward splitting procedure given the non-differentiability of some parts of the functional (ℓ_1 norm) (Combettes and Vũ, 2014; Bolte et al., 2014). The optimization procedure is described in Algorithm 1.

Algorithm 1: Alternating minimization algorithm.
Initialize $\boldsymbol{\alpha}^{(1)}(0), \ldots, \boldsymbol{\alpha}^{(n)}(0), \boldsymbol{w}(0)$

for $t < t_{max}$ **do**
  for $p = 1, \ldots, n$ **do**
    $\boldsymbol{\alpha}^{(p)}(t) \leftarrow$ minimize Problem (7) with $\boldsymbol{w} = \boldsymbol{w}(t-1)$
  end
  $\boldsymbol{w}(t) \leftarrow$ minimize Problem (7) with $\boldsymbol{\alpha} = \boldsymbol{\alpha}(t)$
  if *stop criterion is met* **then**
    return $\boldsymbol{\alpha}^{(1)}(t), \ldots, \boldsymbol{\alpha}^{(n)}(t), \boldsymbol{w}(t)$
  end
end

Minimization of $\boldsymbol{\alpha}$. Fixing $\boldsymbol{w}$, for each patient p the functional w.r.t. $\boldsymbol{\alpha}^{(p)}$ takes the form of a standard logistic regression. Its minimization is then performed by computing the derivative on the logistic loss and then applying the soft-threshold operator (Tibshirani, 1996) on the result of the gradient descent step.

Minimization of $\boldsymbol{w}$. The minimization of $\boldsymbol{w}$ is more tricky, given its non-separability across various patients. Its gradient, computed on the differentiable part of the functional, is a sum of gradient computed for each patient p. Then, we apply the soft-thresholding operator to enforce sparsity in the solution. Also, we project the kernel weights into the positive half-space by applying a threshold on zero. This ensures that a kernel is considered only if its weight is positive, otherwise it is discarded.

Appendix B. Metric scores

With the aim of learning the best decision function, we denote the class of critical areas through the label $y = 1$, and the non-critical one through $y = -1$. The tag is given by medical experts for each channel and we assume it as the ground truth. We evaluate the model performance by quantifying over the validation set. The number of predicted samples which are correctly classified is defined as true negative (TN) and true positive (TP) respectively for the non-critical and critical classes. False negative (FN) quantifies the number of true critical channels which are predicted to belong to the class of non-critical. Vice versa, false positive (FP) corresponds to the number of true non-critical areas which the model predicts to be critical ones.

Given the unbalancedness of our data set, we decide to use several metric scores which are able to reliably attest our model performance.

$$\text{Precision, P} = \frac{\text{TP}}{\text{TP+FP}} \tag{8}$$

$$\text{Recall, R} = \frac{\text{TP}}{\text{TP+FN}} \tag{9}$$

$$\text{True negative rate, TNR} = \frac{\text{TN}}{\text{TN+FP}} \tag{10}$$

$$\text{False positive rate, FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{11}$$

$$\text{False negative rate, FNR} = \frac{\text{FN}}{\text{TP+FN}} \tag{12}$$

$$\text{F1 score, F1} = 2\,\frac{\text{P} \times \text{R}}{\text{P+R}} \tag{13}$$

$$\text{Balanced accuracy, BA} = \frac{1}{2}(\text{R+TNR}) \tag{14}$$

The goodness of the methodology is shown by high values of P, R, TNR, F1 and BA (the higher, the better) and low values of FNR and FPR (the lower, the better).

Appendix C. Results on single patient

We here present the results obtained for each patient. From the top left to the bottom right of each page, the plots represent:

1. **Channels importance** for discriminative task. Recalling Equation (7), we obtain, by repeating the experiment 50 times, the classification coefficients $\boldsymbol{\alpha}^{(p)}$ for each patient p. We remark that each component of $\boldsymbol{\alpha}^{(p)}$ is linked to a specific channel selected in the training phase. This vector is constrained to be sparse and small normed via the $\ell_1\ell_2$ regularization term. For each patient we show the highest values of $\boldsymbol{\alpha}^{(p)}$ across all the repetitions. We also computed the occurrence for each channel as the number of times its coefficient was non-zero on the number of times it was selected in training. In the plot, channels are ordered both in their average coefficient value and occurrence. Note that the algorithm correctly assign positive weights to the critical channels and negative weights to the non critical ones in all of the cases. We noticed that, for some patients, channels frequently used for the classification task during training, correspond to the ones that, in the validation set, are more difficult to discriminate. In our hypothesis these recordings could be more informative for critical area detection, possibly the generator of crucial activity for the discrimination.

2. **Probabilities of each channel** to be critical. The logistic function, quantifies the probability of belonging to the class of critical channels. Therefore, we provide, for each patient, the statistics on the probabilities values for the critical class when selected as validation samples. In red are shown the true critical channels; in cyan the

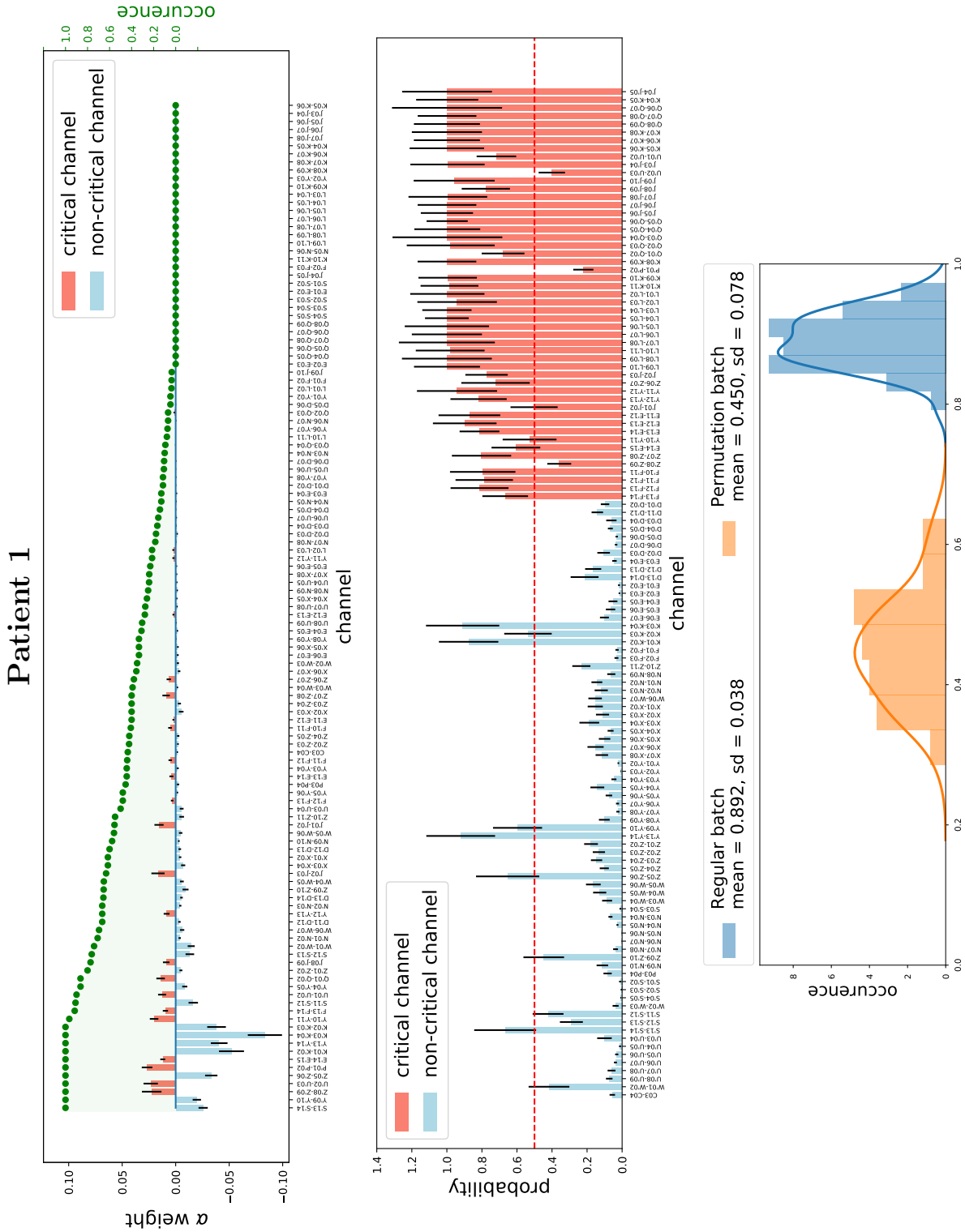
non-critical ones, as tagged by medical experts. The red dashed line, instead, corresponds to 50% probability. This allows to individuate the channels corresponding to less reliable prediction. Note that generally most of the channels are far from chance.

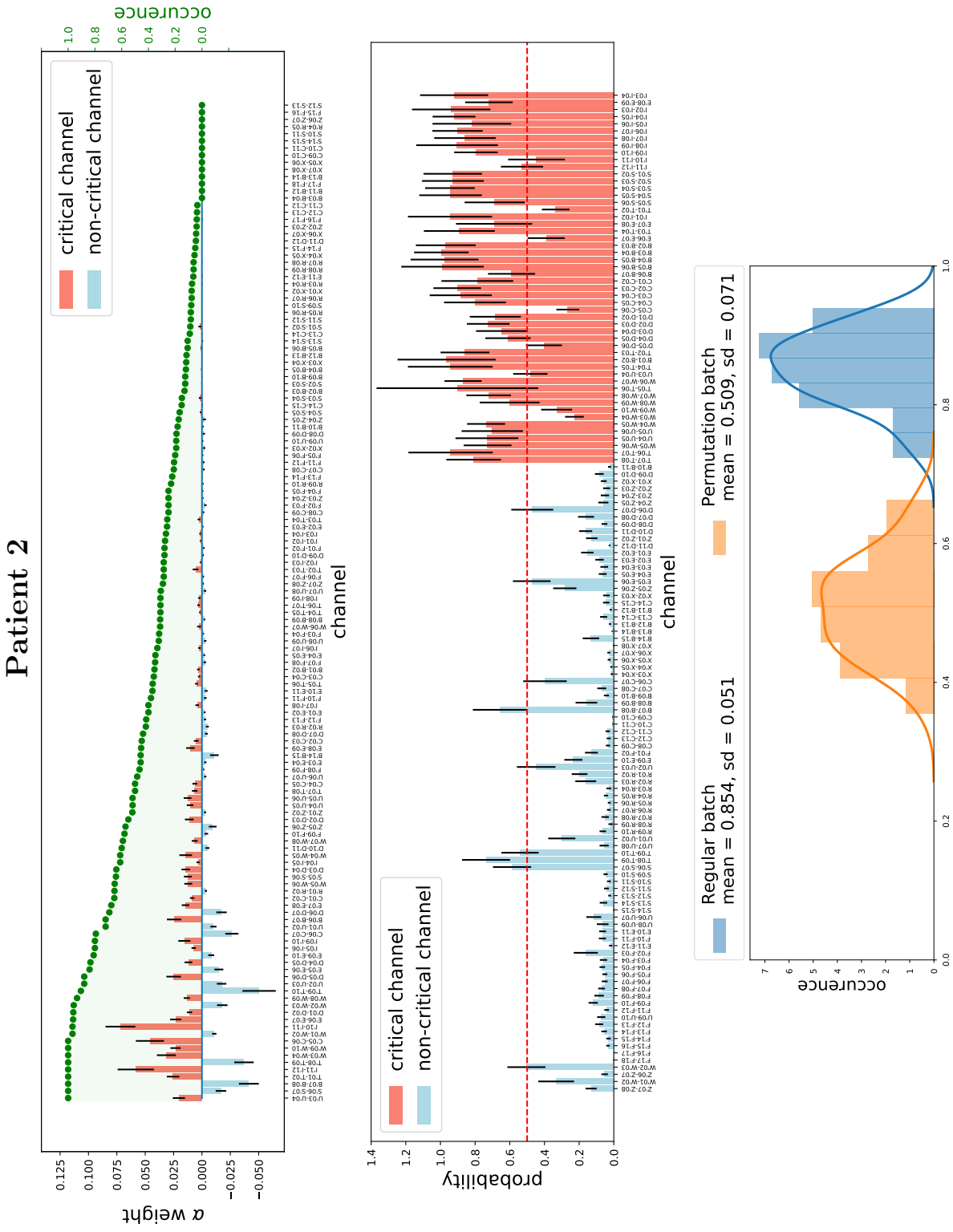
3. **Permutation test** for reproducibility. We performed a random shuffling of the labels throughout 50 repetitions of the experiment, in order to assess the reliability of our model. We use a non-parametric test (Kolmogorov-Smirnov) to reject the null hypothesis. For all the patients we rejected the null hypothesis with high significance values ($p < 10^{-5}$).

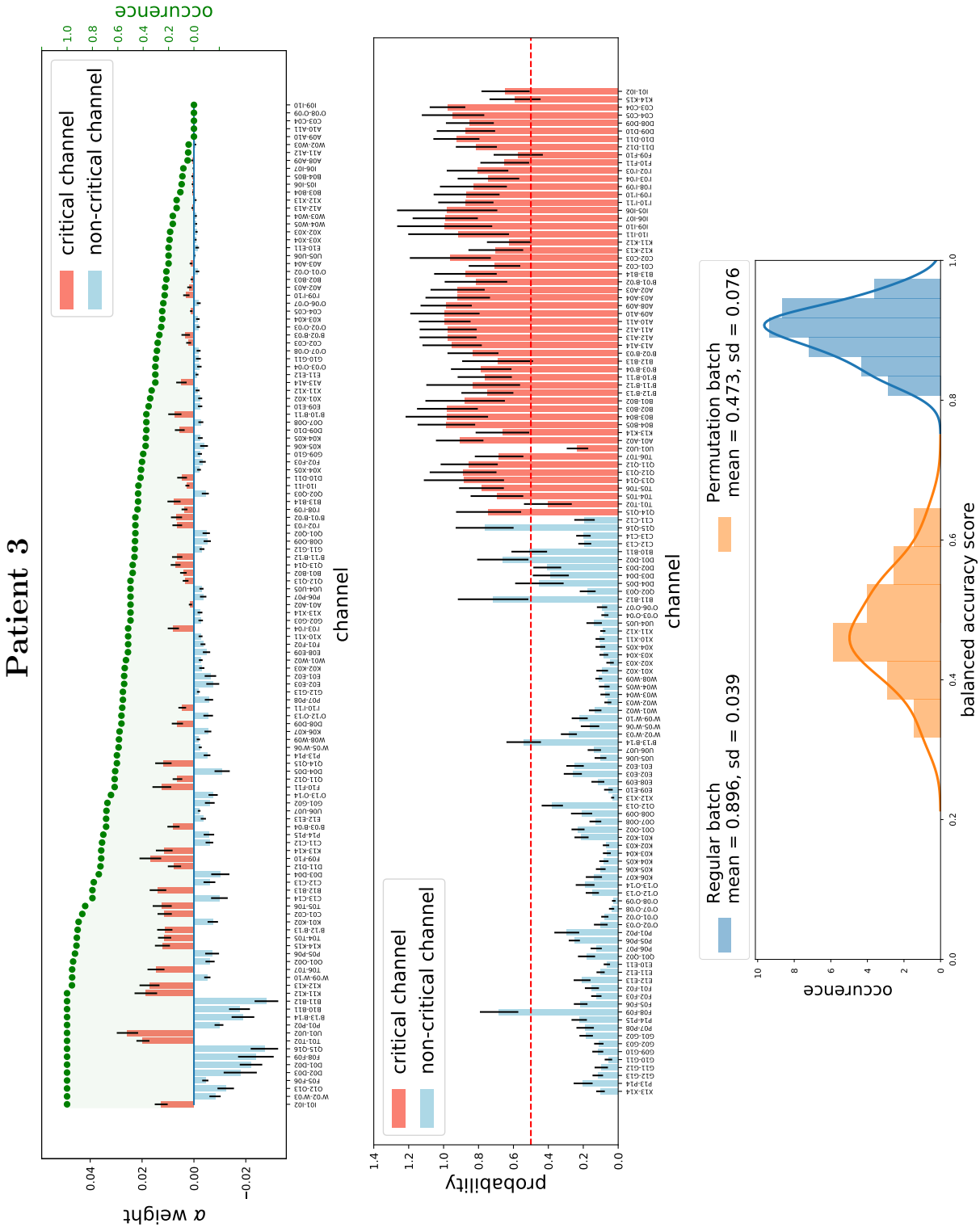
4. **Metric evaluation**. We report in Table 2 the metric scores obtained for each patient across the 50 repetitions in validation. We used several metrics which quantify the model performance taking into account classes unbalancedness. Note that all the results have a low variance, meaning that the learning pipeline is stable across all repetitions.

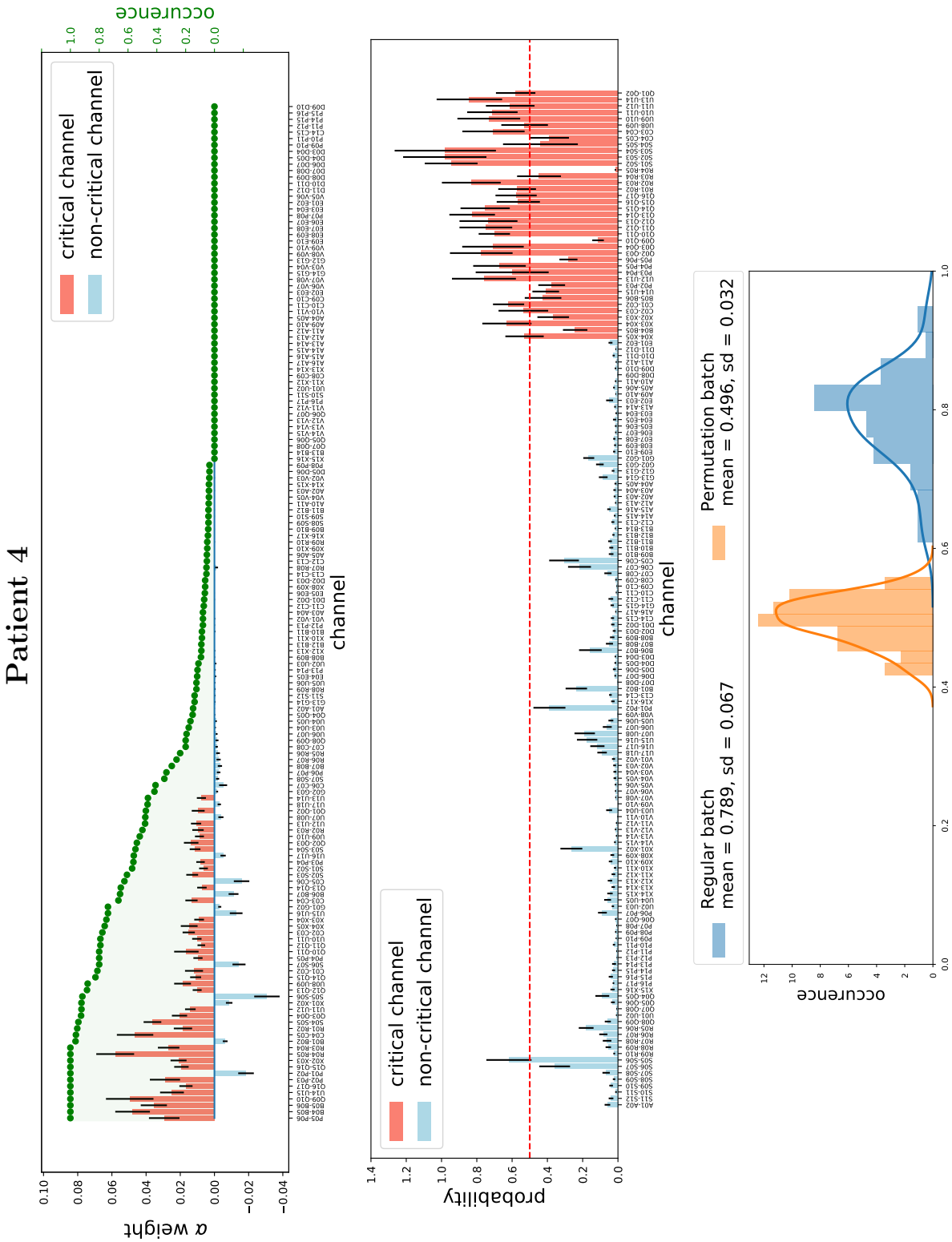
	P	R	TNR	FPR	FNR	F1 score	BA
Patient 1	0.88 ± 0.06	0.86 ± 0.06	0.92 ± 0.04	0.08 ± 0.04	0.14 ± 0.06	0.87 ± 0.04	0.89 ± 0.04
Patient 2	0.78 ± 0.11	0.87 ± 0.06	0.88 ± 0.05	0.12 ± 0.05	0.13 ± 0.06	0.82 ± 0.07	0.88 ± 0.04
Patient 3	0.87 ± 0.08	0.89 ± 0.06	0.91 ± 0.05	0.09 ± 0.05	0.11 ± 0.06	0.88 ± 0.05	0.90 ± 0.04
Patient 4	0.60 ± 0.13	0.91 ± 0.08	0.88 ± 0.03	0.12 ± 0.03	0.09 ± 0.08	0.71 ± 0.11	0.89 ± 0.05
Patient 5	0.92 ± 0.27	0.75 ± 0.22	0.92 ± 0.27	0.0 ± 0.0	0.17 ± 0.05	0.83 ± 0.24	0.84 ± 0.25
Patient 6	0.88 ± 0.06	0.89 ± 0.06	0.89 ± 0.05	0.11 ± 0.05	0.11 ± 0.06	0.88 ± 0.05	0.89 ± 0.04
Patient 7	0.73 ± 0.11	0.93 ± 0.05	0.89 ± 0.04	0.11 ± 0.04	0.07 ± 0.05	0.81 ± 0.07	0.91 ± 0.03
Patient 8	0.92 ± 0.06	0.84 ± 0.05	0.90 ± 0.06	0.10 ± 0.06	0.16 ± 0.05	0.88 ± 0.04	0.87 ± 0.04
Patient 9	0.95 ± 0.06	0.91 ± 0.06	0.92 ± 0.08	0.08 ± 0.08	0.09 ± 0.06	0.92 ± 0.03	0.91 ± 0.04
Patient 10	0.58 ± 0.14	0.85 ± 0.10	0.84 ± 0.05	0.16 ± 0.05	0.15 ± 0.10	0.68 ± 0.11	0.84 ± 0.06
Patient 11	0.89 ± 0.06	0.81 ± 0.06	0.88 ± 0.05	0.12 ± 0.05	0.19 ± 0.06	0.85 ± 0.04	0.85 ± 0.04
Patient 12	0.82 ± 0.11	0.93 ± 0.04	0.90 ± 0.06	0.10 ± 0.06	0.07 ± 0.04	0.87 ± 0.07	0.92 ± 0.04
Patient 13	0.77 ± 0.10	0.92 ± 0.06	0.86 ± 0.05	0.14 ± 0.05	0.08 ± 0.06	0.84 ± 0.06	0.89 ± 0.04
Patient 14	0.87 ± 0.08	0.90 ± 0.05	0.93 ± 0.04	0.07 ± 0.04	0.10 ± 0.05	0.88 ± 0.05	0.92 ± 0.03
Patient 15	0.93 ± 0.04	0.94 ± 0.04	0.93 ± 0.04	0.07 ± 0.04	0.06 ± 0.04	0.94 ± 0.03	0.94 ± 0.03
Patient 16	0.72 ± 0.10	0.87 ± 0.06	0.83 ± 0.05	0.17 ± 0.05	0.13 ± 0.06	0.78 ± 0.07	0.85 ± 0.04
Patient 17	0.89 ± 0.10	0.92 ± 0.05	0.92 ± 0.07	0.08 ± 0.07	0.08 ± 0.05	0.90 ± 0.05	0.92 ± 0.04
Patient 18	0.56 ± 0.14	0.87 ± 0.11	0.81 ± 0.05	0.19 ± 0.05	0.13 ± 0.11	0.67 ± 0.12	0.84 ± 0.06
Average	0.81 ± 0.12	0.88 ± 0.05	0.89 ± 0.03	0.11 ± 0.04	0.11 ± 0.04	0.83 ± 0.08	0.89 ± 0.03

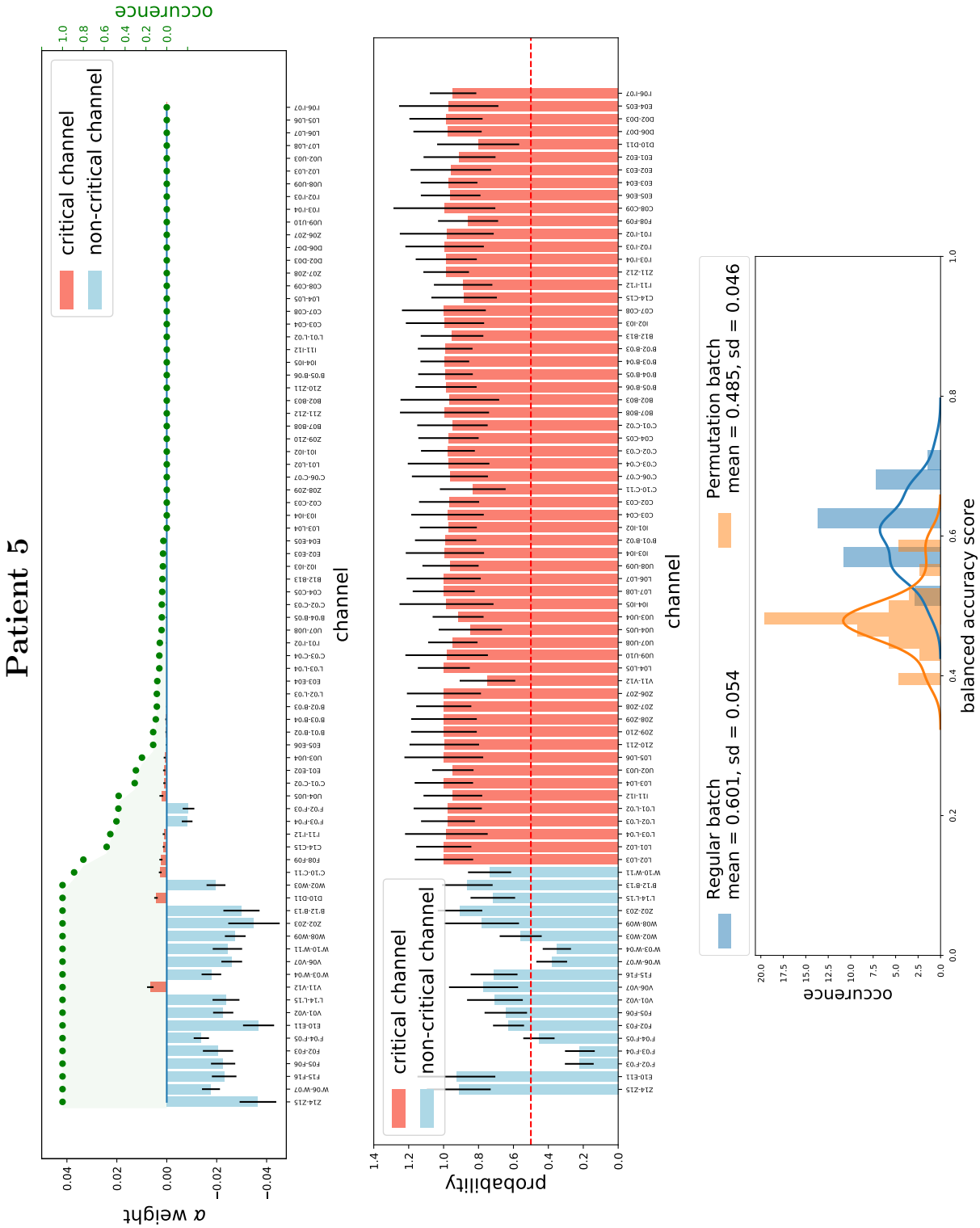
Table 2: Performance of the model obtained by repeating the learning and validation procedure 50 times for each patient. Several scores are used to evaluate the goodness of the method, given the unbalancedness of the dataset.

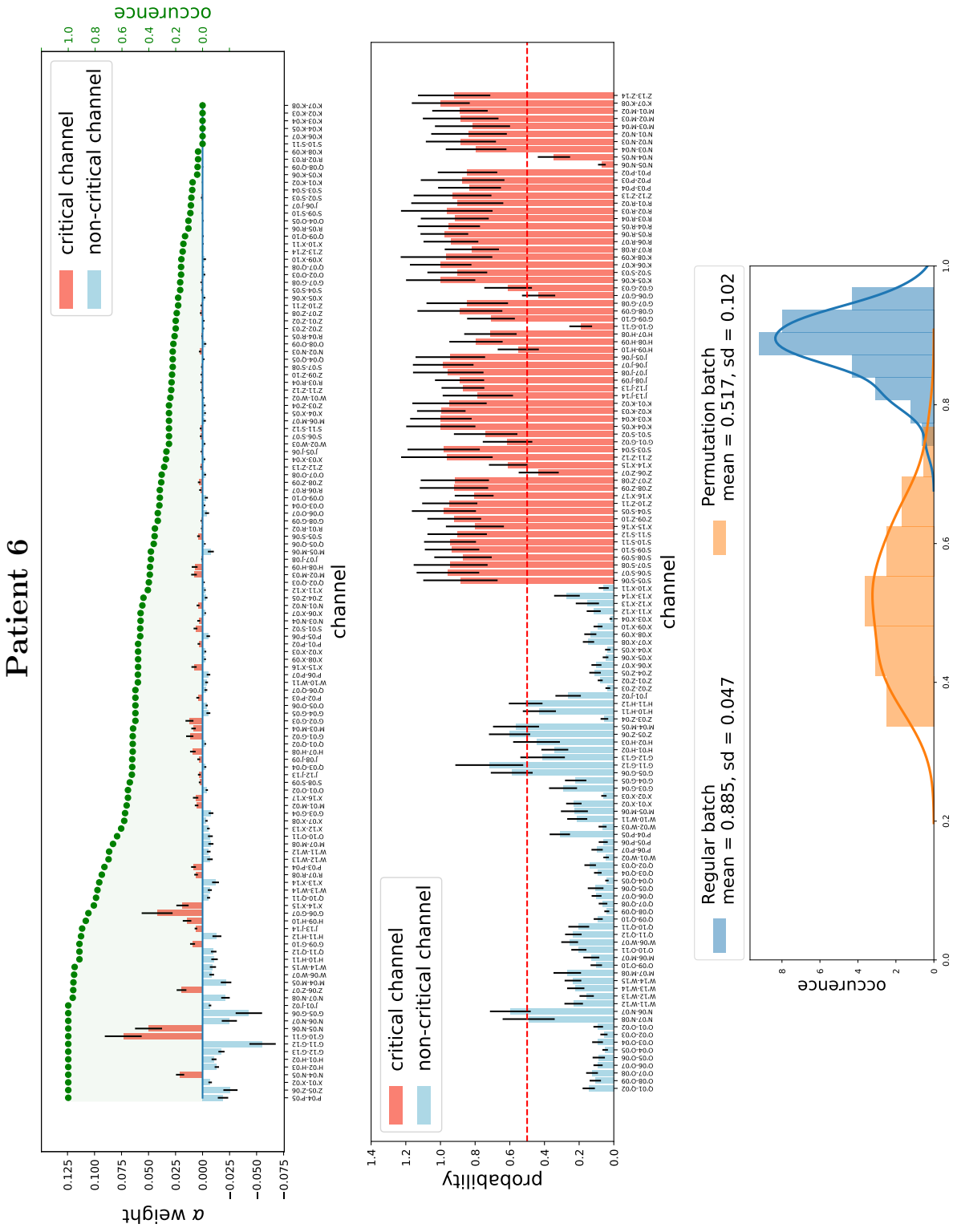

Patient 1
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
occurence
Regular batch
mean = 0.892, sd = 0.038
Permutation batch
mean = 0.450, sd = 0.078
balanced accuracy score

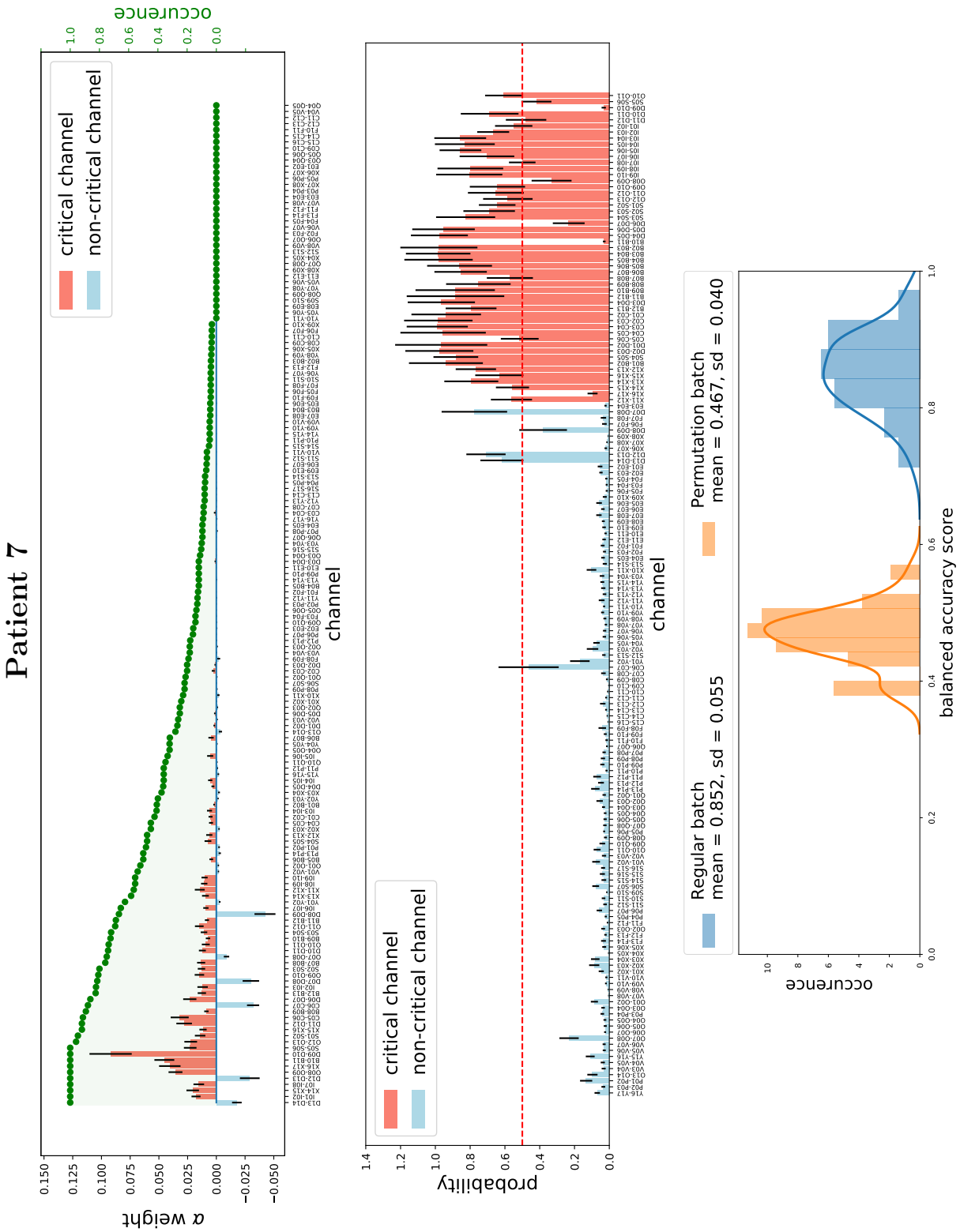

Patient 2
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
Regular batch
mean = 0.854, sd = 0.051
Permutation batch
mean = 0.509, sd = 0.071
occurence
balanced accuracy score

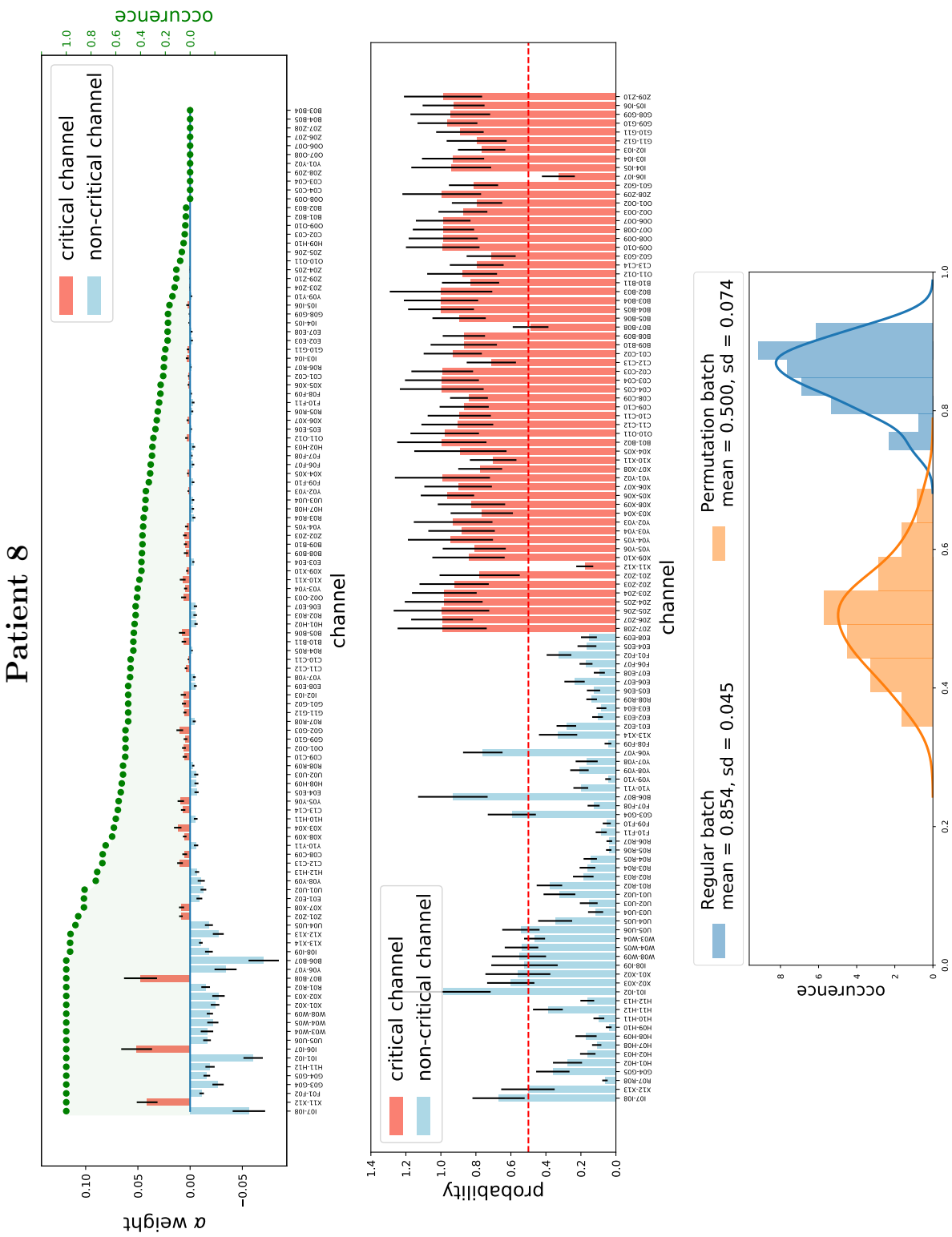

Patient 3
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
occurence
Regular batch
mean = 0.896, sd = 0.039
Permutation batch
mean = 0.473, sd = 0.076
balanced accuracy score

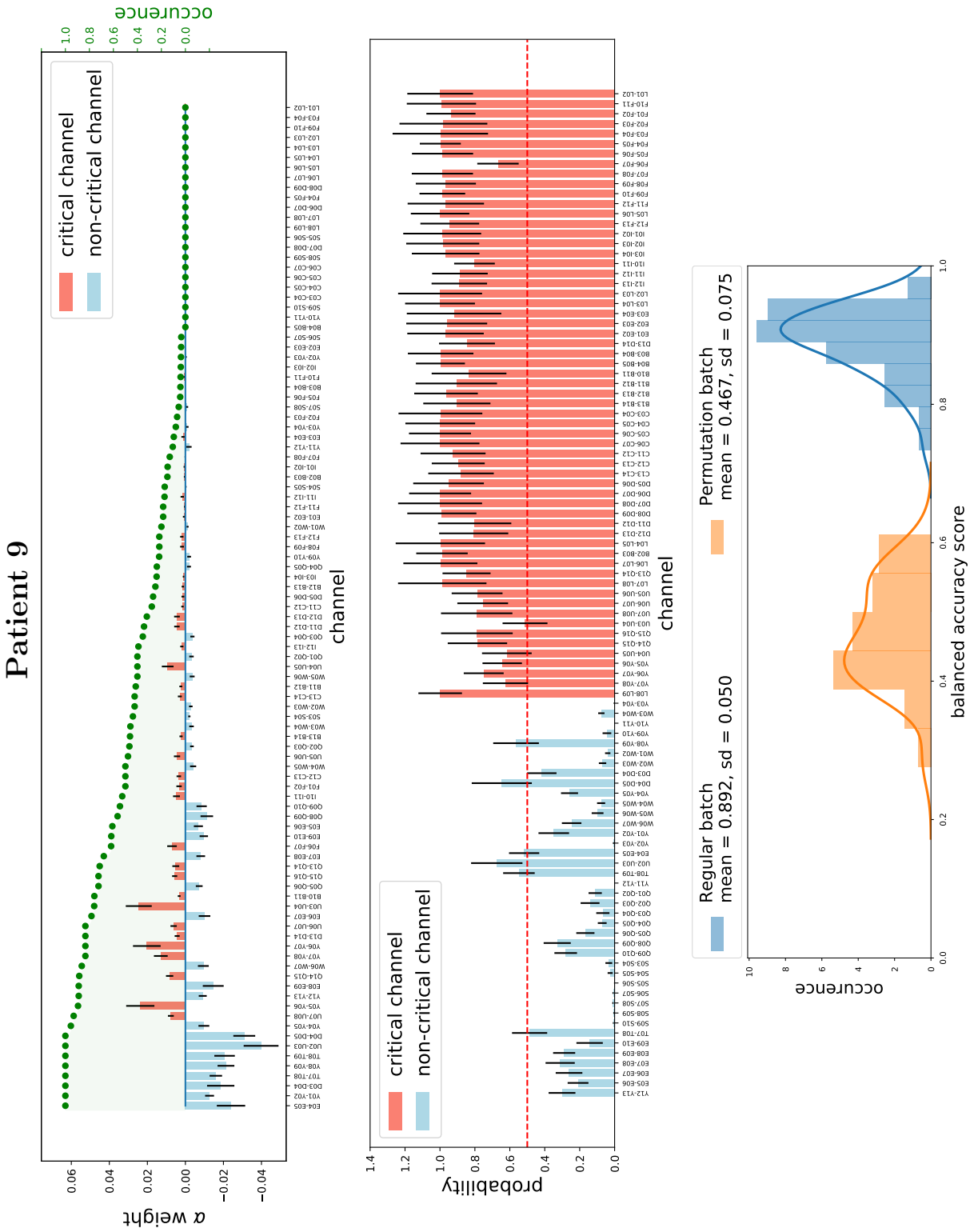

Patient 4
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
occurence
Regular batch
mean = 0.789, sd = 0.067
Permutation batch
mean = 0.496, sd = 0.032
balanced accuracy score

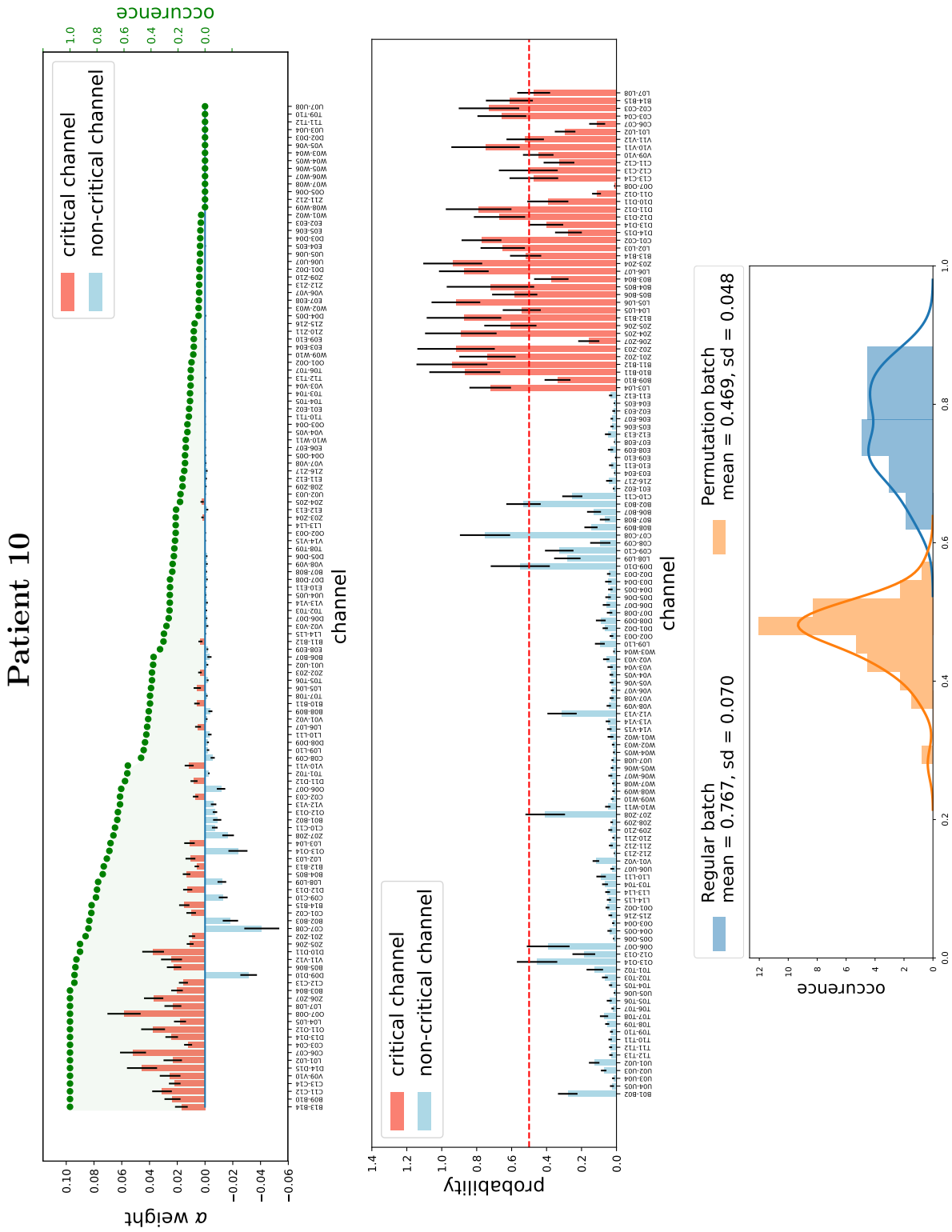

Patient 5
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.601, sd = 0.054
Permutation batch
mean = 0.485, sd = 0.046
balanced accuracy score

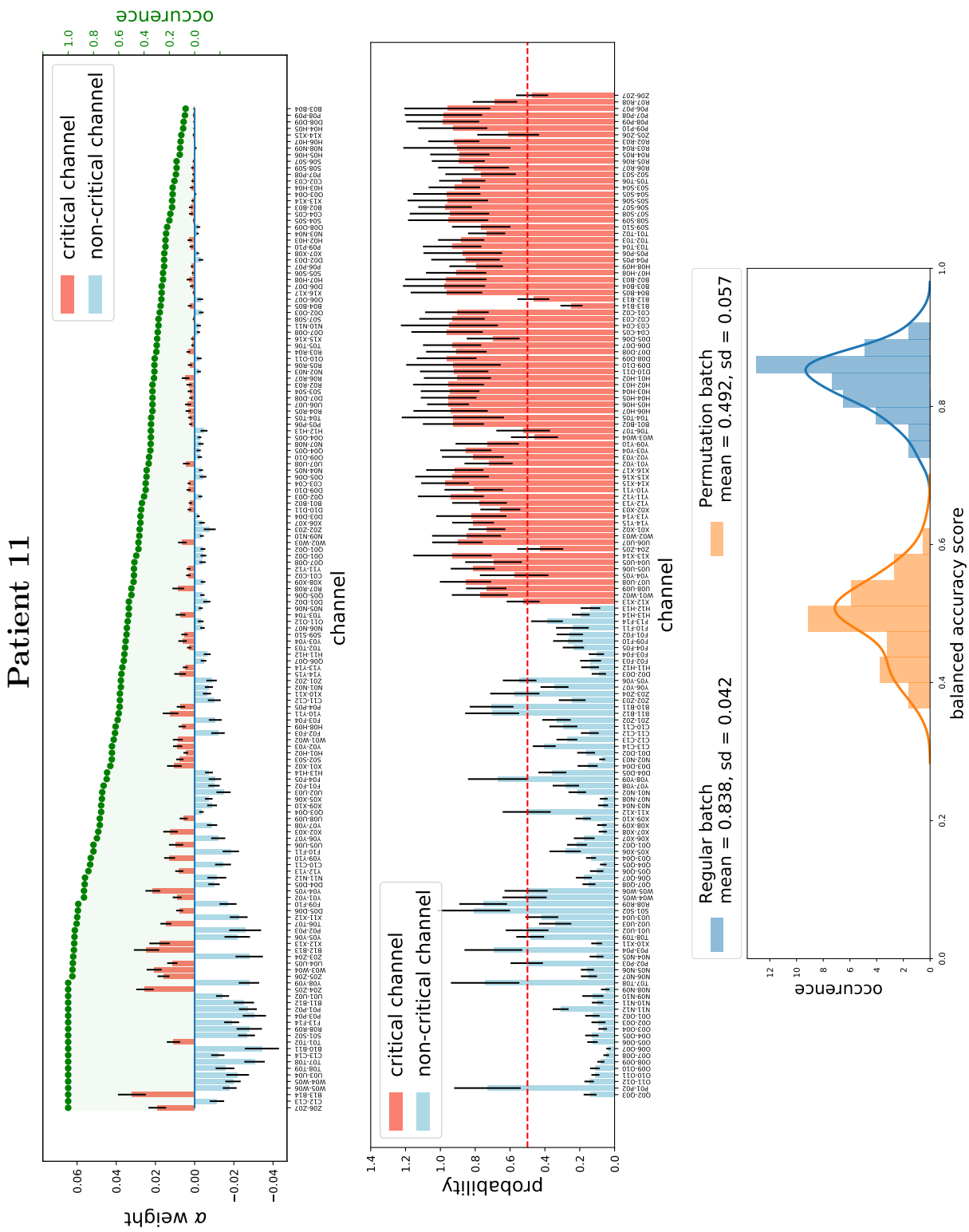

Patient 6
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.885, sd = 0.047
Permutation batch
mean = 0.517, sd = 0.102
balanced accuracy score

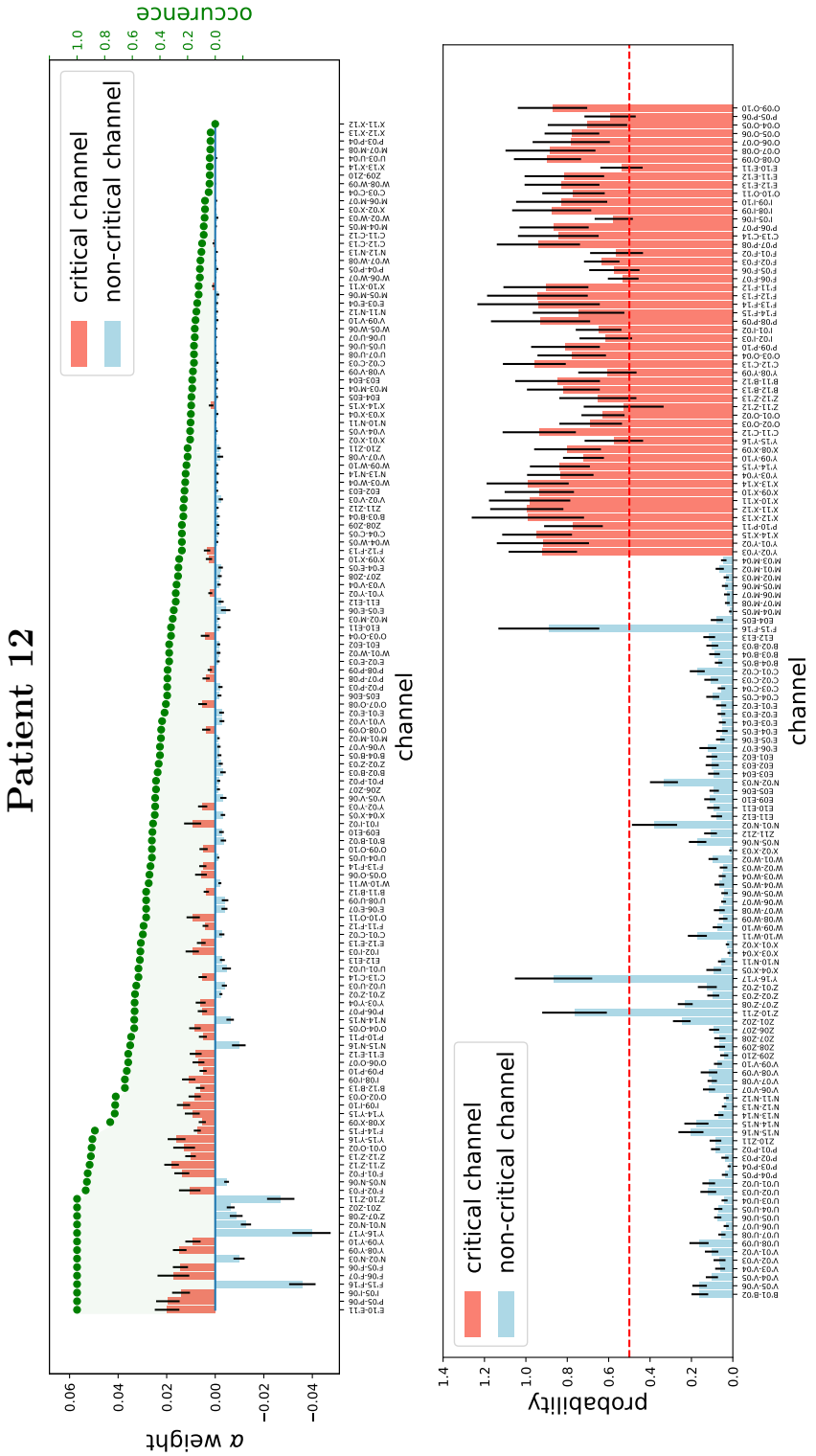

Patient 7
α weight
channel
occurrence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
Regular batch
mean = 0.852, sd = 0.055
Permutation batch
mean = 0.467, sd = 0.040
occurrence
balanced accuracy score

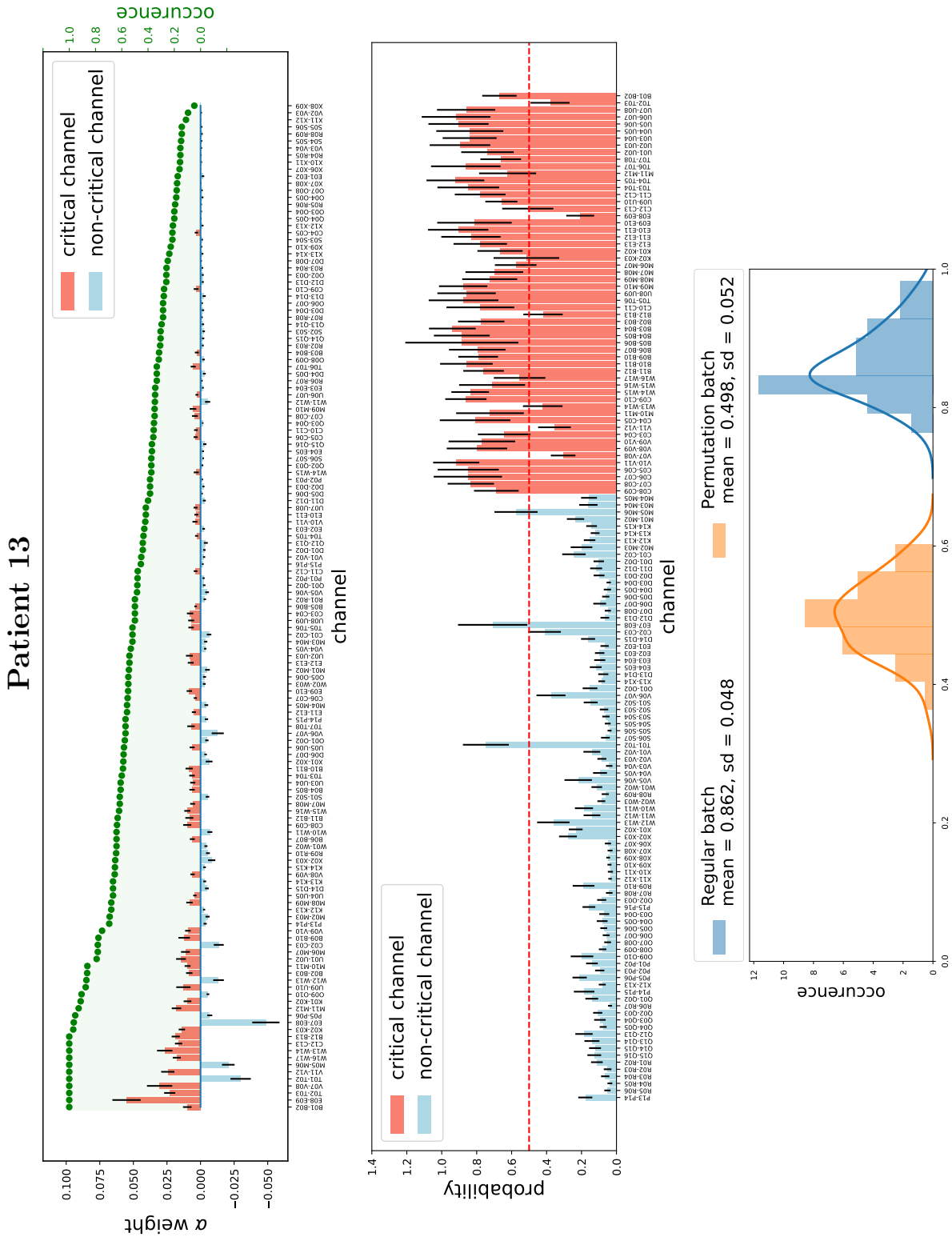

Patient 8
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.854, sd = 0.045
Permutation batch
mean = 0.500, sd = 0.074
balanced accuracy score

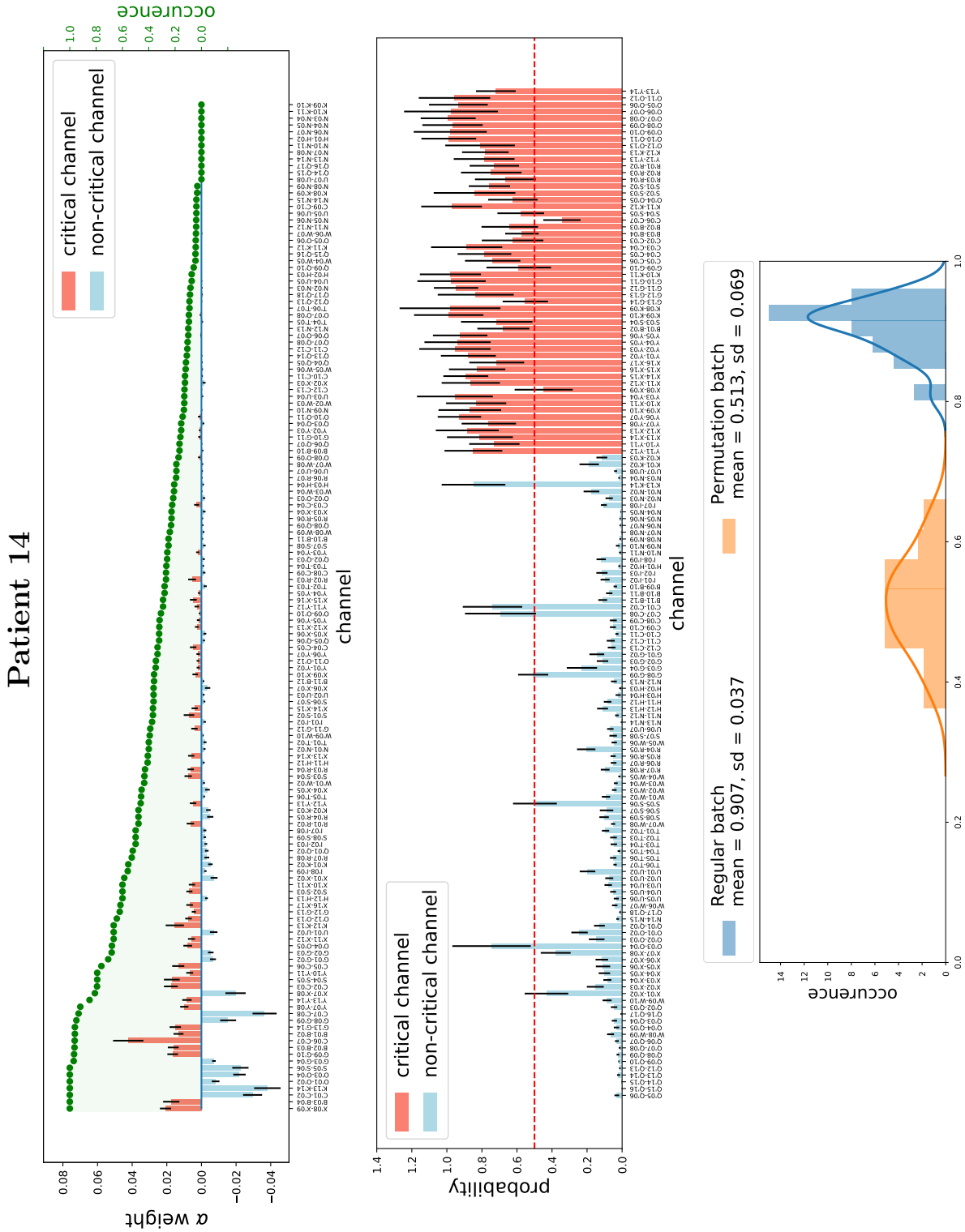

Patient 9
α weight
channel
occurrence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
Regular batch
mean = 0.892, sd = 0.050
Permutation batch
mean = 0.467, sd = 0.075
occurrence
balanced accuracy score

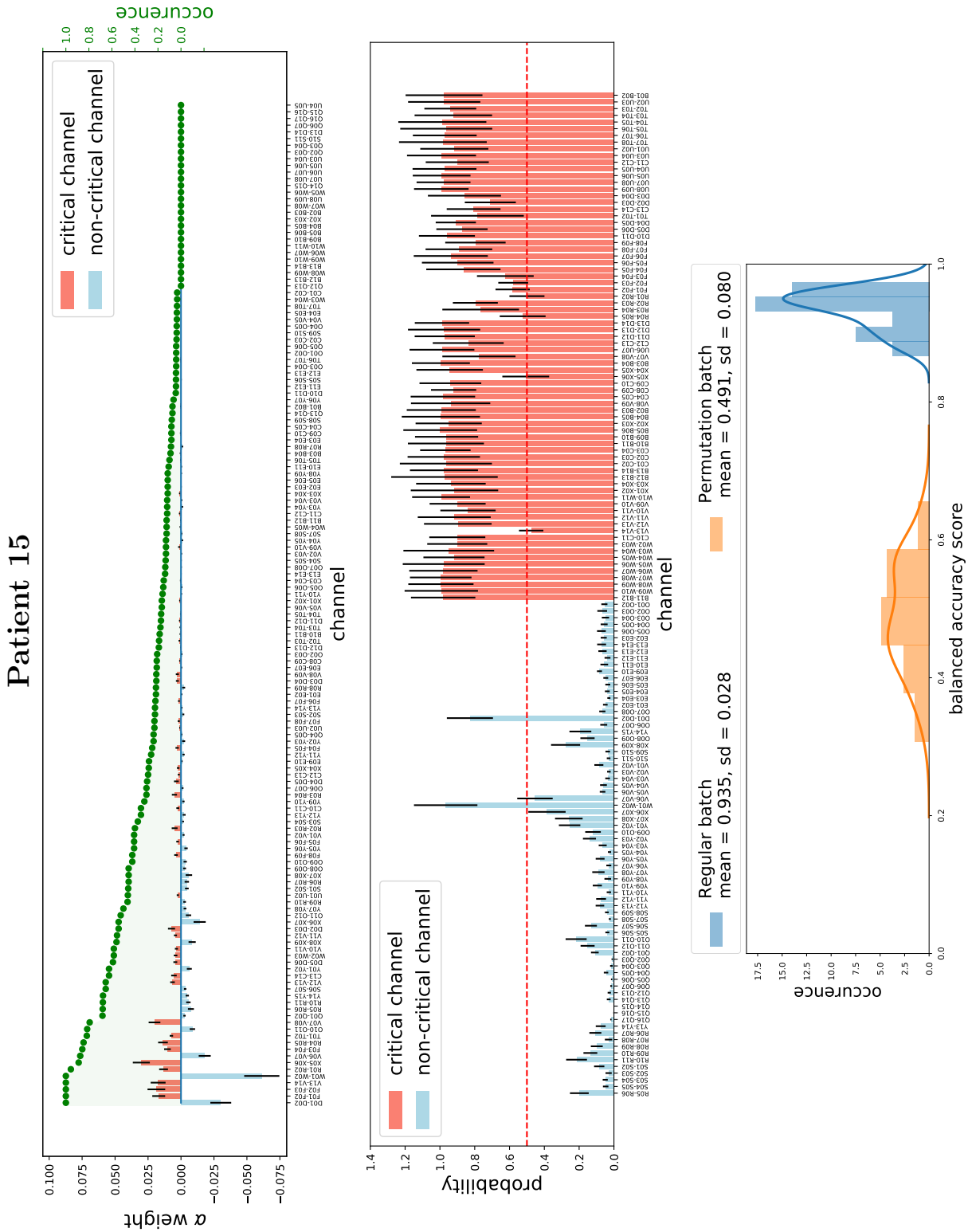

Patient 10
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.767, sd = 0.070
Permutation batch
mean = 0.469, sd = 0.048
balanced accuracy score

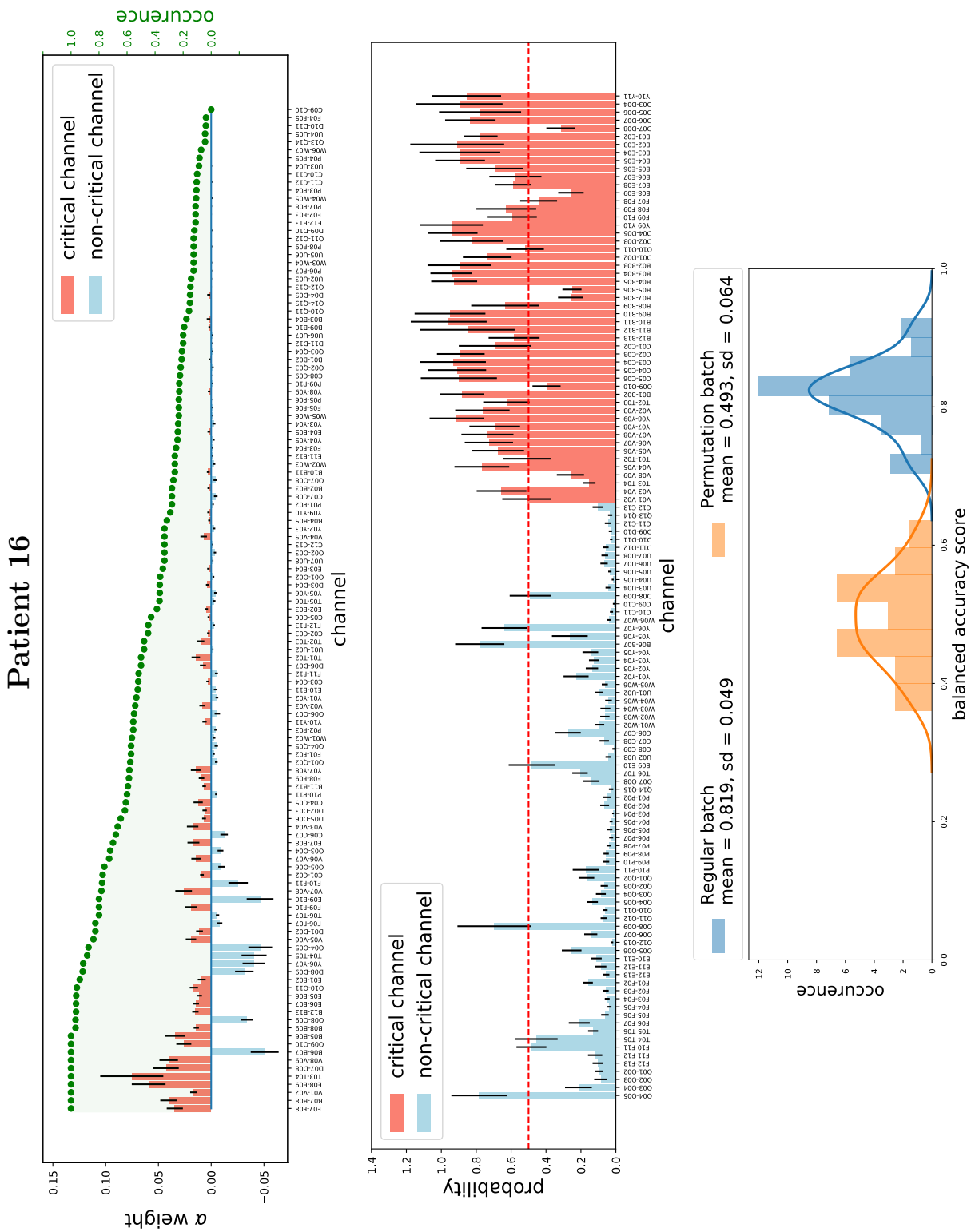

Patient 11
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.838, sd = 0.042
Permutation batch
mean = 0.492, sd = 0.057
balanced accuracy score

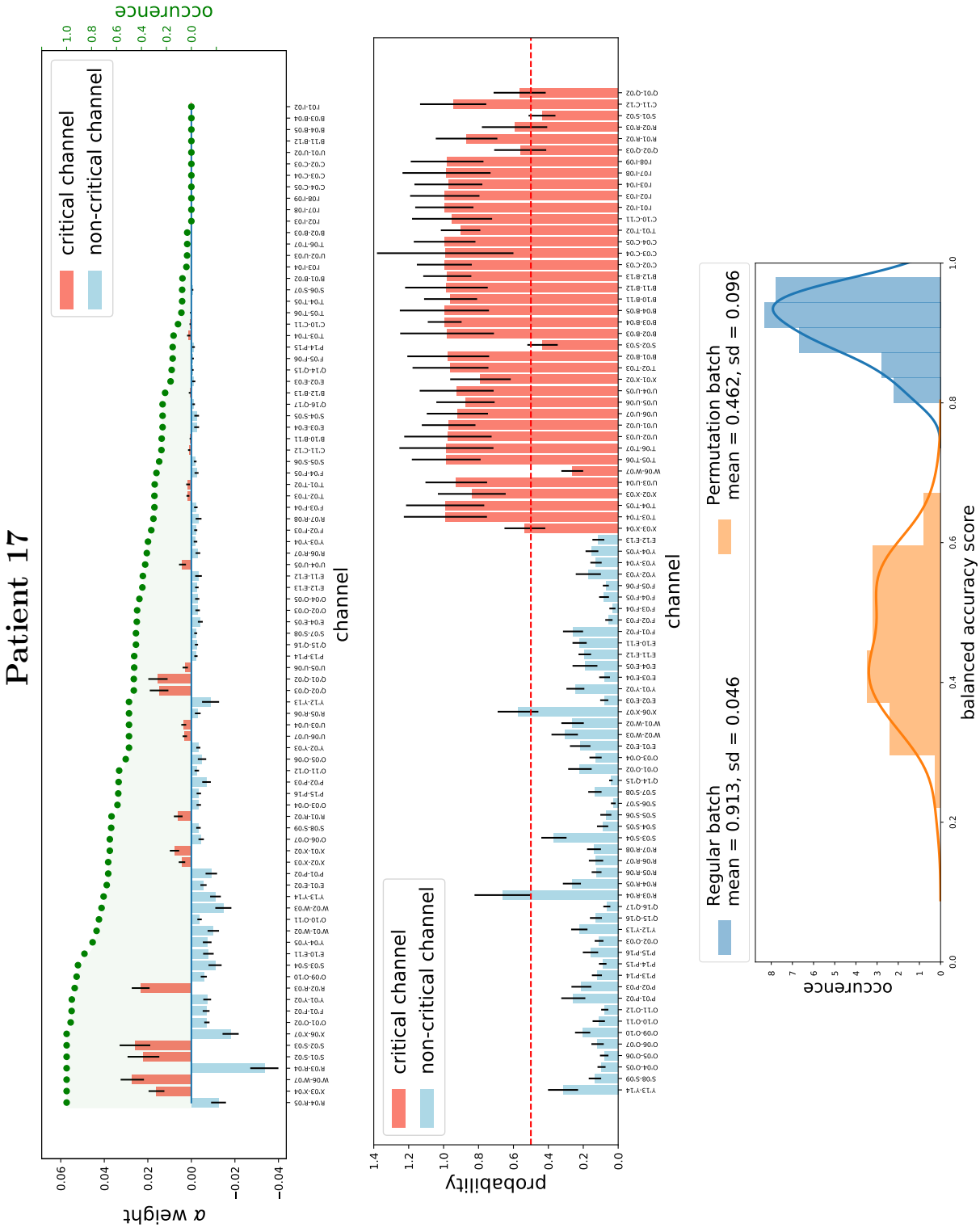

Patient 12
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
Regular batch
mean = 0.891, sd = 0.057
Permutation batch
mean = 0.488, sd = 0.074
occurence
balanced accuracy score

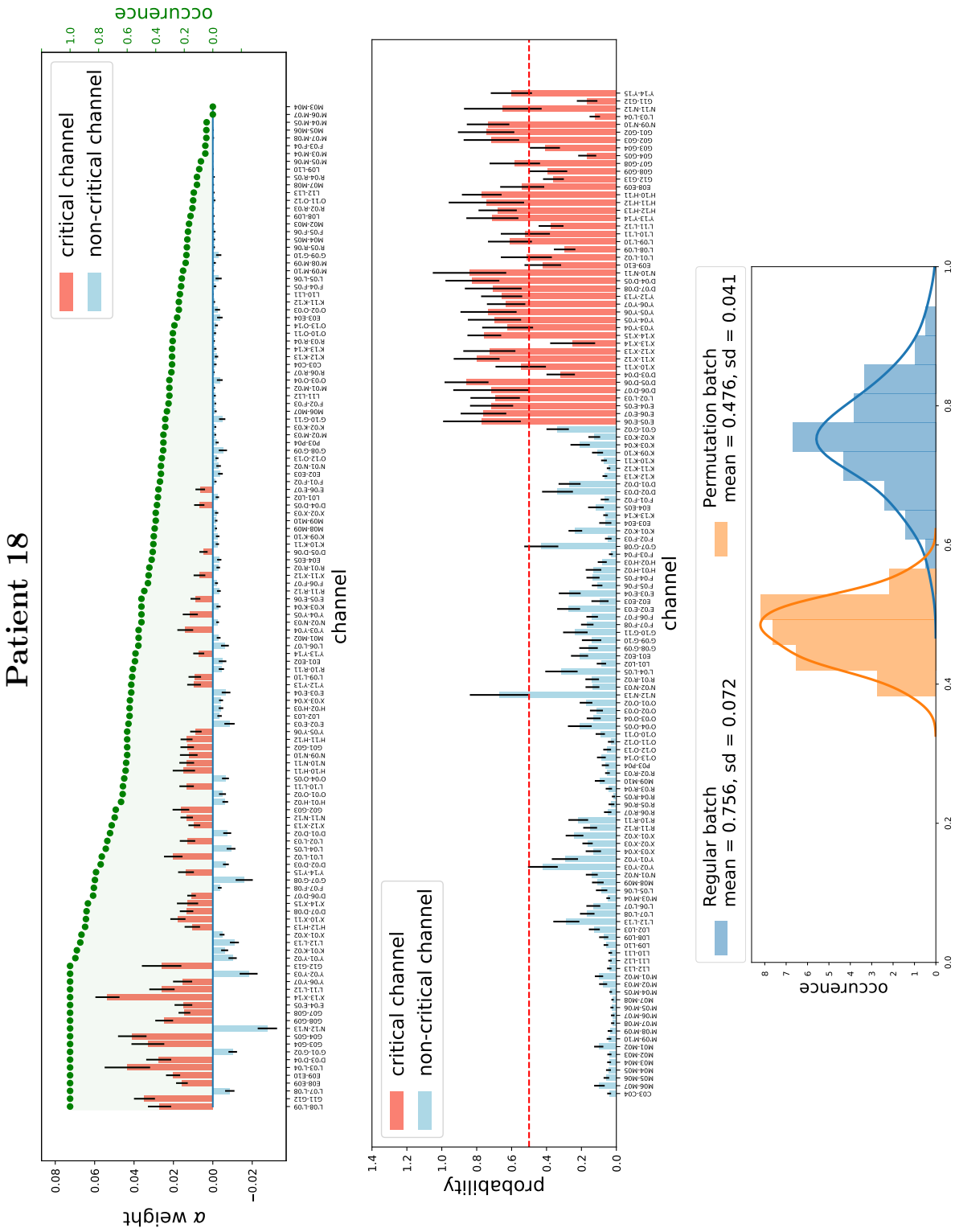

Patient 13
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.862, sd = 0.048
Permutation batch
mean = 0.498, sd = 0.052
balanced accuracy score


Patient 14
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
Regular batch
mean = 0.907, sd = 0.037
Permutation batch
mean = 0.513, sd = 0.069
occurence
balanced accuracy score


Patient 15
α weight
channel
occurrence
critical channel
non-critical channel
probability
Regular batch
mean = 0.935, sd = 0.028
Permutation batch
mean = 0.491, sd = 0.080
balanced accuracy score


Patient 16
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
Regular batch
mean = 0.819, sd = 0.049
Permutation batch
mean = 0.493, sd = 0.064
occurence
balanced accuracy score


Patient 17
α weight
channel
occurence
critical channel
non-critical channel
probability
channel
critical channel
non-critical channel
occurence
Regular batch
mean = 0.913, sd = 0.046
Permutation batch
mean = 0.462, sd = 0.096
balanced accuracy score


Patient 18
α weight
channel
occurence
critical channel
non-critical channel
probability
Regular batch
mean = 0.756, sd = 0.072
Permutation batch
mean = 0.476, sd = 0.041
balanced accuracy score