Chaos

Machine learning evaluates changes in functional connectivity under a prolonged cognitive load

Cite as: Chaos **31**, 101106 (2021); https://doi.org/10.1063/5.0070493
Submitted: 07 September 2021 • Accepted: 04 October 2021 • Published Online: 21 October 2021

Nikita Frolov, Muhammad Salman Kabir, Vladimir Maksimenko, et al.

View Online Export Citation CrossMark

Scilight
Summaries of the latest breakthroughs in the physical sciences
AIP Publishing

Machine learning evaluates changes in functional connectivity under a prolonged cognitive load

Cite as: Chaos **31**, 101106 (2021); doi: 10.1063/5.0070493
Submitted: 7 September 2021 · Accepted: 4 October 2021 ·
Published Online: 21 October 2021

View Online Export Citation CrossMark

Nikita Frolov,[1,2,a)] Muhammad Salman Kabir,[3] Vladimir Maksimenko,[1,2] and Alexander Hramov[1,2,4]

AFFILIATIONS

[1] Neuroscience and Cognitive Technology Laboratory, Center for Technologies in Robotics and Mechatronics Components, Innopolis University, 420500 Innopolis, Russia
[2] Center for Neurotechnology and Machine Learning, Immanuel Kant Baltic Federal University, 236041 Kaliningrad, Russia
[3] Department of Robotics and Computer Vision, Innopolis University, 420500 Innopolis, Russia
[4] Department of Theoretical Cybernetics, Saint Petersburg State University, 199034 St. Petersburg, Russia

[a)] **Author to whom correspondence should be addressed:** n.frolov@innopolis.ru

ABSTRACT

One must be aware of the black-box problem by applying machine learning models to analyze high-dimensional neuroimaging data. It is due to a lack of understanding of the internal algorithms or the input features upon which most models make decisions despite outstanding performance in classification, pattern recognition, and prediction. Here, we approach the fundamentally high-dimensional problem of classifying cognitive brain states based on functional connectivity by selecting and interpreting the most relevant input features. Specifically, we consider the alterations in the cortical synchrony under a prolonged cognitive load. Our study highlights the advances of this machine learning method in building a robust classification model and percept-related prestimulus connectivity changes over the conventional trial-averaged statistical analysis.

Machine learning (ML) is a state-of-the-art computational tool employed at analyzing big data in fundamental and applied science. Recently, it gained popularity in neuroscience due to its ability to recognize hidden patterns and nonlinear relations in large amounts of nonstationary and ambiguous neuroimaging data. Analysis of functional connectivity matrices is a perfect example of such a computational task assigned to machine learning. Since many ML models remain as black-boxes, interpretation of the meaningful data, based on which models make decisions, may potentially shed light on the properties of the analyzed system, i.e., the brain. Here, we evaluate alterations in functional connectivity under a prolonged cognitive load from the ML perspective. We collect the most relevant inputs using the feature engineering (FE) procedure and trial-averaged statistical analyses of functional connectivity. We establish that the features selected via FE provide higher model performance than ones obtained using trial-averaged analyses. Moreover, the interpretation of FE features possesses a less ambiguous explanation of neuronal processes underlying the changes in integrative brain dynamics.

I. INTRODUCTION

The application of machine learning (ML) in fundamental and applied science has attracted considerable attention in recent years.[1] It is due to the ability of intelligent algorithms to generalize poorly structured data and find hidden patterns in high-dimensional inputs.

Regarding the fundamental problems, nonlinear dynamics and chaos control actively implement ML tools. Recent advances in this field demonstrate the efficiency of supervised algorithms in predicting chaotic systems[2–5] and excitable media,[6,7] exploring epidemic spreading,[8] and analyzing complex networks.[9,10] One of the special applications of ML models in nonlinear science is detecting synchronization. Since synchronization is a universal fundamental phenomenon, diverse scientific areas demand advanced methods for its diagnostics.[11] Ibáñez-Soria *et al.* addressed the identification of generalized synchronization in a time-dependent fashion using an echo-state network.[12] Our recent study[13] demonstrated an approach to the same problem but in terms of a stationary feed-forward neural network. Banerjee *et al.* proposed a reservoir-computing-based method to infer causal dependencies and network links.[14]

In neuroscience, synchronization of neuronal networks is well-known to underlie normal functioning and pathological states of the brain.[15] Thus, alternation in large-scale interactions between the remote brain regions and their network properties contain relevant information about the brain's state.[16,17] One can implement this concept as real-time network brain–computer interfaces on quantifying, predicting, or classifying time-varying brain states.[18,19] It seems promising to assign costly and high-dimensional computation to machine learning. Recent studies show a notable progress in recognizing signatures of autism spectrum disorder,[20] Alzheimer's,[21] schizophrenia,[22] etc., by feeding models with functional connectivity features.

The problem behind such studies is that the functional connectivity matrices are giant arrays of data processed mainly by black-box ML models. Understanding the reasons behind predictions is, however, quite important in assessing trust, especially in those areas in which it is a question of the human condition (healthcare, medicine, neuroscience, etc.).[23,24] We argue that the interpretability of ML models, or features upon which they make decisions, is crucial in developing robust trustworthy algorithms. On the other hand, it can also contribute to a better understanding of integrative brain dynamics.

The current study contributes to unraveling the black-box problem. The main objective is to show that reducing a large set of input connectivity features appears helpful in understanding neuronal processes underlying the differences between the brain states aside from optimizing ML performance. As a specific target within this problem, we considered how a prolonged repetition of a cognitive task (or *prolonged cognitive load*) reconfigured the prestimulus functional network structure to adapt to task performance.[25] Finally, we establish advances of interpretable machine learning approach[26] over the trial-averaged statistical analysis.

II. MATERIALS AND METHODS

A. Experimental task and participants

We developed the experiment in such a way as to explore the aspects of neural processing underlying perceptual decision-making while classifying ambiguous visual stimuli. Moreover, it addresses behavioral adaptation and its neural correlates. Our previous studies reported the experimental paradigm in detail.[25,27]

Briefly, the experimental session consisted of 400 repetitions of the same visual classification task. The task required a participant to respond via button clicking as quickly as possible to a short-term presentation of an ambiguous image of the Necker cube. Depending on the contrast of inner edges, one can interpret the presented image as either a left-oriented cube (response using the left button) or a right-oriented cube (response using the right button). Moreover, the contrast of inner edges defined the ambiguity of the stimuli divided into high-ambiguity (HA) and low-ambiguity (LA) ones according to the measured response times (RTs).[25,28] We randomly picked the duration of visual stimulus presentation and the pause between trials in 1.0–1.5 s and 3.0–5.0 s, respectively. Overall, the experimental session for each participant lasted for ≈40 min.

Twenty healthy participants, comprising nine females, aged 25–35 (Mean = 26.1, *SD* = 4.6) were recruited for the experiment. Throughout the session, the EEG recorder Encephalan-EEG-19/26 measured participant's electrical cortical activity using 31 sensors placed according to the extended 10–20 layout [Fig. 1(a)]. All participants were familiar with the experimental task and did not participate in similar experiments in the last six months. The experimental studies were performed under the Declaration of Helsinki and approved by the local Research Ethics Committee of Innopolis University.

B. Epoch sampling

For each participant, we sampled 400 EEG epochs according to the experimental protocol. Each epoch contained 2 s of prestimulus electrical cortical activity. Our previous studies showed the effect of behavioral adaptation, i.e., reducing response time from the beginning to the end of the prolonged experimental session.[25,27] To trace neural correlates of behavioral adaptation, we considered the first and the last 5 min of the experimental session as its Early and Late stages, respectively. For each stage, we collected 40 EEG epochs in such a way as to include an equal number of epochs, corresponding to the presentation of left- and right-oriented stimuli, as well as HA and LA stimuli. These criteria were supposed to exclude a potential bias caused by the properties of presented visual stimuli.

C. Connectivity analysis

To evaluate prestimulus functional connectivity, we used the measure of coherence.[29] Coherence estimate between signals $x_i(t)$ and $x_j(t)$ at single trials was defined in the frequency domain $f = 4$–40 Hz with resolution $\Delta f = 0.5$ Hz as

$$\mathrm{Coh}_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_i(f)P_j(f)}. \tag{1}$$

Here, $P_i(f)$ and $P_j(f)$ are the power spectral densities of $x_i(t)$ and $x_j(t)$, respectively, and P_{ij} is a cross-spectral density of $x_i(t)$ and $x_j(t)$. Coherence is defined on the interval $[0,1]$, where $\mathrm{Coh}_{ij} = 1$ implies perfect correlation between $x_i(t)$ and $x_j(t)$, and vice versa if $\mathrm{Coh}_{ij} = 0$. $\mathrm{Coh}_{ij}(f)$ was evaluated using the `coherence` method implemented in the signal processing package `scipy.signal` for Python. The `coherence` method yielded spectral densities $P_i(f)$, $P_j(f)$, and $P_{ij}(f)$ via Fast Fourier Transform (FFT).

Within-frequency connectivity matrices W_{ij} sized 31 × 31 were filled with the values of coherence averaged over the frequency bands of interest, FOIs, such that $W_{ij} = \langle \mathrm{Coh}_{ij}(f)\rangle\big|_{\mathrm{FOI}}$ and $i,j = 1 \div 31$. For each FOI, we produced single-trial connectivity matrices for 20 participants with 40 trials in two experimental conditions [$20 \times 40 \times 2 = 1600$ in total, Fig. 1(b)].

D. Network-based statistics

Inference of statistically significant differences in the prestimulus functional connectivity between the Late and Early stages of the experimental session on the group level was performed using the following approach.

For each subject, we collected trial-averaged frequency domain representations of coherence in Early and Late conditions. For

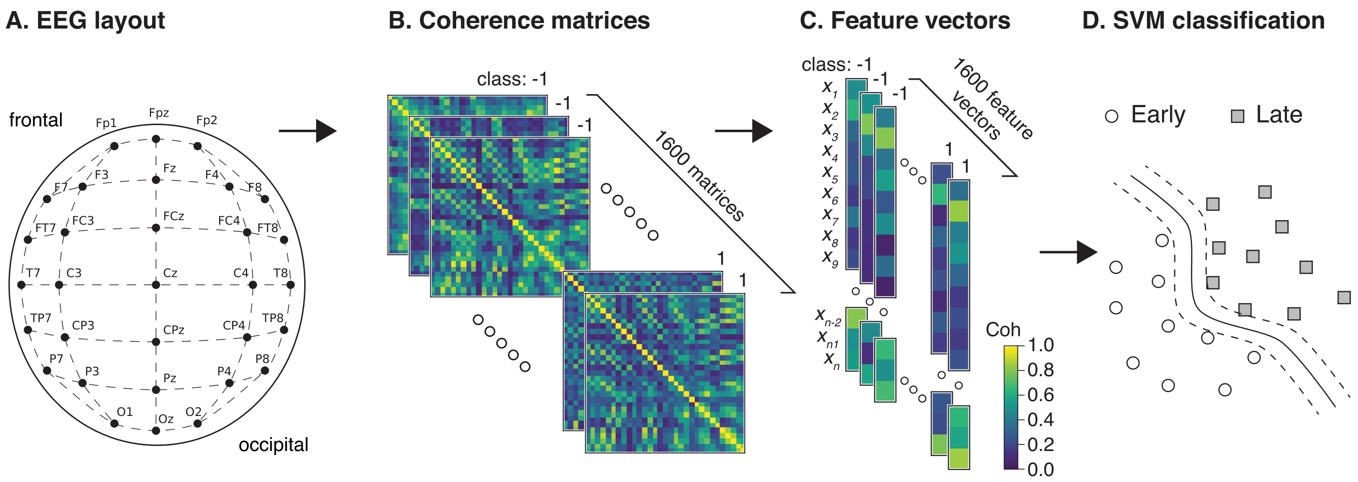

FIG. 1. Pipeline of the electroencephalography (EEG) data analysis. (a) Extended "10–20" 31–sensors EEG layout illustrating the placement of electrode sites measuring electrical cortical activity of the brain. (b) Subject-wise computation of single-trial prestimulus coherence matrices within frequency band of interest (FOI), 20 × 80 = 1600 in total. Conditions "*Early*" and "*Late*" are assigned numerical values of "−1" and "1," respectively. (c) Reducing coherence matrices to the *n*-dimensional feature vectors composed of the most informative connectivity components extracted during the feature engineering procedure. (d) Sketch of discrimination between the brain states in the "*Early*," circles, and "*Late*," squares, conditions using a nonlinear support vector machine (SVM) classifier. Dashed lines indicate $(n-1)$-dimensional boundaries of the considered classes, and the solid line represents a $(n-1)$-dimensional dividing hyperplane.

each triplet (i, j, f), we produced the value of subject-wise t-statistics between conditions Late and Early. Then, for each frequency component f, we computed a number of links (i, j), whose value of t-statistic exceeded a predefined α-level separately for the left- and right-tailed distributions. We considered three α-levels: (i) $\alpha = 0.05$, $|t_{(19)}| \geq 1.729$; (ii) $\alpha = 0.025$, $|t_{(19)}| \geq 2.093$; (iii) $\alpha = 0.01$, $|t_{(19)}| \geq 2.539$. Thresholding the obtained dependencies of the number of links vs frequency, we extracted the frequency bands of interest, FOIs, demonstrating a significant effect.

Finally, the significance of group-level connectivity differences between Late and Early conditions was evaluated using the network-based statistics (NBS) approach.[30]

E. Machine learning

We used ML to infer changes in functional connectivity between experimental conditions. After collecting connectivity matrices within FOIs, we applied feature engineering (FE) to sort connectivity features in descending order of their relevance in discrimination between the experimental conditions. Taking n top features as inputs, $n = 1, \ldots, N$, where N is a maximal number of features, we tested the performance of a nonlinear classifier. Thus, we generated dependency of classification performance vs the number of top features. Then, we selected the optimal number of features providing sufficient accuracy for a small number of inputs. The obtained reduced set of features was further considered a sought connectivity structure, whose coupling change was the most informative for classification. A short flow chart of the entire simulation is shown in Fig. 1, and a detailed description of FE and nonlinear classification is given below.

Feature engineering. FE is one of the core concepts in ML, which greatly impacts the performance of the developed model. Until now, connectivity matrices computed for each FOI contain $N = 31 \times (31 - 1)/2 = 465$ unique functional connections between electrode sites, or features. In generally, it is undesirable to use all features since irrelevant or less relevant features can negatively impact model performance and greatly increase computational cost. In order to select key features, filter-type feature selection algorithm was employed[31] [Fig. 1(c)]. The filter-type feature selection algorithm measures the importance of inputs based on their characteristics, such as feature variance and feature relevance to the response. We extracted important features as part of a data preprocessing step, and then the model is trained using the selected features. We used a chi-squared test as a feature selection criterion. A higher chi-squared test score means that this feature has higher discriminative power than a feature with a lower chi-squared test score. A small feature set—top 20 features (F20) for each FOI—was considered. Top 20 features are an optimal number of features providing a sufficient performance of SVM classifier (see the supplementary material for details).

Nonlinear classifier. We employed a support vector machine (SVM) with radial basis function kernel as a classifier [Fig. 1(d)]. The classification was performed on a single-trial level, i.e., taking into account connectivity features of all 1600 EEG epochs. Classes Early and Late were assigned numerical values of −1 and 1, respectively. We used a k-fold cross-validation to train and test the selected SVM classifier.[32] This validation scheme is suitable when the data set size is not very large. In a k-fold cross-validation scheme, the data set was randomly permuted and split up into $k = 20$ groups or folds (80 samples per group). Then, for each k, the kth group was taken as a test data set, while the remaining $k - 1$ groups comprised the training data set. We trained the model on the training set and evaluated using the test set, resulting in $k - 1$ pieces of the model training. Before training, data were shuffled once so that folds in each case remain the same. To assess the classification performance, we computed the k-fold cross-validation score as

$$\text{score} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{2}$$

Here, TP is true positive, which means the model correctly assigned class "1" to the Late connectivity. TN is true negative, which means the model correctly assigned class "-1" to the Early connectivity. FP is false positive, which means the model erroneously assigned class "1" to the Early connectivity. Finally, FN is false negative, which means the model erroneously assigned class "-1" to the Late connectivity.

ML simulations were fully performed using **sklearn** package for Python.

III. RESULTS AND DISCUSSION

A. Group-level network-based statistic

Group-level statistical analyses using NBS revealed two functional connectivity structures exhibiting significant changes in coherence throughout the experimental session. Figure 3(a) reports the negative effect, i.e., a functional network demonstrating decreased coherence from the Early to Late stages of the experimental session. We observed a negative effect within the frequency band 4–6.5 Hz at $\alpha = 0.01$ [Fig. 2(a), left panel]. Corresponding functional network, $p = 0.034$ via NBS, is presented in the right panel of Fig. 2(a). It shows a reduced theta-band coherence between bilateral frontal and frontocentral electrodes in the Late stage compared to the Early one. There is evidence that frontal theta oscillations reflect realization of cognitive control.[33–35] Several previous studies associate an increased frontal theta activation and phase coupling with

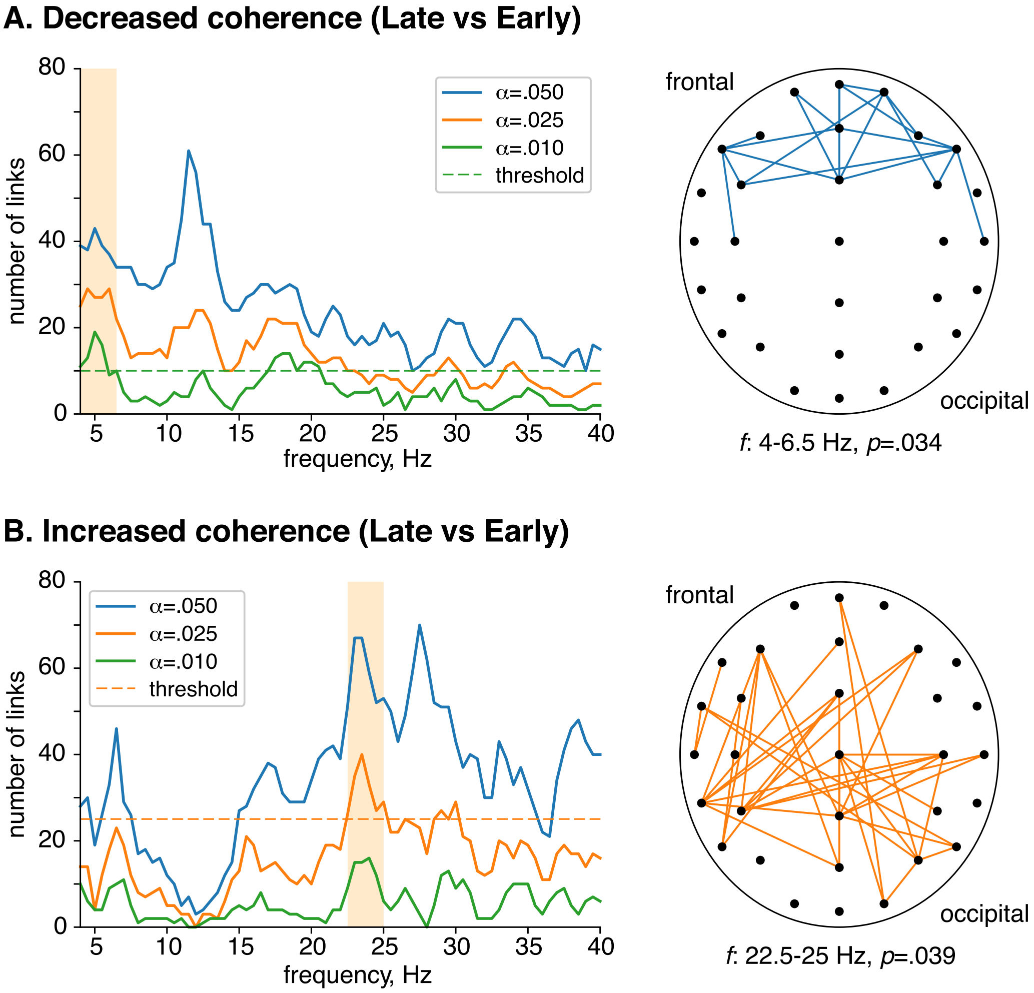

FIG. 2. Connectivity analysis using network-based statistics. (a) Functional network demonstrating a significant *negative* effect, i.e., a decreasing coherence from Early to Late stage. (b) Functional network demonstrating a significant *positive* effect, i.e., an increasing coherence from Early to Late stage. In both subplots, the right panel shows the number of links exceeding a predefined α-level vs frequency, and the left panel illustrates corresponding functional networks demonstrating a significant effect.

the process of reinforcement learning and exploration.[36,37] Cavanagh *et al.* emphasize that frontal theta oscillations, a hallmark of encoding prediction errors, are closely related to behavioral adaptation. Moreover, Cohen and Cavanagh have shown that the level of frontal theta correlates with task performance in single trials.[38] Summarizing the above, we can interpret our result as a meaning that at the beginning of the experimental session, naive participants explore unfamiliar and ambiguous visual stimuli to extract relevant features and maximize their performance by reducing response time and error rate.

A positive effect, i.e., enhancement of coherence throughout the session, was found in the narrow part of the beta-band 22.5–25 Hz at $\alpha = 0.025$ [Fig. 2(b), left panel]. Corresponding functional network, $p = 0.039$ via NBS, is reported in the right panel of Fig. 2(b). It involves enhanced brain-wide beta-band coherence (Late vs Early), including right-lateralized occipital, parietal, and temporal electrodes, midline sensors, as well as the left-lateralized temporal electrode sites. Derivation of the exact implication of prestimulus beta-synchrony from the uncovered connectivity structure is difficult due to the poor spatial localization of the latter. Among the variety of functional roles of beta oscillations in human brain, we suggest protection of "neuronal equilibrium" states[39,40] and "clear-out" of the working memory content carried during trial.[40–42] The former implies the filtering of distractions[43] and maintenance of working memory context preventing ongoing encoding.[44] In this context, we could interpret our results as a superposition of these neural mechanisms in the prestimulus state. Enhanced beta-band interaction between the electrodes covering the prefrontal and sensorimotor cortices may, on the one hand, reflect clearing out a short-term representation of the previous visual stimulus and related button clicking, which should not interfere with the perception of the upcoming stimulus. On the other hand, it may be a signature of maintaining a developed association between the presented stimuli and corresponding motor reactions in long-term memory.

To compare these results with an interpretable machine learning approach, we collected the uncovered functional networks and corresponding frequency bands as a set of features for ML entitled "NBS."

B. Interpretable machine learning

Next, we employ interpretable machine learning to discriminate between the considered brain states and highlight the most informative functional connectivity features. To avoid frequency variability in single trials, we considered connectivity features calculated in broad Theta (4–8 Hz, 465 features) and Beta (15–30 Hz, 465 features) bands along with merged feature vector Theta + Beta ($2 \times 465 = 930$ features). Based on the results of the NBS analysis, we also considered narrow-band Theta (4–6.5 Hz, 465 features) and narrow-band Beta (22.5–25 Hz, 465 features) feature vectors, along with the merged narrow-band Theta + Beta feature vector ($2 \times 465 = 930$ features).

We composed four different configurations of extracted features:

- **All:** feature vector is composed of all 465 features in the case of Theta and Beta bands and of all 930 features in the case of merged Theta + Beta connectivity.
- **F20:** feature vector is composed of top 20 features in the case of Theta, Beta, and merged Theta + Beta connectivity.
- **nAll:** feature vector is composed of all 465 features in the case of narrow Theta and narrow Beta bands and of all 930 features in the case of merged narrow Theta + Beta connectivity.
- **nF20:** feature vector is composed of top 20 features in the case of narrow-band Theta, Beta, and merged Theta + Beta connectivity.

Figure 3 displays top 20 connectivity features extracted via chi-squared test for broad-band FOIs (F20) and narrow-band FOIs (F20n). Connectivity features in Fig. 3 are color-coded with the value of ΔCoh, which is a grand-average difference between coherence in the Late and Early stages of the session,

$$\Delta\mathrm{Coh}_{ij} = \langle \mathrm{Coh}_{ij}^{Late} - \mathrm{Coh}_{ij}^{Early} \rangle, \tag{3}$$

where the operator $\langle \bullet \rangle$ defines averaging across subjects.

In broad-band analysis, the F20-Theta set includes bilateral connections between frontal sensors and occipital, parietal, and temporal electrodes. F20-Beta set includes local left-lateralized coupling between occipital, parietal, temporal, and frontal electrode sites. Extracted features, aside from beta-band O1-Fz and O1-Fp1 connections, are characterized by positive $\Delta\mathrm{Coh}_{ij}$. In the F20-(Theta + Beta) set, 19 of 20 features belong to broad-band beta connectivity, and only one belongs to theta.

In narrow-band analysis, the nF20-Theta feature set contains several left-lateralized connections between frontal, temporal, and parietal sensors, similar to the F20-Theta set. Moreover, nF20-Theta includes several frontal links having negative $\Delta\mathrm{Coh}_{ij}$, which coincides with previous findings obtained via the NBS approach. The nF20-Beta set contains fewer strengthening links in the left hemisphere and more weakening large-scale connections between sensor O1 and frontal sensors (Fz, F3, F9, Fpz, and Fp1). All features in the nF20-(Theta + Beta) set belong to narrow-band beta connectivity.

Although functional connections extracted by the feature selection algorithm differ from NBS results, they share some common structures. First of all, it applies to elevated left-lateralized frontoparietal and temporal beta connectivity. Previously, Hipp *et al.* analyzed the perception of ambiguous visual stimuli and reported similar prestimulus networks.[45] They emphasized that the fluctuations of large-scale beta synchrony over these areas could reflect visual attention changes that modulate the stimulus's perceptual organization. We may conclude that the beta-band structures highlighted in Fig. 3 play an essential role in the prestimulus interaction of neuronal processing streams bouncing perception. At the same, theta-band connectivity features were not as informative as beta-band ones, as follows from the bottom panels in Fig. 3. Based on this observation, we may conclude that either theta-band features or associated neuronal processes do not significantly change the brain state throughout the experiment or exhibit strong inter-trial variability.

Intending to identify which set of features and frequency band, or combination of frequency bands, was the most relevant in single-trial classification, we used them as inputs for nonlinear SVM. We verified its performance via *k*-fold cross-validation. For each (feature set, frequency band) pair, we collected a sample of *k*-fold cross-validation scores, 20 scores per sample, indicating the performance of the nonlinear SVM classifier, see Methods. Group means

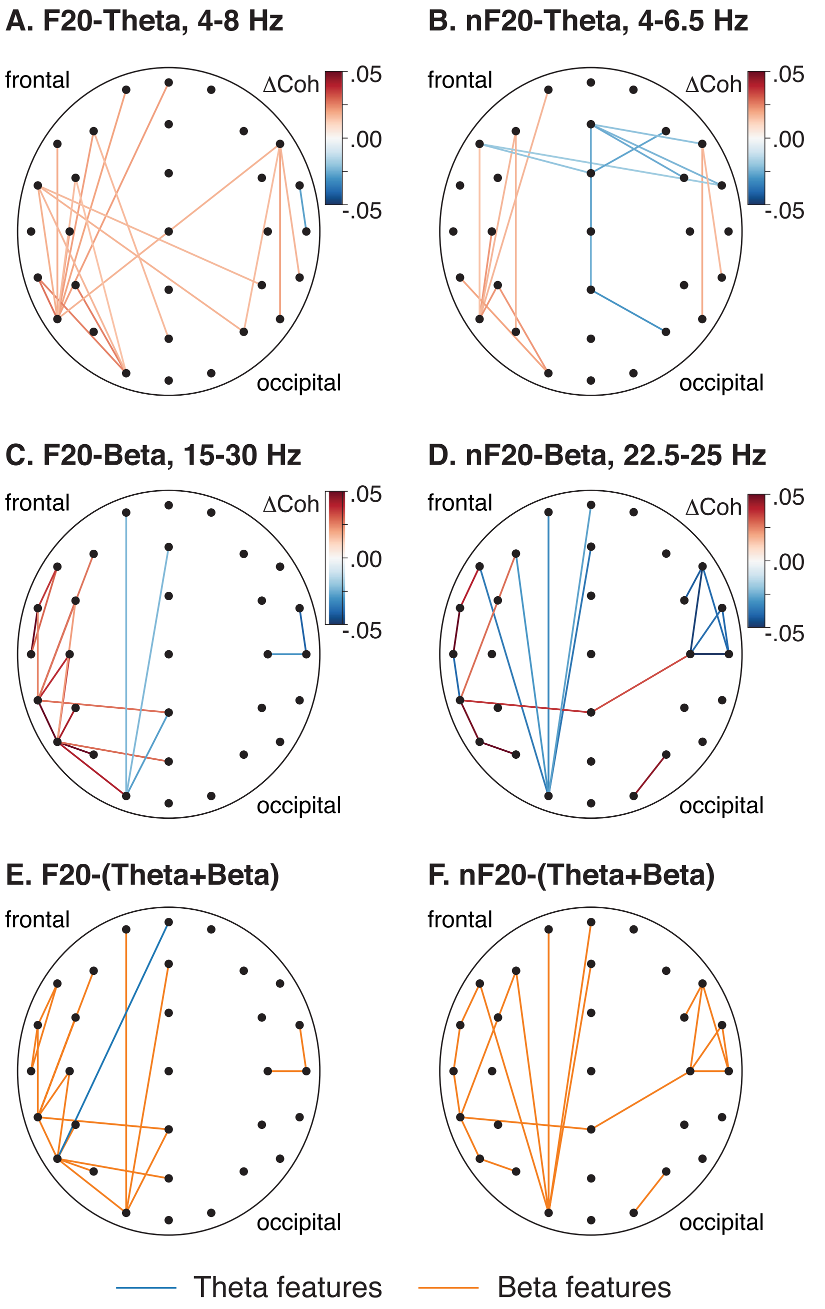

FIG. 3. Features selected using the chi-squared/filter method. Each subplot displays top 20 features in different frequency ranges: broad-band (left column) and narrow-band identified by NBS (right column). (a) Broad-band F20-Theta, 4–8 Hz. (b) Narrow-band nF20-Theta, 4–6.5 Hz. (c) Broad-band F20-Beta, 15–30 Hz. (d) Narrow-band nF20-Beta, 22.5–25 Hz. Connections are color-coded with the value of △Coh, a group-mean difference of coherence between Late and Early conditions. (e) and (f) show top 20 connectivity features in merged Theta (blue) and Beta (orange) sets.

and standard deviations are presented in Fig. 4. We compared these samples using repeated measures (RM) ANOVA with two within-subject factors: (i) frequency band and (ii) feature set. Primarily, RM ANOVA indicated that SVM performance is significantly influenced by a frequency band, within which the connectivity is computed ($F_{1.241,19} = 544.372$, $p < 0.001$, $\eta^2 = .59$). It turned out that theta-band connectivity features were the least informative with an insufficient classification accuracy of 50%–60%, on average. This observation supports our previous conclusion on the contribution of theta-band connectivity to the development of the integrated brain state. Importantly, our comparison shows that broad-band connectivity features provide approximately 10% higher performance than narrow-band ones. We suppose that this is due to the well-known inter-subject variability of oscillatory rhythms of electrical cortical activity.[46]

Post hoc analysis indicated that the feature vector All-Beta was the most informative, providing a mean classification accuracy of 89.63% ($t_{19} > 4.812$, $p < 0.001$ via paired t-test). Despite that,

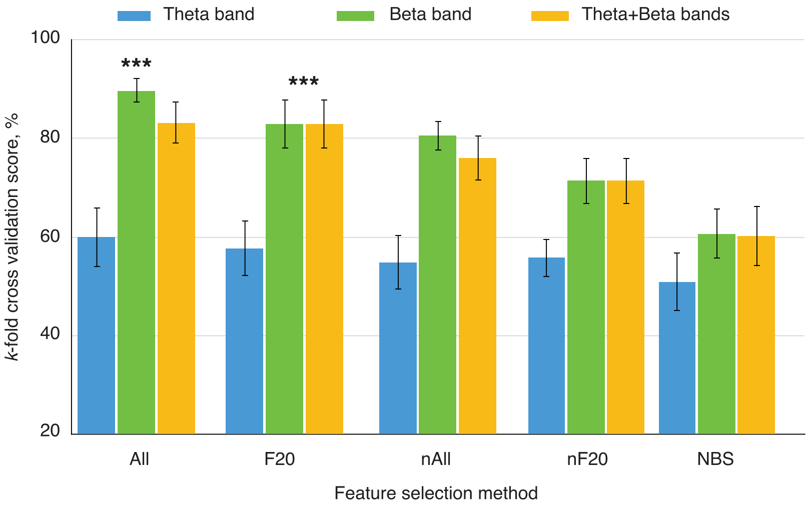

FIG. 4. *k*-fold cross-validation scores comparison. Here, bars and whiskers show means and standard deviations over the folds; "***" indicates the level of significance $p < 0.001$ via repeated measures analysis of variance (RM ANOVA).

F20-Beta and F20-(Theta + Beta) are of special notation. Although supporting less classification accuracy than All-Beta, these sets demonstrated equal suitable performance of 82.94% at lower computational costs. F20-Beta and F20-(Theta + Beta) also significantly outperformed F20-Theta as well as nAll-, nF20, and NBS-pairs, for which the mean cross-validation score did not exceed 80% ($t_{19} > 4.767$, $p < 0.001$ via paired t-test). Noteworthy that NBS features extracted based on the trial-averaged analysis provided the lowest classification performance, which was about 50%–60%, on average. Thus, we suppose that trial-averaged statistical analysis is inappropriate in selecting connectivity features in single-trial classification instead of the considered interpretable machine learning.

IV. CONCLUSIONS

We demonstrated the applicability of interpretable machine learning in evaluating changes in functional connectivity. By employing an FE algorithm, we extracted sets of informative features of functional connectivity. Their relevance was verified using a nonlinear SVM classifier. We tested this approach against the trial-averaged group-level statistical analysis—the test task aimed at finding changes in prestimulus functional connectivity during a prolonged cognitive load.

Our results emphasize the importance of selecting and interpreting inputs for building high-performance ML models. First, we observed that functional connectivity features extracted via FE provided more than 20% higher classification accuracy in single trials compared to features selected using trial-averaged statistical analysis. Second, we found that FE captured functional networks of better spatial localization and rejected several irrelevant features obtained by trial-averaged analysis. The latter helped us to evaluate robust changes in functional connectivity. It also allowed us to draw clear conclusions about the modulation of the prestimulus visual attention network influencing the perception of an ongoing ambiguous stimulus.

We conclude that relying on the results of the trial-averaged analysis may lead to erroneous single-trial classification due to inter-trial variability of the measured variables. We expect that our findings could potentially bring a new perspective on the use of interpreted machine learning in neuroscience and medicine.[23,24] It primarily applies to the development of brain–computer interfaces, where the application of AI tools is extremely demanded.[18,19]

SUPPLEMENTARY MATERIAL

See supplementary material for the detailed report on feature selection.

ACKNOWLEDGMENTS

This study was supported by the Russian Science Foundation (Grant No. 19-12-00050).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the Figshare repository at https://doi.org/10.6084/m9.figshare.12155343.v2, Ref. 47.

REFERENCES

[1] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science **349**, 255–260 (2015).
[2] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, "Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data," Chaos **27**, 121102 (2017).
[3] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," Phys. Rev. Lett. **120**, 024102 (2018).
[4] Z. Lu, B. R. Hunt, and E. Ott, "Attractor reconstruction by machine learning," Chaos **28**, 061104 (2018).
[5] A. Wikner, J. Pathak, B. Hunt, M. Girvan, T. Arcomano, I. Szunyogh, A. Pomerance, and E. Ott, "Combining machine learning with knowledge-based

modeling for scalable forecasting and subgrid-scale closure of large, complex, spatiotemporal systems," Chaos **30**, 053111 (2020).

[6] R. S. Zimmermann and U. Parlitz, "Observing spatio-temporal dynamics of excitable media using reservoir computing," Chaos **28**, 043118 (2018).

[7] S. Saha, A. Mishra, S. Ghosh, S. K. Dana, and C. Hens, "Predicting bursting in a complete graph of mixed population through reservoir computing," Phys. Rev. Res. **2**, 033338 (2020).

[8] S. Ghosh, A. Senapati, A. Mishra, J. Chattopadhyay, S. K. Dana, C. Hens, and D. Ghosh, "Reservoir computing on epidemic spreading: A case study on COVID-19 cases," Phys. Rev. E **104**, 014308 (2021).

[9] A. Panday, W. S. Lee, S. Dutta, and S. Jalan, "Machine learning assisted network classification from symbolic time-series," Chaos **31**, 031106 (2021).

[10] N. Kushwaha, N. K. Mendola, S. Ghosh, A. D. Kachhvah, and S. Jalan, "Machine learning assisted chimera and solitary states in networks," Front. Phys. **9**, 147 (2021).

[11] A. Pikovsky, J. Kurths, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, 2003), p. 12.

[12] D. Ibáñez-Soria, J. García-Ojalvo, A. Soria-Frisch, and G. Ruffini, "Detection of generalized synchronization using echo state networks," Chaos **28**, 033118 (2018).

[13] N. Frolov, V. Maksimenko, A. Lüttjohann, A. Koronovskii, and A. Hramov, "Feed-forward artificial neural network provides data-driven inference of functional connectivity," Chaos **29**, 091101 (2019).

[14] A. Banerjee, J. Pathak, R. Roy, J. G. Restrepo, and E. Ott, "Using machine learning to assess short term causal dependence and infer network links," Chaos **29**, 121104 (2019).

[15] A. Schnitzler and J. Gross, "Normal and pathological oscillatory communication in the brain," Nat. Rev. Neurosci. **6**, 285–296 (2005).

[16] D. S. Bassett and O. Sporns, "Network neuroscience," Nat. Neurosci. **20**, 353–364 (2017).

[17] A. E. Hramov, N. S. Frolov, V. A. Maksimenko, S. A. Kurkin, V. B. Kazantsev, and A. N. Pisarchik, "Functional networks of the brain: From connectivity restoration to dynamic integration," Phys. Usp. **64**, 584 (2021).

[18] V. P. Buch, A. G. Richardson, C. Brandon, J. Stiso, M. N. Khattak, D. S. Bassett, and T. H. Lucas, "Network brain-computer interface (NBCI): An alternative approach for cognitive prosthetics," Front. Neurosci. **12**, 790 (2018).

[19] A. E. Hramov, V. A. Maksimenko, and A. N. Pisarchik, "Physical principles of brain-computer interfaces and their applications for rehabilitation, robotics and control of human brain states," Phys. Rep. **918**, 1–133 (2021).

[20] G. Deshpande, L. Libero, K. R. Sreenivasan, H. Deshpande, and R. K. Kana, "Identification of neural connectivity signatures of autism using machine learning," Front. Hum. Neurosci. **7**, 670 (2013).

[21] L. Cai, X. Wei, J. Liu, L. Zhu, J. Wang, B. Deng, H. Yu, and R. Wang, "Functional integration and segregation in multiplex brain networks for Alzheimer's disease," Front. Neurosci. **14**, 51 (2020).

[22] Z. Zhao, J. Li, Y. Niu, C. Wang, J. Zhao, Q. Yuan, Q. Ren, Y. Xu, and Y. Yu, "Classification of schizophrenia by combination of brain effective and functional connectivity," Front. Neurosci. **15**, 552 (2021).

[23] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," Proc. Natl. Acad. Sci. U.S.A. **116**, 22071–22080 (2019).

[24] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell. **1**, 206–215 (2019).

[25] N. Frolov, V. Maksimenko, and A. Hramov, "Revealing a multiplex brain network through the analysis of recurrences," Chaos **30**, 121108 (2020).

[26] To avoid misunderstanding, we refer to interpretable machine learning in terms of the possibility of explaining its input features.

[27] V. Maksimenko, A. Kuc, N. S. Frolov, A. Hramov, A. Pisarchik, and M. Lebedev, "Neuronal adaptation in the course of the prolonged task improves visual stimuli processing," bioRxiv (2020).

[28] V. A. Maksimenko, A. Kuc, N. S. Frolov, M. V. Khramova, A. N. Pisarchik, and A. E. Hramov, "Dissociating cognitive processes during ambiguous information processing in perceptual decision-making," Front. Behav. Neurosci. **14**, 95 (2020).

[29] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," Front. Syst. Neurosci. **9**, 175 (2016).

[30] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: Identifying differences in brain networks," Neuroimage **53**, 1197–1207 (2010).

[31] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications* (Springer, 2008), Vol. 207.

[32] T. Fushiki, "Estimation of prediction error by using k-fold cross-validation," Stat. Comput. **21**, 137–146 (2011).

[33] J. F. Cavanagh and M. J. Frank, "Frontal theta as a mechanism for cognitive control," Trends. Cogn. Sci. **18**, 414–421 (2014).

[34] P. S. Cooper, F. Karayanidis, M. McKewen, S. McLellan-Hall, A. S. Wong, P. Skippen, and J. F. Cavanagh, "Frontal theta predicts specific cognitive control-induced behavioural changes beyond general reaction time slowing," Neuroimage **189**, 130–140 (2019).

[35] N. A. Herweg, E. A. Solomon, and M. J. Kahana, "Theta oscillations in human memory," Trends. Cogn. Sci. **24**, 208–227 (2020).

[36] J. F. Cavanagh, M. J. Frank, T. J. Klein, and J. J. Allen, "Frontal theta links prediction errors to behavioral adaptation in reinforcement learning," Neuroimage **49**, 3198–3209 (2010).

[37] P. Domenech, S. Rheims, and E. Koechlin, "Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex," Science **369**, eabb0184 (2020).

[38] M. X. Cohen and J. F. Cavanagh, "Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict," Front. Psychol. **2**, 30 (2011).

[39] A. K. Engel and P. Fries, "Beta-band oscillations—Signalling the status quo?," Curr. Opin. Neurobiol. **20**, 156–165 (2010).

[40] R. Schmidt, M. H. Ruiz, B. E. Kilavik, M. Lundqvist, P. A. Starr, and A. R. Aron, "Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function," J. Neurosci. **39**, 8231–8238 (2019).

[41] M. Lundqvist, P. Herman, M. R. Warden, S. L. Brincat, and E. K. Miller, "Gamma and beta bursts during working memory readout suggest roles in its volitional control," Nat. Commun. **9**, 1–12 (2018).

[42] A. M. Bastos, R. Loonis, S. Kornblith, M. Lundqvist, and E. K. Miller, "Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory," Proc. Natl. Acad. Sci. U.S.A. **115**, 1117–1122 (2018).

[43] B. A. Zavala, A. I. Jang, and K. A. Zaghloul, "Human subthalamic nucleus activity during non-motor decision making," eLife **6**, e31007 (2017).

[44] S. Hanslmayr, J. Matuschek, and M.-C. Fellner, "Entrainment of prefrontal beta oscillations induces an endogenous echo and impairs memory formation," Curr. Biol. **24**, 904–909 (2014).

[45] J. F. Hipp, A. K. Engel, and M. Siegel, "Oscillatory synchronization in large-scale cortical networks predicts perception," Neuron **69**, 387–396 (2011).

[46] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," Brain Res. Rev. **29**, 169–195 (1999).

[47] V. Maksimenko, N. Frolov, A. E. Hramov, A. N. Pisarchik, A. Kuc, and M. Lebedev (2020). "EEG and behavioral data for studying neural adaptation during the prolonged visual stimuli classification task," Figshare. https://doi.org/10.6084/m9.figshare.12155343.v2