

Bayesian Topological Learning for Brain State Classification

Farzana Nasrin, Christopher Oballe, David L. Boothe, and Vasileios Maroulas

Abstract—Investigation of human brain states through electroencephalograph (EEG) signals is a crucial step in human-machine communications. However, classifying and analyzing EEG signals are challenging due to their noisy, nonlinear and nonstationary nature. Current methodologies for analyzing these signals often fall short because they have several regularity assumptions baked in. This work provides an effective, flexible and noise-resilient scheme to analyze EEG by extracting pertinent information while abiding by the 3N (noisy, nonlinear and nonstationary) nature of data. We implement a topological tool, namely persistent homology, that tracks the evolution of topological features over time intervals and incorporates individual's expectations as prior knowledge by means of a Bayesian framework to compute posterior distributions. Relying on these posterior distributions, we apply Bayes factor classification to noisy EEG measurements. The performance of this Bayesian classification scheme is then compared with other existing methods for EEG signals.

Index Terms—Bayesian classification, EEG signals, intensity, marked Poisson point processes, persistent homology, topological data analysis

I. INTRODUCTION

The emergence of computational intelligence has led us to an era of excellent communication between users and systems. These human-computer communications do not require any external device or muscle intervention and enable computers to be deliberately controlled via the monitoring of brain state signals. In order to potentially improve human-machine interactions, it is crucial to analyze and interpret physiological measurements effectively to assess individual's states [1], [2]. Brain signals can encode one's expectations as a form of *prior beliefs*, which have an influence on behavior in times of uncertainty [3]. A Bayesian approach that integrates prior knowledge of an individual's innate brain activity with newly

measured data may improve individual's state detection, which can aide to characterize and control one's actions.

EEG signals are 3N—nonstationary, nonlinear and noisy [4]. In particular, they are obscured by various forms of noise, are nonlinear due to the complexity of underlying interaction in the nervous system [4]–[6] and are nonstationary due to the involvement of different time scales in brain activity [7]. The 3N nature of EEG signals requires methods that can encode individual's brain history and draw statistical inferences for these signals.

In this paper, we develop a Bayesian classification scheme relying on the posterior distributions of persistence diagrams, which are pertinent topological descriptors. Persistent homology is a widely used tool for topological data analysis (TDA) that captures topological features at multiple scales and produces summary representations, called persistence diagrams, that encode the lifespan of the topological features in the data. Persistent homology has proved to be promising in the field of data sciences yielding astounding results in a variety of applications [8]–[23]. Indeed, physiological signals' features are defined by the topological changes of the signals across time. Engaging TDA in the study of physiological signals is recently emerging. The authors of [24] measure the topology of EEG data with persistence landscapes to detect differences in the EEG signals during epilepsy attacks versus those produced during healthy brain activity. However, this method does not investigate the distribution of the diagrams themselves and suffers from a loss of pertinent information. Several other studies implement traditional machine learning based on feature extractions [25]–[27]. As selection of an appropriate feature is crucial, these methods rely on summaries of persistence diagrams, which already summarize the underlying data themselves. We develop a Bayesian learning approach that can be applied *directly* on the space of persistence diagrams. However, this learning scheme depends on the estimation of posterior probabilities, which is not straightforward due to the unusual multiset structure of persistence diagrams.

To establish a Bayesian framework, we need to define the prior uncertainty and likelihood through probability distributions of persistence diagrams. By viewing persistence diagrams as finite set samples, the authors of [13] propose a nonparametric estimation of the probability density function of random persistence diagrams. They also show that the probability density function can successfully detect the underlying dynamics of EEG signals and compare it with other pre-existing TDA methods. A prior distribution of persistence diagrams can be obtained through this density function. How-

F. Nasrin, C. Oballe and V. Maroulas are with University of Tennessee, Department of Mathematics, Knoxville, TN, USA (email: fnasrin@utk.edu). D. L. Boothe is with US Army Research Laboratory (ARL), Aberdeen Proving Ground, MD 21005, USA. The authors thank Piotr J. Franaszczuk from US Army Research Laboratory for his insightful comments and the software that generates dataset considered in this work.

FN's research was supported by the ARO W911NF-17-1-0313. CO's research was sponsored by the ARL and was accomplished under Cooperative Agreement Number W911NF-19-2-0302. DB's work was supported by the Combat Capabilities and Development Command US ARL. VM's work was partially supported by the ARO W911NF-17-1-0313, NSF MCB-1715794 and DMS-1821241, Thor Industries/ARL W911NF-17-2-0141, and ARL and was accomplished under Cooperative Agreement Number W911NF-19-2-0328. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ever, computing posteriors entirely through the random set analog of Bayes' rule may have exponential computational complexity [28]. To address this, we model random persistence diagrams as point processes. In particular, we utilize Poisson point processes which can be entirely characterized by their intensity.

We commence the Bayesian framework by modeling random persistence diagrams generated from a Poisson point process with a given intensity, which captures the prior uncertainty. In the case of brain state detection, we can incorporate individual's expectations about the statistical regularities in the environment as prior knowledge [3]. Alternatively, we can choose an uninformative prior intensity when no expert opinion or information about individual's expectations is available. We construct the likelihood component of our framework by utilizing the theory of marked point processes. Indeed, we employ the topological summaries of signals in place of the actual signals. This proves to be useful for a range of physiological signal analysis [1], [24]–[27], [29], [30]. The application considered in this paper is the classification of EEG signals, which allows us to predict individual's brain states and advances human-computer communication techniques. Through these topological summaries, we adopt a substitution likelihood technique [31] rather than considering the full likelihood of the entire signal data.

Next, we develop a learning method by relying on the posterior obtained from the Bayesian framework. This method is remarkably flexible as it abides by the 3N nature of the signals and is extremely powerful as it incorporates individual's expectations or domain experts' knowledge as prior beliefs. Furthermore, the Bayes factor provides a measure of confidence that in turn dictates whether further investigation is feasible. Our model enjoys a closed form of the posterior distribution through a conjugate family of priors, e.g., the Gaussian mixtures. Hence the prior-to-posterior updates yield posterior distributions of the same family. We present a detailed example of our closed form implementation on simulated EEG signals to demonstrate computational tractability and showcase applicability in classification through Bayes factor estimation. Furthermore, we present a detailed comparison with other TDA and non-TDA based learning methods.

This paper is organized as follows. Section II provides a brief overview of persistence diagrams and Poisson point processes. We establish the Bayesian framework for persistence diagrams in Section III. We then develop our Bayesian learning method in Section IV, which is used to quantify the classification outcome. Section V-A introduces a closed form to the posterior intensity utilizing Gaussian mixture models. To assess the capability of our algorithm, we investigate its performance in classifying EEG signals and provide comparisons with several other existing methods in Section V. Finally, we end with the conclusion in Section VI.

II. BACKGROUND

We commence by discussing the background essential for building our Bayesian model. In Subsection II-A, we start with

the formation of persistence diagrams (PDs) by implementing sublevel set filtrations. In order to model the uncertainty present in these persistence diagrams, we consider them as point processes and pertinent definitions from point processes (PPs) are given in Subsection II-B.

A. Persistent Homology for Noisy Signals

Persistent homology is a tool from TDA that provides a robust way to model the topology of real datasets by tracking the evolution of homological features and summarizing these in persistence diagrams. Several methods exist to generate persistence diagrams such as Vietoris Rips or Čech filtrations [32], but such techniques require the transformation of a signal to an appropriate point cloud using Takens's delay embedding theorem. To circumvent this transformation to point clouds, we employ the sublevel set filtration method, which summarizes the shape of signals *directly* in a PD by employing local critical points as tersely outlined next.

Consider a signal as a bounded and continuous function of time $f(t)$ (Fig. 1 (a)). The sublevel set filtration tracks the evolution of connected components in sets $f^{-1}((-\infty, r])$, as r increases. The central idea is that as r increases the connectivity of the set $f^{-1}((-\infty, r])$ remains unchanged except when it passes through a critical point. For a given connected component, we record the value of r at which is born (when r reaches a local minimum), call it b , and the value at which disappears (when r reaches a local maximum), call it d , by merging with a pre-existing connected component. That is to say, whenever two connected components merge, the one born later disappears while the one born earlier persists by the elder rule [32]. Once we reach the value $\max f(t)$ in the filtration, all the sublevel sets have merged into a single connected component, and we terminate the procedure. For every connected component that arises in the filtration, we plot the points (b, d) in $\mathbb{R}^2$ and call the resulting collection a persistence diagram (Fig. 1 (b)). To facilitate computation and preserve the geometric information, we apply the linear transformation $(b, p) = T(b, d) = (b - \min(b), d - b)$ to each point in our persistence diagrams. We refer to the resulting coordinates as birth and persistence, respectively, in $\mathbb{W} := \{(b, p) \in \mathbb{R}^2 \mid b, p \geq 0\}$ and call this transformed persistence diagram a tilted representation (Fig. 1 (c)). Hereafter whenever we refer to persistence diagrams, we imply their tilted representation.

B. Poisson Point Processes

One samples from a finite point process $\mathcal{P}$ on a Polish space $\mathbb{X}$ by generating a random number N according to a cardinality distribution and then for $N = n$ spatially distribute $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{X}$ according to a probability distribution. In other words, a finite point process is characterized by a probability mass function (pmf) of the cardinality and a joint probability density function (pdf) of the elements for a given cardinality. We model random persistence diagrams as Poisson point processes (PPPs). The defining feature of these point processes is that they are solely characterized by a single

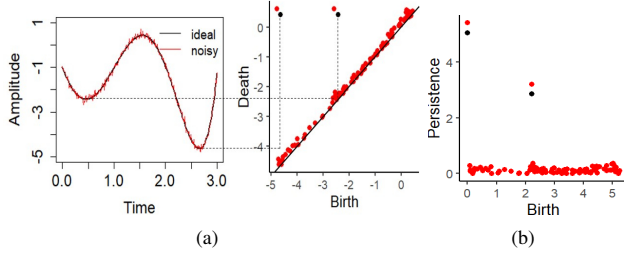


Fig. 1. (a) is illustrating the conversion of signals to the corresponding PDs using sublevel set filtrations. A smooth signal and a noisy version of it are presented in black and red respectively. For persistence diagrams, we make consistent color choices to instantiate the robustness of persistent homology to noise. (b) is the tilted representation of the PD in (a).

parameter known as their intensity. The intensity $\lambda(x)$ of a given $x \in \mathbb{R}^d$ is the density of the expected number of points per unit volume at x . Indeed, the intensity serves as an analog of the first order moment of a random variable. The intensity in a Poisson point process accounts for the joint pdf of elements, and the cardinality is Poisson with mean $\mu = \int \lambda(x)$.

Considering persistence diagrams as modeled by such processes, a link is needed between the prior and the data/likelihood to conduct Bayesian analysis. The *marked point process* provides this connection. Effectively, a marked point process is a special case of bivariate point process where one PP Ψ_M in the Polish space $\mathbb{W}_M$ (containing the marks) is determined given knowledge of the PP Ψ in the Polish space $\mathbb{W}$. A marked Poisson point process Ψ_M is a finite PP on $\mathbb{W} \times \mathbb{W}_M$ such that: (i) $\Psi = (\{p_n\}, \{\mathbb{P}_n(\bullet)\})$ is a PPP on $\mathbb{W}$, and (ii) for a realization $(\mathbf{x}, \mathbf{m}) \in \mathbb{W} \times \mathbb{W}_M$, the marks $m_i \in \mathbf{m}$ of each $x_i \in \mathbf{x}$ are drawn independently from a given stochastic kernel $\ell(\bullet|x_i)$.

III. THE BAYESIAN MODEL

According to Bayes' theorem, the posterior is proportional to the product of a likelihood function and a prior. To investigate Bayesian framework for persistence diagrams, we need to compute the conditional distribution $p(D_X|D_Y)$ by establishing the proposed Bayesian formula for persistence diagrams, $p(D_X|D_Y) \propto \mathcal{L}(D_Y|D_X)p(D_X)$, where the likelihood $\mathcal{L}(D_Y|D_X)$ and the prior $p(D_X)$ need to be defined and computed for random persistence diagrams. We employ a likelihood model for the persistence diagrams generated from the signals which is analogous to the idea of substitution likelihood [31]. Next, we develop the prior and likelihood on the space of persistence diagrams.

Prior: To model prior knowledge for the brain state classification problem, human expectations for statistical regularities in the environment and the uncertainty involved are summarized as a persistence diagram D_X . We assume that the underlying prior uncertainty of a persistence diagram D_X is generated by a Poisson point process $\mathcal{D}_X$ with intensity $\lambda_{\mathcal{D}_X}$. An example of prior persistence diagrams is shown in Fig. 2 (a) as black rectangles. Any point x in a persistence diagram D_X may not be observed in actual data due to the presence of noise, sparsity, and/or other unexpected scenarios. We address this instance by defining a probability function $\alpha(x)$. In particular, if x is not observed in the data, the probability of this event

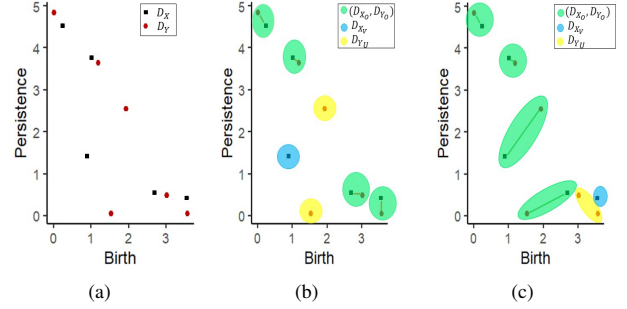


Fig. 2. (a) is a sample D_X from prior Poisson PP $\mathcal{D}_X$ and an observed persistence diagram D_Y . (b) and (c) are the decomposition of D_X into D_{X_O} & D_{X_V} and D_Y into D_{Y_O} & D_{Y_U} .

is $(1 - \alpha(x))$ and similarly $\alpha(x)$ is the probability of x being observed.

Data/Likelihood Model: EEG signals are encoded into the observed PDs, D_Y , using the method discussed in Section II-A. Points $y_i \in D_Y$ are linked to points in PD D_X , generated by the prior PPP. We investigate the linking of these points to the prior PPP by relying on the theory of marked Poisson point processes (MPPP) [33], [34]. The probability density of the MPPP is given by a stochastic kernel, ℓ such that the marks $m(x_i)$ of x_i are drawn independently from $\ell(\cdot|x_i)$, which in our case plays the exact role of the likelihood (see Section II-B for details). One needs to account for all possible marks, with the more likely marks realized as larger likelihood values $\ell(y_i|x_i)$ for all $(x_i, y_i) \in D_X \times D_Y$. In order to accommodate the nature of persistence diagrams, we need to define one last point process that unveils the topological noise in the observed data. Intuitively, this point process consists of the points in the observed diagram that fail to associate with the prior. We define this as a Poisson point process $\mathcal{D}_{Y_U}$ with intensity $\lambda_{\mathcal{D}_{Y_U}}$.

A sample observed persistence diagram is shown in Fig. 2 (a) as red hexagons. Fig. 2 (b) and (c) show different combinations of possible associations between prior and data in the green regions. However, it is evident that the associations in (b) would have higher likelihood values than that in (c) and in turn, would have more impact on posteriors. Also, for every configuration, some of the observed points do not associate with any point $x_i \in D_X$, which is shown with blue regions. We denote the features in blue regions as D_{X_V} , which stands for the features that vanished. If they are not vanished and make associations with features of D_Y , we denote it as D_{X_O} . Samples from $\mathcal{D}_{Y_U}$ are shown in Fig. 2 (b) and (c) as yellow regions.

Posterior: With the above model characterization, the posterior intensity which explicitly show the update of the prior has the following form [22]:

$$\lambda_{\mathcal{D}_X|D_{Y_{1:m}}}(x) = (1 - \alpha(x)) \lambda_{\mathcal{D}_X}(x) + \frac{\alpha(x)}{m} \sum_{i=1}^m \sum_{y \in D_{Y_i}} \frac{\ell(y|x) \lambda_{\mathcal{D}_X}(x)}{\lambda_{\mathcal{D}_{Y_U}}(y) + \int_{\mathbb{W}} \ell(y|u) \alpha(u) \lambda_{\mathcal{D}_X}(u) du}. \quad (1)$$

In the posterior intensity density, the two terms reflect the decomposition in the prior point process. The first term is for the features of prior which may not be observed and hence the intensity is weighted by $(1 - \alpha(x))$. On the other hand, the second term corresponds to the features in prior that may be observed and similarly is weighted by $\alpha(x)$. Here we observe an expression consistent with the traditional Bayes' theorem, specifically a product of prior intensity and likelihood divided by a normalizing constant. The normalizing constant consists of two terms illustrating the two instances of our data model. $\mathcal{D}_{Y_U}$ consists of the features that are not associated to the prior and this is evident in the first term of the normalizing factor. Consequently, the second term provides the contribution of the observed data from $\mathcal{D}_Y$, coupling with prior features to form the marked PPP.

IV. BAYESIAN CLASSIFICATION

In this section, we develop a Bayesian learning approach that discriminates EEG signals from different cognitive states. In particular, we present a classification scheme based on Bayes factors of persistence diagrams generated from physiological signals. We start by extracting fundamental topological features from a collection of EEG signals and record the information in persistence diagrams using the sublevel set filtration discussed in Section II-A.

For a persistence diagram D that needs to be classified, we assume that D is sampled from a Poisson point process $\mathcal{D}$ in $\mathcal{H}$ with prior intensity $\lambda_{\mathcal{D}}$. Consequently, its probability density has the form $p_{\mathcal{D}}(D) = \frac{e^{-\lambda}}{|D|!} \prod_{d \in D} \lambda_{\mathcal{D}}(d)$, where $\lambda = \int_{\mathbb{W}} \lambda_{\mathcal{D}}(u) du$ is the expected number of points in $\mathcal{D}$. For training sets $Q_{Y^k} := D_{Y_{1:n}^k}$ for $k = 1, \dots, K$ from K classes of random diagrams $\mathcal{D}_{Y^k}$, we obtain the posterior intensities by following the estimation process discussed in Section III. The posterior probability density of $\mathcal{D}$ given the training set Q_{Y^k} defined as

$$p_{\mathcal{D}|\mathcal{D}_{Y^k}}(D|Q_{Y^k}) = \frac{e^{-\lambda}}{|D|!} \prod_{d \in D} \lambda_{\mathcal{D}|\mathcal{D}_{Y^k}}(d). \quad (2)$$

The posterior probability densities given the other training sets are obtained by analogous expressions. Consequently, the Bayes factor is defined as

$$BF^{i,j}(Q_{Y^i}, Q_{Y^j}) = \frac{p_{\mathcal{D}|\mathcal{D}_{Y^i}}(D|Q_{Y^i})}{p_{\mathcal{D}|\mathcal{D}_{Y^j}}(D|Q_{Y^j})} \quad (3)$$

For every pair of (i, j) for $1 \leq i, j \leq K$ if $BF^{i,j}(Q_{Y^i}, Q_{Y^j}) > c$, we will assign one vote to class Q_{Y^i} or otherwise for $BF^{i,j}(Q_{Y^i}, Q_{Y^j}) < c$. The final assignment of the class of D to a class is obtained by a majority voting scheme.

V. APPLICATION TO EEG

A. Conjugate family of priors for EEG signals

Here, we present a closed form of the posterior distribution through a conjugate family of priors, e.g., the Gaussian mixtures. Hence the prior-to-posterior updates yield posterior distributions of the same family. We specify the prior intensity density as $\lambda_{\mathcal{D}_X}(x) = \sum_{j=1}^N c_j^{\mathcal{D}_X} \mathcal{N}^*(x; \mu_j^{\mathcal{D}_X}, \sigma_j^{\mathcal{D}_X} I)$, where

N is the number of corresponding mixture components and $\mathcal{N}^*$ is the restricted Gaussian density on the wedge $\mathbb{W}$. In a similar fashion, we define the density of the Poisson point process $\mathcal{D}_{Y_U}$. The likelihood density is also Gaussian as $\ell(y|x) = \mathcal{N}^*(y; x, \sigma^{\mathcal{D}_Y \circ I})$ and $\alpha(x) = \alpha$. With all of these we obtain a Gaussian mixture posterior intensity density of the form

$$\lambda_{\mathcal{D}_X|\mathcal{D}_{Y_{1:m}}}(x) = (1 - \alpha)\lambda_{\mathcal{D}_X}(x) + \frac{\alpha}{m} \sum_{i=1}^m \sum_{y \in \mathcal{D}_{Y_{ij}=1}}^N C_j^{x|y} \mathcal{N}^*(x; \mu_j^{x|y}, \sigma_j^{x|y} I), \quad (4)$$

where $C_j^{x|y}$, $\mu_j^{x|y}$ and $\sigma_j^{x|y}$ are weights, mean and variance of the posterior intensity respectively, corresponding to the second part of (1), and these are pertinent updates of the prior parameters [22].

B. EEG Datasets

US Army Aberdeen Proving Ground (APG) researchers have simulated noisy EEG signals based on different mental activities. We used this dataset for our analysis mainly focusing on two different frequency bands – alpha and beta. Alpha (frequency from 8 to 13 Hz) corresponds to intense mental activity, stress, and tension, and beta (frequency 13–30Hz) correlates with active attentions and focusing on concrete problems or solutions [35]. As the dataset contains several predominant oscillations based EEG signals, a Gaussian conjugate prior produces promising results for estimating the posterior probabilities as well as for Bayes factor classification [36]–[38].

C. Posterior estimation of EEG Datasets

We first converted the EEG signals to persistence diagrams via sublevel set filtrations. In Fig. 3, we present two samples from the EEG dataset of alpha (a) and beta (d) bands respectively along with their persistence diagrams in (b) and (e). Typically EEG signals encode various forms of noise and the simulated EEG dataset accounts for this by corrupting these signals with additive noise. The signals in Fig. 3 have the signal to noise ratio (SNR) 0, which implies equal contribution from signal and noise. In Fig. 3 we illustrate a posterior intensity estimation of a noisy alpha band and a noisy beta band utilizing (4). To demonstrate a data-driven posterior, we employed an uninformative prior of the form $\mathcal{N}((3, 3), 20I)$. To present the intensity maps uniformly, we divide the intensities by their corresponding maxima and call them scaled intensities ranging from 0 to 1.

D. EEG signal classification with Bayesian learning

Detection and classification of specific patterns in the brain activity are crucial steps in understanding functional behaviors for developing human-machine communications. We have taken the first step toward engaging Bayesian learning in EEG signal analysis by implementing Gaussian posterior intensities as explained in Section V-A and using these posteriors for a binary Bayes factor classification. From the dataset provided by APG researchers, we used two instances of additive noise

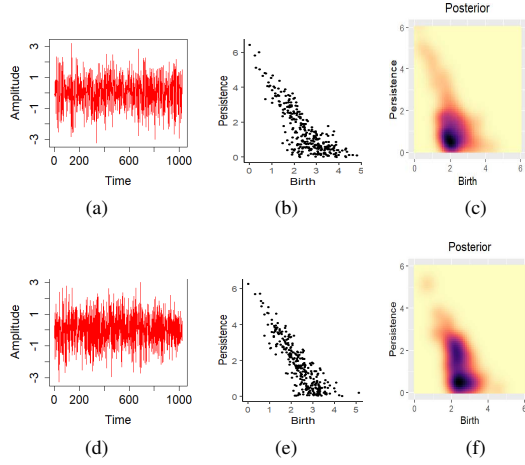


Fig. 3. (a) is an alpha band simulated EEG signal, (b) is the corresponding persistence diagram using sublevel sets and (c) is the posterior intensity map obtained from (4). Similarly (d) is a beta band, (e) is the corresponding persistence diagram, and (f) is the corresponding posterior intensity map.

in order to represent cases with two different SNR. Our considered dataset comprises SNRs such as 3 and 5, where SNR 5 has more contribution from the signal than SNR3.

We followed the process discussed in Section V-A to estimate the posterior intensity of a persistence diagram D in $\mathcal{H}$ given a training set Q_Y , with the goal of identifying the correct class of D . We used our developed R package BayesTDA to obtain posterior intensities. Consequently, the probability density was obtained from (2). After computing the intensities with respect to the training sets from both of the classes, the Bayes factor was computed by (3) as the ratio of the posterior probability densities of the unknown persistence diagram D given each of the two competing training sets from Q_Y or $Q_{Y'}$. For a threshold c , $BF(D) > c$ implies that D belongs to Q_Y and $BF(D) < c$ implies otherwise.

We implemented 10-fold cross validation for estimating the accuracy. For this we partitioned each class into 10 different sets and 9 of them for each class were used for training sets, and 1 was used for testing. We repeated this 10 times so that every partition acts as the testing data exactly once. We then found the average among all partitions. Results from the Gaussian learning scheme are presented in Fig. 4. We compared the results of Gaussian learning scheme with Artificial Neural Networks (ANNs), logistic regression (LR) with features (mean, standard deviation and entropy of the recorded coefficients) extracted from Wavelet Transform (WT). We prefer to use WT rather than Fourier transform (FT) due to its inability to analyze nonstationary nature of EEG signals [39], [40]. Both ANN and LR have been widely applied for physiological signals classification [25], [41]–[46]. We also compared our result with an exiting TDA technique namely, persistence landscape [47]. We extracted the first landscape functions of the persistence diagrams for all considered EEG signals and implemented support vector machine and logistic regression on the extracted landscape function. Our results for classifying these two bands of SNRs outperforms the other

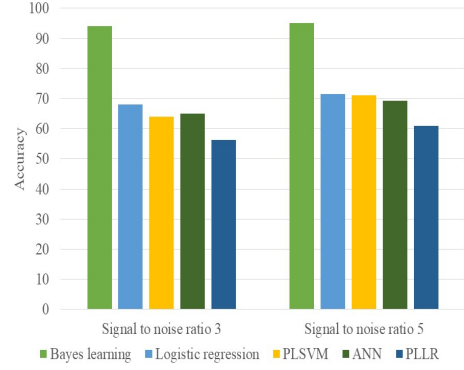


Fig. 4. Comparison of our Bayesian learning method with logistic regression, Artificial Neural Network (ANN), persistence landscape with support vector machine (PLSVM), and persistence landscape with logistic regression (PLLR). ANN was trained by a standard back propagation algorithm, a sigmoid activation function, and some other parameters such as, number of hidden layers = 100, the maximum number of iterations = 1000, error threshold = 0.001 and learning rate = 0.1.

existing TDA and non-TDA based classification methods over all levels of SNRs considered here. Furthermore, the Gaussian learning scheme is able to classify almost perfectly with a high signal to noise ratio.

VI. CONCLUSION

In this work, we have proposed a Bayesian framework for persistence diagrams that incorporates prior beliefs about signals and does not rely on any regularity assumptions such as stationarity or linearity for the computation of posterior distributions. The topological descriptors, e.g., persistence diagrams of EEG signals can decipher essential shape peculiarities by avoiding complex and unwanted geometric features. As required for a Bayesian paradigm, we incorporate prior uncertainty by viewing persistence diagrams as Poisson PPs with a given intensity. We model the connection between prior PP and persistence diagrams of noisy observations through marked PPs. These models the data likelihood component of the Bayesian framework. Additionally, we define the likelihood through topological summaries of a signal rather than using the entire signal. This is analogous to the substitution likelihood discussed by Jeffreys [31].

Relying on the posterior distributions obtained from the Bayesian framework, we develop a Bayesian learning scheme. Furthermore, we present a closed form of the posterior estimation through a conjugate family for priors. In the case of synchronized brain activity, this implementation is useful for analyzing EEG signals. This exhibits the ability of our method to recover the underlying persistence diagram, analogously to the standard Bayesian paradigm for random variables.

We employ this Bayesian learning scheme for EEG signal classification. We provide a detailed comparison with some of the existing methods of signal classification and showcase that our method outperforms them. For comparison purposes, we pursue two directions. Firstly, we compare with two most widely used signal classification algorithms—neural nets and logistic regression. Secondly, we show a comparison between

our method and another topological tool, namely persistence landscape, with traditional machine learning methods such as support vector machine and logistic regression. We exhibit higher accuracy for all considered cases. Thus, the Bayesian inference developed here opens up new avenues for machine learning algorithms for complex signal analysis *directly* on the space of persistence diagrams.

REFERENCES

- [1] M. Z. Ilyas, P. Saad, M. I. Ahmad, and A. R. I. Ghani, "Classification of EEG signals for brain-computer interface applications: Performance comparison," in *2016 International Conference on Robotics, Automation and Sciences (ICORAS)*, Nov 2016, pp. 1–4.
- [2] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors (Basel)*, vol. 19, no. 6, p. 1423, 2019.
- [3] H. Sohn, D. Narain, N. Meirhaeghe, and M. Jazayeri, "Bayesian computation through cortical latent dynamics," *Neuron*, 2019.
- [4] W. Klonowski, "From conformons to human brains: an informal overview of nonlinear dynamics and its applications in biomedicine," *Nonlinear Biomedical Physics*, pp. 1–5, 2007.
- [5] C. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field," *Clinical Neurophysiology*, vol. 116, pp. 2266 – 2301, 2005.
- [6] W. Klonowski, E. Olejarczyk, and R. Stepień, "'Epileptic seizures' in economic organism," *Physica A*, vol. 342, pp. 701–707, 2004.
- [7] P. Indic, R. Pratap, V. Nampoori, and N. Pradhan, "Significance of time scales in nonlinear dynamical analysis of electroencephalogram signals," *Int Journal of Neuroscience*, vol. 99, no. 1–4, pp. 181–194, 1999.
- [8] V. Patrangenaru, P. Bubenik, R. L. Paige, and D. Osborne, "Topological data analysis for object data," *arXiv:1804.10255*, 2018.
- [9] W. Guo, K. Manohar, S. L. Brunton, and A. G. Banerjee, "Sparse-TDA: Sparse realization of topological data analysis for multi-way classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1403 – 1408, 2018.
- [10] A. E. Sizemore, J. E. Phillips-Cremens, R. Ghrist, and D. S. Bassett, "The importance of the whole: Topological data analysis for the network neuroscientist," *Network Neuroscience*, 2018.
- [11] A. Babichev and Y. Dabaghian, "Persistent memories in transient networks," *Emergent Complexity from Nonlinearity, in Physics, Engineering and the Life Sciences*, vol. 191, pp. 179–188, 2017.
- [12] C. A. Biscio and J. Möller, "The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications," *Journal of Computational and Graphical Statistics*, pp. 1537–2715, 2019.
- [13] V. Maroulas, J. L. Mike, and C. Oballe, "Nonparametric estimation of probability density functions of random persistence diagrams," *Journal of Machine Learning Research*, 2019, in press, arXiv:1803.02739.
- [14] I. Sgouralis, A. Nebenführ, and V. Maroulas, "A Bayesian topological framework for the identification and reconstruction of subcellular motion," *SIAM Journal on Imaging Sciences*, vol. 10, no. 2, pp. 871–899, 2017.
- [15] V. Maroulas and A. Nebenführ, "Tracking rapid intracellular movements: a Bayesian random set approach," *The Annals of Applied Statistics*, vol. 9, no. 2, pp. 926–949, 2015.
- [16] J. Mike, C. D. Sumrall, V. Maroulas, and F. Schwartz, "Nonlandmark classification in paleobiology: computational geometry as a tool for species discrimination," *Paleobiology*, vol. 42, no. 4, pp. 696–706, 2016.
- [17] A. Marchese and V. Maroulas, "Signal classification with a point process distance on the space of persistence diagrams," *Advances in Data Analysis and Classification*, vol. 12, no. 3, pp. 657–682, 2018.
- [18] —, "Topological learning for acoustic signal identification," in *2016 19th International Conference on Information Fusion (FUSION)*, 2016, pp. 1377–1381.
- [19] Y. Lee and *et al.*, "Quantifying similarity of pore-geometry in nanoporous materials," *Nature Communications*, vol. 8, p. 15396, 2017.
- [20] T. Ichinomiya, I. Obayashi, and Y. Hiraoka, "Persistent homology analysis of craze formation," *Physical Review E*, vol. 95, no. 1, p. 012504, 2017.
- [21] M. Kimura, I. Obayashi, Y. Takeichi, R. Murao, and Y. Hiraoka, "Non-empirical identification of trigger sites in heterogeneous processes using persistent homology," *Scientific reports*, vol. 8, no. 1, p. 3553, 2018.
- [22] V. Maroulas, F. Nasrin, and C. Oballe, "A Bayesian framework for persistent homology," *arXiv:1901.02034*, 2019.
- [23] D. P. Humphreys, M. R. McGuirl, M. Miyagi, and A. J. Blumberg, "Fast estimation of recombination rates using topological data analysis," *GENETICS*, 2019.
- [24] Y. Wang, H. Ombao, and M. K. Chung, "Topological data analysis of single-trial electroencephalographic signals," *The Annals of Applied Statistics*, vol. 12, no. 3, pp. 1506–1534, 2018.
- [25] M. Dindin, Y. Umeda, and F. Chazal, "Topological data analysis for arrhythmia detection through modular neural networks," *arXiv:1906.05795*, 2019.
- [26] Y. Wang, D. D. M. K. Chung, A. Lutz, and R. Davidson, "Topological network analysis of electroencephalographic power maps," *Connectomics Neuroimaging*, vol. 10511, pp. 134–142, 2017.
- [27] M. Piangerelli, M. Rucco, L. Tesei, and E. Merelli, "Topological classifier for detecting the emergence of epileptic seizures," *BMC Research Notes*, vol. 11, p. 392, 2018.
- [28] I. Goodman, R. Mahler, and T. N. Hung, *Mathematics of Data Fusion*. Dordrecht/Boston/London: Kluwer Academic Publisher, 1997.
- [29] F. Altindis, B. Yilmaz, S. Borisenok, and K. Icoz, "Use of topological data analysis in motor intention based brain-computer interfaces," *26th European Signal Processing Conference*, 2018.
- [30] E. Campbell, A. Phinyomark, and E. Scheme, "Feature extraction and selection for pain recognition using peripheral physiological signals," *Frontiers in Neuroscience*, vol. 13, p. 437, 2019.
- [31] H. Jeffreys, *Theory of Probability*. Clarendon Press, 1961.
- [32] H. Edelsbrunner, *Computational topology: an introduction*. Providence, R.I.: American Mathematical Society, 2010.
- [33] J. F. C. Kingman, *Poisson Processes*. Oxford: Clarendon Press, 1993.
- [34] M. Jacobsen, *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhauser, 2005.
- [35] S. Siuly, Y. Li, and Y. Zhang, *EEG Signal Analysis and Classification: Techniques and Applications*. Cham, Switzerland: Springer, 2016.
- [36] K. H. Norwich, *Information, Sensation, and Perception*. San Diego, CA: Academic Press, Inc, 1993.
- [37] M. J. A. M. van Putten and C. J. Stam, "Application of a neural complexity measure to multichannel EEG," *Physics Letters A*, vol. 281, pp. 131–141, 2001.
- [38] R. A. A. Ince and *et al.*, "A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula," *Human Brain Mapping*, vol. 38, pp. 1541–1573, 2017.
- [39] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN Neuroscience*, 2014. [Online]. Available: <https://doi.org/10.1155/2014/730218>
- [40] G. Fisco and *et al.*, "Combining EEG signal processing with supervised methods for Alzheimer's patients classification," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, p. 35, 2018.
- [41] K. Sivasankari and K. Thanushkodi, "An improved EEG signal classification using neural network with the consequence of ICA and STFT," *J Electr Eng Technol*, vol. 9, no. 3, pp. 1060–1071, 2014.
- [42] A. Subasi and E. Ercelebi, "Classification of EEG signals using neural network and logistic regression," *Computer Methods and Programs in Biomedicine*, vol. 78, pp. 87–99, 2005.
- [43] R. Tomioka, K. Aihara, and K. Muller, "Logistic regression for single trial EEG classification," *Advances in Neural Information Processing Systems*, vol. 19, p. 1377?1384, 2007.
- [44] E. Kabir, Siuly, and Y. Zhang, "Epileptic seizure detection from eeg signals using logistic model trees," *Brain Informatics*, vol. 3, no. 2, pp. 93–100, Jun 2016.
- [45] P. D. Prasad, H. N. Halahalli, J. P. John, and K. K. Majumdar, "Single-trial EEG classification using logistic regression based on ensemble synchronization," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 1074–1080, 2014.
- [46] M. M. E. Bahy, M. Hosny, A. Mohamed, and S. Ibrahim, "EEG signal classification using neural network and support vector machine in brain computer interface," *Advances in Intelligent Systems and Computing*, 2016.
- [47] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, vol. 16, pp. 77–102, 2015.