Gaussian Processes with Physiologically-Inspired Priors for Physical Arousal Recognition

S Ghiasi[1], A Patane[2], A Greco[1], L Laurenti[2], EP Scilingo[1] and M Kwiatkowska[2]

Abstract— While machine learning algorithms are able to detect subtle patterns of interest in data, expert knowledge may contain crucial information that is not easily extracted from a given dataset, especially when the latter is small or noisy. In this paper we investigate the suitability of Gaussian Process Classification (GPC) as an effective model to implement the domain knowledge in an algorithm's training phase. Building on their Bayesian nature, we proceed by injecting problem-specific domain knowledge in the form of an *a-priori* distribution on the GPC latent function. We do this by extracting hand-crafted features from the input data, and correlating them to the logits of the classification problem through fitting a prior function informed by the physiology of the problem. The physiologically-informed prior of the GPC is then updated through the Bayes formula using the available dataset. We apply the methods discussed here to a two-class classification problem associated to a dataset comprising Heart Rate Variability (HRV) and Electrodermal Activity (EDA) signals collected from 26 subjects who were exposed to a physical stressor aimed at altering their autonomic nervous systems dynamics. We provide comparative computational experiments on the selection of appropriate physiologically-inspired GPC prior functions. We find that the recognition of the presence of the physical stressor is significantly enhanced when the physiologically-inspired prior knowledge is injected into the GPC model.

I. INTRODUCTION

Machine learning (ML) models and data-driven algorithms are often the prime choice for tackling recognition/detection problems in bio-engineering applications. While these techniques are largely successful, in many instances outperforming human/expert recognition performance [1], they neglect human-generated expert knowledge and previous problem-specific findings [2]. Incorporating such domain-specific knowledge further informs ML models about the context of the input data, which not only may enhance the recognition performance, but can also provide the user with an interpretable model [3]. Previous studies have considered including domain-specific knowledge into ML algorithms in various forms. Examples include the proposal for physics-informed random forest classifier for predictive turbulence modeling [4] or injecting the domain knowledge in the form of hierarchies into logistic regression models, with applications in hospital readmission prediction [5].

Gaussian Process (GP) models [6], in particular, through their non-parametric Bayesian formulation, facilitate flexible

[1] Department of Information Engineering & Bioengineering and Robotics Research Centre E. Piaggio, School of Engineering, University of Pisa, Italy.
[2] Department of Computer Science, University of Oxford

Corresponding author email: shadi1337@gmail.com

introduction of domain-specific knowledge into their learning algorithms, in the form of an *a-priori* distribution over the prediction model. In fact, several studies have looked at solving recognition problems via the application of GP models to physiological signals. GP regression has been used to design a vital-sign early warning system [7], to model noisy heart rate data [8], and to estimate respiratory rate [9]. The application of GP regression algorithms to physiological data has also been extended to model multivariate physiological time series simultaneously [10], and to propose a patient-specific approach to model time series of manual observational data in the presence of noise, incompleteness and artifacts [7]. Another example is the framework proposed in [11] where the GPs are applied to model the hemodynamics in task fMRI, resulting in the enhancement of the detection of brain activity compared to traditional parametric and non-parametric LTI models. However, among the studies that applied the GPs to physiological data, the design of an appropriate *a-priori* distribution over the GP model that successfully captures expert knowledge remains a challenge.

In this study, we aim to investigate the design and the effectiveness of GP prior functions that rely on physiological expert knowledge and the quantitative information available in the problem-specific training data. We show that integrating interpretable and medically meaningful knowledge into the GPC learning algorithm significantly enhances the recognition accuracy. We apply our GP-based framework to a dataset where two of the most relevant and used Autonomic Nervous System (ANS) correlates, Heart Rate Variability (HRV) and Electrodermal Activity (EDA), are acquired while subjects experience a physical stressor task. Our results demonstrate that introducing previous knowledge of ANS correlates of physical arousal into GP models improves detection accuracy of physical stress.

II. GAUSSIAN PROCESS CLASSIFICATION WITH PHYSIOLOGICAL-INSPIRED PRIORS

GPs are a class of machine learning models that by relying on the Bayes formula enable us to insert prior information into the learning model. In GPC settings we proceed by defining a latent variable $f \in \mathbb{R}$ that represents the classification logit and by putting a prior distribution over the latent space in the form of a Gaussian process. That is, given the input data item x, we define a prior distribution p on f such that $p(f|x) = \mathcal{N}(f; \mu(x|\alpha), k(x, x|\beta))$, for a specific choice of the functional form of the mean function μ and kernel function k and hyper-parameters α and β. Intuitively, the mean function $\mu(x|\alpha)$ and the kernel function $k(x, x|\beta)$ qualitatively encode our prior knowledge of the relationship between the input data and the subject's condition, while the

set of hyper-parameters α and β quantifies this relationship. Throughout this paper we use the squared-exponential kernel for the choice of k, as this provides flexible and smooth GPC models that can adapt to different classification tasks for specific choices of β [12]. The selection of the mean prior function is discussed in the following subsection. In this study we focus on two-class classification problem for simplicity of exposition. However, we can generalise the formulation to multi-class and regression settings [12]. Consider a dataset $\mathcal{D} = \{(x^{(i)}, y^{(i)}) \mid x^{(i)} \in \mathbb{R}^n, y^{(i)} \in \{0,1\}, i = 1, \ldots, N\}$, where N represents the number of data points. Let $\mathbf{x} = [x^{(1)}, \ldots, x^{(N)}]$ be the combined input vector, $\mathbf{y} = [y^{(1)}, \ldots, y^{(N)}]$ be the combined class vector and $\mathbf{f} = [f^{(1)}, \ldots, f^{(N)}]$ be the latent variable vector over $\mathbf{x}$. In Bayesian inference settings, given a test point $x^* \in \mathbb{R}^n$ the output classification probability on x^*, that is, the probability that x^* belongs to class 1, is computed by the GPC model through the following set of equations [13]:

$$p(\mathbf{f}|\mathcal{D}, \alpha, \beta) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x}, \alpha, \beta)}{p(\mathbf{y}|\mathbf{x})} \quad (1)$$

$$p(\bar{f}|\mathcal{D}, x^*, \alpha, \beta) = \int p(\bar{f}|\mathbf{x}, x^*, \mathbf{f}, \alpha, \beta)p(\mathbf{f}|\mathcal{D}, \alpha, \beta)d\mathbf{f} \quad (2)$$

$$p(y=1|\mathcal{D}, x^*) = \int \sigma(\bar{f})p(\bar{f}|\mathcal{D}, x^*, \alpha, \beta)d\bar{f} \quad (3)$$

where $\bar{f} \in \mathbb{R}$, $\sigma(\cdot)$ is the likelihood function (i.e. the sigmoid) and $p(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^{N} \left[\sigma(f^{(i)})^{y^{(i)}}(1 - \sigma(f^{(i)}))^{1-y^{(i)}} \right]$ is the dataset likelihood. The above set of equations first computes the posterior distribution over the training data through the Bayes formula (Eqn 1), which is then used to evaluate the posterior latent distribution over the test point x^* (Eqn 2), and finally the classification probability for x^* is computed (Eqn 3). Unfortunately, because of the non-Gaussian likelihood used in classification, the posterior $p(\mathbf{f}|\mathcal{D}, \alpha, \beta)$ is not Gaussian, and hence the integrals in Eqns (2) and (3) cannot be computed analytically [13]. We apply the Laplace method to approximate the GPC posterior distribution. This proceeds by computing a Gaussian approximation of the latent posterior distribution $p(\bar{f}|\mathcal{D}, x^*, \alpha, \beta)$ of Eqn (2). Details of the computation can be found in [12].

A. Physiologically-inspired prior

We implement prior information within the GPC by designing a prior mean function $\mu(x|\alpha)$ on top of a feature-extraction pipeline informed by the problem's physiology. Given the raw signal $s \in \mathbb{R}^t$, where t is the number of time points, we map it to the feature space through the function $QANS(s) : \mathbb{R}^t \to \mathbb{R}^k$, which encodes the signal processing and feature-extraction routines, where k is the dimension of the feature space and $QANS(s) = [QANS_1(s), QANS_2(s), \ldots, QANS_k(s)]$. We investigate different functional forms to combine the features obtained from the $QANS(s)$ transformation in the prior space for the *a-priori* mean function $\mu(x|\alpha)$, namely, zero mean, constant mean, linear mean, polynomial function and weighted sum of projected cosines:

$$\mu(x|\alpha) = 0 \quad (4)$$
$$\mu(x|\alpha) = \alpha \quad (5)$$

$$\mu(x|\alpha) = \sum_{j=1}^{k} \alpha_j QANS_j(s) \quad (6)$$

$$\mu(x|\alpha) = \sum_{p=1}^{d} \sum_{j=1}^{k} \alpha_{pj} QANS_j(s)^p \quad (7)$$

$$\mu(x|\alpha) = \sum_{p=1}^{d} \alpha_p^{(1)} \cos\left(\sum_{j=1}^{k} \alpha_{pj}^{(2)} QANS_j(s) + \alpha_p^{(3)} \right) \quad (8)$$

where d in Eqn (7) is the degree of the polynomial function used and in Eqn (8) is the number of projected cosine components. The vector α in Eqns (4)–(8) is used to represent all the hyper-parameters involved in each specific mean function definition. In the experiments, we learn α from the data using a maximum likelihood estimation approach [12].

III. MATERIALS AND METHODS

A. Experimental protocol

We apply the GPC classification framework integrated with a physiologically-inspired prior to a dataset collected from 26 healthy subjects who were exposed to a physical stressor. During the experiment, two diverse physiological conditions were experienced by the subjects, Condition 1, which is the resting state to ensure hemodynamic stabilization, and Condition 2, which is the presence of a physical stressor known to alternate the ANS dynamics, mostly by increasing the arousal level perceived by subjects. During the resting state, the subjects were asked to sit in a comfortable chair for 4 minutes while watching a black screen. The physical stressor was administered to the subjects through a well-known standardised ANS clinic test, the cold pressor test. In this test, the subjects submerged their left hand up to the wrist into a tank filled with ice and water with a temperature of 0-4 degrees centigrade for a period of 3 minutes. During the experiment, the ECG and EDA signals were acquired using BIOPAC MP35 device with a sampling rate of 500 Hz. The experiment was approved by the "Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana". The recordings were carried out in agreement with the Declaration of Helsinki. Written informed consent was obtained from all subjects. More details on the dataset can be found in [14].

B. Definition of physiologically-inspired priors

We formulate the problem of the recognition of physical arousal as a two-class classification problem, where, given a vector of ANS physiological correlates $QANS(s) = [QANS_1(s), QANS_2(s), \ldots, QANS_k(s)]$, the goal is to estimate the condition y of the subject, where $y = 0$ stands for resting state and $y = 1$ for the presence of arousal. Particularly, in this work, we extracted features from both the HRV and EDA signals, which are known to be highly sensitive to ANS dynamics alternations [15], [16].

1) HRV-based feature space ($QANS_{HRV}$): The HRV time series, defined as the beat to beat variation of the instantaneous heart rate, has been demonstrated to have strong correlation with the ANS activity [15]. Through the time domain analysis of HRV, first and second order statistical moments, μ_{RR} and σ_{RR}^2, are obtained. Moreover, the spectral

quantification of HRV in the low-frequency (LF) and high-frequency (HF) power bands has been shown to provide useful means for understanding the interplay between the two main branches of the ANS activity, i.e. the sympathetic and the parasympathetic outflows. In fact, the LF and the HF indices as well as the ratio between them, i.e., (LF/HF), have been extensively used to interpret the physiological neural correlates behind cardiovascular oscillations [15].

2) EDA-based feature space ($QANS_{EDA}$): The electrical properties of the skin, induced by the activity of the sweat glands in response to a physiological stimulus, are reflected in the EDA signal. High amplitudes in the EDA signal occur during the presence of an elicitation with a high intensity, therefore demonstrating an indication of sympathetic nervous system activity [16]. The EDA signal is decomposed into two main components, the tonic and phasic activities, that have different relationships with the evoked stimulus [17]. In the literature, characterizing the ANS dynamics through the quantification of these two main components has been thoroughly investigated [16], [18]. The tonic activity representing the overall psycho-physiological state of the subject is the low-frequency component of skin conductance, while the high-frequency component is the phasic activity showing rapid alternations in the electrodermal activity. Mainly, the phasic component is quantified by the number of significant phasic driver peaks (nSCR), the sum of SCR amplitudes (SumAmpSCR), the maximum value of SCR amplitudes (MaxAmpSCR), the mean value of phasic activity (PhasicMean) and the standard deviation of phasic activity (PhasicStd). The quantification of tonic activity involves the mean, standard deviation and the maximum value of tonic activity leading to the indices, TonicMean, TonicStd and TonicMax, respectively. Another index ($EDASymp$) showing high correlation with the activity of the sympathetic nervous system is obtained from the spectral analysis of the EDA signal. Particularly, we obtain this index by applying a time-frequency analysis, using short-time Fourier transform and Welch periodogram, to the preprocessed EDA signal and integrating the resulting spectrum within a specific frequency band $(0.045 - 0.25 Hz)$ [19].

3) HRV, EDA-combined feature space ($QANS_{HE}$): The correlation between the high-frequency power of the HRV with the parasympathetic activity and the EDA spectral power with the sympathetic outflow led to the proposal of a new index reflecting the sympathovagal balance in our recent studies [14], [20]. This, $QANS_{HE}$, obtained through a spectral analysis of EDA and HRV signals, has been shown to outperform the previous indicator of sympathovagal activity, leading to a better characterization of ANS dynamics.

C. Model training

We group these aforementioned QANS indices into three sets of features to be fed into the classification algorithm for the recognition task. Indices calculated from the HRV signal form *Feature Set 1*, while *Feature Set 2* contains information from the EDA signal. The indices obtained from both EDA and HRV signals are grouped into *Feature Set 3*. In detail:

- *Feature Set 1* = $QANS_{HRV}$ = [μ_{RR}, σ_{RR}^2 , LF, HF, LF/HF]
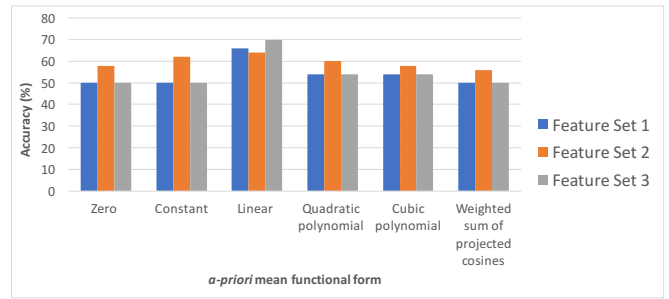

Fig. 1. Classification results in terms of accuracy (%) using three sets of features with six different choices of mean prior function. For the definition of the Feature Sets and the functions refer to the text in Section IV.

- *Feature Set 2* = $QANS_{EDA}$ =[nSCR, SumAmpSCR, MaxAmpSCR, PhasicMean, PhasicStd, Tonicmean, TonicStd, TonicMax, EDA_{Symp}]
- *Feature Set 3* = [$QANS_{HRV}$, $QANS_{EDA}$, $QANS_{HE}$].

Each of the feature sets was in turn chosen to define a candidate GPC prior function. During the training phase, the six different *a-priori* functions described in Section III were selected as the mean prior function, $\mu(x|\alpha)$. The classification results are given in terms of the recognition accuracy after implementing the standard Leave One Subject Out (LOSO) validation algorithm. Matlab software (R2017b version) and the Gaussian Processes for Machine Learning (GPML) toolbox [12] were used to implement GPC model training and prior function estimation.

IV. RESULTS

The classification results are depicted in Figure 1 for the three different *Feature Sets* and the five *a-priori* mean functional forms. The results for all feature sets demonstrate that the GP classifier is not able to satisfactorily distinguish the resting and arousal sessions when a non-informative prior function is used. On the other hand, when including mean prior functions, the GP model increases the recognition accuracy. More specifically, the highest accuracy (70%) is achieved when *Feature Set 3* is used in combination with the linear prior function. Interestingly, the results for the other feature sets also demonstrate that the linear function outperforms all mean functional forms. There is a significant increase in accuracy (up to +20%) when the linear function is chosen for *Feature Set 3* compared to the case when no *a-priori* mean function is considered. Specifically, this accuracy is 20% higher than the constant function and weighted sum of projected cosines, and 16% higher than the quadratic and cubic polynomial functions. The recognition accuracy for *Feature Set 2* without the choice of prior function is 58%. Although this value is 2% higher compared to the form of weighted sum of projected cosines for the mean function, it is still 4% lower than the constant mean function and 6% lower than the linear function. The classification results for *Feature Set 1* also show a significant increase in recognition (16%) when the appropriate prior function is chosen.

V. DISCUSSION

In this preliminary study, we propose a new approach for the development of machine learning models that can

efficiently incorporate both data and domain knowledge. This approach not only enhances the recognition accuracy, but also potentially provides more understandable models for the biomedical engineering community and the clinicians.

We focus on GP models due to their non-parametric Bayesian formulation, which provides a perfect environment for the inclusion of previously acquired knowledge or assumptions in the form of prior functions. Particularly, thanks to specific prior models based on the combination of well-known quantifiers of ANS dynamics, we can impose physiologically sound knowledge on the learning model, which can help in the recognition of different physiological states. We applied the proposed methodology to the problem of identifying an arousal state induced by a physical stressor such as the cold pressor test. The physiological state of 26 healthy volunteers was characterized through the analysis of HRV and EDA signals. The selected prior models showed an improved recognition accuracy compared to a zero mean function, which is commonly adopted in the GPC applications. Interestingly, but somewhat unsurprisingly, the linear combination of the selected ANS quantifiers significantly outperforms the other tested (nonlinear) models. Indeed, given the fact that our knowledge of the relationship between the central arousal state and the information provided by each HRV and EDA metric and their combination is still ill-defined, only a multi-feature approach can consider all the possibly relevant information. However, this increases the complexity of the model by introducing a deterministic noise which affects learning adversely and increases the risk of overfitting [21]. A linear model helps to mitigate overfitting by constraining the model complexity when the number of variables added to the mean prior function increases. In addition, the use of a maximum likelihood estimation of the prior parameters may induce overfitting. This could be solved by using a full Bayesian approach also for selecting the parameters of the mean prior function. However, the results discussed are given for LOSO cross validation and we did not observe overfitting, in terms of significant differences in train/test accuracy, in the models obtained. This might be due to the Bayesian nature of our framework that provides automatic model regularisation [13].

A further confirmation that the information provided by a multisignal and multivariate approach is more effective is evident from the comparison of the results associated with the three feature sets, which show better performance when EDA and HRV were both considered.

In conclusion, the proposed approach provides a framework for the design and evaluation of the suitability of a variety of prior models, which encode problem-specific knowledge into the GPC model. This can be used to quantify the relative importance of prior models and observations (i.e., data). This enables the use of GPC-based interpretability measures [22], in the form of an understanding of the relative contributions of the prior function w.r.t. the data, in the final classification result. As future work, we plan to extend the methodology by injecting problem-based prior functions into the GP regression formulation. This approach will allow us to consider the dynamics of the physiological signals for designing the effective form of prior function for each input.

REFERENCES

[1] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.

[2] D. Ong, H. Soh, J. Zaki, and N. Goodman, "Applying probabilistic programming to affective computing," *IEEE Transactions on Affective Computing*, 2019.

[3] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 1135–1144.

[4] J.-X. Wang, J.-L. Wu, and H. Xiao, "Physics-informed machine learning approach for reconstructing reynolds stress modeling discrepancies based on dns data," *Physical Review Fluids*, vol. 2, no. 3, p. 034603, 2017.

[5] S. Radovanović, B. Delibašić, M. Jovanović, M. Vukićević, and M. Suknović, "A framework for integrating domain knowledge in logistic regression with application to hospital readmission prediction," *International Journal on Artificial Intelligence Tools*, vol. 28, no. 06, p. 1960006, 2019.

[6] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[7] e. a. Clifton, Lei, "Gaussian process regression in vital-sign early warning systems," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 6161–6164.

[8] O. Stegle, S. V. Fallert, D. J. MacKay, and S. Brage, "Gaussian process robust regression for noisy heart rate data," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 9, pp. 2143–2151, 2008.

[9] e. a. Pimentel, Marco AF, "Probabilistic estimation of respiratory rate using gaussian processes," in *2013 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 2902–2905.

[10] R. Dürichen, M. A. Pimentel, L. Clifton, A. Schweikard, and D. A. Clifton, "Multitask gaussian processes for multivariate physiological time-series analysis," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 314–322, 2014.

[11] J. Wilzen, A. Eklund, and M. Villani, "Physiological gaussian process priors for the hemodynamics in fmri analysis," *arXiv preprint arXiv:1708.06152*, 2017.

[12] C. E. Rasmussen and H. Nickisch, "Gaussian processes for machine learning (gpml) toolbox," *Journal of machine learning research*, vol. 11, no. Nov, pp. 3011–3015, 2010.

[13] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Springer, 2003, pp. 63–71.

[14] S. Ghiasi, A. Greco, R. Barbieri, E. P. Scilingo, and G. Valenza, "Assessing autonomic function from electrodermal activity and heart rate variability during cold-pressor test and emotional challenge," *Scientific Reports*, vol. 10, no. 1, pp. 1–13, 2020.

[15] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[16] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[17] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.

[18] A. Greco, G. Valenza, and E. P. Scilingo, *Advances in Electrodermal activity processing with applications for mental health*. Springer, 2016.

[19] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Annals of biomedical engineering*, vol. 44, no. 10, pp. 3124–3135, 2016.

[20] S. Ghiasi, A. Greco, M. Nardelli, V. Catrambone, R. Barbieri, E. P. Scilingo, and G. Valenza, "Investigating phasic activity of time-varying high-order spectra: A heartbeat dynamics study during cold-pressor test," in *2018 Computing in Cardiology Conference (CinC)*, vol. 45. IEEE, 2018, pp. 1–4.

[21] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning from data*. AMLBook New York, NY, USA:, 2012, vol. 4.

[22] A. Blaas, L. Laurenti, A. Patane, L. Cardelli, M. Kwiatkowska, and S. Roberts, "Robustness quantification for classification with gaussian processes," *arXiv preprint arXiv:1905.11876*, 2019.