Joint data imputation and mechanistic modelling for simulating heart-brain interactions in incomplete datasets

Jaume Banus, Maxime Sermesant, Oscar Camara, Marco Lorenzi

Joint data imputation and mechanistic modelling for simulating heart-brain interactions in incomplete datasets

Jaume Banus[1], Maxime Sermesant[1], Oscar Camara[2], and Marco Lorenzi[1]

[1] Inria, Epione team, Université Côte d'Azur, Sophia Antipolis, France
`jaume.banus-cobo@inria.fr`
[2] PhySense, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

Abstract. The use of mechanistic models in clinical studies is limited by the lack of multi-modal patients data representing different anatomical and physiological processes. For example, neuroimaging datasets do not provide a sufficient representation of heart features for the modeling of cardiovascular factors in brain disorders. To tackle this problem we introduce a probabilistic framework for joint cardiac data imputation and personalisation of cardiovascular mechanistic models, with application to brain studies with incomplete heart data. Our approach is based on a variational framework for the joint inference of an imputation model of cardiac information from the available features, along with a Gaussian Process emulator that can faithfully reproduce personalised cardiovascular dynamics. Experimental results on UK Biobank show that our model allows accurate imputation of missing cardiac features in datasets containing minimal heart information, e.g. systolic and diastolic blood pressures only, while jointly estimating the emulated parameters of the lumped model. This allows a novel exploration of the heart-brain joint relationship through simulation of realistic cardiac dynamics corresponding to different conditions of brain anatomy.

Keywords: Gaussian Process · Variational Inference · Lumped model · Missing features · biomechanical simulation

1 Introduction

Heart and brain are characterized by several common physiological and pathophysiological mechanisms [6]. The study of this multi-organ relationship is of great interest, in particular to better understand neurological diseases such as vascular dementia or Alzheimer's disease. The development of computational models simulating heart and brain dynamics is currently limited by the lack of databases containing information for both organs. Neuroimaging datasets often provide a limited number of cardiac-related measurements, usually restricted to brachial diastolic and systolic blood pressure (DBP, SBP) [7]. These quantities provide a limited assessment of the cardiac function, thus compromising the possibility of further analysis of cardiovascular factors in brain disorders.

On the contrary, the availability of a rich set of cardiac information in heart studies allows the use of cardiovascular models to estimate descriptors of the cardiac function that are not possible to measure in-vivo, such as contractility or stiffness of the heart fibers. These models optimize the parameters through data-assimilation procedures to reproduce the observed clinical measurements [11], but usually do not include neurological factors. The ability to jointly account for cardiovascular descriptors and brain information is key to gather novel insights about the relationship between heart dynamics and brain conditions.

Most of current studies relating heart and brain are based on statistical association models, such as based on multivariate regression [5]. While this kind of analysis allows to easily formulate and test association hypotheses, it usually offers a limited interpretation of the complex relationship between organs. This issue is generally addressed by mechanistic modeling, allowing deeper insights on physiological and biomechanical aspects. These models allow for example to describe the brain vasculature, and to quantify physiological aspects such as blood flow auto-regulation effects in the brain [1], up to the simulation of the whole-body circulation with detailed compartmental components [3]. Although these approaches offer a high level of interpretability, they are usually severely ill-posed and require large data samples and arrays of measurements to opportunely tune their parameters.

To bridge the gap between data-driven and mechanistic approaches to heart-brain analysis, in this work we propose to learn cardiovascular dynamics from brain imaging and clinical data by leveraging on large-scale datasets with missing cardiac information. This is achieved through an inference framework composed of two nested models accounting respectively for the imputation of missing cardiac information conditioned on the available cardiac and brain features, and for a Gaussian Process emulator that mimics the behavior of a lumped cardiovascular model. This setting allow us to formulate a probabilistic end-to-end generative model enabling imputation of missing measurements and estimation of cardiovascular parameters given a subset of observed heart and brain features.

Results on real data from the UK Biobank show that our framework can be used to reliably estimate and simulate cardiac function from datasets in which we have minimal cardiovascular information, such as SBP and DBP only. Moreover, the proposed framework allows novel exploration of the joint heart-brain relationship through the simulation of realistic cardiac dynamics corresponding to different scenarios of brain anatomy and damage.

2 Methods

2.1 Problem statement

We denote by ν the vector representing brain image-derived phenotypes (IDPs) and clinical information such as age or body surface area (BSA), and by x the vector of cardiac IDPs and blood pressure measurements. The vector x can be represented as $x = \{\hat{x}, x_{obs}\}$ where $\hat{x}$ represents the unobserved information

we wish to impute and x_{obs} the observed one. Moreover, we assume that for each observation x a corresponding set of parameters y of the associated lumped model is available. We would like to learn a generative model in which we assume that the unobserved measurements are generated by a latent random variable z conditioned on the variables ν. Hence, our generative process can be seen as sampling from a distribution $p(z|\nu)$ and then obtaining $\hat{x}$ with probability $p(\hat{x}|z, \nu)$. Due to the association between cardiac IDPs, x, and lumped model parameters y, we also assume that the latter are dependent from ν through z. Our graphical model is shown in Figure 1, while the evidence lower bound (ELBO) of the joint data marginal $p(y, \hat{x}|x_{obs}, \nu)$ writes as:

$$
\begin{aligned}
\log p(y, \hat{x}|x_{obs}, \nu) &= \int \log p(y, \hat{x}|x_{obs}, z, \nu) p(z|\nu) dz \\
&\geq \mathbb{E}_{q_\phi(z|x_{obs}, \nu)} \log p(y, \hat{x}|x_{obs}, z, \nu) \\
&\quad - KL(q_\phi(z|x_{obs}, \nu) || p_\theta(z|\nu)) \\
&= \mathbb{E}_{q_\phi(z|x_{obs}, \nu)} \log p_\omega(y|x_{obs}, z) \\
&\quad + \mathbb{E}_{q_\phi(z|x_{obs}, \nu)} \log p_\theta(\hat{x}|\nu, z) \\
&\quad - KL(q_\phi(z|x_{obs}, \nu) || p_\theta(z|\nu)) \equiv \mathcal{L}(\theta, \phi, \omega; x_{obs}, \nu)
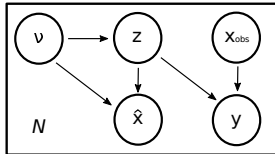\end{aligned}
\tag{1}
$$

Fig. 1. Graphical model of our framework. From a latent variable z we generate the unobserved features $\hat{x}$ conditioned on the variables ν and we estimate y via Gaussian process regression. During inference x_{obs} and ν are used to estimate the approximated posterior $q_\phi(z|x_{obs}, \nu)$.

The approximation of the posterior distribution by $q_\phi(z|x_{obs}, \nu)$ defines the optimization of the ELBO through variational inference. The variational distributions $q_\phi(z|x_{obs}, \nu)$ and $p_\theta(\hat{x}|\nu, z)$ are parametrized respectively with parameters ϕ and θ. The term $\log p_\omega(y|x_{obs}, z)$ in the ELBO denotes the emulator which approximates the mechanistic behavior of the lumped model via Gaussian Process regression parametrized by ω. The choice of a GP as emulator is motivated by the uncertainty of the data, hence it is desirable to obtain a distribution of interpolating functions rather than a single deterministic function. Moreover, GPs have already been proved to be valid emulators of 1D mechanistic vascular models [10]. The GP allow us to sample functions $f(x)$ from a given prior parameterized by a mean $\mu(x)$ and covariance $\Sigma(x)$, i.e: $f(x) \sim \mathcal{N}(\mu(x), \Sigma(x))$ to

obtain the marginal $y \sim \mathcal{N}(\mu(x), \Sigma(x) + \sigma^2 I)$. The prior mean $\mu(x)$ is here set to 0 and we use a radial-basis function (RBF) kernel for the covariance:

$$k^j(x_i, x_i') = \alpha_j^2 exp\left(-\frac{(x_i - x_i')^2}{2\beta_i^2}\right),\tag{2}$$

Since our data is multi-dimensional, $k^j(x_i, x_i')$ is the kernel for the jth target and the ith predictor. The hyper-parameters of the kernel $\omega = \{\alpha, \beta\}$ represent the output amplitude α and length scale β of the sampled functions. The goal during training is to learn the hyper-parameters that maximize the marginal likelihood of the observed data y .

The second term of the ELBO is related to the imputation of $\hat{x}$. The term $\log p_\theta(\hat{x}|\nu, z)$ denotes the log-likelihood of the imputed features, which can be seen as the reconstruction error. The last term $KL(q_\phi(z|x_{obs}, \nu)||p_\theta(z|\nu))$ is the Kullback–Leibler divergence between variational approximation and prior for z, that can be expressed in a closed form given that both distributions are Gaussians. The imputation scheme is equivalent to a conditional variational autoencoder (CVAE) which has become a popular approach to feature imputation [8]. Fast and efficient optimization in our model is possible by means of stochastic gradient descent thanks to the closed form for data fit and KL terms, the use of the reparametrisation trick and Monte Carlo sampling.

2.2   Data processing and cardiovascular model

From UK Biobank we selected a subset of 3445 subjects for which T1, T2 FLAIR magnetic resonance images (MRI) and several brain and cardiac IDPs were available. Among the brain IDPs we used total grey matter (GM), total white matter (WM) and ventricles volumes. We used T1 and T2 FLAIR images to obtain the number of white matter hyper-intensities (WMHs) and their total volume relying on the lesion prediction algorithm (LPA), available from the lesion segmentation toolbox (LST) [14] of SPM [3]. WMHs are a common indicator of brain damage of presumably vascular origin [15]. We combined WM and GM volumes into a single measurement that we denoted as brain volume. The WMHs total volume and number of lesions presented a skewed distribution, and were Box-Cox transformed prior to the analysis. Regarding the cardiac IDPs we selected stroke volume (SV), ejection fraction (EF) and end-diastolic volume (EDV) for the left ventricle. All brain-related volumes were normalized by head size. Besides IDPs, we had access to blood pressure measurements (DBP, SBP) and sociodemographic features such as age and body surface area (BSA). We used DBP and SBP to compute MBP as $MBP = DP + (SP - DP)/3$. Next, we used the lumped cardiovascular model derived in [4] to obtain additional indicators of the cardiac function. In particular we estimated the contractility of the main systemic arteries (τ), peripheral resistance (R_p) , the radius of the left ventricle (R_0), contractility of the cardiac fibers (σ_0) and their stiffness (C_1). The parameters of the model were selected based on the available clinical data. To obtain

---

[3] https://www.fil.ion.ucl.ac.uk/spm

the target values for the emulator, the data-assimilation procedure was carried out according to the approach presented in [2, 12].

2.3   Experiments

The data was split in two sets: one containing the full-information (2309 subjects), and one in which cardiac IDPs ($\hat{x}$) and the estimated model parameters (y) were removed (1136 subjects, Table 1). The quality of imputation was compared to conventional methods such as mean, median and k-nearest neighbors (KNN) imputation. We subsequently assessed the relationship learnt by our model between brain features and cardiovascular parameters through simulation. Starting from the mean values of each parameter we sampled along the dimension of the different conditional variables ν that we used to parameterize the prior. This procedure allowed us to assess their influence in the inferred simulation parameters.

Table 1. Variables used in our framework. Mean blood pressure (MBP), diastolic blood pressure (DBP), stroke volume (SV), end-diastolic volume (EDV), ejection fraction (EF), heart fibers contractility (σ_0), ventricle size (R_0), heart fibers stiffness (c_1), peripheral resistance (R_p) and aortic compliance (τ)

Input (x_{obs})	Condition (ν)	Imputed ($\hat{x}$)	Predicted (y)
MBP	Brain volume	SV	σ_0
DBP	Ventricles volume	EDV	R_0
	WMHs volume	EF	c_1
	Num. WMHs		R_p
	BSA		τ
	Age		

3   Results

Data imputation and regression. We assessed the performance of our model by measuring the mean squared error (MSE) on the testing data, for both the imputation of missing cardiac information and the emulation of the lumped model parameters. The results in Figure 2a show that our method gives significantly better estimates than mean and median imputation, and comparable predictions to KNN, for which the optimal number of neighbors was optimized through 10-fold cross-validation and corresponds to $K = 10$. At the same time the emulator consistently gives low errors for the parameters' estimation (Figure 2b). In supp. Figure 3 we present the most relevant predictors for each emulated feature based on their β values, while a qualitative comparison between the distributions of imputed and emulated features compared to the ground truth data is available in in supp. Figure 4.
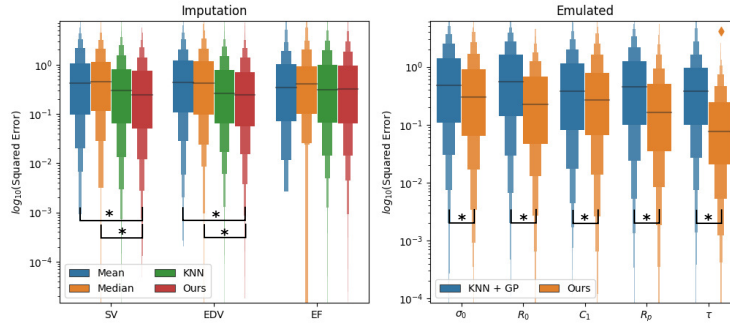
Fig. 2. Mean squared error (MSE) a) of the imputation of missing cardiac measurements ($\hat{x}$) b) of the estimated parameters of the emulated cardiovascular lumped model (y). * denotes that the MSE distributions are significantly different with respect to our method according to the Wilcoxon rank-sum test using a significance level of $a = 0.05$, Bonferroni corrected by multiple comparison.

Cardiovascular dynamics simulation and model plausibility. In Figure 3a we observe the change in the predicted parameters of the cardiovascular model as we sample along the range of values of the different conditional variables ν. Figure 3b shows the pressure-volume (PV) loops generated by the lumped model using the inferred parameters. The simulated PV loops highlight meaningful relationships:

- An increase in the volume of WMHs is associated to decreased SV and EDV, together with a smaller reduction of ESV, leading to a decrease of EF which is related to reduced contractility.
- Similar dynamics are associated to brain volume loss.
- The number of WMHs exhibits different dynamics than WMHs volume, associated to the increase in afterload and the increase of EDV.

From the plots showing the estimation of the parameters we can observe the ones driving the observed dynamics. For example, WMHs and brain volumes changes are mainly driven by the joint evolution of peripheral resistance (R_p), contractility (σ_0) and size of the left ventricle (R_0). The changes in the number of WMHs are related to heart-remodelling changes, driven by R_0 and by a decrease of σ_0 and the ventricular stiffness c_1. In supp. Figure 2 we provide analysis and discussion of the remaining conditional features (age, BSA and brain ventricles volume). Overall, in the simulated dynamics we can identify several physiological responses in line with the clinical literature. Moreover, our model may be useful to give insights in currently controversial topics such as concerning the pathogenesis of WMHs. Our results suggest that the evolution of cardiac function with respect to brain and WMHs volumes is similar to the one due to aging (see supp. Figure 2), while the effect induced by the number of WMHs is
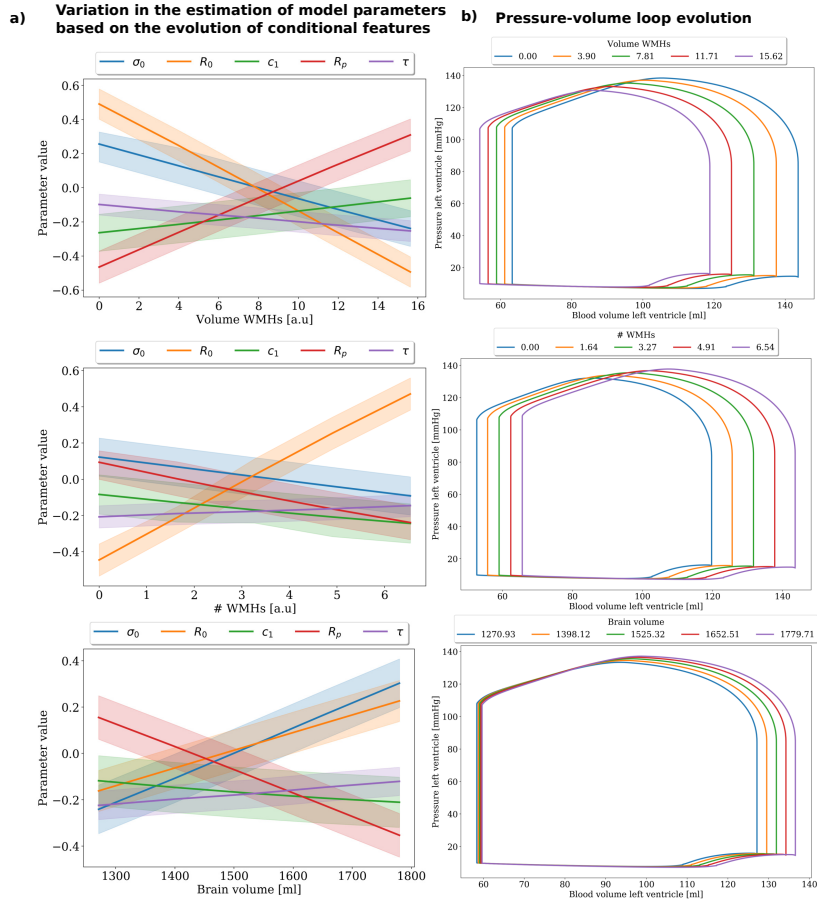
Fig. 3. a) Inferred model parameters and their respective confidence interval as we sample along the dimension of the different conditional variables ν while keeping the other elements of the generative framework constant. b) Pressure-volume loops generated by the the cardiovascular lumped model given the mean inferred parameters.

similar to the one related to ventricles enlargement. In both cases the changes in R_0, σ_0 and C_1 suggest that the increase in the number of WMHs and the enlargement of the brain ventricles are related to heart-remodelling processes, such as loss of contractility or decrease in compliance. These findings are in line with clinical observations [9] relating lower cardiac output with higher burden of WMHs and reduced brain volume.

4   Conclusion

We presented a generative model that enables the analysis of complex physiological relationships between heart and brain in datasets where we have minimal

available features. The framework allow us to emulate a lumped cardiovascular model through data-driven inference of mechanistic parameters, and provides us a generative model to explore hypothetical scenarios of heart and brain relationships. In the future, the model will allow to potentially transfer the knowledge learnt in UK Biobank to datasets where we have minimal cardiac information, to explore the relationship between brain conditions and cardiovascular factors in specific clinical contexts, such as in neurodegeneration. Our approach could also be extended to account for deep learning architectures, and the framework could be further improved by jointly accounting for multiple outputs, which are currently modelled independently, or by including spatial information from imaging data, beyond the modelling of scalar volumetric features. Furthermore, while the cardiovascular features considered in this study are rather general, more complex features will allow to study more realistic cardiovascular models. For example, while the mechanistic model used in this study does not simulate cerebral blood flow, previous studies suggested that WMHs may be due to local vascular impairment [13]. Hence, by selecting appropriate clinical features we could constrain the imputation by means of any biophysical model representing the desired aspect of systems biology. This could represent an innovative tool in real world scenarios, for which multi-modal patient data is often limited or not available.

Acknowledgments

References

1. Acosta, S., et al.: An effective model of cerebrovascular pressure reactivity and blood flow autoregulation. Microvasc. Res. **115**(November 2016), 34–43 (2018)
2. Banus, J., et al.: Large Scale Cardiovascular Model Personalisation for Mechanistic Analysis of Heart and Brain Interactions **3504**, 285–293 (2019)
3. Blanco, P., et al.: An anatomically detailed arterial network model for one-dimensional computational hemodynamics. T-BME **62(2)**, 736–753 (2015)
4. Caruel, M., et al.: Dimensional reductions of a cardiac model for effective validation and calibration. Biomech. Model. Mechanobiol. **13**, 897–914 (2014)

5. Cox, S.R., et al.: Associations between vascular risk factors and brain MRI indices in UK Biobank. European Heart Journal **44**, 1–11 (2019)
6. Doehner, W., et al.: Heart and brain interaction in patients with heart failure: overview and proposal for a taxonomy. Europ. Jour. of H. Failure **20**(2) (2018)
7. Epstein, N.U., Lane, K.A., Farlow, M.R., Risacher, S.L., Saykin, A.J., Gao, S., Initiative, A.D.N.: Cognitive dysfunction and greater visit-to-visit systolic blood pressure variability. Journal of the American Geriatrics Society **61**(12), 2168–2173 (2013)
8. Ivanov, O., et al.: Variational autoencoder with arbitrary conditioning. 7th, ICLR 2019 pp. 1–25 (2019)
9. Jefferson, A.L., et al.: Lower Cardiac Output Is Associated with Greater White Matter Hyperintensities in Older Adults with Cardiovascular Disease. J Am Geriatr Soc **55**, 1044–1048 (2009)
10. Melis, A., et al.: Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators. INT J NUMER METH BIO. **33**(12), 1–11 (2017)
11. Molléro, R., Pennec, X., Delingette, H., Garny, A., Ayache, N., Sermesant, M.: Multifidelity-cma: a multifidelity approach for efficient personalisation of 3d cardiac electromechanical models. Biomechanics & Modeling in Mechanobiology **17(1)**, 285–300 (2017)
12. Molléro, R., et al.: Population-based priors in cardiac model personalisation for consistent parameter estimation in heterogeneous databases. INT J NUMER METH BIO. (2018)
13. Müller, L.O., Toro, E.F.: Enhanced global mathematical model for studying cerebral venous blood flow. Journal of Biomechanics **47(13)**, 3361–3372 (2014)
14. Schmidt, P.: Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging. PhD (January 2017)
15. Wardlaw, J.M., Valdés Hernández, M.C., Muñoz-Maniega, S.: What are white matter hyperintensities made of? relevance to vascular cognitive impairment. Journal of the American Heart Association **4(6)** (2015)