Data-driven modelling of subjective pain/pleasure assessments as responses to vaginal dilation stimuli

Damiano Varagnolo, Steffi Knorn, Ernesto Oliver-Chiva,
Reinhilde Melles and Marieke Dewitte

***Abstract*— Women affected by pain during penetrative sexual intercourse are often treated using fixed-size vaginal dilators that are regularly perceived as uncomfortable and leading to premature treatment drop-outs. These dilators could be improved by making them adaptive, i.e., able to exert dynamically different pressures on the vaginal duct to simultaneously guarantee comfort levels and achieve the medical dilation objectives. Implementing feedback control would then benefit from models that connect the patients' comfort levels with their experienced physiological stimuli.**

Here we address the problem of data-driven quantitative modelling of pain/pleasure self-assessments obtained through medical trials. More precisely, we consider time-series records of Pelvic Floor Muscles (PFM) pressure, vaginal dilation, and pain/pleasure evaluations, and model the relations among these quantities using statistical analysis tools. Besides this, we also address the important issue of the individualization of these models: different persons may respond differently, but these variations may sometimes be so small that it may be beneficial to learn from several individuals simultaneously. We here numerically validate the previous claim by verifying that clustering patients in groups may lead, from a data-driven point of view, to models with a significantly improved statistical performance.

***Index Terms*— dyspareunia, modelling of psychological systems, support vector classification, clustering**

I. Introduction

Motivations: Pain during penetrative sexual intercourse for prolonged periods of time as a consequence of Genito-pelvic pain / penetration disorders (GPPPD) or other conditions is estimated to affect 30-40% of women at least once in their life [1, Chap. 2]. The pain can be caused by physiological causes (e.g., complications after cervix cancer surgeries, vaginal radiotherapies, Mayer-Rokitansky-Küster-Hauser syndromes, male-to-female gender confirmation surgeries) and psychosocial causes (e.g., traumatic sexual experiences) [1, Chap. 3]. Observations from practitioners indicate also that psychological mechanisms (e.g., anxiety, catastrophising pain and avoidance of sexual intimacy) and interpersonal factors (e.g., hostile partner responses, relationship conflict) may maintain, prolong and exacerbate the suffering.

Treatments may combine psychological (e.g., Cognitive Behavioral Therapies (CBTs)) and physiological treatments, the latter potentially including stretching the vaginal duct, desensitizing the vestibulum, and relaxing the pelvic floor muscles [2]–[4] through vaginal dilators. However, since these therapies are perceived as invasive, lengthy and uncomfortable, patients often delay, avoid or stop treatment and hence prolong their suffering [5].

Improving these treatments is however non-trivial: on the one hand patients shall feel sufficiently comfortable to avoid dropping out. On the other hand there is the need to dilate the duct and stimulate the pelvic floor as much as possible to achieve the medical target while making the treatment as temporally short as possible. A primary issue is thus to design individualized optimal dilation strategies that account for this intrinsic tradeoff and that can be executed using vaginal dilators that are adaptable in size.

Several adaptable dilators have been developed: The Vaginal Pressure Inducer (VPI), developed at Maastricht University hospital, consists of a flexible balloon whose size can be gradually adjusted by inflating it with warm water [6]. Another example is the *Milli* dilator [7], a dildo that can expand its width (controlled by buttons on its base) and hence responds to the need for a more gradual and gentle stretching of the duct. These solutions, however, do not implement *feedback* concepts, i.e., do not adapt the dilation patterns to the patient's response starting from measurements, references, and –potentially– quantitative models connecting stimuli with physiological and psychological outcomes. In contrast, feedback control may simultaneously accommodate a patient's physiological response and medical needs while preventing pain and anxiety. An expected positive side effect is increased motivation due to increasing the patients' self-efficacy.

To be able to compute individual and adaptive vaginal dilation patterns, there is the need for individualized quantitative models that describe how patients will most likely respond to vaginal dilation and possibly other stimuli in conjunction with measurements of physiological and subjective signals. Ideally, these models should enable not only implementing model-based control strategies, but also interpretation by both medical personnel and patients.

Existing models in the literature: The medical literature comprises several physiological models that analyse some

D. Varagnolo and E. Oliver-Chiva are with the Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden. S. Knorn is with the Department of Engineering Sciences, Uppsala University, Sweden. R. Melles and M. Dewitte are with the Department of Psychology and Neuroscience, Clinical Psychological Science, Behavioural Medicine, Maastricht Universitair Medisch Centrum, Maastricht, The Netherlands. Emails: damiano.varagnolo@ltu.se, steffi.knorn@angstrom.uu.se, ernoli-6@student.ltu.se, marieke.dewitte@maastrichtuniversity.nl, reinhilde.melles@mumc.nl

The research leading to these results has received funding from the Swedish research council Norrbottens Forskningsråd.

cause-effect implications (e.g., [2], [8]–[17]). But all these models describe static cause-effect relationships and lack describing the *dynamics* of the processes. Towards closing this gap, data-driven dynamical models of female response to vaginal dilation were derived in [18], where time-series of pelvic floor pressure collected from healthy patients during ad-hoc medical trials were used to investigate which type of dynamical models can accurately describe the recorded data. [18], however, focused on physiological responses, leaving the psychological side completely unexplored.

Only few models the psychological or subjective response of women to vaginal dilation are available. The relations between sexual arousal and sexual desire seem to be complex and the existing literature orbits around the Basson's non-linear model of the female sexual response [19], that states that the sexual desire is affected by several psychological inputs (e.g., satisfaction with the relationship, self-image, previous sexual experiences), so that the desire is not just governed by biological factors. Indeed in this model the goal of sexual activity for women is not necessarily orgasm, but rather personal satisfaction, which can manifest itself as physical satisfaction (orgasm) and/or emotional satisfaction (e.g., a feeling of intimacy and connection with a partner). As for how the sexual desire relates to sexual arousal, some psychological factors (e.g., desire for increased emotional closeness and intimacy, etc.) may trigger a predisposition to participate in sexual activity. Sexual arousal may be triggered by conversations, music, reading or viewing erotic materials, or direct stimulation, which may lead to an increasing desire to continue the activity. Nonetheless, some other psychological factors may work as turn-off factors and diminish (up to vanishing) experienced sexual arousal and desire.

The alternative Masters & Johnson's sexual response model [20] distinguishes between various phases ("excitement / arousal", "plateau", "orgasm" and "resolution"): it describes the physiological responses of the female body in all these phases, but does not include quantitative descriptions of the dynamics of the system. Both Basson's and Masters & Johnson's models moreover focus on sexual responses of healthy women that do not suffer from pain during penetrative sexual intercourses. Hence, variables such as perceived fear or pain are not included.

For now the unique dynamical model describing the interplay of several key variables seems to be published in [21] and consists of two distinct loops, named the Circle Of Fear (COF) and Circle Of Pleasure (COP). The COF describes the facts that: *i)* pelvic muscle activity before or at the beginning of penetration may lead to pain; *ii)* fear induces muscular tension; and *iii)* inducing positive erotic stimuli may reduce fear. The COP instead relies on the Basson's model and describes that *i)* the physiological arousal increases if the patient is sexually stimulated and subjectively aroused; *ii)* the subjective arousal increases with sexually stimulation and pleasurable physical sensations; and *iii)* physiological arousal affects the subjective arousal indirectly via the intermediate state variable of physical pleasure. The model in [21] is solely based on known cause-effect relationships from the medical literature, informed guesses from experts in the field, and the objective of striking a balance between accuracy and simplicity to enable mathematical analysis. However, the model in [21] is neither directly based on specific medical tests nor measurement data, and is hence not validated from field experiments.

Contributions: Towards obtaining individualized quantitative models of the psychological responses to vaginal dilation stimuli, we analyze data-driven learning strategies based on experimental data recorded at Maastricht university hospital (described in Section II). We thus: *i)* cast the learning problem using a Support Vector (SV) framework that enables implementing dilation-control strategies and interpretations by medical personnel and patients, analyze the predictive performance of Support Vector Machine (SVM) on the available data, and draw some practical conclusions from these performance; *ii)* consider that, as often happens, we face a big constraint on the amount of available data. We thus consider the additional problem of understanding if (and how) grouping different patients into clusters and learning the models from different "clustered" datasets may help improving the learning process.

After describing the medical data set in Section II, we summarize the modelling choices in Section III. Sections IV and V contain the strategies for modelling individual patients and for extending these models to groups of patients and our quantitative results. Conclusions are drawn in Section VI.

II. Medical data set

This study is based on medical data recorded at Maastricht University Hospital and described in more detail in [6]. The data include participants' responses to a gradual vaginal dilation that is forced by the VPI, an inflatable balloon to be inserted at the introitus as graphically summarized in Figure 1. Patients undergoing the trial were also watching sequences of 5-minutes long erotic or non-erotic movies in the (tentatively) neutral environment.

The study included 36 women without sexual problems, aged between 18 and 45 years, in a steady heterosexual rela-

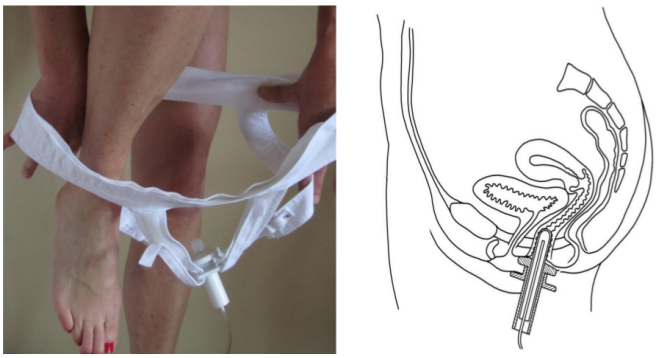

Fig. 1. Picture of the VPI (left) and schematic description of its usage (right). A pump can fill the balloon with water at body temperature; the length of the inflated area is up to 6 cm. When filled, the balloon gives an outward omnidirectional pressure to the surrounding tissues.

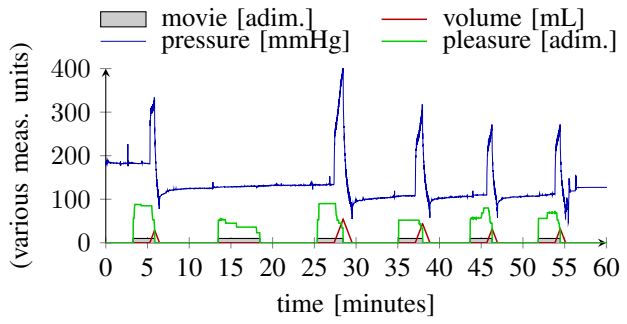

Fig. 2. Example of a typical dataset of time-series from one of the patients in the considered clinical trial. The six movie clips described above are in this case started at minutes 3, 13, 26, 35, 43 and 52. The VPI was inserted in the duct during the whole trial but inflated only while watching the movies (but the second one). The original sampling rate of the system is 10Hz; these raw signals were then downsampled to 1Hz in our following derivations.

tionship for at least 3 months, and being sexually active including coitus. Each individual participated in single sessions where, while using the VPI and watching movies sequences, they recorded their perceived level of comfort/pleasure on a scale from 0 to 100 with an opportune slider. As soon as the pressure felt unbearable, participants could end the experiment and force the deflation of the balloon by pressing an emergency button. Sessions started with the presentation of a neutral acclimatization movie with pressure induction using the VPI. This was followed by one high-arousal sexual movie without inducing vaginal pressure, then followed by four randomized movies with inducing pressure (one high-arousal and sexual, one low-arousal and sexual, one high-arousal and nonsexual, and one neutral movie), see Figure 2.

Since the pressure is measured at the pump, it should be considered an aggregated indication of the force exerted by the Pelvic Floor Muscles (PFM). Due to the mechanics of the system, the pressure data are subject to noise and measurement inaccuracies. Since the perceived pleasure was measured by a simple slider during the experiments, they are also subject to noise. First of all, it is visible in the data that women sometimes did not change their pleasure levels for several minutes followed by sometimes rapid changes or ending the inflation of the balloon. It is hence assumed that they sometimes simply "forgot" to update their pleasure levels through the slider. Further, one must keep in mind that a subjective measure such as the perceived pleasure also greatly depends on the individual expectation and definition of pleasurable sensation (aspects that constitute a further human-induced measurement noise). Statistically modelling this noise in an accurate way is probably a formidably complex problem, and thus we leave this issue for future work. In this paper we then do not take these considerations explicitly into account, and derive our models choosing the simple model structures indicated in the following Section III.

III. The Modelling Problem

Our focus is to obtain data-driven models that can describe and predict changes in pleasure levels in women as a response to vaginal dilation stimuli. While dynamical models such as Hammerstein-Wiener are found to be suitable to describe physiological models in several applications (see [18], [22], [23]), as hinted in the previous section our datasets for modelling the pleasure levels contains recorded pleasure data that appear to be heavily affected by human-induced noise. We thus avoid considering dynamics and formulate a function estimation problem, i.e., assume that there is a static map between the inputs and the output.

In details, the inputs of the system will be the measurable physiological quantities (i.e., the volume of the dilator, denoted with u_{volume}, the pressure of the pelvic floor muscles, u_{pressure}, and their time derivatives). The output will be the subjectively assessed (and typically non-measured) pleasantness level of the treatment, y_{pleasure}. Before identifying maps of the type $y_{\text{pleasure}} = \psi\left(u_{\text{volume}}, u_{\text{pressure}}\right)$ there is the need to discuss the structure of the model. To this purpose, visually inspecting Figure 2 we notice that there exist:

- **positive and negative jumps** in the measured pleasure level y_{pleasure}, i.e., sudden increases or decreases (above a chosen threshold $\bar{y}_{\text{pleasure}}$) indicating that the subject has been experiencing something pleasurable or unpleasant that motivated or reminded her to report this;
- **continuations**, i.e., periods where users do not change the perceived pleasure level y_{pleasure}. Note that the available information does not allow to differentiate between possible reasons for such "continuations": e.g., the subjects may be experiencing changes that are too small to be worth recording, or simply forget to update their indications;
- **stops**, i.e., situations where the patients press an emergency stop button to indicate that the experienced pressure or other sensations were considered unbearable. "Stops" are in a sense the limit case of "negative jumps".

Intuitively, high volumes, pressures, and positive derivatives of these signals should increase the likelihood that the user will press the stop button or set a negative jump in y_{pleasure}. Letting $u(t-T:t)$ indicate a signal u in the time window $[t-T, t]$, this intuition says that if there is a "stop" event happening at time t_a and a "continuation" happening at time t_b, then the norms of $u_{\text{pressure}}(t_a - T : t_a)$, $\dot{u}_{\text{pressure}}(t_a - T : t_a)$, $u_{\text{volume}}(t_a - T : t_a)$ and $\dot{u}_{\text{volume}}(t_a - T : t_a)$ should be statistically higher than the norms of $u_{\text{pressure}}(t_b - T : t_b)$, $\dot{u}_{\text{pressure}}(t_b - T : t_b)$, $u_{\text{volume}}(t_b - T : t_b)$ and $\dot{u}_{\text{volume}}(t_b - T : t_b)$ for an opportune (and to be determined from the data) window length T whose physical meaning is a particular type of human reaction time. As Figure 3 shows, the collected datasets confirm this intuition, indicating that there are zones of volume, pressure and their derivatives that are clearly associated to specific events.

Recall then that the original and foreseen control problem is to design u_{volume}, u_{pressure}, $\dot{u}_{\text{volume}}$ and $\dot{u}_{\text{pressure}}$ so that *i)* the user does not experience feelings considered unbearable (i.e., avoid stop events), and *ii)* to minimize/avoid unpleasant experiences (i.e., avoid negative jumps events). Thus, the modelling problem is not to find a model for y_{pleasure}, i.e., a ψ s.t. $y_{\text{pleasure}} = \psi\left(u_{\text{volume}}, u_{\text{pressure}}\right)$, but rather to find a model that can predict positive and negative jumps, continuations

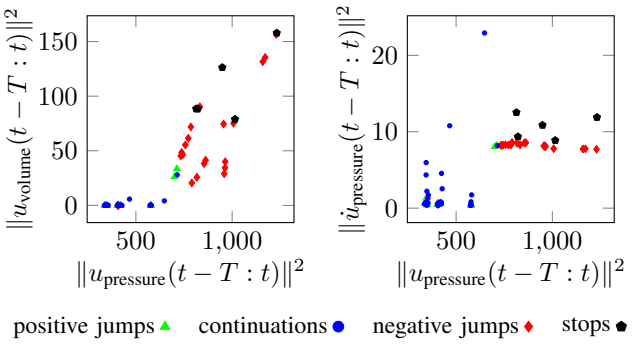

Fig. 3. Illustration of the features corresponding to the time-series signals shown in Figure 2 (due to space limitations we omit plotting the couples u_{volume}-$\dot{u}_{\text{volume}}$ and $\dot{u}_{\text{pressure}}$-$\dot{u}_{\text{volume}}$). In our experiments we empirically determined the time window $T = 10$ seconds and the threshold $\bar{y}_{\text{pleasure}} = 2$ by manually minimizing the Frobenius norm of the cross classification matrix $\mathcal{M}$ defined in the following Section V.

and stops. In other words, we cast the modelling problem as finding a ϕ s.t.

$$y = \phi\left(u_{\text{volume}}, u_{\text{pressure}}\right), \quad \text{with}$$
$$y \in \{\text{pos. jump, continuation, neg. jump, stop}\}. \quad (1)$$

Following the intuition developed above, this problem corresponds to finding partitions of the features space, a problem that can be cast naturally as a Support Vector Classification (SVC) problem.

IV. Individual Models of the Patients Reactions

Assume the availability of P individual time-series datasets of P women (denoted, for simplicity, with the IDs $1, \ldots, P$) as, e.g., the one in Figure 2. Given the features defined in Section III, for each $p \in \{1, \ldots, P\}$ it is possible to transform the associated time-series dataset into a features-oriented dataset $\mathcal{D}_p$ similar to the one represented in Figure 3. For each $p \in \{1, \ldots, P\}$ we can moreover divide $\mathcal{D}_p$ into a training set $\mathcal{D}_p^{\text{train}}$ and a test set $\mathcal{D}_p^{\text{test}}$. In our setup, $\mathcal{D}_p^{\text{train}}$ corresponds to the periods relative to the third and fourth movie, and $\mathcal{D}_p^{\text{test}}$ corresponds to the last two movies. The first two movies are disregarded since they correspond to an acclimatisation period. Given $\mathcal{D}_p^{\text{train}}$ we can train an individual SVC machine $\phi_p(\cdot)$. Each classifier $\phi_p(\cdot)$ can then be applied on any dataset $\mathcal{D}_j^{\text{train}}$, which corresponds to testing how well the model of patient p can classify the features of patient j. Testing all possible combinations of p and j leads to a cross-Classification Error (CE) matrix $\mathcal{M}^{\text{train}}$ whose (p, j)-th element is the classification error when using ϕ_p to classify $\mathcal{D}_j^{\text{train}}$:

$$\mathcal{M}^{\text{train}}: \quad \begin{array}{c|cccc} & \mathcal{D}_1^{\text{train}} & \mathcal{D}_2^{\text{train}} & \cdots & \mathcal{D}_P^{\text{train}} \\ \hline \phi_1 & \text{CE}_{11} & \text{CE}_{12} & \cdots & \text{CE}_{1P} \\ \phi_2 & \text{CE}_{21} & \text{CE}_{22} & \cdots & \text{CE}_{2P} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_P & \text{CE}_{P1} & \text{CE}_{P2} & \cdots & \text{CE}_{PP} \end{array}$$

The problem of determining the structure of ϕ in (1) becomes thus the problem of finding the best SVC type, kernel and hyperparameters, that, for our specific problem, can be cast in several ways. The extremes are: *a)* select individual and potentially different optimal structures ϕ_p through individual Leave-One-Out (LOO) Cross Validation (CV) strategies for each patient p; *b)* constrain all ϕ_p's to share the same type and kernel, and choose them by minimizing the Frobenius norm of $\mathcal{M}^{\text{train}}$ (the hyperparameters being again potentially tunable in an individualized manner). Here we choose strategy b, since it increases the possibilities of introducing concepts of distances between different machines, e.g., by comparing them by the respective support vectors. For completeness, for our dataset the best SVC structure (among linear, polynomial up to degree 4 and radial basis kernels) was empirically determined as linear and based on the four features u_{pressure}, u_{volume}, $\dot{u}_{\text{pressure}}$, and $\dot{u}_{\text{volume}}$ defined over time windows of 10 seconds. The actual values of the cross-classification error matrix $\mathcal{M}^{\text{train}}$ are graphically reported in Figure 4. Finally, note that up to now the quantities have been defined using the training sets $\mathcal{D}_p^{\text{train}}$. The test sets $\mathcal{D}_p^{\text{test}}$ will indeed be used in Section V to assess the predictive performance of the final classifiers.

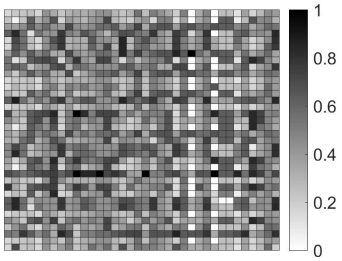

Fig. 4. Cross-classification error matrix $\mathcal{M}^{\text{train}}$ relative to the considered clinical trials, expressed as a greyscale image. Note that the pixels on the diagonal of this image correspond to classification errors in the training set for each individual patient. The minimum, average, and maximum cross classification errors were respectively 0, 0.446, and 1.

V. From Individual Models to Group Models

Inspecting the cross-classification error matrix shown in Figure 4 we can find couples of patients p and j that have low cross-classification errors, i.e., such that ϕ_p classifies well $\mathcal{D}_j^{\text{train}}$ and ϕ_j classifies well $\mathcal{D}_p^{\text{train}}$. Intuitively, this is an indication that these patients share similar models, and that thus they may be in a sense "clustered" together (see also [24] for other applications of the concepts developed hereafter).

The natural questions are then:

1) Do patients tend to fall into a finite set of well defined categories?
2) If so, assuming that patients p and j belong to the same category, would a "group" classifier ϕ_{pj} trained with the dataset $\mathcal{D}_{pj}^{\text{train}} = \mathcal{D}_p^{\text{train}} \cup \mathcal{D}_j^{\text{train}}$ have better predictive capabilities than the individual machines ϕ_p and ϕ_j, i.e., better performance in classifying the test sets $\mathcal{D}_p^{\text{test}}$ and

$\mathcal{D}_j^{\text{test}}$? (Note that this concept can obviously be extended to groups of an arbitrary number of patients.)

The questions above may be answered through data-driven methodologies that check if partitioning the set $\{1,\dots,P\}$ into K disjoint groups leads to models with greater approximation capabilities, and thus a better usage of the available datasets. Determining the groups $\{G_1,\dots,G_K\}$ can then naturally be performed through first introducing an opportune concept of "distance" between the various patients, and then using classical clustering approaches based on the set of so-defined distances.

Defining these distances can be made by exploiting the structure of the classifiers (e.g., the different positions of the various support vectors across different machines), or intuitions based on the performance of the classification. E.g., the more the classification errors on the training sets CE_{pj}, CE_{pp}, CE_{jp}, and CE_{jj} are similar (i.e., the more the machines ϕ_p and ϕ_j can be swapped) the less the two patients p and j may be considered different (at least from a training sets perspective). Among the various possibilities, due to limitations in space we focus only on this last strategy, that we prefer over the others since it is more prone to intuitive interpretability for medical personnel.

To define the concept of distance between two patients from the cross-classification error matrix $\mathcal{M}^{\text{train}}$ in Section IV, we start by verifying from Figure 4 that $\mathcal{M}^{\text{train}}$ is not guaranteed to be symmetric. This means that $\mathcal{M}^{\text{train}}$ does not define a metric (i.e., a function satisfying non-negativity, symmetry and the triangle inequality). It is however possible to transform $\mathcal{M}^{\text{train}}$ into a dissimilarity matrix whose element (p,j) is given by

$$d_{ij}^{\text{train}} := \frac{\max\left(0,\text{CE}_{pj}-\text{CE}_{pp}\right)+\max\left(0,\text{CE}_{jp}-\text{CE}_{jj}\right)}{2}. \quad (2)$$

Even if the d_{pj}^{train}'s in general do not satisfy the triangle inequality, (2) is a proxy for how much the models of patients p and j differ (or, more precisely, how much the models based on their training sets differ). The d_{pj}^{train} constitute indeed a pseudo-distance, and can be used to run k-medoids [25, Sec. 14.3.10], an opportune generalization of k-means for the case of clustering through pseudo-distances.

To use k-medoids, though, there is the need to define the number of groups K. In compliance with classical clustering approaches, we thus propose to cast the problem as a numerically optimization problem where the solution minimizes within-cluster variances, i.e., to let

$$\{G_1^*,\dots,G_{K^*}^*\} := \arg\min_{\widetilde{K},\widetilde{G}_1,\dots,\widetilde{G}_K} \sum_{k=1}^{\widetilde{K}} \left(\sum_{p,j\in\widetilde{G}_k} d_{pj}^{\text{train}} \right) \quad (3)$$

where the superscript $*$ denotes optimality w.r.t. the just introduced cost function. Because this is a notoriously NP-hard problem, for which obtaining the optimal solution becomes rapidly numerically infeasible even for small dataset sizes, we solve (3) in an approximate way leveraging on the existing clustering algorithms available in the literature. More precisely, for every plausible number of groups K, we propose to:

1) compute, starting from the set of dissimilarity indexes d_{pj}^{train}'s, a corresponding clustering of the patients in K groups $\{G_1,\dots,G_K\}$ using a K-medoids clustering strategy;
2) for each group $G_k=\{p_1,\dots,p_{|G_k|}\}$ (whose physical meaning is "persons with similar models") form the group-wide training set $\mathcal{D}_{G_k}^{\text{train}} = \cup_{j\in G_k}\mathcal{D}_j^{\text{train}}$, train the group-wide SVC ϕ_{G_k}, and compute the classification error $\text{CE}_{G_kG_k}$ that the classifier ϕ_{G_k} commits in classifying its own training set;
3) compute the weighted average of the training errors $\text{CE}_{G_kG_k}$'s, where the weights correspond to the cardinalities of the various groups G_k's, and denote this average with $\overline{CE}_K$. (Recall that K indicates in how many clusters the original set of patients was divided and the clusters are indexed by k.)

Choosing that number of groups K^* that minimizes the average training errors $\overline{CE}_K$ means thus choosing that K^* (and that composition of the groups G_k^*) that maximizes the statistical performance of the classifiers from a training perspectives. The question is whether K^* and the corresponding groups compositions G_k^* lead to good prediction performance, i.e., if they perform better on the test datasets $\mathcal{D}_p^{\text{test}}$ than the individually trained machines. The answer to the question is plotted in Figure 5, where we compare the weighted average errors both in training and in test.

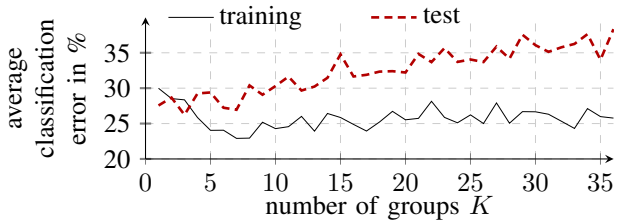

Fig. 5. Weighted average classification errors for the learned classifiers in the training sets and in the test sets as a function of the number of patients groups K.

Our strategy indicates that, for the available datasets, the best number of groups is 7, corresponding to an average classification error in the training sets of 22.9%, and in the test sets of 26.9%. Inspecting the graph, some conclusions follow: From a test sets perspective, the best would seem to be 3 groups, for a classification error of 26.2%. These numbers, though, slightly depend (even if not very heavily) on the hyperparameters of the estimation scheme (e.g., the length of the time window T chosen for computing the features, cf. Figure 3). Nonetheless, varying these hyperparameters does not change the trends for which the curves have minima around 5 to 10 groups (training case) and 3 to 5 groups (test case). The chosen K^* tends thus to overestimate the one that is optimal according to the performance in the test sets. Yet, that K^* leads also to considerable improvements of the performance against learning each patients individually (a

strategy that would lead, as shown in Figure 5, to an average classification error of 38.3% in the test sets).

Summarizing, even if we have been using a rather small dataset whose statistical significance is insufficient to make claims valid for the whole human population, the evidence seems to indicate that grouping patients together and performing joint learning is, in this particular medical framework, beneficial from statistical perspectives.

VI. Conclusions

Towards the technological goal of developing vaginal dilators that can autonomously adapt to patients and maximize the medical exercising while respecting the comfort levels, we studied how to derive quantitative models for effectively forecasting changes in pain/pleasure levels in patients subject to measurable vaginal dilation inputs. Specifically, we focused on the very important concept of how to automatically determine individualization levels of the aforementioned models: Since it is known that persons tend to respond differently, one should learn individual models. At the same time, differences among specific persons may sometimes be very small. Hence, it may be beneficial to learn from several individuals simultaneously.

We thus investigated how to structure such type of models as opportune support vector classifiers, and how to group patients starting from physiological measurements and subjective indications of pleasure / pain together. We then applied our strategies to data from 36 patients collected through ad-hoc medical tests, and obtained numerical results that seem to indicate that grouping patients is, in our specific medical context, actually improving the overall statistical performance of the models. Average classification errors on test sets passed indeed from 38.3% in the "non-grouped-patients" case to a 26.9% in the "grouped-patients" one.

Despite obtaining results that are in accordance with what intuition would suggest, we encountered technical and theoretical problems that deserve dedicated future investigations. For example, we have completely neglected discussing the case where a new patient is added to the dataset, and thus how to assign her to a group and potentially adapt the composition of the groups in a recursive (and non-naïve) way. Moreover, the used measurements are subjective, and thus contain psychological factors (e.g., forgetting about updating) that are non-observable in the settings considered in this paper. Modelling these factors corresponds in other words to modelling psychological systems, a subject that is poised to be very challenging and probably requiring the development of new ad-hoc mathematical tools for data-driven learning of psychological responses. We nonetheless consider these important topics and aim at investigating them in deep in our future works.

References

[1] A. Goldstein, C. Pukall, and I. Goldstein, *Female Sexual Pain Disorders*. John Wiley & Sons, 2009.

[2] Y. M. Binik, S. Bergeron, and S. Khalifé, "Dyspareunia and vaginismus: so-called sexual pain," in *Principles and practice of sex therapy*. The Guilford Press, 2006, ch. 5, pp. 124–156.

[3] S. Bergeron, S. Khalifé, H. I. Glazer, and Y. Binik, "Surgical and Behavioral Treatments for Vestibulodynia: Two-and-One-Half-Year Follow-up and Predictors of Outcome," *Obstetrics & Gynecology*, vol. 111, no. 1, pp. 159–166, 2008.

[4] A. T. Goldstein, D. Klingman, K. Christopher, C. Johnson, and S. C. Marinoff, "Surgical Treatment of Vulvar Vestibulitis Syndrome: Outcome Assessment Derived from a Postoperative Questionnaire," *The Journal of Sexual Medicine*, vol. 3, no. 5, pp. 923–931, 2011.

[5] K. Macey, A. Gregory, D. Nunns, and R. das Nair, "Women's experiences of using vaginal trainers (dilators) to treat vaginal penetration difficulties diagnosed as vaginismus: A qualitative interview study," *BMC Women's Health*, vol. 15, no. 1, pp. 1–12, 2015.

[6] R. Melles, M. Dewitte, M. M. ter Kuile, C. Bonnemayer, and M. Peters, "The vaginal pressure inducer: A new device to test the (un)pleasurableness and tolerance of vaginal pressure and the influence of sexual stimuli," *Journal of Sex & Marital Therapy*, vol. 44, no. 2, pp. 189–200, 2018.

[7] Milli-medical, "Milli 263," 2017.

[8] V. Puppo, "Anatomy and physiology of the clitoris, vestibular bulbs, and labia minora with a review of the female orgasm and the prevention of female sexual dysfunction," *Clinical Anatomy*, vol. 26, no. 1, pp. 134–152, 2013.

[9] R. J. Levin, "The Physiology of Sexual Arousal in the Human Female: A Recreational and Procreational Synthesis," *Archives of Sexual Behavior*, vol. 31, no. 5, pp. 405–411, 2002.

[10] S. C. Boyer, "The relationship between pain sensitivity and vasocongestion due to sexual arousal in women with provoked vestibulodynia," Ph.D. dissertation, Queen's University, 2009.

[11] V. Puppo, "Embryology and anatomy of the vulva: the female orgasm and women's sexual health," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 154, no. 1, pp. 3–8, 2011.

[12] J. van der Velde, E. Laan, and W. Everaerd, "Vaginismus, a Component of a General Defensive Reaction. An Investigation of Pelvic Floor Muscle Activity during Exposure to Emotion-Inducing Film Excerpts in Women with and without Vaginismus," *International Urogynecology Journal*, vol. 12, no. 5, pp. 328–331, 2001.

[13] S. Both, R. van Lunsen, P. Weijenborg, and E. Laan, "A New Device for Simultaneous Measurement of Pelvic Floor Muscle Activity and Vaginal Blood Flow: A Test in a Nonclinical Sample," *The Journal of Sexual Medicine*, vol. 9, no. 11, pp. 2888–2902, 2012.

[14] H. Van Lunsen and M. Ramakers, "The hyperactive pelvic floor syndrome," *Acta endoscopica*, vol. 32, no. 3, pp. 275–285, 2002.

[15] M. Brauer, E. Laan, and M. M. ter Kuile, "Sexual Arousal in Women with Superficial Dyspareunia," *Archives of Sexual Behavior*, vol. 35, no. 2, pp. 187–196, 2006.

[16] M. A. Farmer and C. M. Meston, "Predictors of genital pain in young women." *Arch. of sexual behavior*, vol. 36, no. 6, pp. 831–43, 2007.

[17] M. M. ter Kuile, S. Both, and J. van Lankveld, "Cognitive Behavioral Therapy for Sexual Dysfunctions in Women," *Psychiatric Clinics of North America*, vol. 33, no. 3, pp. 595–610, 2010.

[18] S. Knorn, D. Varagnolo, E. Oliver-Chiva, R. Melles, and M. Dewitte, "Data-driven modelling of pelvic floor muscles dynamics," in *under review for IFAC Symposium on Biological and Medical Systems (BMS)*, 2018.

[19] R. Basson, "The Female Sexual Response: A Different Model," *Journal of Sex & Marital Therapy*, vol. 26, no. 1, 2000.

[20] Masters and Johnson, *Human Sexual Response*. Ishi Press, 1966.

[21] D. Varagnolo, S. Knorn, R. Melles, and M. Dewitte, "Qualitative modeling of pressure vs. pain relations in women suffering from dyspareunia," in *Proceedings of the 20th IFAC Worldcongress*, 2017.

[22] V. Bro and A. Medvedev, "Nonlinear dynamics of the human smooth pursuit system in health and disease: Model structure and parameter estimation," in *Conference on Decision and Control (CDC)*, 2017.

[23] R. Langdon, P. Docherty, Y.-S. Chiew, K. Möller, and J. Chase, "Use of basis functions within a non-linear autoregressive model of pulmonary mechanics," *Biomedical Signal Processing and Control*, vol. 27, 2016.

[24] X. Li and K.-C. Wong, "Evolutionary multiobjective clustering and its applications to patient stratification," *IEEE Transactions on Cybernetics*, 2018.

[25] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.