Daily Pain Prediction in Workplace Using Gaussian Processes

Chetanya Puri[1][a], Stijn Keyaerts[2,3][b], Maxwell Szymanski[4,5][c], Lode Godderis[2,3][d], Katrien Verbert[4][e], Stijn Luca[6,*][f] and Bart Vanrumste[1,*][g]

[1]*eMedia Lab and STADIUS, Department of Electrical Engineering (ESAT), KU Leuven, Belgium*

[2]*Knowledge, Information and Research Center (KIR), Group Idewe (External Service for Prevention and Protection at Work), Leuven, Belgium*

[3]*Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium*

[4]*Department of Computer Science, KU Leuven, Belgium*

[5]*Human-computer Interaction research group, Department of Computer Science, KU Leuven, Belgium*

[6]*Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium*

{*chetanya.puri, maxwell.szymanski, lode.godderis, katrien.verbert, bart.vanrumste*}*@kuleuven.be, stijn.keyaerts@idewe.be, stijn.luca@ugent.be*

Keywords: Pain Management, Public Health Informatics, Time Series Forecasting, Bayesian Prediction.

Abstract: Work-related Musculoskeletal disorders (MSDs) account for 60% of sickness-related absences and even permanent inability to work in the Europe. Long term impacts of MSDs include "Pain chronification" which is the transition of temporary pain into persistent pain. Preventive pain management can lower the risk of chronic pain. It is therefore important to appropriately assess pain in advance, which can assist a person in improving their fear of returning to work. In this study, we analysed pain data acquired over time by a smartphone application from a number of participants. We attempt to forecast a person's future pain levels based on his or her prior pain data. Due to the self-reported nature of the data, modelling daily pain is challenging due to the large number of missing values. For pain prediction modelling of a test subject, we employ a subset selection strategy that dynamically selects a closest subset of individuals from the training data. The similarity between the test subject and the training subjects is determined via dynamic time warping-based dissimilarity measure based on the time limited historical data until a given point in time. The pain trends of these selected subset subjects is more similar to that of the individual of interest. Then, we employ a Gaussian processes regression model for modelling the pain. We empirically test our model using a leave-one-subject-out cross validation to attain 20% improvement over state-of-the-art results in early prediction of pain.

1 INTRODUCTION

Musculoskeletal disorders (MSD) are presently a widespread type of work-related health problem and a leading cause for absenteeism from work across all sectors and occupations. Around 60% of all the health related problems in Europe (EU) are work-related MSDs that account for 60% of sickness related absences and even permanent inability to work (Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2017). This creates a financial burden on individuals, businesses, and society (Kok et al., 2020). Prevention of MSDs from the outset of a person's career will allow for an extended work life and better job satisfaction (Kim, 2018). MSD prevention can also address the long-term implications of demographic ageing, as outlined in the objectives of the Europe 2020 strategy for smart, sustainable, and inclusive growth. Consequently, MSDs are not only an occupational burden, but also a public health and societal challenge (Kok et al., 2020).

Long-term impacts of MSDs include "Pain chronification", which is the transformation of tran-

[a] https://orcid.org/0000-0002-3474-4898
[b] https://orcid.org/0000-0001-6431-092X
[c] https://orcid.org/0000-0002-6506-3198
[d] https://orcid.org/0000-0003-4764-8835
[e] https://orcid.org/0000-0001-6699-7710
[f] https://orcid.org/0000-0002-6781-7870
[g] https://orcid.org/0000-0002-9409-935X
[*] These Authors Contributed Equally.

sient pain into permanent pain as a result of recurrent physical strain sustained while doing work-related activities (Morlion et al., 2018). Other than physical pain experience, there is vast amount of evidence on the importance of pain coping strategies, cognitive appraisals (e.g. catastrophizing, high threat values, and fear-avoidance believes), negative emotions and expectations (Moseley and Arntz, 2007), (Nijs et al., 2011), (Edwards et al., 2016). These factors influence how sensory information is processed in the spinal cord and the brain. There are several models that integrate different biopsychosocial factors to the perception of pain such as the fear-avoidance model (Vlaeyen and Linton, 2000), avoidance-endurance model (Hasenbring et al., 2012), and the common sense model (Bunzli et al., 2017). These models illustrate how various persons experience pain, which results in the subjectivity of pain assessments. Thus, predicting pain in a personalised manner for early intervention is essential for preventing pain persistence. It is crucial that both patients and medical practitioners have the education and abilities necessary to manage pain correctly (Morlion et al., 2018).

Currently, pain management is done based on the initial patient evaluation (history, physical examination) which is followed by prompt treatment based on the level of the pain (Morlion et al., 2018). This is especially true for the acute stages of pain. In cases of chronic pain, evidence suggests that therapies should be directed less by current pain levels and more by participation in valued activities despite discomfort (McCracken and Eccleston, 2005). Therefore, appropriately measuring pain in early stages can aid in pain management by evaluating medication efficacy, comprehending the complicated relationship between pain and personal/contextual factors, and preparing patients and healthcare providers for a challenging period with flare-ups. Preventative pain management can also reduce the likelihood of it becoming chronic. Multiple pain management applications exists but are only limited to maintaining logs of the level/intensity of the pain (Lalloo et al., 2015). However, for more successful pain management, it is essential to accurately estimate the pain in advance, preferably several days ahead, which can help a person moderate his or her expectations and anxieties about returning to work. Additionally, pain prediction can provide healthcare practitioners with a better understanding of the required treatment and assist with individualised planning.

This study attempts to forecast the pain experienced by workers from various industries several days ahead based on the daily recorded history of pain. To this end, workers were asked to record their daily levels of satisfaction and pain in a smartphone application on a scale from 0 (no pain) to 100 (worst possible pain). Then, we attempt to forecast future pain levels by modelling individual pain levels recorded until a certain day. This is technically difficult because the data is self-reported, and app users did not always indicate their pain levels on a daily basis, resulting in missing data. Fig. 1 shows how an individual's pain data looks like with respect to time. With the study data of over 300 days, more than 70% of the participants lack 66% of the daily data ($<$ 100 entries).

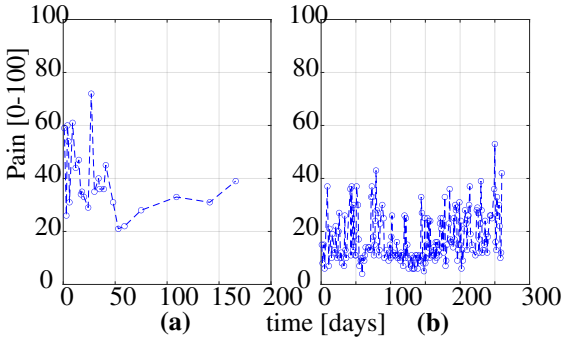

Figure 1: Two distinct users documented their pain levels over a 300-day period using (a) as few as 22 samples and (b) as many as 228 samples.

The lack of observations in an individual's time series data restricts the application of traditional time series modelling techniques such as the autoregressive integrated moving average model (ARIMA), which requires uniformly sampled data (Shumway and Stoffer, 2017).

Authors in (Lötsch and Ultsch, 2018) provide an extensive survey of the application of modern machine learning techniques for the estimation or detection of pain. The majority of the pain experiences discussed in the literature (Lötsch and Ultsch, 2018) are related to a hospital or post-operative scenario, rather than persistent workplace-related pain. Furthermore, pain forecasting models that use machine learning are built on clinical data (e.g. drugs administered, patient comorbidity data) collected during pain experiences post a surgical operation, enriching the information available for modelling (Tighe et al., 2015), (Lee et al., 2019).

Deep learning (DL) is a branch of machine learning that, when given massive volumes of data, may automatically learn representations from raw data to achieve a specific objective, such as classification or regression (Lai et al., 2018), (Laptev et al., 2017), (De Brouwer et al., 2019), (Liu et al., 2021). DL has been used in multiple healthcare related applications that can predict the health of an individ-

ual from the time series data. For example, detecting cardiac abnormality (Strodthoff and Wagner, 2020) or forecasting glucose levels (Li et al., 2019) in individuals. Works such as (Lipton et al., 2016), (Futoma et al., 2017) developed deep learning techniques that can address the non-uniformly sampled time-series data when a large training dataset is available. Traditional machine learning strategies, however, outperform deep learning strategies when training data is insufficient (Makridakis et al., 2018).

In this work, we follow the method proposed in (Puri et al., 2019), where a subset of the training data is selected followed by learning a regression model based on Gaussian processes (GP). Here, we would like to showcase the efficacy of the subset selection approach followed by GP based regression to model an individual's pain measurements until a certain day. The subset selection approach works by first selecting individuals from training data that resemble closely the progression of pain over time to that of the target individual. The number of individuals to be selected is chosen dynamically based on the similarities across individuals. The dynamically chosen subset along with the avaiable data from the target subject is then used to train a regression model for improved prediction performance. However, directly applying the method of (Puri et al., 2019) doesn't give the best results owing to the subjective nature of the pain measurements. Hence, we add a pre-processing treatment of the data prior to subset selection and learning the GP model. We explain the need to do so as follows,

1. Pain measurements of an individual vary a lot across time. This might be because of the pain persistence over time or by the number of individual days with more stress resulting in more pain. Hence, unlike a general increasing trend in gestational weight gain (Puri et al., 2019), it is difficult to find a pattern in the pain measurements over time. Thus, there are anomalous instances in the pain measurements that can result in an inaccurate general model.
2. Pain measurements are self-reported and are highly subjective in nature. This means that individuals have certain biases to only rate their pain (scored between [0-100]) around a fixed baseline, e.g., a person with a baseline *reported* pain of 20 will seldom report a pain of 80. Thus, scaling individual pain measurements for modelling is a necessary step.

The objective of this work is to study if:

- It is possible to estimate an individual's pain levels from a small number of non-uniformly collected historical pain measurements.
- Given the subjective nature of pain data, is it possible to use previous pain measurements of other individuals in a training dataset to enhance pain prediction?

The main contributions of this paper are:

1. We develop models of daily pain data to forecast and manage pain level trends over time.
2. We propose a two-step pre-processing strategy to enhance pain prediction modelling. This is accomplished by smoothing the pain time series in training data and self-normalising the target individual's pain data with the few measurements provided.
3. We use a subset-selection strategy to generate the most informative subset of training data for a given target individual. Individuals in this closest subset exhibit similar pain trends to the individual of interest.
4. We devise modelling based on the selected subset using Gaussian processes for *multi-step* forecasting of pain up to n-days ahead in time.

The dataset is described in section 2, followed by the proposed methodology in section 3. In section 4, we describe the experiments conducted to generate the results. Results and their implications are discussed in greater detail in section 5, and concluding remarks are presented in section 6. Section 7 concludes the paper by discussing potential future directions and constraints.

2 DATA

In this study, 340 participants were recruited from various work sectors. At the start of the study, participants were asked about different work-related factors, their pain complaints, pain-related perceptions, coping strategies, and other contextual factors (indeed, such as physical activity and time spent sitting). From January 2021 to May 2022, they were required to maintain a daily journal in which they recorded their overall pain levels, mood (not with yes/no questions), stress levels, and satisfaction along with baseline questions such as age, gender, height, weight, and industry of employment. The pain levels were recorded on a scale of 0 (best) to 100 (worst) using an mHealth smartphone application.[1]. Yes/No questions such as mood (sad, angry, happy, fatigued, cheerful) were also part of the daily journal.

190 participants were excluded because they did not record daily pain values at all. In addition, 51

[1] https://www.idewe.be/health-empower

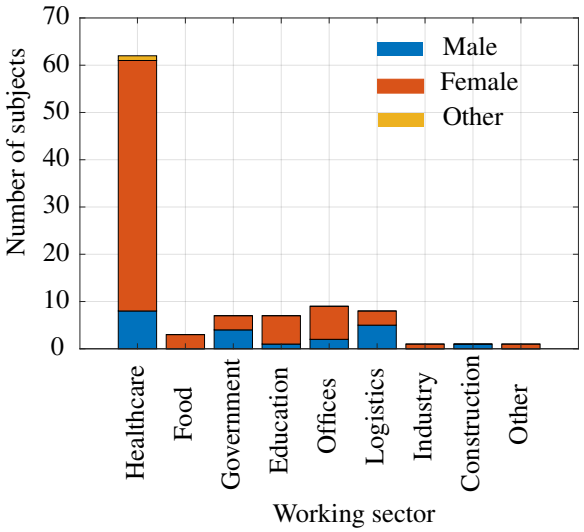

Figure 2: 21 males and 77 females participated in the study with majority (62 out of 99) working in the healthcare industry providing care.

more individuals were removed based on the criterion of not having more than 10 daily pain values recorded, with more than 2 values separated by 1 week. The remaining 99 participants' data were used to develop pain prediction models. Fig. 2 presents the gender-wise distribution of participants in different industries.

This study was conducted within the context of the Personal Health Empowerment project, which focused on investigating and developing new monitoring and treatment options for employees with MSDs. The PHE project and corresponding studies were approved by the Social Ethics Commission of KU Leuven (G-2019081713) and carried out according to the Belgian and international privacy and ethical legislation. The Belgian occupational service for protection and prevention at work (IDEWE) was responsible for the recruitment. They distributed the information about the project amongst their clients and employees. Interested employees had to provide informed consent to participate.

3 METHODOLOGY

Let's assume pain levels measured across time are available for N subjects as 'training data' $\mathcal{D} = \{(\mathbf{x}^1,\mathbf{y}^1),...(\mathbf{x}^N,\mathbf{y}^N)\}$, where $\mathbf{x}^i = [t^i_1 t^i_2 t^i_3 \cdots t^i_m]$ represents the input variable 'time' up to a certain day t^i_m and $\mathbf{y}^i = [y^i_1 y^i_2 y^i_3 \cdots y^i_{m^i}]$ represents the output variable 'pain' for the i^{th} subject, where $y^i_k = y(t^i_k)$.

In addition, data from a person of interest, henceforth referred to as the target individual, are provided till a certain day t^+_d as $\mathcal{S} = \{(t^+_1,y^+_1),(t^+_2,y^+_2),\cdots,(t^+_d,y^+_d)\}$.

We try to learn a mapping f from the training and target data, such that,

$$y^+ = f(t^+) + \varepsilon. \tag{1}$$

where $\varepsilon \sim \mathcal{N}(0,\sigma^2)$ is independent and identically distributed (i.i.d) gaussian.

Using the learnt model f, the target individual's pain measurements are then predicted at time $t^+_{m^i}$ as $y(t^+_{m^i}) = f(t^+_{m^i})$.

3.1 Smoothing

Let's begin by discussing the smoothing operation. Given a time series in training data $\mathbf{y}^i = [y^i_1 y^i_2 y^i_3 \cdots y^i_{m^i}]$, a moving average (MA) of order w can be used to obtain a smoothed time series $\hat{\mathbf{y}}^i = [\hat{y}^i_1 \hat{y}^i_2 \hat{y}^i_3 \cdots \hat{y}^i_{m^i}]$. This w-MA can be written as

$$\hat{y}^i_t = \frac{1}{w} \sum_{j=-\frac{w-1}{2}}^{\frac{w-1}{2}} y^i_{t+j}, \tag{2}$$

where w is an odd integer. Moreover, $\lfloor \frac{w}{2} \rfloor$ zeros are padded to the beginning and end of the given time series $\mathbf{y}^i$ to obtain same m number of observations in the derived w-MA time series in eq. 2. If the w-length time window contains missing observations for a given non-uniformly sampled time series, just the available points are used to calculate the moving average.

3.2 Self-Normalisation

We normalise a given time-series with its available individual information. A time-series $\mathbf{y}^i = [y^i_1 y^i_2 y^i_3 \cdots y^i_{m^i}]$ is normalised using mean $\mu_{\mathbf{y}^i}$ and standard deviation $\sigma_{\mathbf{y}^i}$ calculated as follows:

$$\begin{aligned} \mu_{\mathbf{y}^i} &= \frac{1}{m^i} \sum_{j=1}^{m^i} y^i_j \\ \sigma_{\mathbf{y}^i} &= \sqrt{\frac{1}{m^i} \sum_{j=1}^{m^i} (y^i_j - \mu_{\mathbf{y}^i})^2}. \end{aligned} \tag{3}$$

The j^{th} observation $(\bar{y}^i_j)$ of normalised time-series $\bar{\mathbf{y}}^i = [\bar{y}^i_1 \bar{y}^i_2 \bar{y}^i_3 \cdots \bar{y}^i_{m^i}]$ is obtained from the time-series $\bar{\mathbf{y}}^i$ as

$$\bar{y}^i_j = \frac{y^i_j - \mu_{\mathbf{y}^i}}{\sigma_{\mathbf{y}^i}} \tag{4}$$

The normalised data can be rescaled to original scale as $y^i_j = \bar{y}^i_j \times \sigma_{\mathbf{y}^i} + \mu_{\mathbf{y}^i}$.

3.3 Regression

We use Gaussian Processes (GP) as they are the state-of-the-art time series modelling methods when dealing with missing data. GP is defined as a set of random variables, such that any finite number of them have a joint Gaussian distribution (Rasmussen, 2004). 'f' from eq. (1) is defined as a GP $f(t) \sim \mathcal{GP}(m(t), k(t,t'))$, with mean function $m(t)$ and covariance function $k(t,t')$. We assume the data is noisy with i.i.d gaussian noise, having noise covariance σ_n^2, and choose a squared exponential kernel as the gaussian covariance function to model the closeness of two observations,

$$k(t,t') = \sigma_f^2 exp\left[\frac{-(t-t')^2}{2l^2}\right]. \tag{5}$$

As is evident from eqn. 5, the similarity between two observations decreases exponentially as t begins to differ from t', i.e the similarity is highest when $t = t'$. Thus, when two observations are far apart in time, the kernel considers them more dissimilar than when they are closer together in time.

Given $\acute{\mathbf{y}} = [y_1^1, \cdots, y_m^1, \cdots, y_N^1, \cdots, y_N^N]^T$ and $\mathbf{K}$ as a matrix of entries $K_{p,q} = k(t_p, t_q), \forall t_p, t_q \in \mathcal{D}$. We optimise the hyper-parameters $\{\sigma_f, l, \sigma_n\}$ by maximising the marginal likelihood $p(\acute{\mathbf{y}}|\mathcal{D}; \{\sigma_f, l, \sigma_n\})$ (Rasmussen, 2004). The prediction at time $t_{m^i}^+$ is given as a gaussian distribution whose mean, μ and variance, σ^2 are given by

$$\begin{aligned} \mu(t_{m^i}^+) &= \mathbf{k}_+{}^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\acute{\mathbf{y}} \\ \sigma(t_m^+) &= k(t_m^+, t_m^+) - \mathbf{k}_+{}^T(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{k}_+, \end{aligned} \tag{6}$$

where $\mathbf{k}_+ = \mathbf{k}(t_m^+)$, $\mathbf{k}(t_m^+) = [k(t_m^+, t_1^1), \cdots, k(t_m^+, t_m^N)]^T$.

Gaussian process prediction is hampered by the fact that the computing complexity of inference and likelihood evaluation is $O(n^3)$, where n is the input size, making it impractical for bigger data sets. Next, we will explore subset selection, which can minimise computing complexity while enhancing prediction accuracy.

3.4 Subset Selection

We follow the subset selection approach from (Puri et al., 2019) to find a smaller but informative subset ($\hat{\mathcal{D}}$) of the training data for a given target individual's data. Particularly, a subset $\hat{\mathcal{D}}$ with $M(<<N)$ individuals' data is found from the given training data $\mathcal{D}$,

$$\begin{aligned} \hat{\mathcal{D}} &= \{(\mathbf{x}^1, \mathbf{y}^1), \cdots, (\mathbf{x}^M, \mathbf{y}^M)\} \\ &= \left\{(t_1^1, y_1^1), \cdots, (t_1^M, y_1^M), \cdots, (t_{m^M}^M, y_{m^M}^M)\right\}, \end{aligned} \tag{7}$$

such that the individuals selected in the subset are similar to target individual's pain trend.

Using a subset $\hat{\mathcal{D}}$ with $M(<<N)$ individuals' data gives a computational advantage over considering N subjects, as the time complexity of GPs training and inference is proportional to the cubic power of the number of observations. Furthermore, if the most informative subset is selected, the prediction capability is improved. This is due to the fact that, during training, observations from M patients with a similar trend in pain are close to each other and have less variability at any given time t. Due to inter-subject variances, this variability (at time t) is high when all N individuals are considered for training Gaussian processes.

To find the closeness between two time series, we use the Dynamic Time warping (DTW) as the distance metric. The choice of DTW metric as a distance measure is due to its capability to index time series with unequal lengths (Keogh and Ratanamahatana, 2005).

The subset selection is a two-step process in which (i) distances between the target time series and time series in training data is calculated, and then (ii) the nearest subset is *dynamically* selected based on the calculated distances.

Distances between the target data $\mathcal{S} = \{(t_1^+, y_1^+), (t_2^+, y_2^+)..., (t_d^+, y_d^+)\}$ and individual time series in training data $\mathcal{D}$ are calculated using the dynamic time warping (DTW) distance metric. Let's denote the DTW distance between target time series (denoted by $+$) and i^{th} time series in training data by λ_{i+}. Remark that that target data is only available until t_d^+ but the time-series in training data are present until $t_m^i (>> t_d^+)$. Therefore, the data for time series in training data are considered only until day t_d^+ to calculate the distance λ_{i+}. If the data at t_d^i is not available, the nearest time point $< t_d^+$ is chosen. The distance vector $\mathbf{\Lambda}_+ = [\lambda_{1+}\, \lambda_{2+} \cdots \lambda_{N+}]$ is calculated between target time series and all the time series in training data.

Subset selection is dynamically done based on the distance vector $\mathbf{\Lambda}_+$. First, $\mathbf{\Lambda}_+$ is sorted in ascending order. This ensures that the subjects are arranged in order of their closeness to the target subject, $\hat{\mathbf{\Lambda}}_+ = [\hat{\lambda}_{1+}\, \hat{\lambda}_{2+} \cdots \hat{\lambda}_{N+}]$, such that $\hat{\lambda}_{k+} \leq \hat{\lambda}_{(k+1)+} \forall k = 1, 2, \cdots, N$. Second, *turning* points at index 'k' are calculated, such that,

$$\left(\hat{\lambda}_{(k-1)+} - \hat{\lambda}_{(k-2)+}\right) \leq \left(\hat{\lambda}_{k+} - \hat{\lambda}_{(k-1)+}\right) \geq \left(\hat{\lambda}_{(k+1)+} - \hat{\lambda}_{k+}\right),$$

Multiple such turning points can exist at different indexes in $\hat{\mathbf{\Lambda}}_+$ vector. Third, the value at the distance value at the index k where the first turning point occurs (λ_k) is chosen as the distance threshold to calculate the closest subset. i^{th} time series in $\mathcal{D}$ is selected in the subset if $\lambda_{i+} < \lambda_k$. Choosing the first turning point enables the dynamic selection of the smallest

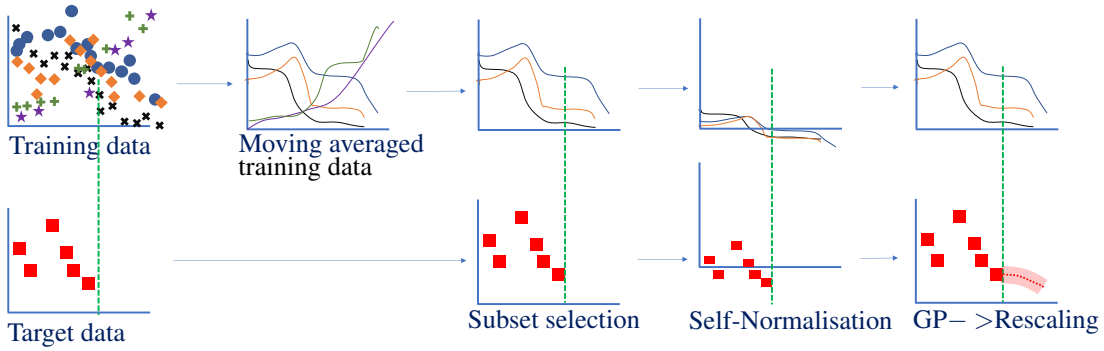

Figure 3: An illustration of our methodology. Moving averaging is performed on the training data to smoothen it. Target data is available until a day t_d^+ (dotted green line). Subset selection is performed on moving-averaged training data that shares similar temporal pattern to the target observations. Each time series (target or training) is self-normalised with its available observations before being fed to Gaussian Processes. A prediction on target data is made (red dotted line).

and most informative subset. Fig. 3 showcases the processing pipeline where the moving average based smoothing is done on the training data before subset selection. Since pain levels in each individual series are normalized using self-data, all training time series are scaled to the same level prior to being fed into the Gaussian process model. We will observe that this enhances the reliability of the predictions.

4 EXPERIMENTS

We perform leave-one-subject-out (LOSO) cross-validation to evaluate the performance of our proposed approach. In each iteration, a unique individual's data is treated as target data and rest of the subjects' data are the training data. We first smoothen the training data and target data using a moving averaging of order five ($w = 5$). Then, in each iteration, a closest subset is evaluated dynamically with respect to the target data followed by self-normalising each time series (target data and selected subset) using eq. 3 and 4. Note that our subset selection approach dynamically selects a threshold in each iteration (i.e for each target data). A GP based regression is performed to forecast the future values for the target subject. The performance of regression was computed using Mean Absolute Error (MAE) averaged over N subjects. MAE for prediction at a time t_h is given as $MAE(t_h) = \frac{1}{N}\sum_{i=1}^{N} |y^{pred}(t_h^i) - y^{orig}(t_h^i)|$.

4.1 State-of-the-Art

- **Baseline:** A baseline was created to judge the performance of the algorithms. This baseline was created by using the last available value of the target subject as future prediction of the daily pain value.
- **ARIMA:** Auto-Regressive Integrated Moving Average (ARIMA) has remained a state-of-the-art time series forecasting approach with uniformly spaced samples of time series (Box et al., 2015). Through linear interpolation, uniformity was introduced into the sparsely sampled pain time series of the subject of interest. Then, an ARIMA(p,d,q) model was fit on the uniformly sampled target time series. In order to find the optimal autoregressive order (p), degree of differencing (d), and moving average order (q), a grid-search was performed to find the optimal hyperparameters following (Shibata, 1976). The learned model is then used to make a multi-step-ahead prediction of pain levels using the optimized hyperparameters.
- **LSTM:** Long short-Term Memory networks (LSTM) are deep learning techniques that can produce exceptional prediction performance by implementing gates (forget, memory, and output) that regulate the flow of information during training (Hochreiter and Schmidhuber, 1997). We follow a similar approach as with ARIMA approach where the avaiable data from a target subject is uniformly sampled by linear interpolation. We evaluate an LSTM network with 10 hidden units and the training is done using ADAM's optimisation to minimise mean absolute error (Kingma and Ba, 2015).
- **Maximum-a-Posteriori (MAP) Estimation:** A l^{th} order polynomial can be fit using available target data to estimate the polynomial coefficients $\theta_i, \forall i \in \{1,2,\cdots,l\}$ (Puri et al., 2019). Moreover, subjects from training data can be used to create priors over the polynomial coefficients to get a better estimate known as *maximum-a-posterior* (MAP) estimate (Puri et al., 2019). We

test with polynomial of different orders (order 1 to 5) to find that the first order polynomial produces the least mean absolute error in LOSO cross-validation.

5 RESULTS & DISCUSSION

In this research, we investigate whether it is possible to estimate a person's pain levels using a small number of non-uniformly collected historical pain measurements. As pain data is subjective and varies amongst individuals, we also intended to determine if we might improve pain prediction by incorporating the prior pain measurements of other individuals into the training dataset. For this reason, we study the performance of various algorithms presented in this paper when predicting pain levels of an individual in future. In Fig. 4, we present the Mean Absoute Error (MAE) when predicting the pain 7 days ahead on the y-axis. On the x-axis in Fig.4, the availability of target data until a certain day is presented. Subset selection (SS) along with moving averaging (MA) and/or self-Normalisation (SN) were performed and Gaussian processes was used as a regression model. It is evident from Fig. 4 that the performance of subset selection (SS) followed by Gaussian processes (GP) is demonstrably superior to that of Gaussian processes alone. This is a result of the inclusion of an informative subset of participants in training who exhibit a comparable trend in pain to the target data. Additionally, subset selection on the moving averaged (MA) time series of the training data, followed by self-normalisation and subsequently the Gaussian process, performed the best, particularly when predicting for less available target data.

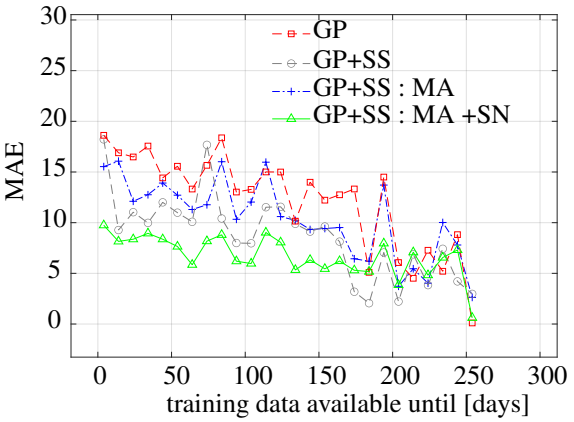

Figure 4: Mean absolute error (MAE) is measured with respect to availability of target data. Different combinations of subset selection (SS) followed by Gaussian processes (GP) were performed with proposed pre-processing components such as moving averaging (MA) and/or self-Normalisation (SN).

We hypothesised that pain data is subjective and that self-reported pain measurements are biased because individuals can only compare their current pain feelings to their past pain experiences. Therefore, self-normalisation with respect to the historical pain measurements of an individual provides this significant performance improvement. In addition, as the availability of personal pain data increases over time, so does the accuracy of prediction. We believe that as more training data becomes available from an individual, the selected subset will consist of subjects whose patterns resemble that of the target subject more closely than when there are only a few data points. Thus, the variance in the training data available for regression is less and thus the prediction improves. This is evident by the decreasing trend in MAE when more training data becomes available. We also tested with self-normalization prior to subset selection and found no significant performance differences. This may be due to the fact that the DTW distance comparison for subset selection compares the relative difference in distances between two time series and picks more or less similar individuals with or without self-normalisation.

Next, we present the comparison of the proposed approach (GP+SS:MA+SN) with state-of-the-art approaches presented in section 4.1 when predicting pain values $[0-100]$. Fig. 5 shows that the proposed approach's performance is best when it comes to early prediction using only few available data points (until day 100).

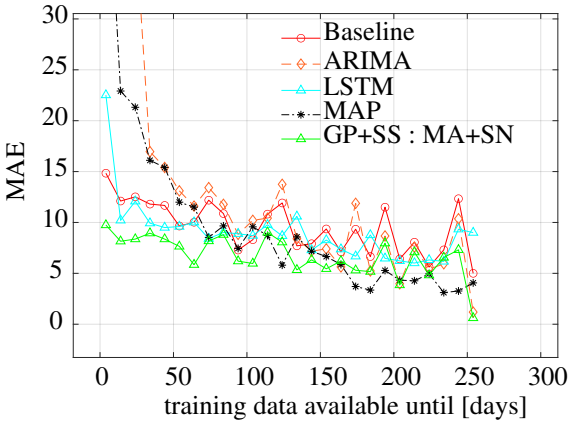

Figure 5: Comparison of the proposed approach with state-of-the-art approaches. When little training data is available (until day 100), the proposed method beats SOTA, and when more training data becomes available, it performs comparably or even better.

The performance becomes comparable (if not better) with the state-of-the-art approaches (MAP) as

more data in time becomes available for a given individual. On the basis of a paired t-test with equal variances, the performance differences between the proposed approach and other SOTA methods are statistically significant at 5% level of significance (until day 50). We discovered no statistically significant difference between the proposed method and MAP-based polynomial estimate when training with data for more than 100 days. Given the simplicity of the dataset, it seems intuitive that when more pain data becomes available, simple polynomial-based estimating algorithms will perform better.

We also observed that the state-of-the-art approaches (except LSTM) perform worse than the baseline when the availability of individual training data is limited (at least until day 50). Remark that the baseline is simply the previous observed value of pain carried forward for the prediction of future values. This is due to the difficulty of modelling sparsely sampled time series with few observations. Our proposed method, on the other hand, overcomes this difficulty by incorporating the subjective nature of pain experience and modelling information rich subset selection along with personal data.

6 CONCLUSION

We proposed a novel Gaussian processes estimator and information-rich preprocessing to model an individual's workplace-related pain experiences. When time series data is irregularly sampled, the proposed approach outperforms state-of-the-art time-series forecasting algorithms for early prediction. This can aid in the development of interventions for managing pain in the workplace, thereby reducing the possibility of 'pain chronification'.

7 LIMITATIONS & FUTURE WORK

A limitation of our approach is the scalability of Gaussian processes as we believe that considering a large number of subjects ($N > 10^4$) will result in a larger subset (high value of M) of training data, increasing the computational complexity of our method. Sparse GPs are model approximation techniques that, when applied to a large number of subjects, can further reduce complexity (Rasmussen, 2004).

In the future, we hope to broaden the modality of the input data in order to obtain more objective feedback on pain experiences. Finding an association of pain with physical activity data measured by a wearable, for example, can help as another meaningful feature to improve prediction performance. Similar to the maximum-a-posteriori approach, priors on the normalisation constants can be generated from training data and used to adjust the self-normalisation mean and standard deviation.

ACKNOWLEDGMENT

Chetanya Puri has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 766139. This work is part of the research project Personal Health Empowerment (PHE) with project number HBC.2018.2012, financed by Flanders Innovation & Entrepreneurship. This publication reflects only the authors' view and the REA is not responsible for any use that may be made of the information it contains.

REFERENCES

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bunzli, S., Smith, A., Schütze, R., Lin, I., and O'Sullivan, P. (2017). Making sense of low back pain and pain-related fear. *journal of orthopaedic & sports physical therapy*, 47(9):628–636.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (2017). *Safer and Healthier Work for All — Modernisation of the EU Occupational Safety and Health Legislation and Policy*. COM.

De Brouwer, E., Simm, J., Arany, A., and Moreau, Y. (2019). GRU-ODE-bayes: Continuous modeling of sporadically-observed time series. In *Advances in Neural Information Processing Systems*, pages 7379–7390.

Edwards, R. R., Dworkin, R. H., Sullivan, M. D., Turk, D. C., and Wasan, A. D. (2016). The role of psychosocial processes in the development and maintenance of chronic pain. *The Journal of Pain*, 17(9):T70–T92.

Futoma, J., Hariharan, S., and Heller, K. (2017). Learning to detect sepsis with a multitask gaussian process rnn classifier. In *International conference on machine learning*, pages 1174–1182. PMLR.

Hasenbring, M. I., Hallner, D., Klasen, B., Streitlein-Böhme, I., Willburger, R., and Rusche, H. (2012). Pain-related avoidance versus endurance in primary care patients with subacute back pain: psychological characteristics and outcome at a 6-month follow-up. *Pain*, 153(1):211–217.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Keogh, E. and Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3):358–386.

Kim, D. (2018). Effect of musculoskeletal pain of care workers on job satisfaction. *Journal of Physical Therapy Science*, 30:164–168.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. *ICLR*.

Kok, J., Vroonhof, P., Snijders, J., Roullis, G., Clarke, M., Peereboom, K., Dorst, P., Isusi, I., and European Agency for Safety and Health at Work (2020). *Work-related musculoskeletal disorders : prevalence, costs and demographics in the EU*. Publications Office.

Lai, G., Chang, W.-C., Yang, Y., and Liu, H. (2018). Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 95–104.

Lalloo, C., Jibb, L. A., Rivera, J., Agarwal, A., and Stinson, J. N. (2015). There's a pain app for that. *The Clinical journal of pain*, 31(6):557–563.

Laptev, N., Yosinski, J., Li, L. E., and Smyl, S. (2017). Time-series extreme event forecasting with neural networks at uber. In *International Conference on Machine Learning*, volume 34, pages 1–5.

Lee, J., Mawla, I., Kim, J., Loggia, M. L., Ortiz, A., Jung, C., Chan, S.-T., Gerber, J., Schmithorst, V. J., Edwards, R. R., et al. (2019). Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. *Pain*, 160(3):550.

Li, K., Liu, C., Zhu, T., Herrero, P., and Georgiou, P. (2019). GluNet: A deep learning framework for accurate glucose forecasting. *IEEE journal of biomedical and health informatics*, 24(2):414–423.

Lipton, Z. C., Kale, D., and Wetzel, R. (2016). Directly modeling missing data in sequences with RNNs: Improved classification of clinical time series. In *Machine Learning for Healthcare Conference*, pages 253–270.

Liu, M., Zeng, A., Xu, Z., Lai, Q., and Xu, Q. (2021). Time series is a special sequence: Forecasting with sample convolution and interaction. *arXiv preprint arXiv:2106.09305*.

Lötsch, J. and Ultsch, A. (2018). Machine learning in pain research. *Pain*, 159(4):623.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3):e0194889.

McCracken, L. M. and Eccleston, C. (2005). A prospective study of acceptance of pain and patient functioning with chronic pain. *Pain*, 118(1-2):164–169.

Morlion, B., Coluzzi, F., Aldington, D., Kocot-Kepska, M., Pergolizzi, J., Mangas, A. C., Ahlbeck, K., and Kalso, E. (2018). Pain chronification: what should a non-pain medicine specialist know? *Current Medical Research and Opinion*, 34(7):1169–1178.

Moseley, G. L. and Arntz, A. (2007). The context of a noxious stimulus affects the pain it evokes. *PAIN®*, 133(1-3):64–71.

Nijs, J., Van Wilgen, C. P., Van Oosterwijck, J., van Ittersum, M., and Meeus, M. (2011). How to explain central sensitization to patients with 'unexplained'chronic musculoskeletal pain: practice guidelines. *Manual therapy*, 16(5):413–418.

Puri, C., Kooijman, G., Masculo, F., Van Sambeek, S., Den Boer, S., Luca, S., and Vanrumste, B. (2019). PREgDICT: Early prediction of gestational weight gain for pregnancy care. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4274–4278. IEEE.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer.

Shibata, R. (1976). Selection of the order of an autoregressive model by akaike's information criterion. *Biometrika*, 63(1):117–126.

Shumway, R. H. and Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.

Strodthoff, N. and Wagner, P. e. (2020). Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1519–1528.

Tighe, P. J., Harle, C. A., Hurley, R. W., Aytug, H., Boezaart, A. P., and Fillingim, R. B. (2015). Teaching a machine to feel postoperative pain: combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain. *Pain Medicine*, 16(7):1386–1401.

Vlaeyen, J. W. and Linton, S. J. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85(3):317–332.