

# Demo Abstract: Privacy preserving Pregnancy weight gain management

Chetanya Puri<sup>1</sup>, Koustabh Dolui<sup>2</sup>, Gerben Kooijman<sup>3</sup>, Felipe Masculo<sup>3</sup>, Shannon Van Sambeek<sup>3</sup>, Sebastiaan Den Boer<sup>3</sup>, Sam Michiels<sup>2</sup>, Hans Hallez<sup>2</sup>, Stijn Luca<sup>4\*</sup>, Bart Vanrumste<sup>1\*</sup>

<sup>1</sup>e-Media Lab, KU Leuven

<sup>2</sup>imec-DistriNet, KU Leuven

<sup>3</sup>Philips Research, Eindhoven

<sup>4</sup>Department of Data Analysis and Mathematical Modelling, Ghent University

<chetanya.puri,hans.hallez,bart.vanrumste>@kuleuven.be,<koustabh.dolui,sam.michiels>@cs.kuleuven.be

<gerben.kooijman,felipe.masculo,shannon.van.sambeek,sebastian.den.boer>@philips.com

stijn.luca@ugent.be

## ABSTRACT

Early gestational weight gain prediction can help expecting women overcome several associated risks. However, training the model requires access to centrally stored privacy sensitive weight and other meta-data. In this demo, we present a privacy preserving federated learning approach where we train a global weight gain prediction model by aggregating client models trained locally on their personal data. We showcase a software data-exploration tool that exhibits local model generation, sharing and updating across users and server for proposed collaborative learning. Our proposed model predicts the final weight category with 61.3% accuracy on day 140, with a 8.8% compromise on the centralized training accuracy.

### ACM Reference Format:

Chetanya Puri<sup>1</sup>, Koustabh Dolui<sup>2</sup>, Gerben Kooijman<sup>3</sup>, Felipe Masculo<sup>3</sup>, Shannon Van Sambeek<sup>3</sup>, Sebastiaan Den Boer<sup>3</sup>, Sam Michiels<sup>2</sup>, Hans Hallez<sup>2</sup>, Stijn Luca<sup>4\*</sup>, Bart Vanrumste<sup>1</sup>. 2019. Demo Abstract: Privacy preserving Pregnancy weight gain management. In *The 17th ACM Conference on Embedded Networked Sensor Systems (SenSys '19)*, November 10–13, 2019, New York, NY, USA. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3356250.3361941>

## 1 INTRODUCTION

Gestational weight gain is a key research topic in the area of health informatics affecting about 70% of pregnant women who end up either gaining too much or too less weight during their pregnancy. Early prediction of gestational weight gain can mitigate this problem by estimating the end-of-pregnancy weight gain given reliable weight measurements taken in uniform intervals are available. In [Puri et al. 2019], the proposed approach predicts the end-of-pregnancy weight category among underweight, normal and obese with an accuracy 63.75% midway through pregnancy and a mean absolute error of 2.572 kgs.

The traditional implementation of the aforementioned approaches would require interaction among three parties, one to collect the data from users, second to pre-process the data while the third to integrate the data and build the model. However, such sensitive

personal data sharing with the service providers is risky as well as uncomfortable to the users. Even though the model learned on this data improves the experience of users, storing and processing the data centrally on a server leads to privacy concerns especially in the light of the General Data Protection Regulation (GDPR) imposed by the European Union (EU).

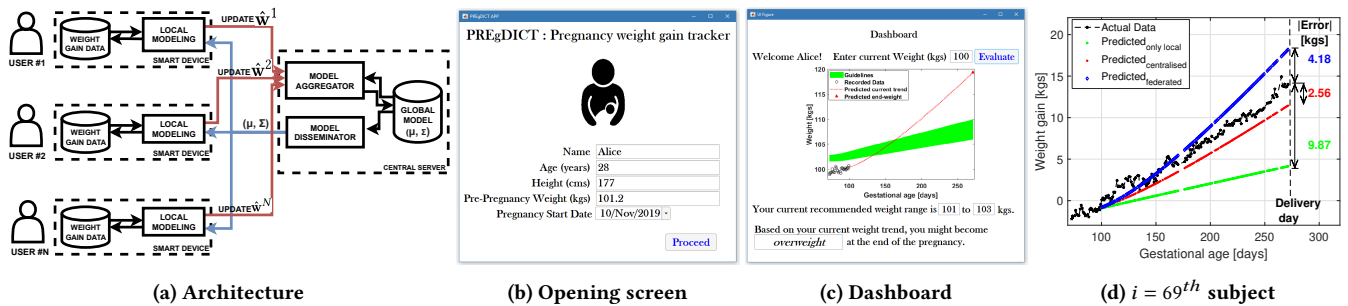
Mobile devices have become ubiquitous and resourceful in terms of available storage and processing capacity. Leveraging this for on-device learning, Google proposed federated learning [McMahan et al. 2016] where devices only share updates on models learned from their local data. This methodology decouples the need to access the training data directly on-device for training a central model. Federated learning is particularly applicable to use-cases where the data is collected from user devices and the data is sensitive in nature. To this end, we have designed and implemented a privacy-preserving federated approach for the prediction of gestational weight gain. The key contributions of this demo is to illustrate: (i) a software tool that generates local models and a server that aggregates these updates (ii) the comparison of accuracy in early prediction of gestational weight gain in the centralized and federated approaches and (iii) the minimization of data sharing between the user-devices and the server.

## 2 METHODOLOGY

There are  $N$  subjects (80 in this study) that are going to participate in the collaborative learning of gestational-weight-gain-prediction model denoted by  $i = 1, 2, 3, \dots, N$ . Each subject by means of self-reporting tools collects and stores its weight-gain measurements at a local device. We denote the the input gestational days up to delivery day  $t_{m_i}^i$  as  $\mathbf{x}^i = \{t_1^i, t_2^i, t_3^i, \dots, t_{m_i}^i\}$ .  $\mathbf{y}^i = \{y_1^i, y_2^i, y_3^i, \dots, y_{m_i}^i\}$  represents the output weight gain for  $i^{th}$  subject, where  $y_k^i = y(t_k^i) \in \mathbb{R}$ . Additionally, pre-pregnancy weight and height are also collected. The objective is to try to collaboratively learn function  $f$  from given data, such that  $y^+ = f(t^+) + \epsilon$ . We can fit a  $p^{th}$ -order polynomial with  $f = w_0 + w_1 t + w_2 t^2 + \dots + w_p t^p$  and estimate the coefficients  $\mathbf{w} = [w_0, w_1, \dots, w_p]^T$  by maximizing the likelihood ( $\mathcal{L}$ ) over individual's personal-training data  $\mathcal{D}$ ,  $\mathcal{L}(\mathbf{w}) = P(\mathcal{D}|\mathbf{w})$ ,  $\hat{\mathbf{w}}_{MLE} = \operatorname{argmax}_{\mathbf{w}} P(\mathcal{D}|\mathbf{w}) = \prod_{i=1}^d p(y_i^+ | t_i^+; \mathbf{w})$ . This equation refers to the model learnt from the individual's sparse limited observations upto given  $t_d$  days.

\*Both authors contributed equally to this research.

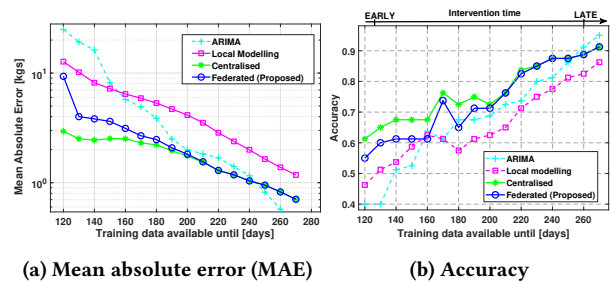
This is author's pre-print. Final version is published in SenSys 2019 and is available at the ACM website.



**Figure 1:** (a) Architecture for our federated approach (b) Opening screen for obtaining meta-data information (c) Dashboard to show the predicted end-of-pregnancy weight gain locally using collaborative model (d) Prediction using federated model is less erroneous than the local model using training weight gain data up to 100 days.

The federated learning process (Fig. 1a) that we utilised is as follows. a) The centralized server sends the meta-data, (for example, order  $p$  of the polynomial, current global model estimate) to the participating subjects, once all subjects agree upon it, b) the local subjects estimate model coefficients  $\hat{\mathbf{w}}_{MLE}$  based on maximising the likelihood of the local data, c) these local model updates are then shared to the server and aggregated there to create global model, d) the global model is shared with the participating subjects. This process is repeated as new subjects participate or the already participating subjects gather more data to push updated local models to the server.

We use a bayesian approach to aggregate these local model updates. We assume  $\mathbf{w} \sim \mathcal{N}(\mu_{\hat{\mathbf{w}}}, \Sigma_{\hat{\mathbf{w}}})$ , such that  $\mu_{\hat{\mathbf{w}}} = \text{mean}([\hat{\mathbf{w}}^1, \hat{\mathbf{w}}^2, \dots, \hat{\mathbf{w}}^N]^T)$  is the mean and  $\Sigma_{\hat{\mathbf{w}}} = \text{cov}([\hat{\mathbf{w}}^1, \hat{\mathbf{w}}^2, \dots, \hat{\mathbf{w}}^N]^T)$  is the covariance matrix of the polynomial coefficients  $\hat{\mathbf{w}}^1, \hat{\mathbf{w}}^2, \dots, \hat{\mathbf{w}}^N$  that are each obtained using the individual gestational weight gain data from each of the  $N$  participating subjects. The distribution over the MLE estimates of the coefficients,  $p(\mathbf{w})$  is acquired from the  $N$  subjects as an *a-priori* estimate. The likelihood learnt from the personal-training data and the *a-priori* distribution learnt from the global updates are then combined to calculate the maximum-a-posteriori (MAP) estimate of the coefficients  $p(\mathbf{w}|\mathcal{D})$ ,  $\hat{\mathbf{w}}_{MAP} = \text{argmax}_{\mathbf{w}} p(\mathbf{w}|\mathcal{D}) \propto \text{argmax}_{\mathbf{w}} P(\mathcal{D}|\mathbf{w})p(\mathbf{w})$ . Fig. 1d shows that prediction using centralised, local and federated approach. Centralised approach performs the best as it trains a global model on the complete pregnancy data stored at the server from the users that subscribed in the past. However, in our federated scenario, we assume that no prior-model is present for the current participating subjects and all the available local models being pushed to the server are from women currently experiencing pregnancies.



**Figure 2:** As expected, (a) MAE reduces and (b) Accuracy increases as more data becomes available locally for training as the pregnancy advances.

### 3 DEMONSTRATION

**Interface.** We realized this demo in MATLAB, where each subject has their own wizard (Fig. 1b) for local data storage and meta-data entry. On clicking ‘proceed’ the user is taken to a dashboard (Fig. 1c) where they can see the weight-gain progression trend.

**Dataset.** We evaluated our model on 80 subjects from European population by collecting their weight measurements throughout the pregnancy using wi-fi connected weighing scale.

**Performance.** Fig. 2 shows the mean absolute error (MAE) and the 3-class classification accuracy for end-of-pregnancy prediction. Our federated approach learns better than the state-of-the-art ARIMA approach [Box et al. 2015] and performs closely with the centralised approach in early gestational weight prediction.

**Objective.** We will allow users to participate in our demo as any of the above 80 subjects during various stages of pregnancies. Our primary objectives are two-fold, (i) realization of the privacy-preserving model sharing between the users and the centralized server instead of user weight-gain and (ii) to demonstrate the improvements in user-experience in gestational weight gain management with our early prediction model. We want to engage the conference audience into discussions regarding their experiences and know-how on gestational weight gain management to further improve our solution. In future work, we aim to reduce the communication overhead and memory footprint of our model.

## REFERENCES

- George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
- H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. 2016. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint arXiv:1602.05629* (2016).
- Chetanya Puri, Gerben Kooijman, Felipe Mascuro, Shannon Van Sambeek, Sebastiaan Den Boer, Stijn Luca, and Bart Vanrumste. 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)*. IEEE,

(in print).

## ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 766139. This article reflects only the author's view and the REA is not responsible for any use that may be made of the information it contains.