Machine-learning prediction of early postpartum prediabetes in women with gestational diabetes mellitus

Durga Parkhi
Advisor: Prof Ponnusamy Saravanan

University of Warwick
Novo Nordisk

25/11/2022

The Global Health Network Conference 2022, University of Cape Town, South Africa
Gestational Diabetes Mellitus is “Glucose Intolerance first diagnosed during pregnancy”.

- ~90% GDM found in LMICs
- GDM women have 10-12 times higher risk of T2D
- Follow-up rates very low (~60% even in high-income countries)
Background

Vision

To improve quality of life of two individuals – the mother and her baby – in one go, & work towards prevention of inter-generational propagation of diabetes

Objective

Antenatal prediction of postpartum prediabetes in GDM women using advanced ML Machine Learning
Methods

Part A. Prediction model development
• 394 samples for 21 features
• Algorithm: Logistic Regression and compare with tree-based algorithms
• Model evaluation metric: area under the ROC curve

Part B. Optimal cut-off selection to prioritize high-risk women depending upon resource availability
• Kullback-Leibler Divergence theory and Information graphs
**Methods**

**Step 1:** Divide full data into n-1 training and 1 testing

**Step 2:** Feature selection using Lasso shrinkage hyperparameter optimization

**Step 3:** Model training using Logistic regression

**Step 4:** Model evaluation (aggregated test predictions)
Methods

The key ideas are:

1. **Identify features** with the potential of prediction of GDM out of a pool of all possible collected features

2. **Create new predictive features** from existing ones

3. Use selected predictors to **build a prediction model** using ML algorithms and a well-designed **model architecture**

4. **Represent the model mathematically** in the form of a CRS

5. **Study the optimal thresholds** for classifying women based on their individual risk

6. Convert all this into a **simple software tool for practical use**
Process and Challenges

1. Small Data size - Only 394 (64.91%) out of 607 had postpartum GTT data available
   • Data augmentation using synthetically generated data

2. Ethics - Data privacy
   • Data replacement using synthetically generated data

3. Data Incompleteness - Failure to achieve 100% follow-up & Missing data
   • Targeted follow-up

4. Data Imbalance - Only 92 (23.35%) out of 394 women had prediabetes
Results and Conclusions

Postpartum Prediabetes Prediction

<table>
<thead>
<tr>
<th>Antenatal Fasting Glucose (mmol/L)</th>
<th>Postpartum Prediabetes Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0944903689501594</td>
</tr>
<tr>
<td>Antenatal HbA1c (mmol/mol)</td>
<td>Low Risk</td>
</tr>
<tr>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Antenatal Fasting Glucose (mmol/L)</th>
<th>Postpartum Prediabetes Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8</td>
<td>0.2697285955868185</td>
</tr>
<tr>
<td>Antenatal HbA1c (mmol/mol)</td>
<td>High Risk</td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgements

Prof Ponnusamy Saravanan
Prof Leelavati Narlikar
Prof Rahul Siddharthan
Dr Uma Ram
Dr Yonas Weldeselassie
Dr Nithya Sukumar
Shivashri Chockalingam
Swetha Sampathkumar
Nishanthi Periyathambi