

# Machine Learning Approaches for Prediction of Preterm Birth

Iliia Vovsha, Ashwath Rajan, Ansaif Salleb-Aouissi (Principal Investigator),

Axinia Radeva, Hatim Diab, Ashish Tomar

<http://www1.ccls.columbia.edu/ansaf/CING/>

Columbia University Center for Computational Learning Systems

Center for Health Analytics  
INSTITUTE FOR DATA SCIENCES AND ENGINEERING

Institute for Data Sciences and Engineering  
COLUMBIA UNIVERSITY

Ronald Wapner

Department of Obstetrics and gynecology  
Columbia University Medical Center

Anita Raja

Department of Software and Information Systems  
University of North Carolina Charlotte

Mary McCord

Medical College  
Children's Hospital of Wisconsin

Tara Randis

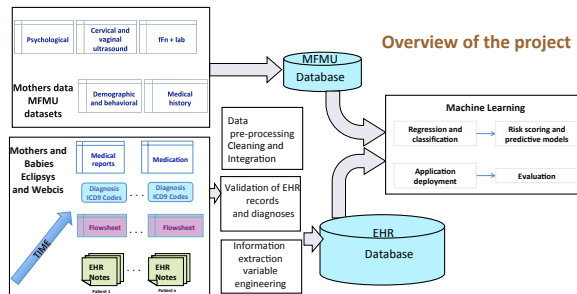
Department of Neonatology  
Columbia University Medical Center

## The Problem of Preterm Birth (PTB)

- Birth of a baby before 37 completed weeks of gestation
- Long-lasting public health problem with heavy emotional and financial consequences to families and society
- Over 26 billion dollars are spent annually on the delivery and care of the 12-13% of infants born preterm in the US.
- Previous research: Focused on individual risk factors
- Goal: Develop a novel prediction system that combines well-known risk factors using machine learning on large scale and high-dimensional data

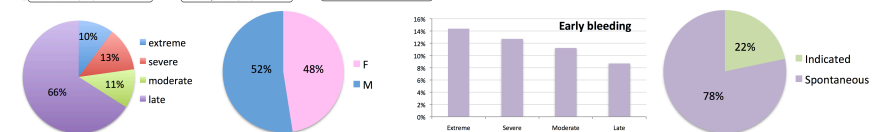
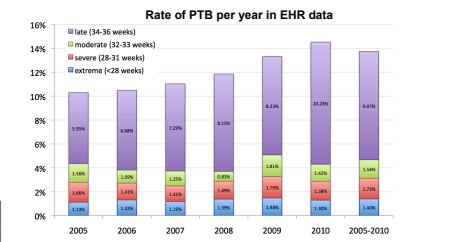
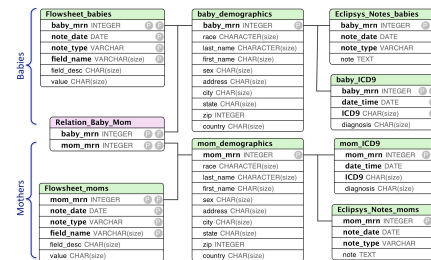


Picture of a 23 weeks preemie in an incubator (source: March of Dimes)



## Electronic Health Records (EHRs)

A 5-year snapshot of the EHR data from the New York Presbyterian Hospital Period from 01/2005 to 10/2011. Population: 43K mothers and 35K babies. 5,113 premature babies identified with ICD9 codes.

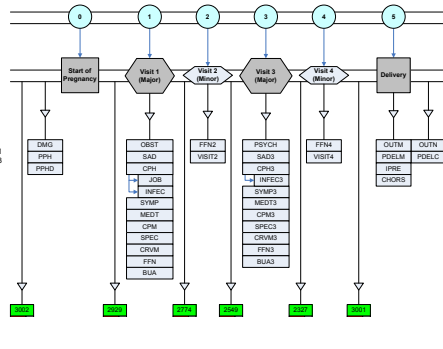


From Left to Right: Rate of PTB in each group; Gender distribution; Early bleeding rates in each PTB group; Rate of indicated versus spontaneous PTB in EHR data.

## Preterm Prediction Study data Maternal-Fetal Medicine Units Network

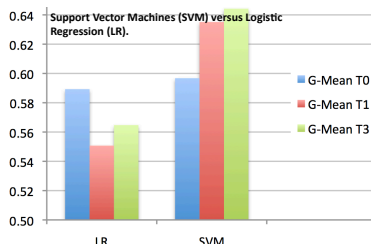
Observational prospective study Performed by NICHD. 2,929 of participating women were followed at 24, 26, 28 and 30 weeks gestation. Detailed clinical information, ultrasound measurements and biochemical samples were obtained.

- DMG: Demographics and Home Life
- PPH: Previous Pregnancy History
- PHD: Previous Pregnancy History Detail
- OBST: Obstetrical & Medical Complications
- SAD: Substance Use V1
- BAU3: Blood & Urine Analysis V3
- CPH: Current Pregnancy History V1
- CPH3: Current Pregnancy History V3
- CPH4: Current or Last Job
- JOB: Infections During This Pregnancy V1
- INFE3: Infections During This Pregnancy V3
- MEDT: Medications and Treatments V1
- MEDT3: Medications and Treatments V3
- SYMP: Symptoms During Previous Week V1
- SYMP3: Symptoms During Previous Week V3
- CPM: Current Pregnancy Measurements V1
- CPM3: Current Pregnancy Measurements V3
- SPEC: Specimen Collection V1
- SPEC3: Specimen Collection V3
- CPM4: Cervical Measurements V1
- CPM5: Cervical Measurements V3
- FFN: Fetal Fibronectin Analysis V1
- FFN2: Fetal Fibronectin Analysis V2
- FFN3: Fetal Fibronectin Analysis V3
- FFN4: Fetal Fibronectin Analysis V4
- BLA: Blood & Urine Analysis V1
- BLA2: Blood & Urine Analysis V2
- BLA3: Blood & Urine Analysis V3
- BLA4: Blood & Urine Analysis V4
- PSYCH: Psychological Questionnaire
- VEI2: Yeast & Intercourse Variables V2
- VEI4: Yeast & Intercourse Variables V4
- OUTM: Pregnancy Outcome, Maternal Data
- OUTM3: Pregnancy Outcome, Neonatal Data
- OUTS: Pregnancy Outcome Status
- INDPT: Indicated Preterm Birth Reasons
- CHROM: Chromosomalis Suspected
- PDEL3M: Preterm Delivery, Maternal Data
- PDEL3C: Preterm Delivery, Clinical Data



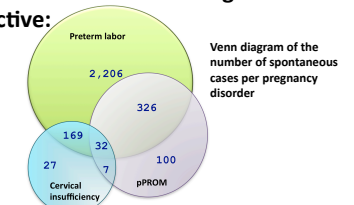
Outcome	N	%
Spontaneous PTB <32 weeks	50	2
Spont PTD <35 weeks	129	4
Spont PTD <37 weeks	309	10
Indicated PTD <37 weeks	124	4
Fetal growth retardation	163	5
Low birth weight	361	12

Our results demonstrate the superiority of non-linear methods in predicting preterm birth. Besides the fact that no time-dependent prediction was ever used on the MFMU dataset, We obtained an average of sensitivity and specificity in predicting PTB of 56% and 68% respectively, well above the 21% for sensitivity and 30% for specificity reported in the literature on this data.



## Summary & Future Work

- 1- Prediction of Preterm Birth is not elusive if we have:
  - Enough quality data to learn from;
  - The effort is multidisciplinary;
  - Do not dilute! focus on spontaneous PTB and nulliparous (the hard cases).
- 2- EHR is a rich information source, but the ability to harness it is forthcoming...
- 3- Exciting application from a machine learning perspective:
  - Multiple class learning;
  - Learning using privileged information;
  - Learning in presence of missing features;
  - Interpretability is also important.



## Acknowledgments

Our warmest thanks go the Office of Vice President of Research for supporting this work through a Research Initiative in Science Engineering (RISE) Grant (2011-2013).

## References

Clinical Informatics Group (2013). Data pre-processing for the preterm prediction study dataset. NICHD (2012). Scientific Vision: The Next Decade (13-7940). Washington, DC.

Template Credit: <http://www.writing.engr.psu.edu/posters.html>