Machine Learning Approaches for Prediction of Preterm Birth
Ilia Vovsha, Ashwath Rajan, Ansaf Salleb-Aouissi (Principal Investigator), Axinia Radeva, Hatim Diab, Ashish Tomar
http://www1.ccls.columbia.edu/ansaf/CING/
Columbia University Center for Computational Learning Systems

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

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.

Electronic Health Records (EHRs)

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.