Disease diagnoses can be represented as relationship graphs between variables observed from the patient. i.e., Presence of a certain graph indicates a disease (state of the individual).

These graphs can be discovered through Data Mining.

Using kernel-based distances, early predictive signals can be derived from sub-graphs of the diagnostic graph.

Machine Learning can enhance discrimination between diagnoses.
Disease diagnoses can be represented as relationship graphs between variables observed from the patient. i.e., Presence of a certain graph indicates a disease (state of the individual).

These graphs can be discovered through Data Mining.

Using kernel-based distances, early predictive signals can be derived from sub-graphs of the diagnostic graph.

Machine Learning can enhance discrimination between diagnoses.

BD f LB example.
Sepsis Diagnosis

Variable 1: Temp > 38.3°C OR Temp < 36.0°C

Variable 2: Heart Rate > 90 beats/min

Variable 3: Respiratory Rate > 20 breaths/min OR PaCO2 < 32 mm/Hg

Variable 4: WBC < 4,000 cells/mL OR WBC > 12,000 cells/mL
Early Prediction: Events Without Therapy

- Necrotizing enterocolitis (NEC).
- Intraventricular Hemorrhage (IVH).
- Periventricular Leucomalacia (PVL).
- Liver Failure.
- Some forms of severe conditions on the next slide.
- We are unable to predict these events.
- There are thus, no preventive therapies available.
Early Prediction: Events With Available Therapy

- Pneumothorax.
- Sepsis, pneumonia, UTI, meningitis.
- Neonatal Seizures.
- Respiratory, Cardiac, Kidney Failure.
- Extubation Failure.
- Pulmonary Interstitial Emphysema.
- Bronchopulmonary Dysplasia.
- Can we predict these events early?
Data Collection (Wishful Thinking)

Time-synched data at 1 sec precision
**BD LB: Technical Themes**

i) Real-time data collection and storage

ii) Extraction of relevant information and their time-evolution

iii) Optimal display of information

iv) **Early prediction of events-of-interest using Machine Learning and Data Mining**

v) Non-linear adaptive control

*Create holistic understanding of the entire neonate*
Big Data in NICU

• Many variables are continuously collected at sampling rates of 10 - 100 Hz.

• None of the caretakers utilize the real-time nature of this Big Data. There is no mechanism to synchronize time.

• There has been no attempt to utilize the interaction among variables.

• Most importantly, there has been no work to look at the interaction between Big Data variables and what is happening to the baby.
Baby M; Case of NEC

- **14:00:** NEC was found. Nursing found abdomen full, tense, no BS.

- **5 hrs prior:** Fellow’s note - ‘Tolerating feeds and stooling, will continue to advance. Does not breathe over the vent very often’.

- **21 hrs prior:** ‘Abdomen is slightly full but he is voiding and stooling’. Personal communication: ‘felt queasy, not enough to call doctors’.

- **42 hrs prior:** ‘noted small area of duskininess (~1 cm circle) near umbilicus. MD notified and at bedside to assess, no changes.’
Baby M: 5 days before NEC
Significant Graphs with 5-events

A

Days

Redness “Queasy” NEC

B

Low RR, Low Temp, High O₂sat
High HR, Low O₂sat, High Temp, High RR

Temp-Lo  Temp-Hi  RR-Lo  RR-Hi  O₂Sat-Lo  O₂Sat-Hi  HR-Lo  HR-Hi
Occurrence of Graphs and Sub-graphs
5-events

4-events

Baby M: Entire 37 days of stay
Distance between graphs $\varepsilon$ and $\gamma$:
(frequency of $\varepsilon$ x frequency of $\gamma$)
x number of common sub-graphs between $\varepsilon$ and $\gamma$

From the distance, we can calculate a predictive score when say $\varepsilon$ is a diagnostic graph and $\gamma$ is one of its subgraphs.

From all sub-graphs that appear, we can calculate a cumulative predictive score.
Log scores for Identification of NEC Patterns and their Prediction

Days to NEC diagnosis
Is it Specific?

- Nature of Temporal Data Mining is such that false-positives are rare as the patterns are in very high dimensions.

- We looked also another premature patient with isolated UTI and found no statistically significant 5-, 6- or 7-event sequences in the entire hospital stay of 133 days.
What did Unni Say?

Incorporating “real” time into Medical Data is extremely useful.
There is currently very little “time” in Medicine.
Let us bring “time” to Detroit.

Discussion items:
- Are real-time data centers realistic?
- How do we combine Data Mining and Machine Learning to create real-time Big Data analytics?
Most Significant Patterns (6-events)

- Temp-Lo
- Temp-Hi
- RR-Lo
- RR-Hi
- O2Sat-Lo
- O2Sat-Hi
- HR-Lo
- HR-Hi

Days

Redness "Queasy" NEC

Low Temp → High HR → High RR → Low O2sat → Low RR → Low HR → High O2sat → High Temp
6 Event Patterns: Entire Hospital Stay

- 6-event sequences which include the time of ‘Queasy’.
Baby M: Significant 7-event Sequences during 37-day stay
Baby M, Case of NEC: Heart Rate

6 hours before Queasy:
Mean: 149
Variance: 98

3 hours during Queasy:
Mean: 131
Variance: 8
Baby M, Case of NEC: Heart Rate

6 hours before Queasy:
Mean: 149
Variance: 98

3 hours during Queasy:
Mean: 131
Variance: 8