ABSTRACT
We present an approach for the analysis of clinical data from extremely preterm infants, in order to determine if they are ready to be removed from endotracheal tube-invasive mechanical ventilation. The data includes outcomes and covariates, with the subject population is naturally quite small. To address this problem, we use feature selection, specifically mutual information (MI), to determine if they are ready to be removed from endotracheal tube-invasive mechanical ventilation (ETT-IMV) in order to survive. Complications associated with prolonged ETT-IMV include pneumonia, airway trauma, air leaks and bronchopulmonary dysplasia. Early extubation carries its own hazards, including compromised gas exchange and ultimately the need for reintubation (extubation failure). Tradeoff between limiting duration of ETT-IMV and avoiding reintubation. Decision to extubate usually physician-driven and subjective. Goal: develop tool to assist physicians in predicting extubation readiness

INTRODUCTION
Motivation
• Most preterm infants (gestational age < 28 weeks) need to undergo endotracheal tube-invasive mechanical ventilation (ETT-IMV) in order to survive
• Complications associated with prolonged ETT-IMV include pneumonia, airway trauma, air leaks and bronchopulmonary dysplasia.
• Early extubation carries its own hazards, including compromised gas exchange and ultimately the need for reintubation (extubation failure).
• Tradeoff between limiting duration of ETT-IMV and avoiding reintubation.
• Decision to extubate usually physician-driven and subjective.
• Goal: develop tool to assist physicians in predicting extubation readiness.

Challenges
• Small dataset (common for clinical studies): 120 babies
• Many features of interest (>100)
• Good feature selection is critical

Class imbalance:
• relatively few pathological examples (the minority class represents ~25% of the data)
• identifying those examples is crucial

DATA
Patient population
• Infants admitted to neonatal intensive care units (NICUs) in hospitals in the Montreal area, Rhode Island and Detroit.
• The database includes:
  - patient demographics (birth weight, gestational age, ...)
  - peri-extubation characteristics (blood gases, ventilator settings, ...)
  - clinical outcomes (extubation failure)
• Extubation failure is defined as the need for reintubation within 7 days following extubation.

RESULTS
Feature Selection
1) Physicians’ input:
• Chose possibly clinically relevant features
2) Automated FS: mutual information (MI)
• MI quantifies how much we know about a random variable given another variable, and KL-divergence of joint distr. $p(x,y)$ w.r.t. $p(x)p(y)$ if $X,Y$ were independent:
$$MI(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$
• We keep features with MI > 0.3 (determined experimentally)

Correcting Class Imbalance – SMOTE
• SMOTE addresses the undersized population by creating synthetic examples.
• Issues with standard approaches:
  - Undersampling the majority class is not feasible on small datasets.
  - Random oversampling can introduce bias.
• SMOTE: For each e.g. $i$ in the minority class, we choose $\xi$ at random from $\{i\}$ nearest neighbours. To create the new sample $\xi'$, let $\xi$ and $\xi'$ represent $i$’s and $\xi$’s feature vectors, respectively. Then
$$\xi' = \xi + \alpha(\xi - x_i)$$
where $\alpha \in (0,1)$ is chosen at random.
• We used SMOTE with $m = 5$ to double the minority class, going from 30 to 60 failures.

Classification
• Tool: open-source ScikitLearn library
• Crossvalidation: leave-one-out.
• - train on n-1 examples
• - test on remaining example
• Algorithms:
  - Logistic Regression (LR): uses logistic function to find the relationship between features and labels.
  - Decision Trees (DT): tree with internal nodes as tests on features and leaves as labels.
  - Support Vector Machines (SVM): creates a boundary between classes by using kernel functions. We used 2 different kernels: linear (LSVM) and Gaussian (GSMV).

DISCUSSION & CONCLUSION
• SMOTE is responsible for:
  - Increase in reliability
  - Decrease of FPR
  - It classifies better outside of training data.
• Oversampling and Classification:
  - Random oversampling can introduce bias
  - We keep features with MI > 0.3

MAIN REFERENCES