Specific Aims

Rationale: Acute respiratory distress syndrome (ARDS) is a severe form of acute hypoxemic respiratory failure affecting 10% of patients admitted to the intensive care unit (ICU).(1-5) ARDS remains under-recognized with correspondingly low rates of life-saving, lung-protective mechanical ventilation (MV) use, and mortality rates of 35-46%.(2,6) Attempts to improve ARDS diagnosis, including rule-based computer algorithms, have seen limited success and require data types that limit automation, timeliness, and reproducibility.(7-15)

To improve automated ARDS detection technology, we propose to use machine learning (ML) to 1) accelerate the diagnosis of ARDS, and 2) identify early ARDS patients most likely to benefit from ARDS-specific treatments.(16-19) ML is a sub-discipline of computer science that uses algorithms to learn complex classification rules based on patterns in the relationships between predictor variables (referred to as features) and outcomes. ML offers advantages over traditional statistical methods including the ability to detect nonlinear relationships, the ability to detect relationships between features in high-dimensionality models, and even the ability to create unique features from patterns in the sample data.(20,21)

We have collected a dataset of VWD from over 500 patients receiving MV, including 156 patients with dual clinician-adjudicated ARDS.(22) Our preliminary analysis of 30 patients using standard ML models and physiologic features derived solely from VWD(23) suggests that ARDS can be detected in the absence of a chest radiograph or other medical history. In the first phase of proposed research, we will extend our preliminary work and develop improved models for automated ARDS detection. In the second phase of research, we will develop models to predict clinical trajectories amongst patients with early ARDS. For this, we will use Hidden Markov Models and advanced ML models amenable to time series modeling such as "deep learning" models that offer advantages over traditional ML algorithms when the input feature set is large and multidimensional. Deep learning models will be used to discover novel features from *a priori* pre-specified input feature sets, subsequently using both pre-specified and novel features for classification.(21) <u>We hypothesize that further model development using additional objective data derived from VWD analysis and the EMR, along with advanced machine learning techniques, will further improve our ARDS detection models and enable the prediction of clinical trajectories in patients with ARDS.</u>

Aim 1. Develop an improved machine learning (ML) model to discriminate between ARDS and other causes of acute respiratory failure using retrospective data. We hypothesize that inclusion of additional VWD-derived physiologic features will improve the accuracy of ARDS detection in a broader cohort of ARDS patients compared to our existing model. We further hypothesize that inclusion of a limited set of objective, EMR-derived features will improve model performance compared to a model based on VWD-derived features alone.

1A. Compare the performance of our current ML ARDS detection model with a model that includes additional VWD-derived physiologic information.
1B. Compare the performance of the best performing model from Aim 1A to a model that also includes objective, EMR-derived features.

Aim 2. Develop a time-series machine learning model to predict clinical deterioration in patients with early ARDS. We hypothesize that a Hidden Markov Model,(24-26) using both time-independent (static) and time-dependent (dynamic) features in the first 24 hours after ARDS onset, will predict clinical worsening of ARDS with greater accuracy than a traditional model using static features alone. We also hypothesize that a model using recurrent neural networks and advanced ML techniques (i.e. "deep learning") will more accurately predict clinical deterioration than a Hidden Markov Model.

2A. Compare the performance of a Hidden Markov Model using using static features to one using both static and dynamic features for predicting clinical deterioration of patients with early ARDS.
2B. Compare the accuracy of the best performing model from Aim 2A to a competing model based on deep learning with recurrent neural networks for predicting clinical deterioration of patients with early ARDS.

Successful completion of this research will advance the use of computational approaches to ARDS diagnosis and prognostication, and may contribute to the development and translation of clinically useful automated systems for rapid ARDS diagnosis and treatment.