**Research Strategy:**

## 1. Background and Significance

**1.1 Existing automated approaches to diagnosing ARDS have not been widely validated and may not be appropriate for timely diagnosis of ARDS:** Automated solutions for ARDS detection have been developed since 2009 with the advent of ARDS "sniffers."(9,10,27) These "sniffers" are algorithms that attempt to detect when a patient meets ARDS diagnostic criteria by determining: 1) when a patient has a PaO2:FiO2 (P/F) below 300 on an arterial blood gas (ABG), and 2) that a radiologist has noted bilateral opacities on a chest X-ray (CXR).(28) However, external validation of the "sniffers" suggests that they may not scale well beyond a single center.(11) Both performance and generalizability of these sniffer systems are limited given sniffers rely on text of radiologists' reports of chest radiographs and institution- or radiologist-specific practice patterns may limit the sensitivity and specificity of detecting bilateral opacities due to differences in terminology.(11,29) Infrequent performance of chest radiographs and ABG's, and delays in interpretation may also limit automation and timeliness of diagnoses.(30,31) All existing sniffers have been marked by limited specificity when validated externally from their originating institutions and,(11) due to limitations on the timing of component diagnostic tests, they may preclude early detection or prognostication of ARDS trajectory.(9,10)

**1.2 Ventilator waveform data shows promise as a source of unbiased information on underlying respiratory physiology:** Previous studies suggest that analyzing high-frequency ventilator waveform data (VWD) may be feasible for detecting ARDS,(32-34) but VWD has yet to be investigated in a substantive way for its ability to accurately diagnose ARDS. There are distinct advantages to using high-frequency VWD to detect ARDS: it is continuously sampled, ubiquitous for intubated patients, unbiased, and rich in physiologic information. Using computational methods, VWD can provide information on respiratory system compliance, a metric that can track the stiffening of the lungs seen with the development of ARDS, and airway resistance which can be used to help distinguish between non-ARDS conditions such as COPD, and asthma, where dynamic hyperinflation may also lead to poor respiratory system compliance.(33-39). Our preliminary data suggest that the physiologic information embedded in VWD can be used to diagnose ARDS. Leveraging previously completed work(22,23) we used VWD from the first 24 hours of MV in 30 patients, including 10 with ARDS, to develop a preliminary ML model for ARDS diagnosis. Using a restricted feature set (e.g. sets of parameters) (Figure 1) with variables derived from VWD (Table 2, Table 3) we were able to correctly identify patients with ARDS with a sensitivity of 70% and specificity of 100% (Table 4). With a less restricted feature set, our model identified ARDS with a sensitivity of 80% and specificity of 95% (Table 5). *Importantly, both models showed sensitivity comparable or superior to clinician recognition*(2,11,40) *without need for a chest radiograph, ABG, or other clinical information.*

**1.3 Our models satisfy a need for automated diagnostic tests in critical care medicine:** With the increasing digitization of medicine, we are entering an age when healthcare providers will be constantly presented with diverse types of patient data without the right tools to translate these data into the information necessary to create effective diagnoses and treatment plans.(30,41,42) Automating the detection of difficult to diagnose diseases stands to reduce cognitive burden on providers, prevent medical errors, and save patient lives. Our proposed research aims to improve this problem with respect to the diagnosis of ARDS, and to provide insights into the use of computational approaches to disease diagnosis and prognostication. With further research, our ARDS detection models could develop into a diagnostic test for ARDS that could be translated into a clinical decision support system that could alert providers to the development of ARDS. Finally, our proposed research will add to the body of science in clinical decision support systems to help ensure that similar systems become more effective and prevalent in the future.

**1.4 Predictive models for disease severity can have clinical utility:** Diseases such as cancer, sepsis, and ARDS all have increased mortality rates as disease severity increases.(2,21)The ability to predict increases in disease severity could identify patients appropriate for more aggressive treatments before a disease worsens to the point where the clinical course is no longer modifiable. In this regard, there are a variety of options available for the management of patients with ARDS.(18,19,43) Attention has thus focused on limiting ARDS progression through early recognition and implementation of lung-protective MV delivery.(44,45) As a result, our proposed work in predicting clinical trajectories in ARDS could lead to more appropriate treatment of patients with ARDS, and it could also contribute to the science of using new predictive models for other diseases such as chronic obstructive pulmonary disease (COPD), sepsis, or cancer.(46-48) These types of predictive models will become increasingly important as medicine becomes more personalized by utilizing high frequency data collection and machine learning analytics to generate highly accurate and actionable prognostic information from a wealth of data.

## 2. Approach

**2.1 Aim 1. Develop an improved machine learning (ML) model to discriminate between ARDS and other causes of acute respiratory failure using retrospective data.**

**2.1.1 Rationale:** Previous studies and our own initial results (Table 4,Table 5) suggest that we may be able to discriminate ARDS from other pathophysiologic conditions using VWD.(32-34) We will add additional predictive features extracted from VWD and the EMR to improve sensitivity, specificity, and generalizability of the existing model.

**2.1.2 Research Design:**

2.1.2.a Aim 1A Methods: Using patient diagnoses adjudicated by two Pulmonary and Critical Care physicians, we will first construct a training and validation set from 502 patients whose data we collected within the first 7 days of intubation, including 156 patients with ARDS. For each ARDS patient we will determine when they first met Berlin ARDS criteria.(28) Patients with other causes of acute respiratory failure will be classified as non-ARDS. We will then use the first 24 hours of VWD from ARDS patients starting at the time each patient satisfied ARDS criteria, or starting in the first 24 hours of mechanical ventilation in patients that did not develop ARDS. We will use these data from these time frames to generate model features, perform model development, and test model performance. Our patient cohort will be divided 70:30. 70% of the cohort will be used for model development. The remaining 30% of the cohort will be used to validate model performance.

Respiratory system compliance ($C_{RS}$) is typically measured by performing the end-inspiratory pause technique to obtain a plateau pressure used to calculate $C_{RS}$.(49) However, this technique is primarily used in volume controlled ventilation, is not appropriate for use in all patients, and is not always reliably performed by clinicians. (50) An alternative to the end-inspiratory pause is to utilize computational algorithms to non-invasively determine lung resistance, and airway compliance from analysis of VWD.(34-36,38,51) These algorithms all rely on the single chamber model of the lung which has been generally accepted as providing a good approximation for the human respiratory system.(35,38) We have recently implemented the aforementioned methods to estimate airway resistance, and $C_{RS}$ from automated analysis of VWD. We will use these features with our existing models to test the hypothesis that inclusion of airway resistance, and $C_{RS}$ will allow our model to improve the sensitivity of ARDS detection without significantly worsening specificity.

2.1.2.b Aim 1B Methods: In Aim 1B, we will gather additional data from the electronic medical record (EMR) to test the hypothesis that inclusion of objective non-VWD data will improve sensitivity and maintain specificity of ARDS detection. To improve model sensitivity, we will integrate available P/F ratios, SpO2/FiO2 ratios, and objective components of the lung injury prediction score (LIPS) such as use of vasopressors (shock), body mass index, and albumin levels.(8,52) To help maintain specificity, we plan to use previously identified types of information that may be used to identify patients in whom hypoxemia is primarily from congestive heart failure such as brain natriuretic peptide levels and left ventricular ejection fraction. (9,10) Our focus on structured, objective data is intended to avoid the use of features whose accuracy, timeliness, or inter-rater reliability may compromise model performance in the first 24 hours of MV. Furthermore, the use of structured data elements (in contrast to radiologists' text reports or other clinical notes) allows feature extraction without need for additional pre-processing or verification of accuracy. For this retrospective study, all EMR-derived features will be extracted automatically from the UC Davis Medical Center's Epic EMR reporting data warehouse using structured query language (SQL) algorithms by a member of the UC Davis IT Health Informatics research IT team. Each patient's identifiers will be replaced with his or her unique study identifier in accordance with our IRB-approved study protocol (#647002).

**2.1.3 Expected outcomes:** We hypothesize that the addition of physiologically significant information such as $C_{RS}$ and airway resistance to our model will enable us to detect ARDS with greater sensitivity and specificity than our current model, and in a more generalizable set of patients.

**2.1.4 Potential Complications:** Although ARDS is associated with decreases in $C_{RS}$, there is a possibility that airway resistance and respiratory compliance measurements derived from VWD using the single chamber model may be adversely affected if a patient is spontaneously breathing or is under light sedation.(50,53,54) To mitigate this we will include Richmond Agitation-Sedation Scale (RASS) scores(55), a well validated measurement of the depth of sedation, in our models to quantify the level of patient sedation when VWD was collected. We will also perform sensitivity analyses, excluding data from patients that are deeply sedated and paralyzed to see if model performance is maintained. We will evaluate our model's performance on the three separate sedation level datasets of paralyzed, deeply sedated, and lightly sedated subjects, and on the combined dataset. As part of the study in sub aim 1A we will also seek to utilize different methods of analytically calculating compliance and resistance that may be more robust to variable patient effort,(35-

37,50,51,53,54,56,57) and we will determine accuracy of these methods under differing circumstances, subject pathophysiology, and sedation levels.

**2.2 Aim 2. Develop a time-series machine learning model to predict clinical deterioration in patients with early ARDS.**

**2.2.1 Rationale:** Heightened ARDS severity is associated with significant increases in patient mortality.(28) ARDS treatment protocols may be most effective when used early, especially in patients with moderate-severe disease, and some therapies may not work or even cause harm in patients with less severe ARDS.(17,18) Models that can distinguish ARDS patients likely to worsen in severity from those likely to rapidly improve would help to individualize the early management of ARDS patients and improve clinical outcomes.(17-19,43)

**2.2.2 Research Design**

2.2.2.a Aim 2A Methods: Using the retrospectively collected first 24 hours of data for ARDS patients, we will annotate whether a subject's P/F measurements decreased, improved, or stayed static, in the next 24-48 hour period post-ARDS onset. Specifically, the single lowest P/F score taken during the first 24 hours will be compared to the lowest P/F in the period of 24-48 hours after ARDS diagnosis, and a change in Berlin ARDS severity class will be used to differentiate between those that worsened, remained stable, or improved. If no P/F score is taken in this time then the SpO2:FiO2 (S/F) ratio will be used as a surrogate measurement.(58-60) We will then utilize VWD and EMR derived features identified to be predictive of ARDS diagnosis in Aim 1 (static features), and add a temporal dimension to them to by utilizing Calvert's and Kam's methodology of determining when a feature value is increasing, decreasing, or not changing (dynamic features).(47,61,62) In addition, we will also investigate whether magnitude and rate of change in features such as respiratory compliance will be predictive of worsening ARDS severity.(63-65)

Using this parameterization approach, we will develop a Hidden Markov Model (HMM) to predict if a subject's ARDS severity class will decrease, increase, or stay static in the next 24 hours. Hidden Markov Modeling is a statistical technique commonly used when developing predictive time series models and has been utilized broadly in medicine.(24-26,66,67) HHM's work by establishing multiple system states that are connected to each other by state transition probabilities. These links dictate at what probability a system will move from one state to another.(68,69) In our case, a patient will represent our system, and they will start in one of three ARDS states based on Berlin severity criteria: mild, moderate, or severe.(28) Using a HMM, we will establish a probabilistic model that determines whether a subject will change state in the next 24 hours, and if so to which ARDS disease state, testing the accuracy of model predictions against known patient trajectories.

2.2.2.b Aim 2B Methods: Next, we will create a model for predicting change in ARDS severity class using a type of ML known as deep learning. Deep learning models use specialized neural networks that are used to generate novel features from the available "raw" input feature set, thus precluding the necessity of feature extraction via specialized expert knowledge.(70-73) Deep learning models offer clear advantages over traditional ML algorithms like Random Forest, and simple neural networks, especially when the input feature set is large and multidimensional, and when complex pattern recognition over time is required for classification.(21,73,74) For model training we will use two deep learning models, recurrent neural networks, and long short-term memory networks, that have shown success in extraction of temporal patterns and in the early detection of sepsis.(21,47,75) Given that important physiologic information such as respiratory compliance is embedded in VWD, we hypothesize that these algorithms will be able to identify pertinent features including temporal patterns diagnostic of worsening severity of ARDS after the patient has met Berlin criteria. As described above, we will use EMR P/F or S/F data to group ARDS patients in our cohort according to whether they worsened, remained stable, or improved in ARDS severity class from day 1 to day 2 after ARDS onset. A "raw" feature set of static and temporally dynamic features derived from VWD and EMR data will be used for model development. Once completed, we will test the hypothesis that our deep learning model will have significantly better sensitivity and specificity for predicting changes in ARDS severity class than the Hidden Markov Model developed in sub aim 2A to determine which analytic method is best for predicting change in ARDS severity.

**2.2.3 Expected Outcomes:** We hypothesize that a Hidden Markov Model, using the best features identified in Aim 1 and additional temporal information, will be able to accurately predict changes in ARDS severity. We hypothesize that the addition of deep learning will allow us to computationally derive additional predictive information that will result in superior predictive performance of our model compared to a Hidden Markov Model.

**2.2.4 Potential Complications:** It is possible that neither HMM nor deep learning models will accurately predict changes in ARDS severity. Despite this possibility, the promising performance of our preliminary model in distinguishing between patients with ARDS, and those without the syndrome strongly suggests that use of physiologic features will be useful in predicting changes in clinical trajectories in ARDS patients. Our use of static and dynamic features create a novel temporal framework for predicting changes in ARDS, and the use of advanced ML methods in Aim 2B that can discover useful temporal patterns in a raw dataset to develop novel predictive features may further mitigate risk of model failure. Finally, the use of a 24-hour window in which to assess for change in ARDS severity during model development may be overly restrictive and compromise model training. It possible that a longer time frame (e.g., 72 or 96 hours from ARDS onset) could better separate patients into groups with distinct trajectories, leading to better model training and thus prognostic performance, which may require further model development and validation experiments with different cutoffs.

## 3. Preliminary Data

### 3.1 Data Collected and Used for Preliminary Models

We previously developed data collection and aggregation architecture to support a clinical study of patient-ventilator interactions. This system and our methodology was approved by the University of California Davis (UCD) institutional review board (IRB) (protocol number 647002). All subjects or their surrogates provided informed consent per the requirement of the study protocol. Our data collection system enables passive, continuous, and automated data collection from multiple mechanical ventilators simultaneously. The success of this system has resulted in the collection of VWD for over 500 patients who have been intubated in the ICUs at UC Davis Medical Center from 2015 to now.(22) Table 1 represents summary statistics resulting from this ongoing study.

| Pathophysiology | Patients | Breaths | Hours |
|---|---|---|---|
| *ARDS* | 156 | 24,111,617 | 17,629 |
| *COPD* | 84 | 8,213,941 | 3,791 |
| *Other* | 262 | 18,670,606 | 14,779 |
| *Total* | 502 | 47,931,228 | 36,199 |

*Table 1: Summary statistics for the ventilator waveform data collected in our ongoing study. Abbreviations: ARDS – acute respiratory distress syndrome. COPD – chronic obstructive pulmonary disease*

To ensure our preliminary dataset used a variety of patients, 10 ARDS, 10 COPD, and 10 patients with other pathophysiology were chosen to train our ARDS detection model. For the subset of ARDS patients, we chose patients in which ARDS was present at the time of intubation, and chose patients with a range of severity levels from mild to moderate. Specifically, 5 patients with severe, 2 patients with moderate, and 3 patients with mild ARDS were chosen. Patients without ARDS included 10 patients with acute exacerbations of COPD and 10 with a range of other indications for MV, all of whom did not meet criteria for ARDS.

### 3.2 Feature Extraction

VWD is collected from the ventilator as a time series of flow and pressure observations. These data are then processed using specialized software developed in our lab to derive clinically relevant features (e.g. parameters) that we can use to populate our model.(23) For example, information such as tidal volume of air inspired (TVi), and expiratory time (E-time) were calculated and used as features in our model. With this approach, we created two sets of features: one utilizing only flow, and dynamic compliance (Table 2), and the other utilizing a broader feature set including pressure and volume related data (Table 3).

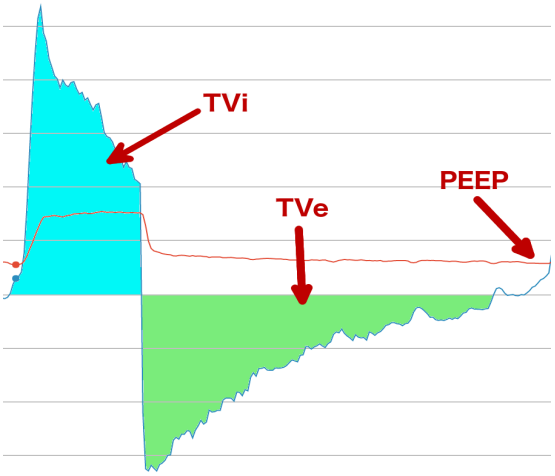


*Figure 1: An illustrative example of how we can extract breath-level features (e.g. parameters) from VWD during pressure controlled ventilation. Flow is measured via the dark blue line, and pressure is measured on the red time series. We can calculate the tidal volume inspired (TVi) by taking the summative area under the curve of the flow waveform for inspiration denoted by the teal coloring. Tidal volume expired (TVe) can be performed by the same process during expiration, an area denoted by the green coloring. Positive end expiratory pressure (PEEP) can be found by identifying the measurement of the pressure curve close to the end of the breath. Additional variables beyond these can also be extracted from VWD and we have extracted close to 50 thus far. The variables used in our model are listed in Table 2 and Table 3.*

#### 3.2.1 Feature Subsets

3.2.1.a Flow and Dynamic Compliance subset:

| Variable Name | Units | Description |
|---|---|---|
| I-time | Seconds | The time from breath start to [x0 minus 1 time point] |
| E-time | Seconds | The time from x0 to breath end |
| I:E ratio | Unitless | The ratio of the I-time to the E-time |
| RR | Unitless | Instantaneous respiratory rate, defined as 60/(I-time+E-time) |
| Mean flow from PEF | Milliliters | The mean flow observation from the point in time peak expiratory flow (PEF) occurred to the point where the breath terminated and a new one begins |
| Tau | seconds | An expiratory time constant we defined by taking the slope of the expiratory flow from .16 seconds after peak expiratory flow to the end of the breath. |
| Dynamic Compliance ($C_{dyn}$) | Unitless | This measure is derived via : $C_{dyn} = \frac{TVi}{PIP-PEEP}$ where TVi is the inspiratory tidal volume. PIP is peak inspiratory pressure, and PEEP is positive end expiratory pressure. |

*Table 2: List of flow and dynamic compliance feature subset with descriptions. These variables were all processed from raw ventilator waveform data. x0 - The place where the flow waveform crosses 0. Normally corresponds with the start of exhalation, but may not in case of patient ventilator asynchrony. PIP - peak inspiratory pressure. PEEP – positive end expiratory pressure*

3.2.1.b Pressure, Volume, Flow, and Dynamic Compliance subset:

In addition to the features described above, the largest set of features has the following features added:

| Variable Name | Units | Description |
|---|---|---|
| TVi | Milliliters/second | Inspiratory tidal volume, defined as the integral of the flow-time curve values from breath start to point where flow crosses 0 (x0). |
| TVe | Milliliters/second | Expiratory tidal volume, defined as the integral of the flow-time curve values from x0 to breath end |
| PIP | cm H2O | Peak inspiratory pressure, defined as the maximum recorded pressure from breath start to [x0 minus 1 time point] |
| ipAUC | cm H2O | The inspiratory pressure area under the curve, defined as the integral of the pressure-time curve from breath start to [x0 minus 1 time point] |
| epAUC | cm H2O | The expiratory pressure area under the curve, defined as the integral of the pressure-time curve from x0 to breath end |
| PEEP | cm H2O | Positive end-expiratory pressure, defined as the average of the last 5 data points from the pressure-time curve of each breath |
| Paw | cm H2O | Mean airway pressure |

*Table 3: List of pressure, volume, flow, and dynamic compliance feature subset with descriptions. Note that the features mentioned here are added to the features from Table 2 to create a larger set of features. These variables were all processed from raw ventilator waveform data. x0 - The place where the flow waveform crosses 0.*

## 3.3 Current ARDS classifier

### 3.3.1 Model Construction

Machine learning (ML) was then used to build a model that could perform classification on pathophysiologic state for future patients. For our purpose, we used supervised machine learning (Random Forest algorithm) because we knew the patient state at the time an observation was taken, enabling us to train the model to recognize ARDS or non-ARDS pathophysiology. To perform supervised ML, we define the classifier function $f$ (e.g. our model) as follows: (76)

$$y = f(X)$$

Where X represents the input observations and $y$ indicates the classification result. We defined X to be a matrix such that $X = \{x_1, x_2, \dots, x_n\}$, where each $x_i \in X$ corresponds with an observation consisting of 20 consecutive breaths (Figure 2). Each $x_i$ takes form $x_i = \{b_{i1}, b_{i2}, \dots, b_{im}\}$ where $b_{ij}$ can be defined as the average value of a feature over the window of 20 consecutive breaths. We define $y$ as a 1-dimensional vector where $y = \{y_1, y_2, \dots, y_n\}$. In our binary classification problem $y_i \in \{0,1\}$, where 0 corresponds with non-ARDS and 1 corresponds with ARDS.

### 3.3.2 Model Performance

To perform classification of patient pathophysiology based on the VWD derived features, a model was constructed using the processed patient data. Learning of model parameters was performed using a Random Forest with 10 classifier trees.(77) The Random Forest algorithm was ideal for our circumstances because it is fast to train, can be used on large datasets, produces performant models, is easy to understand, and is resistant to overtraining.(78) Our Random Forest-derived model was then trained on 27 patients and validated using the remaining 3 patients. We performed this same process in a 10-fold kfold validation, meaning we performed training and validation again 10 times over, but used a different set of 27 patients for training and a different set of 3 patients for validation each time. We performed 10-fold validation to reduce risk of overtraining by ensuring a single patient's data is not included in both the training and test sets, and because we wanted to evaluate our model's performance broadly over all patients in our dataset.(79,80)

Window-level performance of each model was evaluated in terms of whether a window of 20 breaths was classified correctly with corresponding patient pathophysiology. We determined patient-level model performance by aggregating all window-level predictions made for a single patient. A patient was then predicted to have a given pathophysiology based on the most commonly represented disease window.

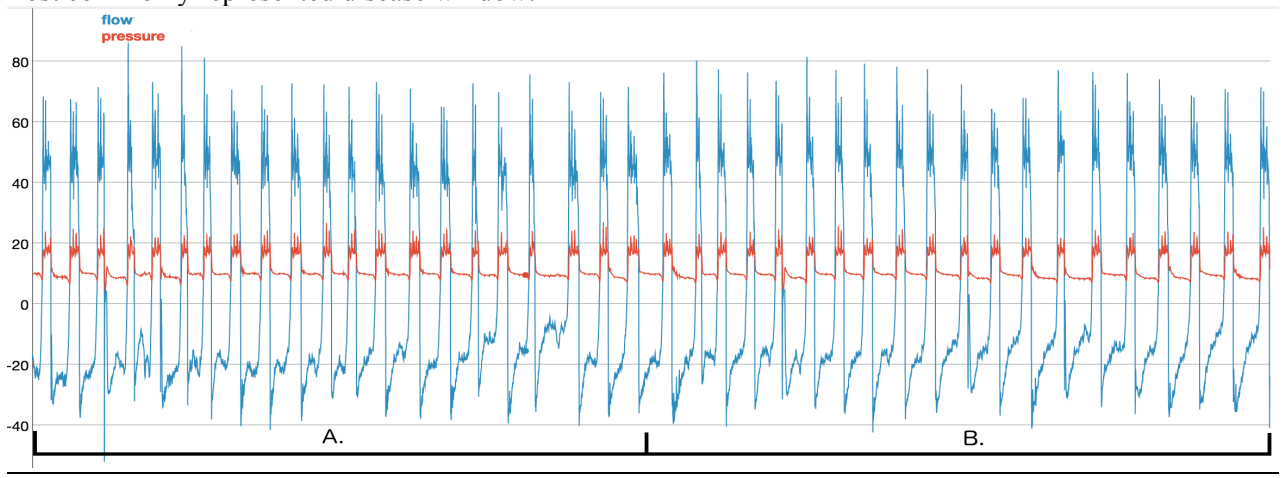


*Figure 2: A sequence of 40 breaths extracted during our study. Flow measurements are denoted in blue, pressure measurements in red. Here, we display two 20 breath windows, A and B, each corresponding with a distinct observation in our model. Each consecutive length of 20 breaths has features like TVi and PEEP extracted from each breath, and then each feature is averaged over each 20 breath sequence. The averaged features are then aggregated to comprise a single observation $x_i$, where $x_i = \{b_{i1}, b_{i2}, \dots, b_{im}\}$ where $b_{ij}$ is a single feature, such as the averaged TVi.*

3.3.2.a Patient-Level ARDS Classifier Performance

*Flow and Dynamic Compliance Model Performance:*

| Pathophysiology | Accuracy | Sensitivity | Specificity | Precision (PPV) | AUC |
|---|---|---|---|---|---|
| *non-ARDS* | 0.9 | 1.0 | 0.7 | 0.87 | N/A |
| *ARDS* | 0.9 | 0.7 | 1.0 | 1.0 | 0.85 |

*Table 4: The per-patient performance for the flow and dynamic compliance model. This model utilizes parameters least likely to be confounding to the prediction of ARDS.*

*Pressure, Volume, Flow, and Dynamic Compliance Model Performance:*

| Pathophysiology | Accuracy | Sensitivity | Specificity | Precision (PPV) | AUC |
|---|---|---|---|---|---|
| *non-ARDS* | 0.9 | 0.95 | 0.8 | 0.91 | N/A |
| *ARDS* | 0.9 | 0.8 | 0.95 | 0.88 | .875 |

*Table 5: The per-patient performance for the pressure, volume, flow, and dynamic compliance model. This model utilizes parameters that may be confounding to the prediction of ARDS. Overall its predictive accuracy is improved but we may be fitting our model too closely to ARDS specific treatment methods such as limiting tidal volumes and increasing PEEP.*

Performance of the model that integrated pressure and volume variables (Table 5) appeared more accurate overall compared to the model that only used flow and dynamic compliance (Table 4). This may have resulted from the latter model's use of features such as PEEP and TVi that are associated with ARDS treatment. Using such features for model training may result in poor model performance in subsequent patients who have unrecognized ARDS, which is why we plan to focus on improving the flow, time, and dynamic compliance model to minimize the risk of training on confounding parameters.