AUTHORS

Petru Anica-Popa: M.P.AnicaPopa@student.tudelft.nl

Optimizing Mechanical Ventilation Support for Patients in Intensive Care Units

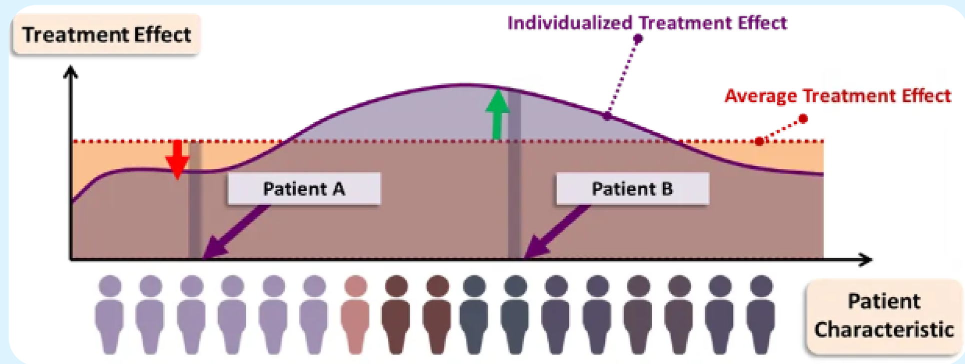
An Analysis of Deep Learning Methods for Personalizing Positive End-Expiratory Pressure Regime

AFFILIATIONS

Responsible Professor: Jesse Krijthe
Supervisors: Rickard Karlsson, Jim Smit
Institution: Delft University of Technology

01. Background knowledge

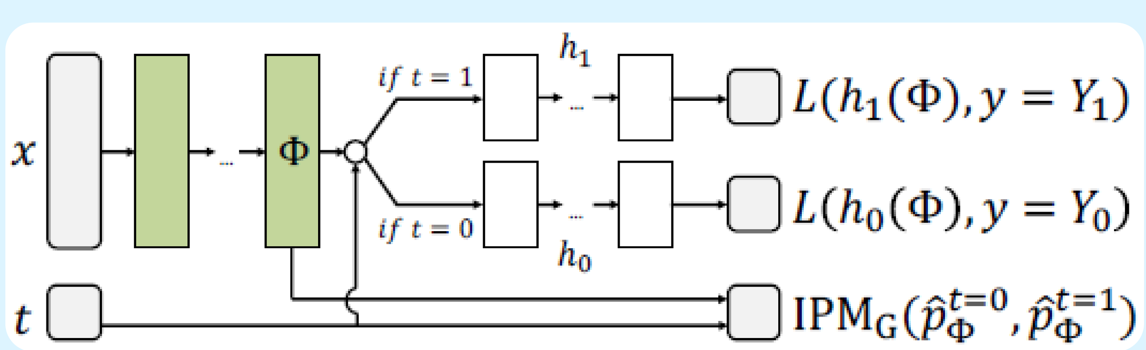
- **Mechanical ventilation**, a critical therapy for ICU patients with acute respiratory failure, involves a key setting known as **Positive End-Expiratory Pressure (PEEP)**.
- While **higher PEEP** can alleviate lung stress and strain, it also carries potential risks, including death. The **optimal level of PEEP** remains **controversial**, with numerous trials comparing low and high PEEP regimes yet failing to reach a consensus.
- The hypothesis suggests that the **benefits of PEEP may vary based on patient characteristics**, implying that a personalized approach to setting PEEP could improve treatment outcomes.



- **Causal inference** is the process of determining and estimating the impact of certain actions on outcomes in a population (**treatment effect**).
- **PEEP selection** can be modeled as a causal inference task of estimating **Individualized Treatment Effects (ITE)** ⇔ estimating **the effect on the survival of patients** from following a **low vs high PEEP regime**
- Multiple **machine learning methods** have been developed for **estimating ITE**. These methods could potentially show **promising results** in determining the **optimal PEEP regime for individual patients**.

Algorithms:

- **TARNet / CFR: an ITE estimator based on neural networks**
- **two metalearners: the S- and T-learners, both implemented using neural networks**



TARNet/CFR neural network architecture

procedure T-LEARNER(X, Y, W)

$$\hat{\mu}_0 = M_0(Y^0 \sim X^0)$$
$$\hat{\mu}_1 = M_1(Y^1 \sim X^1)$$
$$\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$$

T-learner pseudocode

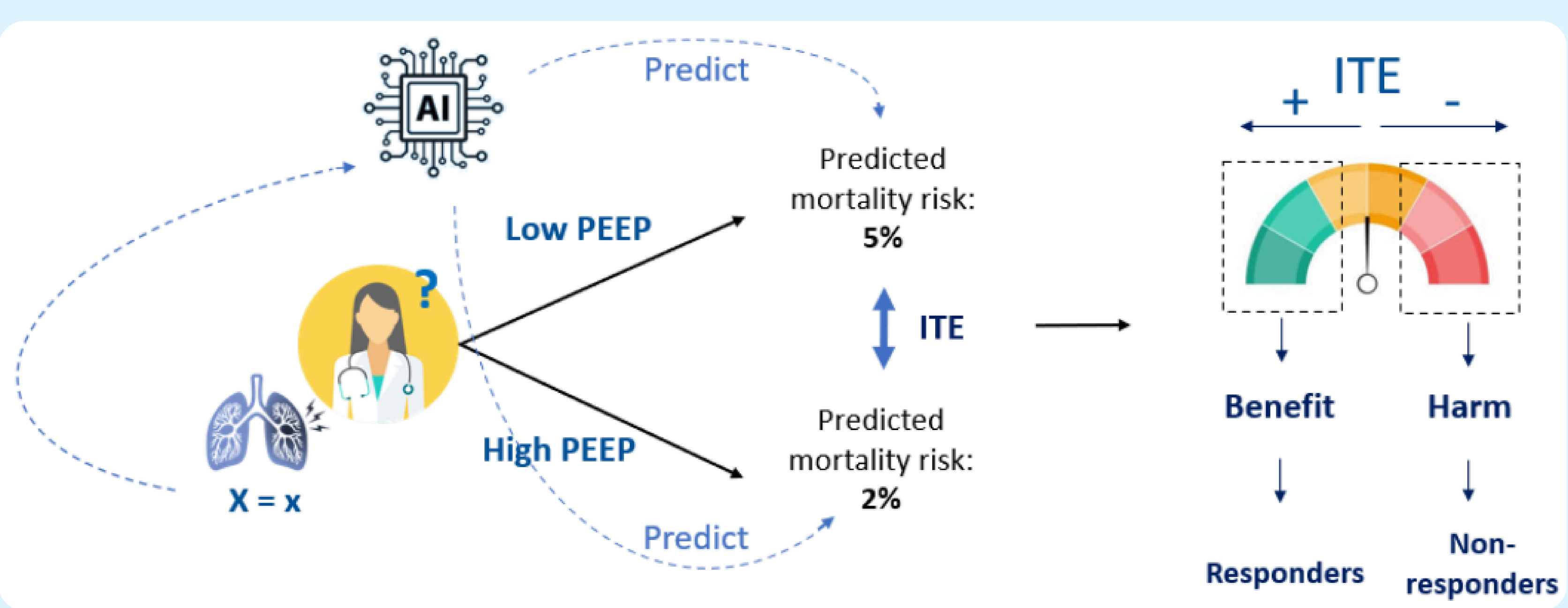
procedure S-LEARNER(X, Y, W)

$$\hat{\mu} = M(Y \sim (X, W))$$
$$\hat{\tau}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)$$

S-learner pseudocode

02. Research question

How can neural network-based machine learning methods such as TARNet and CFR be used to personalize treatment strategies in the ICU by estimating the individualized treatment effects of lower versus higher PEEP regimes for mechanical ventilation on patient outcomes?



03. Methodology

Literature study

Experiment setup
- generate simulation datasets
- preprocess MIMIC-IV dataset
- identify possible confounders

Train and evaluate models
- S- and T-learner
- TARNet and CFR

Analyze results and tune parameters
- Mean Squared Error (MSE)
- Area under Qini curve (AUC)

Validate results on RCT dataset

04a. Results - Simulations

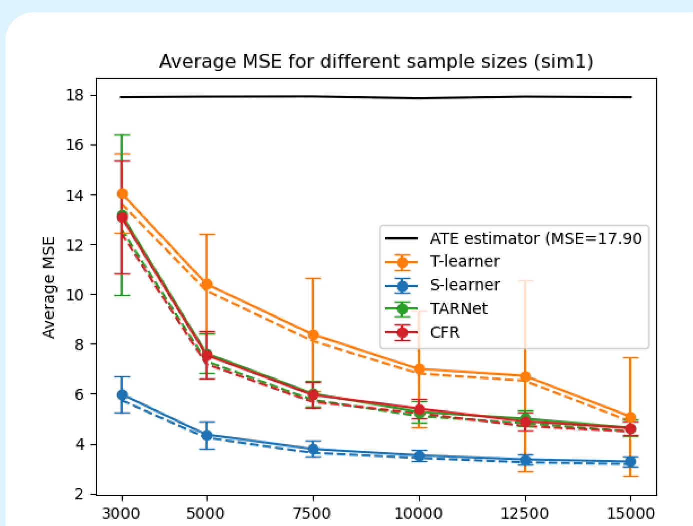


Figure 1: Unbalanced case without confounding (10% treated), simple ITE.

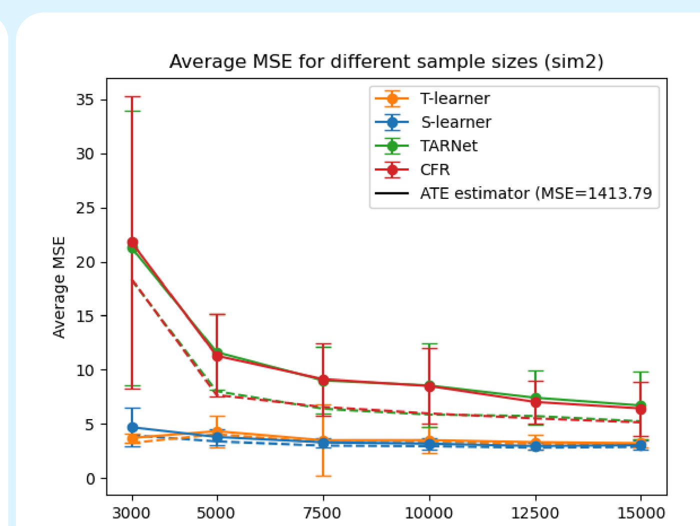


Figure 2: Balanced case without confounding, complex linear ITE.

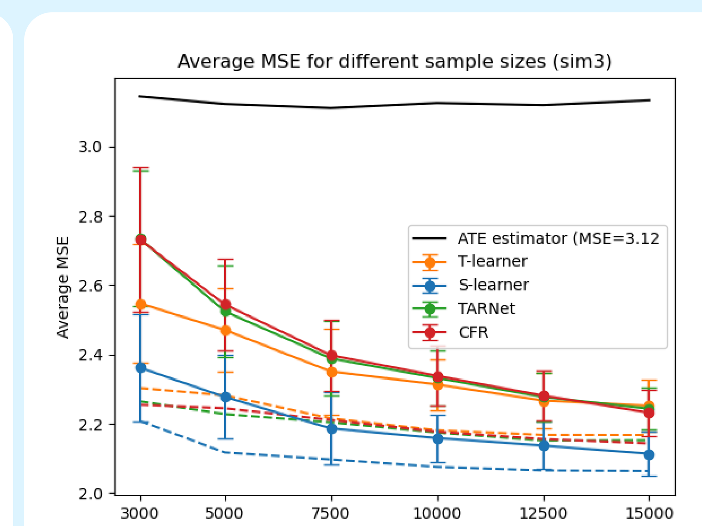


Figure 3: Balanced case without confounding, complex non-linear ITE.

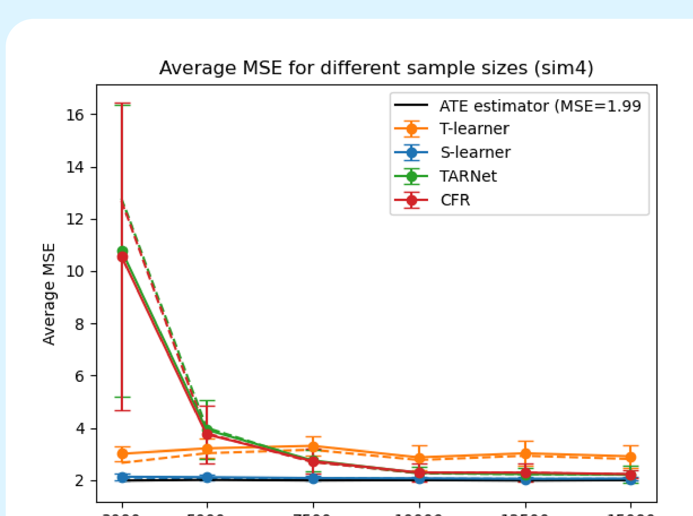


Figure 4: Balanced case without confounding, global linear response, ITE = 0.

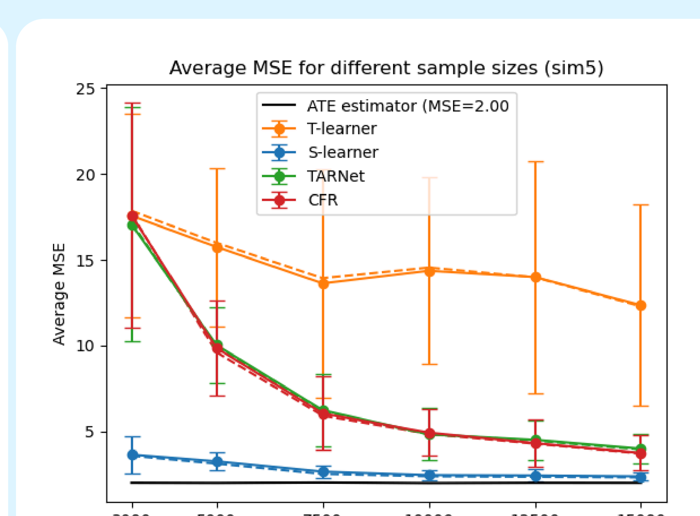


Figure 5: Balanced case without confounding, piecewise linear response, ITE = 0.

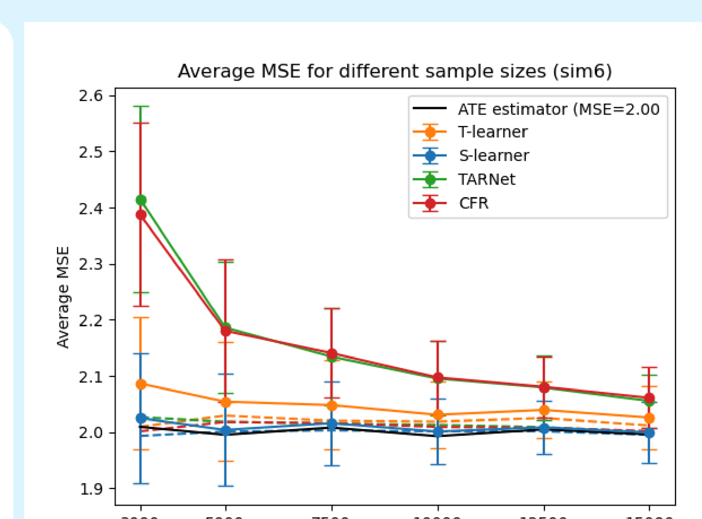


Figure 6: Balanced case with measured confounding, ITE = 0.

04b. Results - MIMIC-IV

Model	Mean	Std	Max
S-learner	*3.33*	*2.28*	*9.41*
T-learner	*3.26*	*2.05*	*10.02*
TARNet	*0.31*	*1.78*	*5.75*
CFR	*0.69*	*1.62*	*5.28*

Table 1: The mean, standard deviation, and maximum AUC for each model on the MIMIC-IV dataset (rounded to the nearest hundredth).

04c. Results - Randomized Control Trials

Model	Area	95% CI
S-learner	*0.22*	*[-2.31, 2.71]*
T-learner	*0.58*	*[-1.89, 2.97]*
TARNet	*0.40*	*[-2.16, 2.91]*
CFR	*0.04*	*[-2.40, 2.43]*

Table 2: AUC with 95% confidence intervals for each model on the RCT dataset (rounded to the nearest hundredth).

05. Conclusions

S-learner
Consistently effective across all simulations and robust in MIMIC-IV and RCT experiments.

T-learner
Struggled during the simulations but performed well in MIMIC-IV and RCT experiments.

TARNet
Outperformed by S-learner across all simulations but improved as sample size increased; had the worst performance on MIMIC-IV but showed better results in RCT experiments.

CFR
Very similar to TARNet across all simulations; improved on TARNet in MIMIC-IV experiments but had significantly worse performance in RCT experiment.

Key findings

- All estimators exhibited **high variance**, likely due to the **limited samples** in MIMIC-IV and RCT experiments. This led to an **overreliance** on data splits and neural network initialization, resulting in **poor generalization**.
- **S- and T-learners** are more suitable for predicting the appropriate PEEP regime for a patient's survival outcome when the **training data is limited**.
- **TARNet and CFR** may perform than the metalearners when there are ample training samples.

06. Limitations and Recommendations

Limitation: Uncertainty in confounder selection.
Recommendation: Investigate diverse confounder sets.

Limitation: Metrics may not fully reflect model performance.
Recommendation: Use a broader set of evaluation metrics.

Limitation: Limited samples hinder learning and generalization.
Recommendation: Test models on larger real-world datasets.

Limitation: TARNet / CFR's theoretical design mismatch with data setup.
Recommendation: Evaluate models on real-world datasets with continuous outcomes.