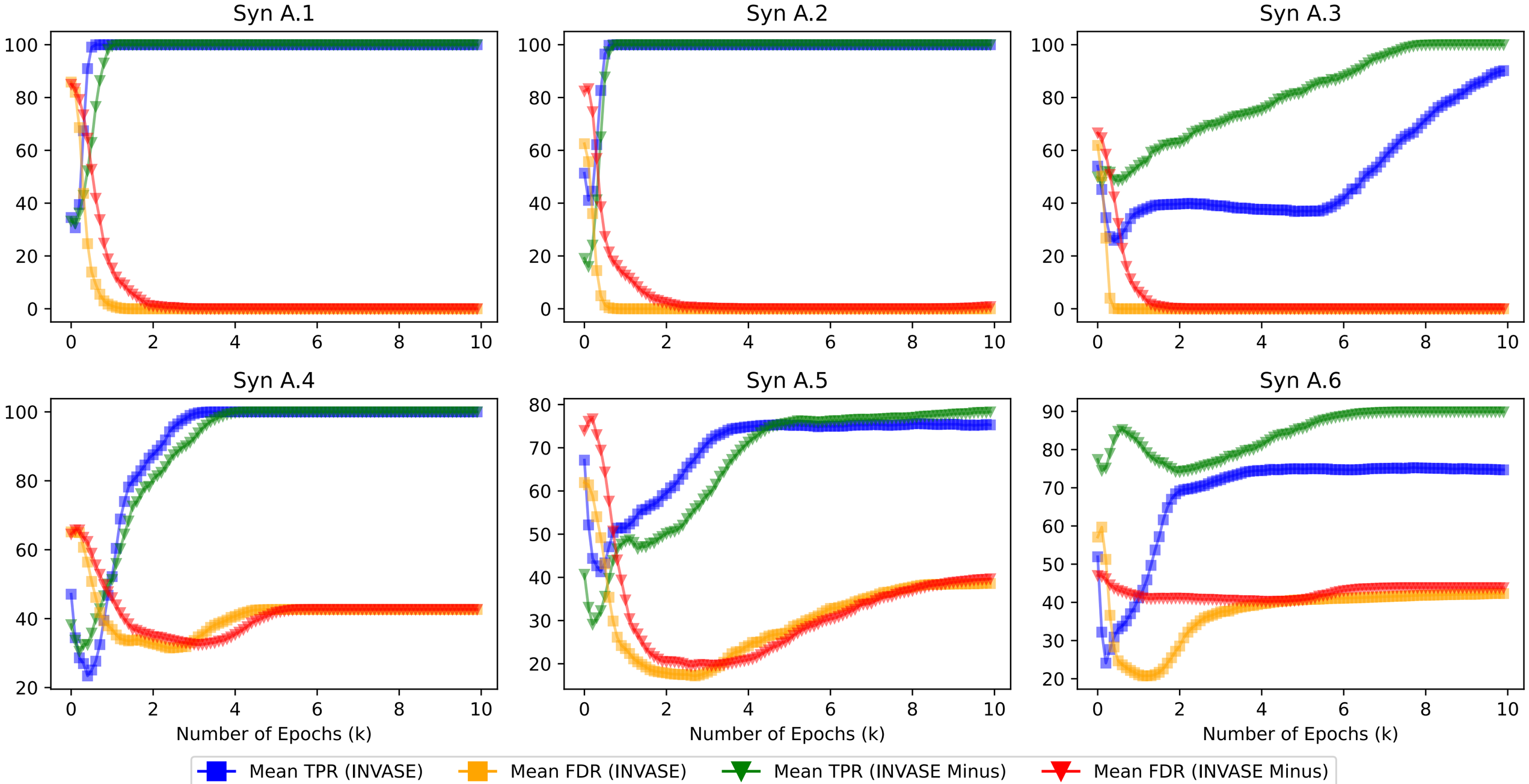
Analysis of Performance Influencing Factors in INVASE

The Impact of Baseline Model



Individual Settings: $\lambda = 0.1$, Activation: ReLU.

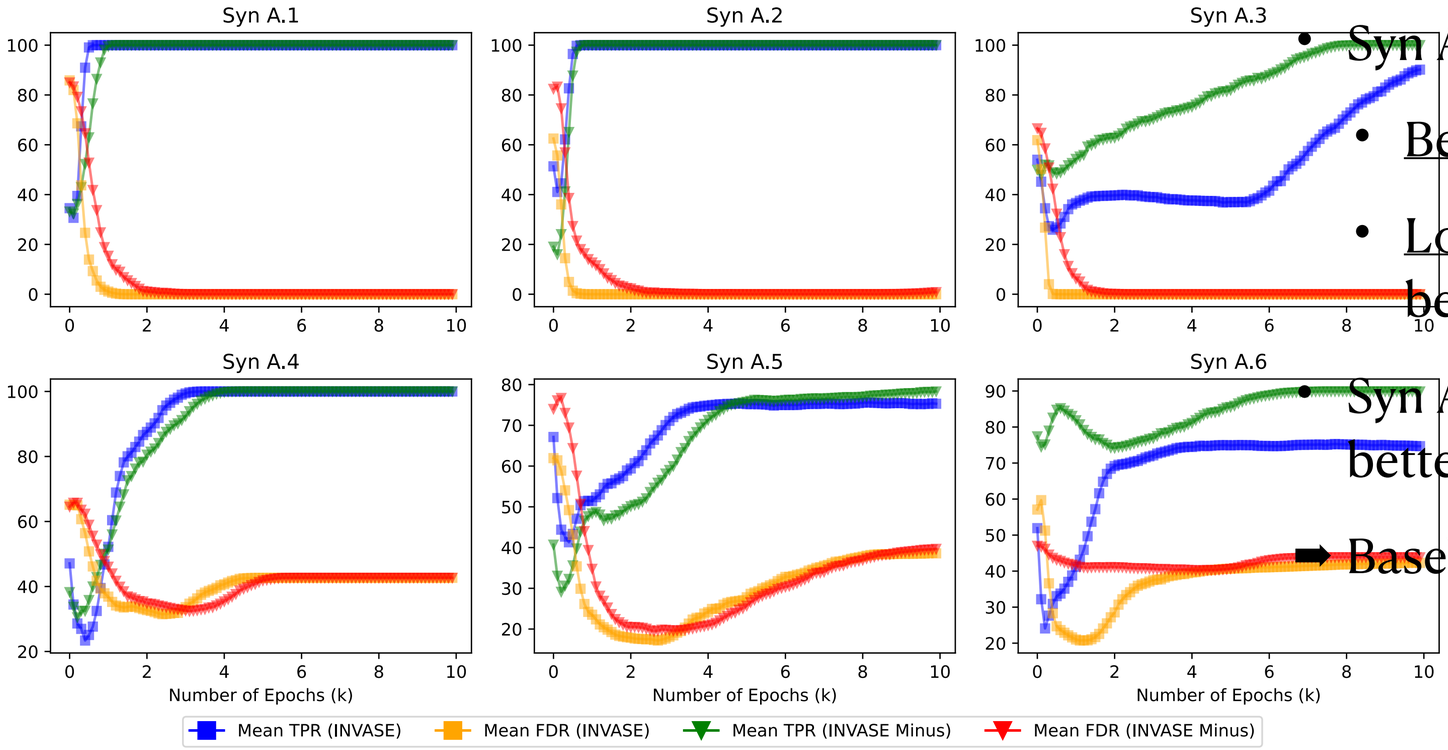
INVASE vs INVASE Minus

- Syn A.1, A.2, A.4: Roughly same
- Syn A.5: Very close
  - Beginning: INVASE slighlty better
  - Longer training: INVASE Minus slightly better
- Syn A.3 & A6: INVASE Minus ever perform better, especially on TPR

➡ Baseline model may not be redundant

Analysis of Performance Influencing Factors in INVASE

The Impact of Baseline Model ⟶ INVASE vs INVASE Minus



Individual Settings: $\lambda = 0.1$, Activation: ReLU.

- Syn A.1, A.2, A.4: Roughly same
- Syn A.5: Very close
  - Beginning: INVASE slighlty better
  - Longer training: INVASE Minus slightly better
- Syn A.3 & A6: INVASE Minus ever perform better, especially on TPR

➡ Baseline model may not be redundant

Analysis of Performance Influencing Factors in INVASE

Conclusion

1. Within the framework of INVASE, how do the following factors influence the explainer's ability to identify relevant features while discarding irrelevant ones?

   C. How sensitive is the explainer's performance to the choice of hyperparameter λ?