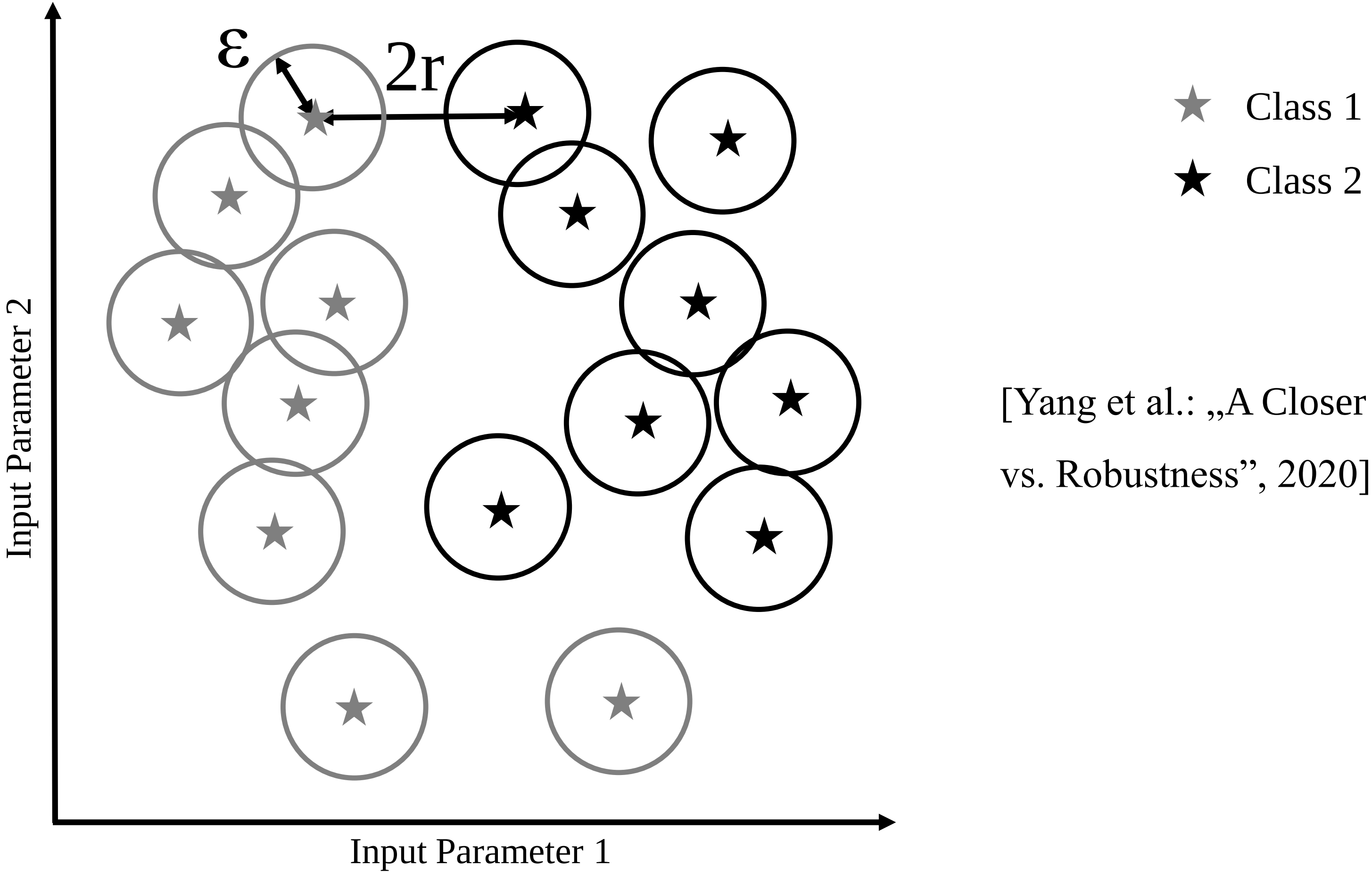
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

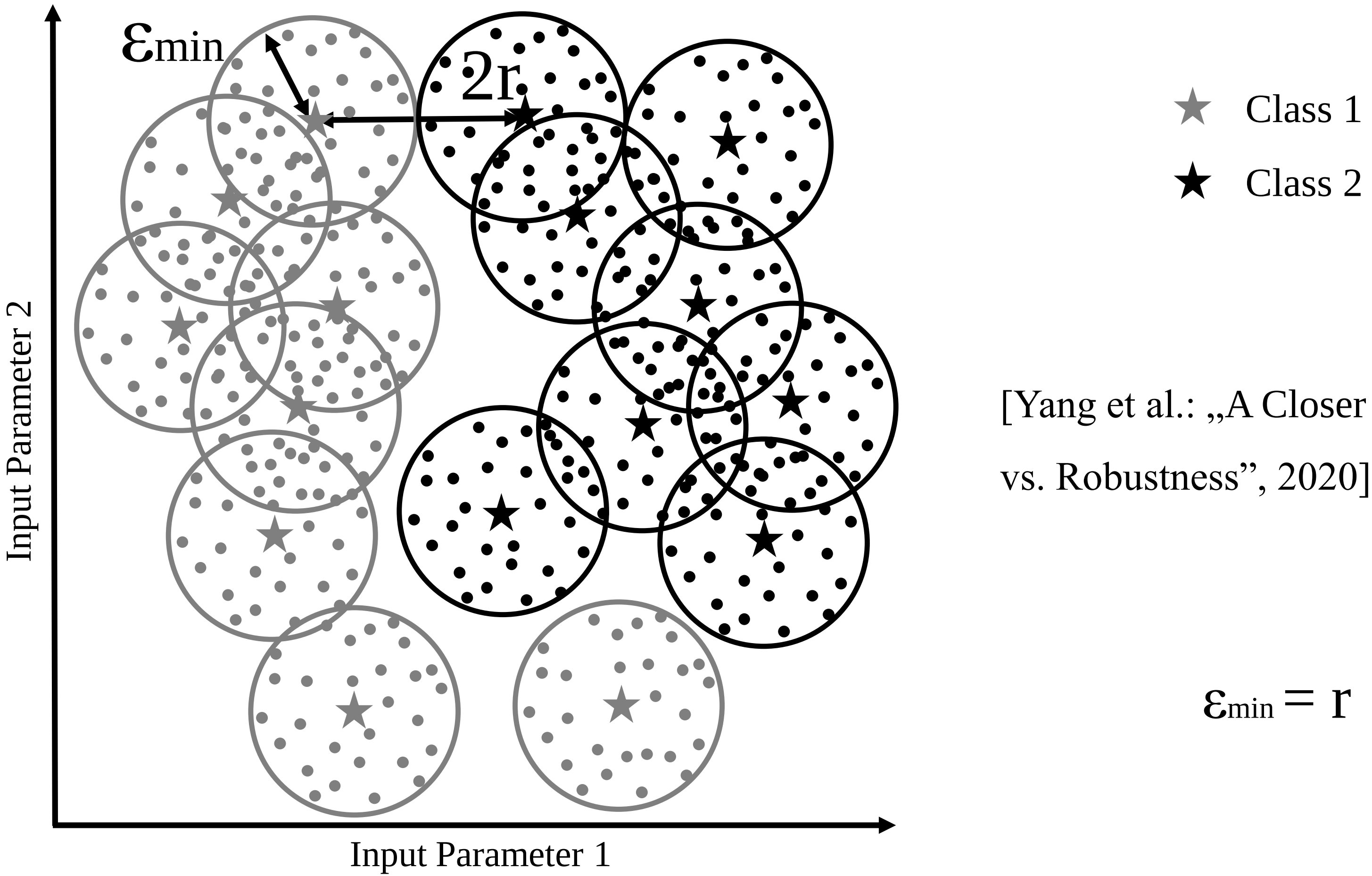
Utilizing Class Separation Distance for the Evaluation of Corruption Robustness of Machine Learning Classifiers

Georg Siedel, Silvia Vock, Andrey Morozov, Stefan Voß

Corruption Robustness Intuition

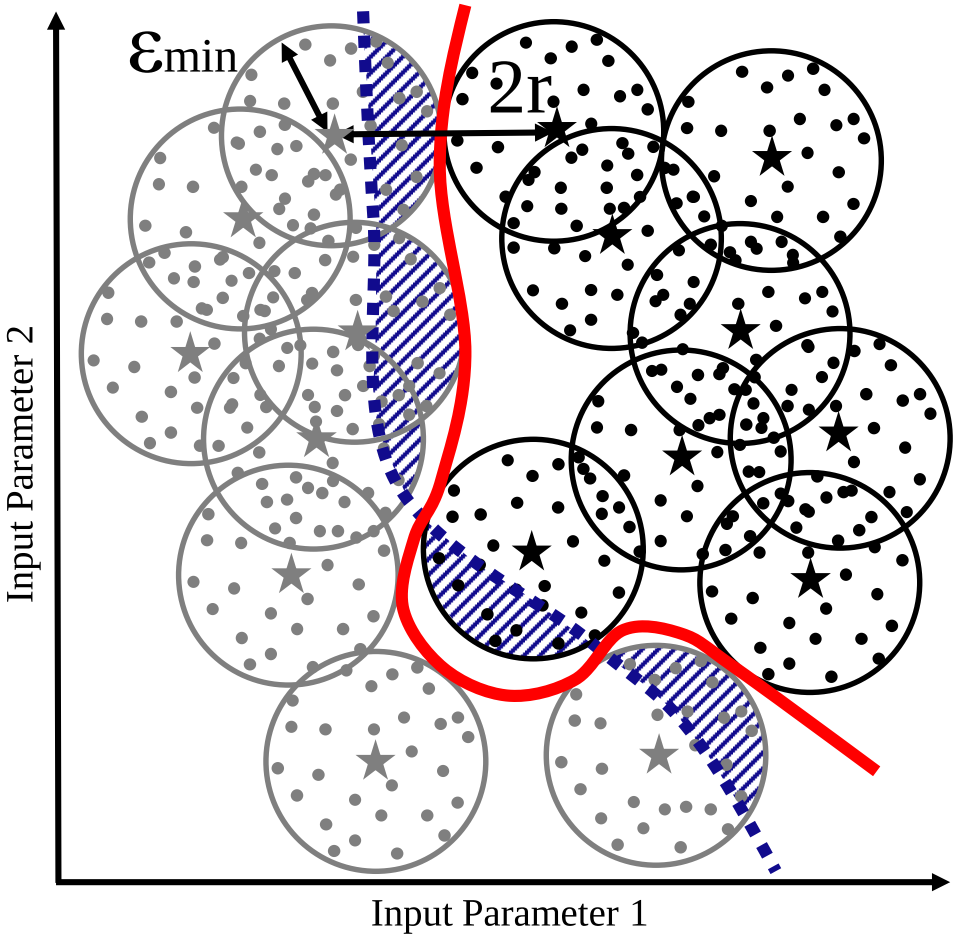


[Yang et al.: „A Closer Look at Accuracy vs. Robustness", 2020]



[Yang et al.: „A Closer Look at Accuracy vs. Robustness”, 2020]

$\varepsilon_{min} = r$



MSCR… Minimal Separation Corruption Robustness

$$MSCR = \frac{(Acc_{robust} - Acc_{original})}{Acc_{original}}$$

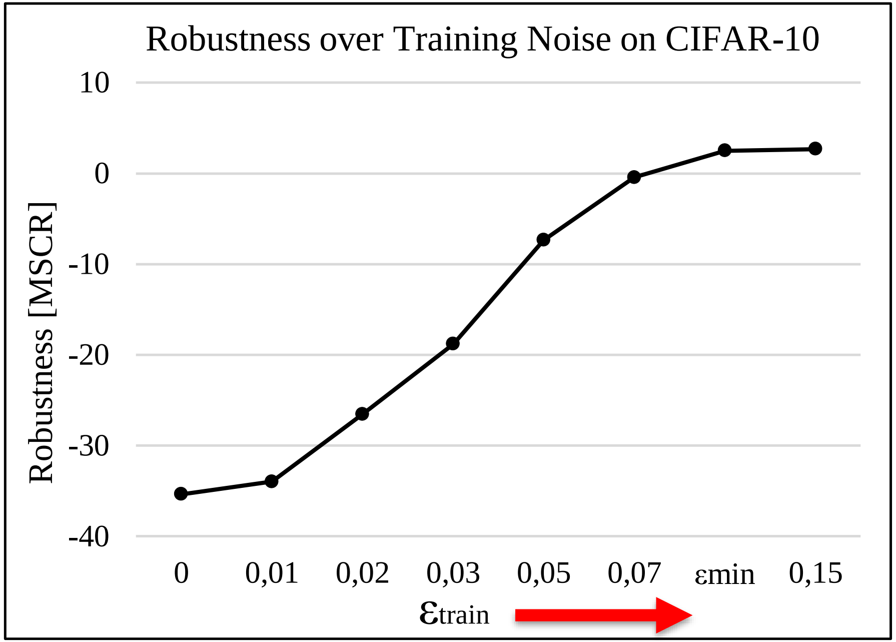
100% $Acc_{original} > Acc_{robust}$

→ MSCR < 0

100% $Acc_{original}$ = 100% Acc_{robust}

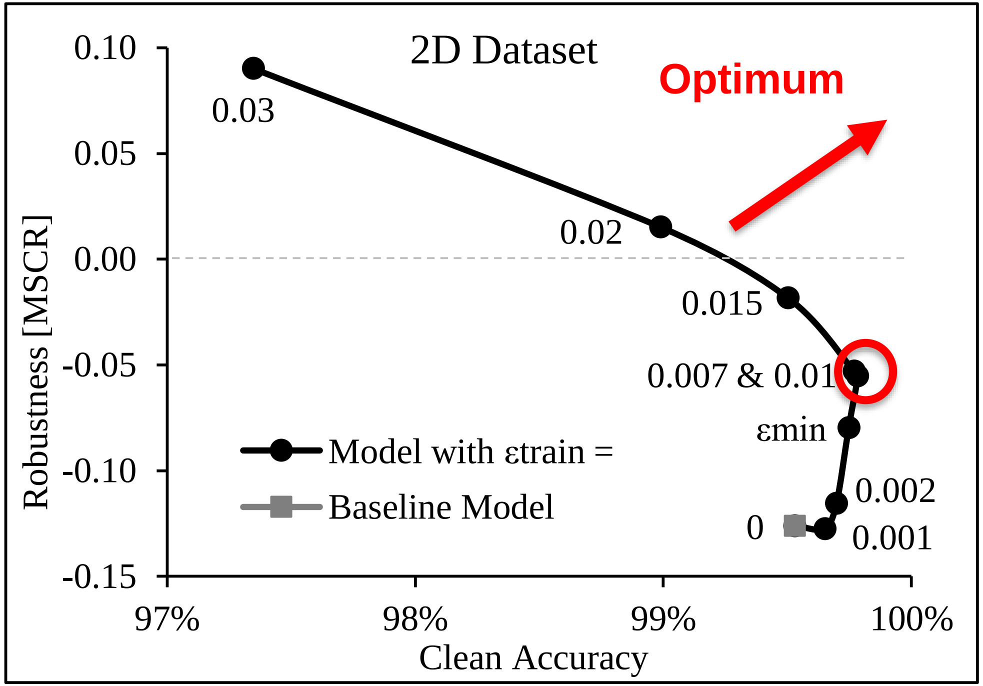
→ MSCR = 0

MSCR Metric Applicability



Trained for higher robustness

Accuracy-Robustness-Tradeoff?



What to take away?

MSCR metric:

- High interpretability

Accuracy-Robustness-Tradeoff:

- Not inherent in our experiments
- Sweetspots via Data Augmentation

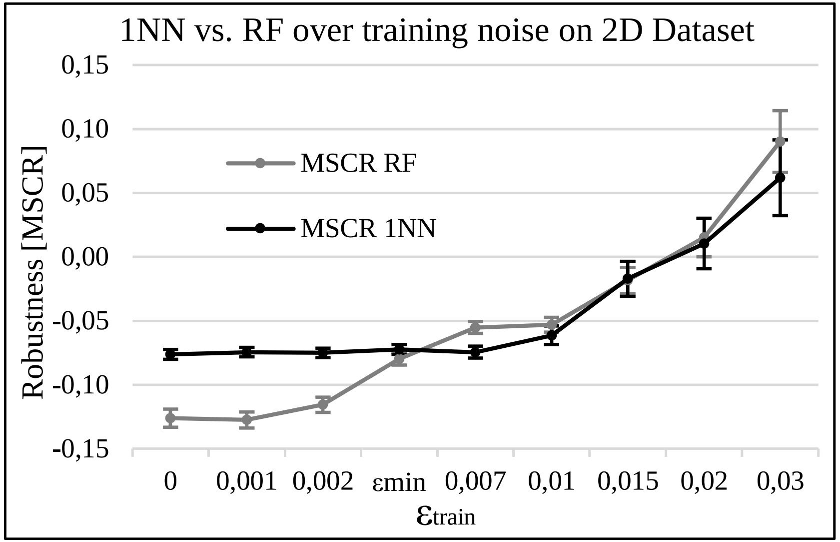
Get into contact:

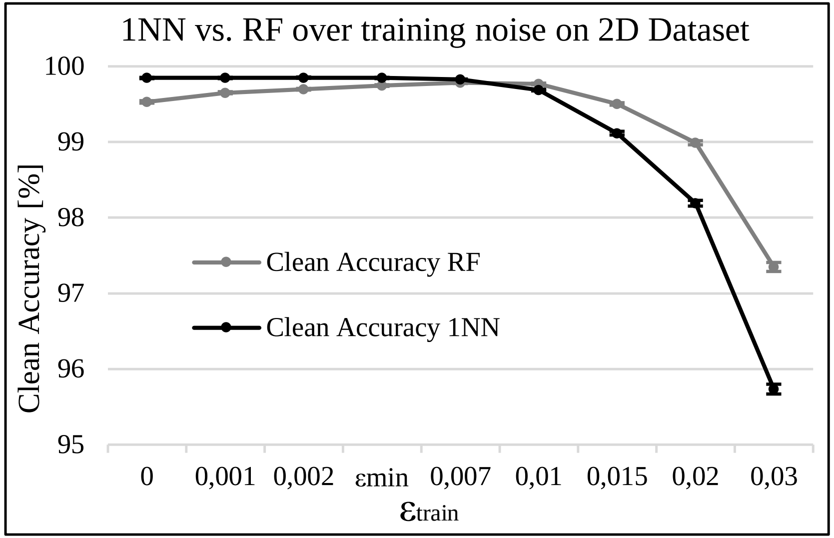
siedel.georg@baua.bund.de

Siedel et al.: Utilizing Class Separation Distance for the Evaluation of Corruption Robustness of Machine Learning Classifiers, The IJCAI-ECAI-22 Workshop on Artificial Intelligence Safety (AISafety 2022).

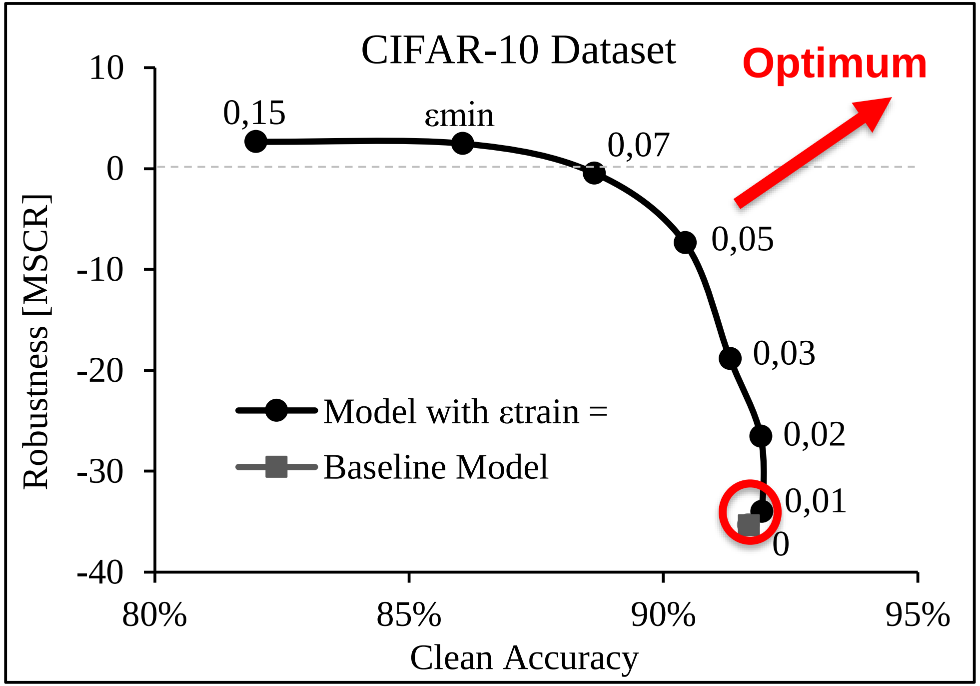
Backup

2D data: 1NN vs Random Forest





Accuracy-Robustness-Tradeoff?



CIFAR-10: Optima of Training vs. Test Noise