AutoML Project: DEHB for Constrained Multi-Objective Optimization

Guri Zabërgja

University of Freiburg

September 22, 2021

Motivation

- Differential evolution is conceptually simple yet a very powerful HPO method

Motivation

- Differential evolution is conceptually simple yet a very powerful HPO method
- Combining Differential Evolution and Hyperband has shown great results

Motivation

- Differential evolution is conceptually simple yet a very powerful HPO method
- Combining Differential Evolution and Hyperband has shown great results
- DEHB fulfills all the desiderata of a good HPO method [Awad et al., 2021]

Motivation

- Differential evolution is conceptually simple yet a very powerful HPO method
- Combining Differential Evolution and Hyperband has shown great results
- DEHB fulfills all the desiderata of a good HPO method [Awad et al., 2021]
- Stronger final performance compared to BOHB especially in high-dimensional discrete valued problems

Motivation

- Differential evolution is conceptually simple yet a very powerful HPO method
- Combining Differential Evolution and Hyperband has shown great results
- DEHB fulfills all the desiderata of a good HPO method [Awad et al., 2021]
- Stronger final performance compared to BOHB especially in high-dimensional discrete valued problems
- DEHB is computationally cheaper compared to most of BO methods

Motivation

- Differential evolution is conceptually simple yet a very powerful HPO method
- Combining Differential Evolution and Hyperband has shown great results
- DEHB fulfills all the desiderata of a good HPO method [Awad et al., 2021]
- Stronger final performance compared to BOHB especially in high-dimensional discrete valued problems
- DEHB is computationally cheaper compared to most of BO methods
- Idea: Extend DEHB to support constraints and multi-objective optimization

From DE to constrained MO-DE

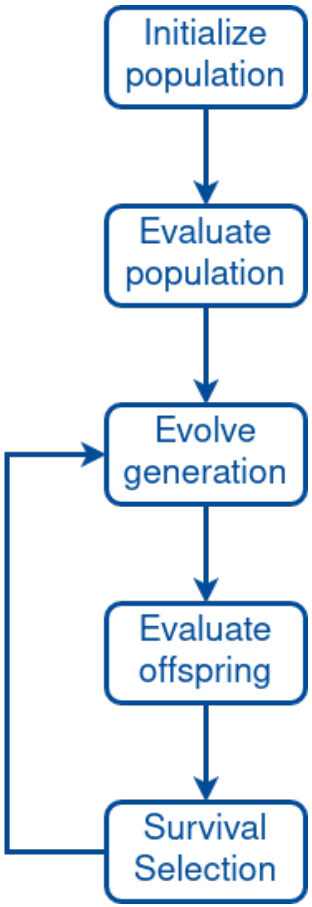


Figure: DE Procedure

From DE to constrained MO-DE

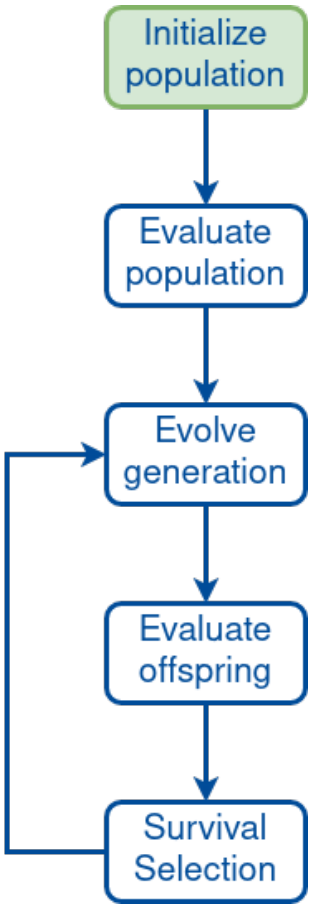


Figure: DE Procedure

From DE to constrained MO-DE

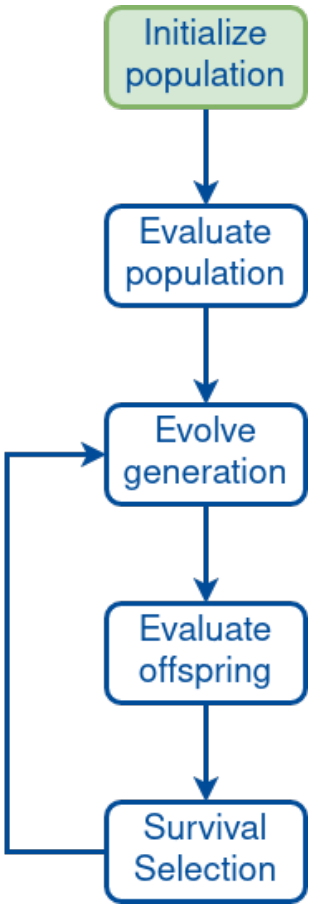


Figure: DE Procedure

- Randomly initialize population according to the maximum model size constraint

From DE to constrained MO-DE

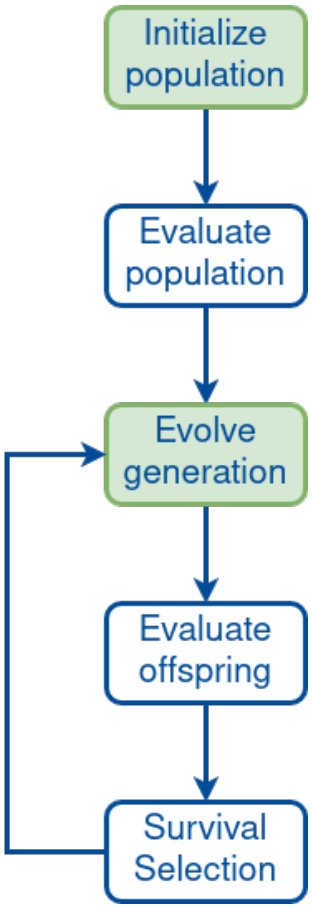


Figure: DE Procedure

- Randomly initialize population according to the maximum model size constraint

From DE to constrained MO-DE

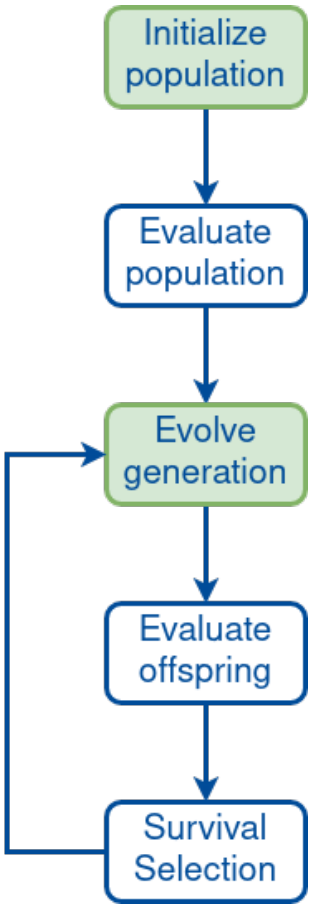


Figure: DE Procedure

- Randomly initialize population according to the maximum model size constraint
- Repeat mutation and crossover until offspring satisfies the maximum model size constraint

From DE to constrained MO-DE

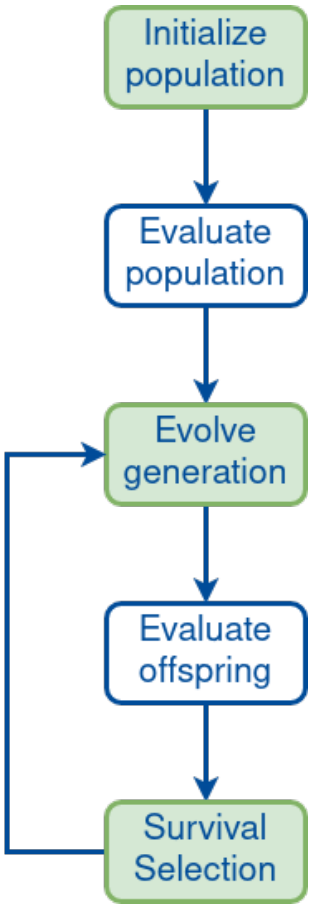


Figure: DE Procedure

- Randomly initialize population according to the maximum model size constraint
- Repeat mutation and crossover until offspring satisfies the maximum model size constraint

From DE to constrained MO-DE

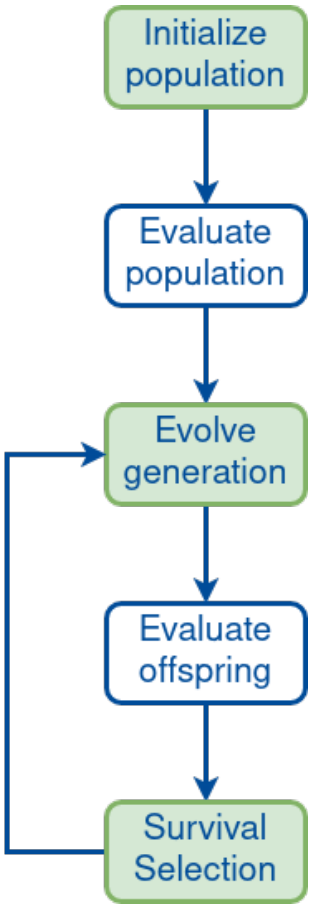


Figure: DE Procedure

- Randomly initialize population according to the maximum model size constraint
- Repeat mutation and crossover until offspring satisfies the maximum model size constraint
- Integrate precision maximization as an additional objective, to guide the search towards regions satisfying the minimum precision constraint

From DE to constrained MO-DE

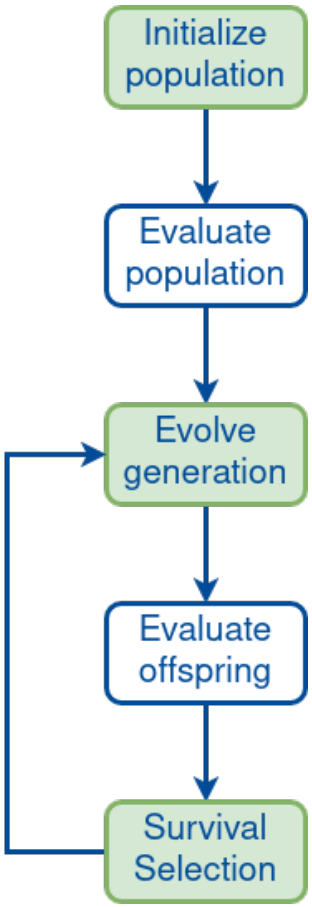


Figure: DE Procedure

- Randomly initialize population according to the maximum model size constraint
- Repeat mutation and crossover until offspring satisfies the maximum model size constraint
- Integrate precision maximization as an additional objective, to guide the search towards regions satisfying the minimum precision constraint
- Select offspring that improve upon the Top-3 Accuracy objective and the additional Precision objective

Constrained MO-DEHB: Precision objective

- Goal: Encourage search towards regions that satisfy the precision constraint

Constrained MO-DEHB: Precision objective

- Goal: Encourage search towards regions that satisfy the precision constraint

$$c_2(\mathbf{x}) = \begin{cases} C - \textit{precision}(\mathbf{x}), & \textit{if} \quad \textit{precision}(\mathbf{x}) < C \\ 0, & \textit{otherwise} \end{cases}$$

where C is the precision constraint value

Constrained MO-DEHB: Precision objective

- Goal: Encourage search towards regions that satisfy the precision constraint

$$c_2(\mathbf{x}) = \begin{cases} C - \textit{precision}(\mathbf{x}), & \textit{if} \quad \textit{precision}(\mathbf{x}) < C \\ 0, & \textit{otherwise} \end{cases}$$

where C is the precision constraint value

- Ensuring exploration: Keep an archive of offspring that don't survive selection[Wang and Cai, 2012]

Constrained MO-DEHB: Precision objective

- Goal: Encourage search towards regions that satisfy the precision constraint

$$c_2(\mathbf{x}) = \begin{cases} C - \textit{precision}(\mathbf{x}), & \textit{if} \quad \textit{precision}(\mathbf{x}) < C \\ 0, & \textit{otherwise} \end{cases}$$

where C is the precision constraint value

- Ensuring exploration: Keep an archive of offspring that don't survive selection[Wang and Cai, 2012]
- Periodically empty the archive by replacing the individuals in the population except the best one

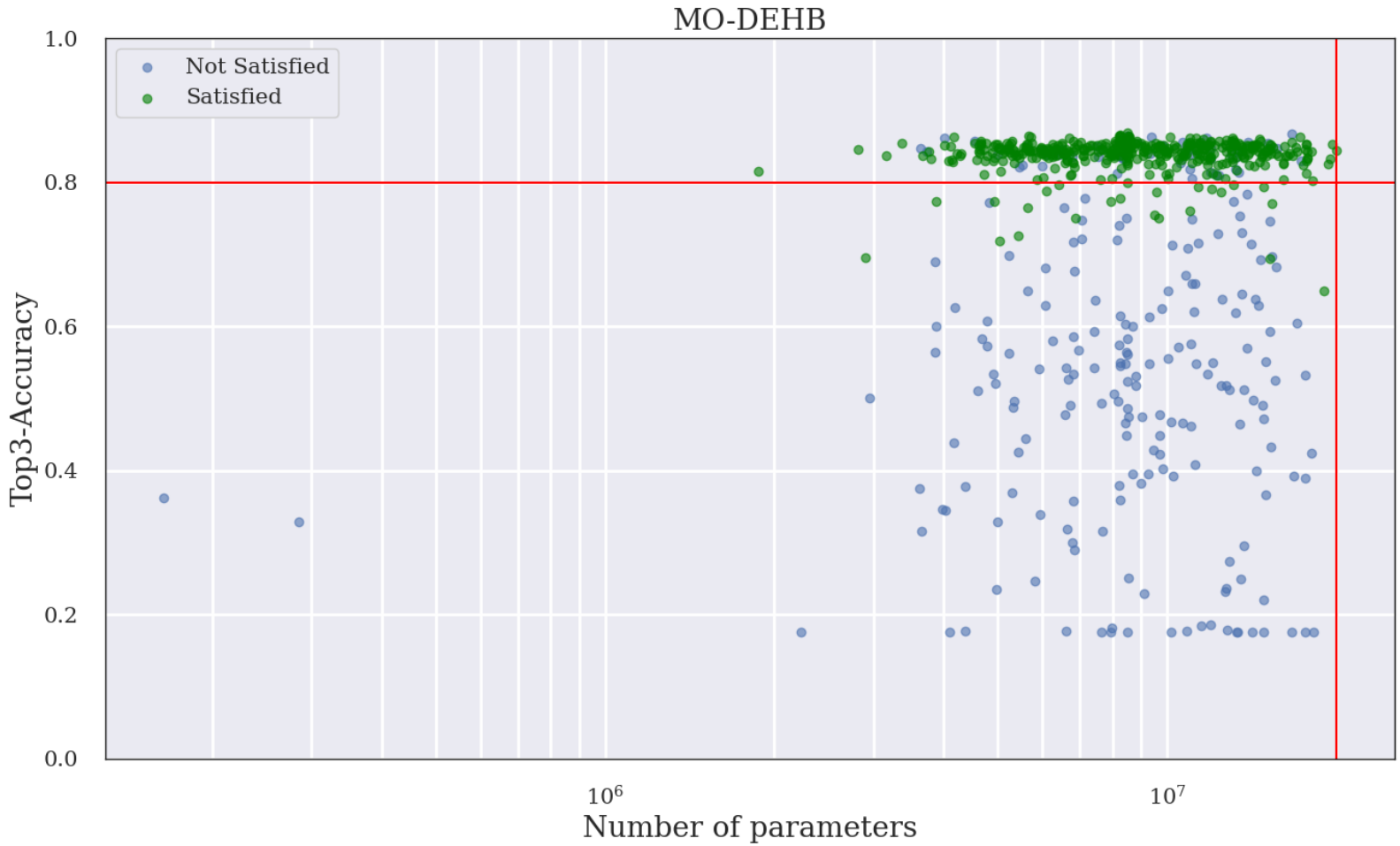
Configuration Space

Configuration Space			
Parameter Name	Range	Parameter Name	Range
n_conv_layers	(1, 3)	n_channels_conv_0	(512, 2048)
n_channels_conv_1	(512, 2048)	n_channels_conv_2	(512, 2048)
n_fc_layers	(1, 3)	n_channels_fc_0	(64, 256)
n_channels_fc_1	(64, 256)	n_channels_fc_2	(64, 256)
batch_size	(128, 512)	global_avg_pooling	[True, False]
kernel_size	[2, 3, 4, 5]	learning_rate	(0.00001, 0.1)
use_BN	[True, False]	dropout_rate	(0.0, 1.0)
optimizer	[SGD, Adam, AdamW]	sgd_momentum	(0.0, 0.99)
weight_decay	(0.00001, 0.1)	/	/

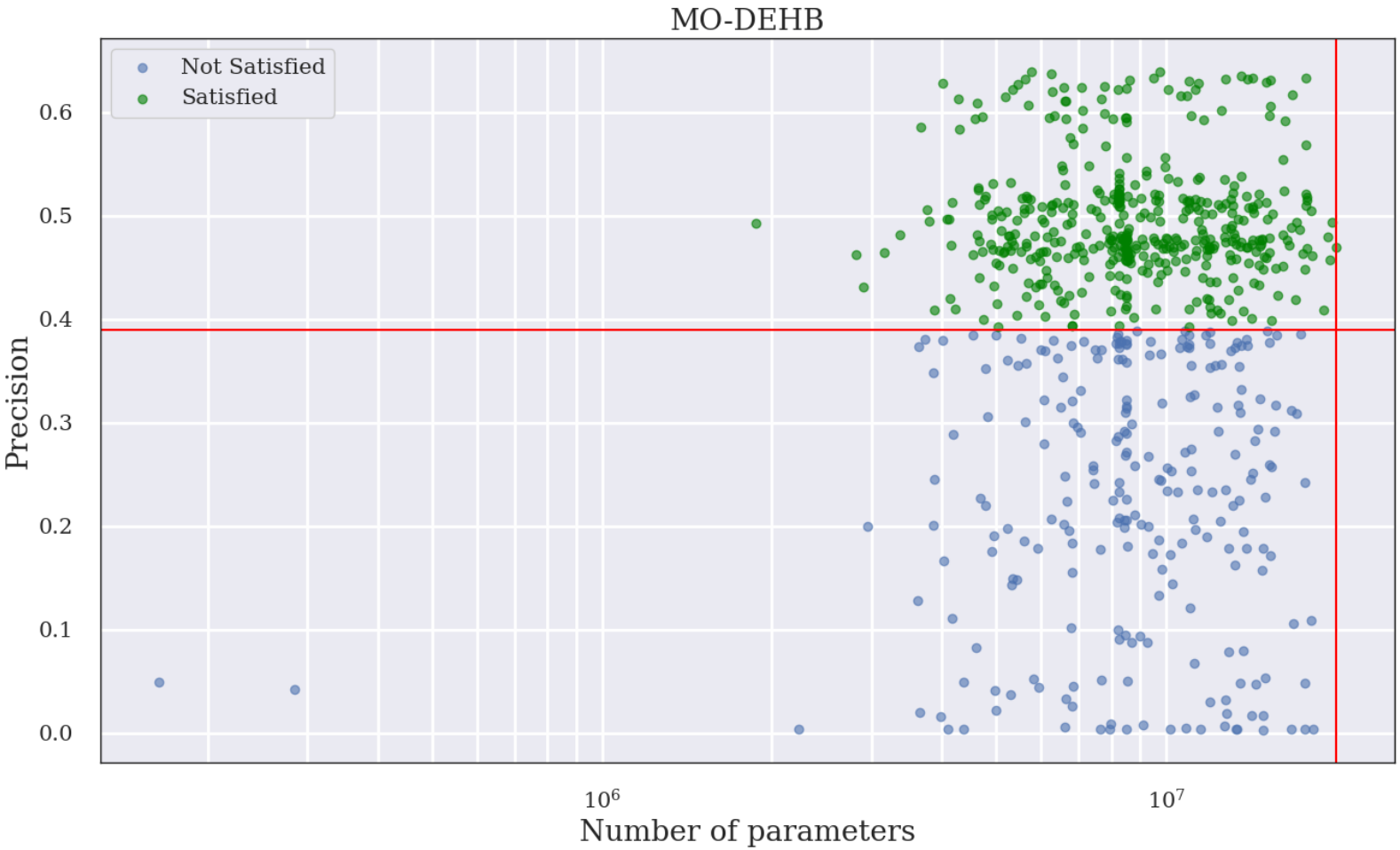
Hardware used

- Personal Computer
- Nvidia RTX 2080
- 16 GB RAM

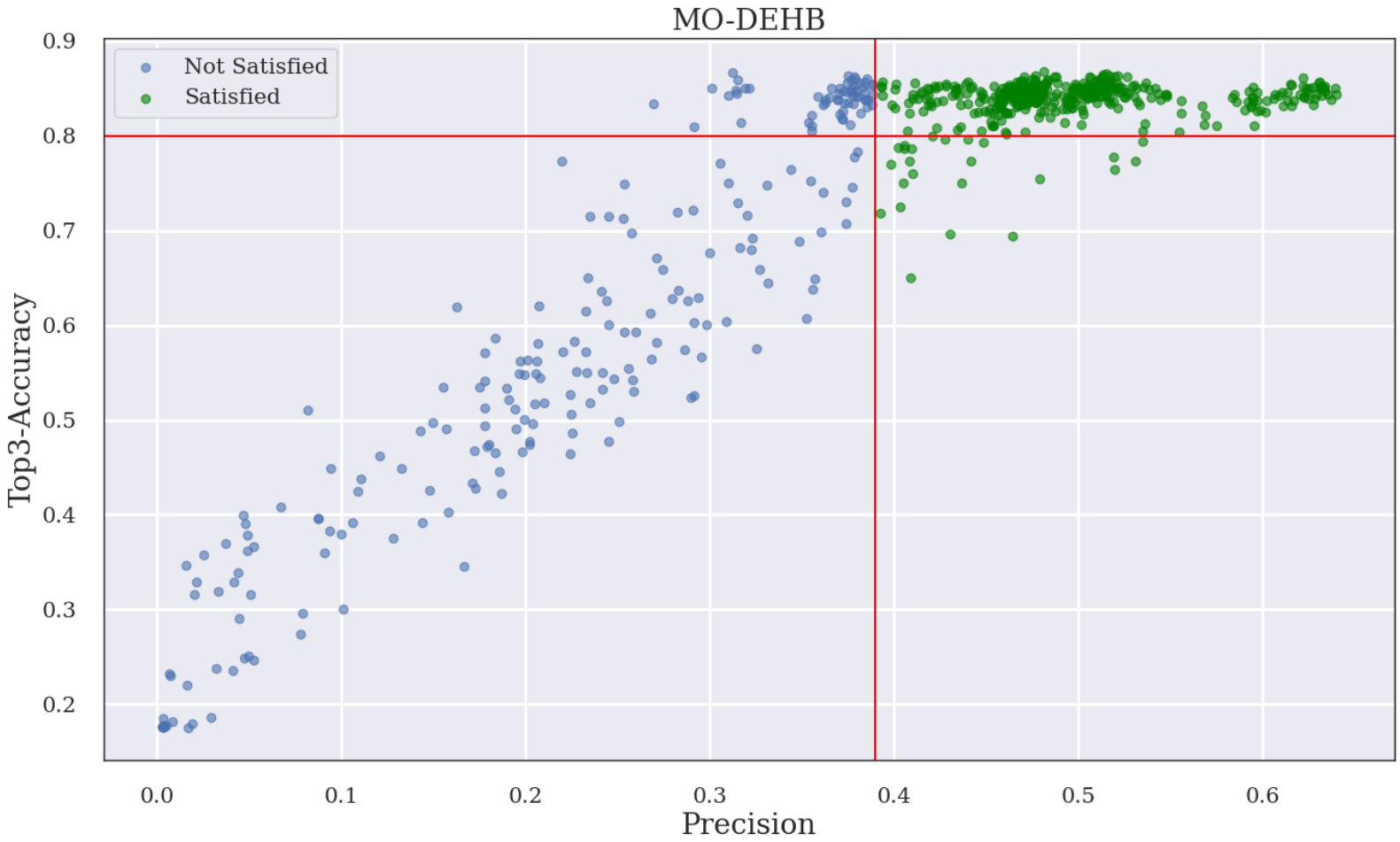
Results: Top3-Accuracy vs Model Size



Results: Precision vs Model Size



Results: Top3-Accuracy vs Precision

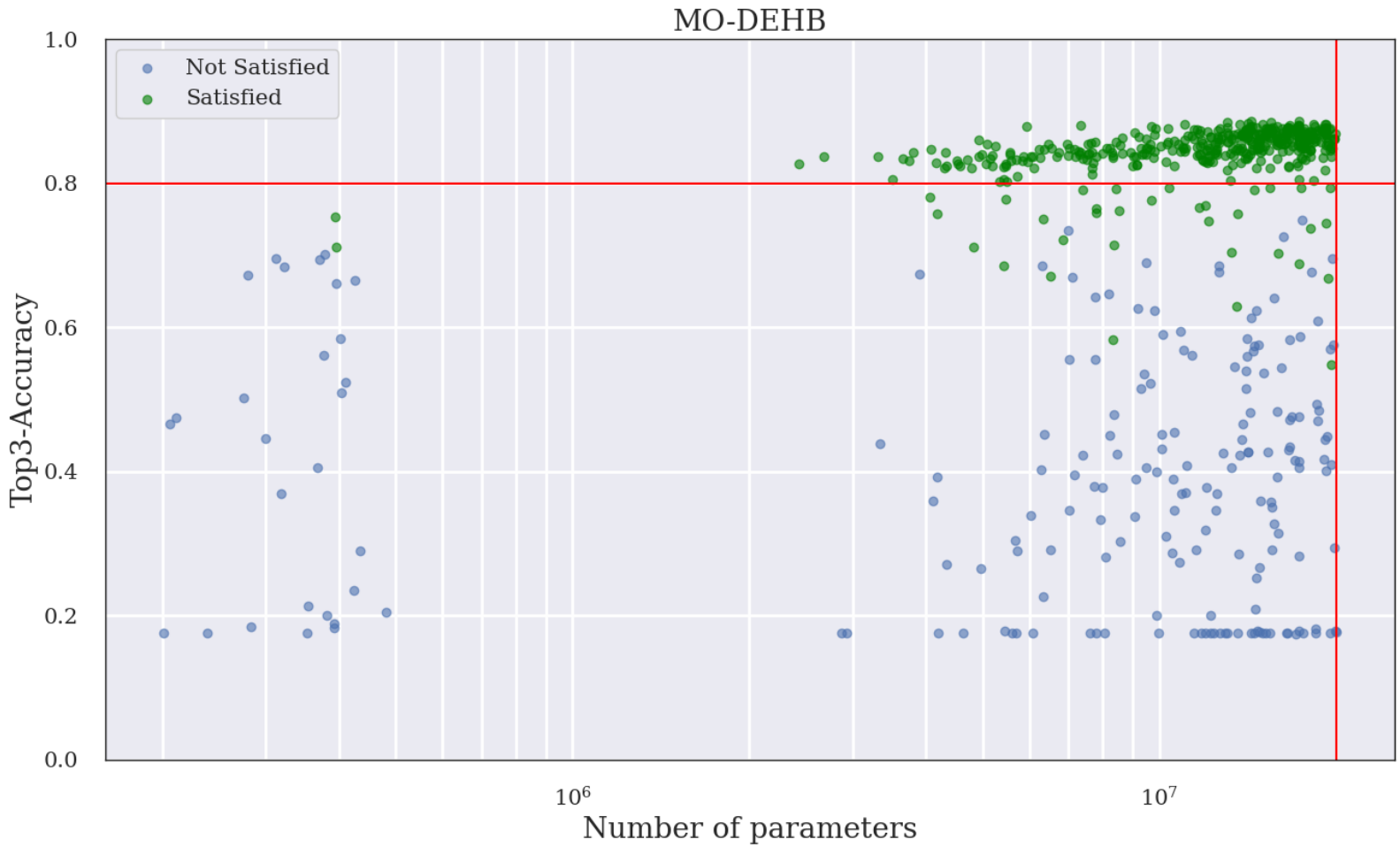


Future Work

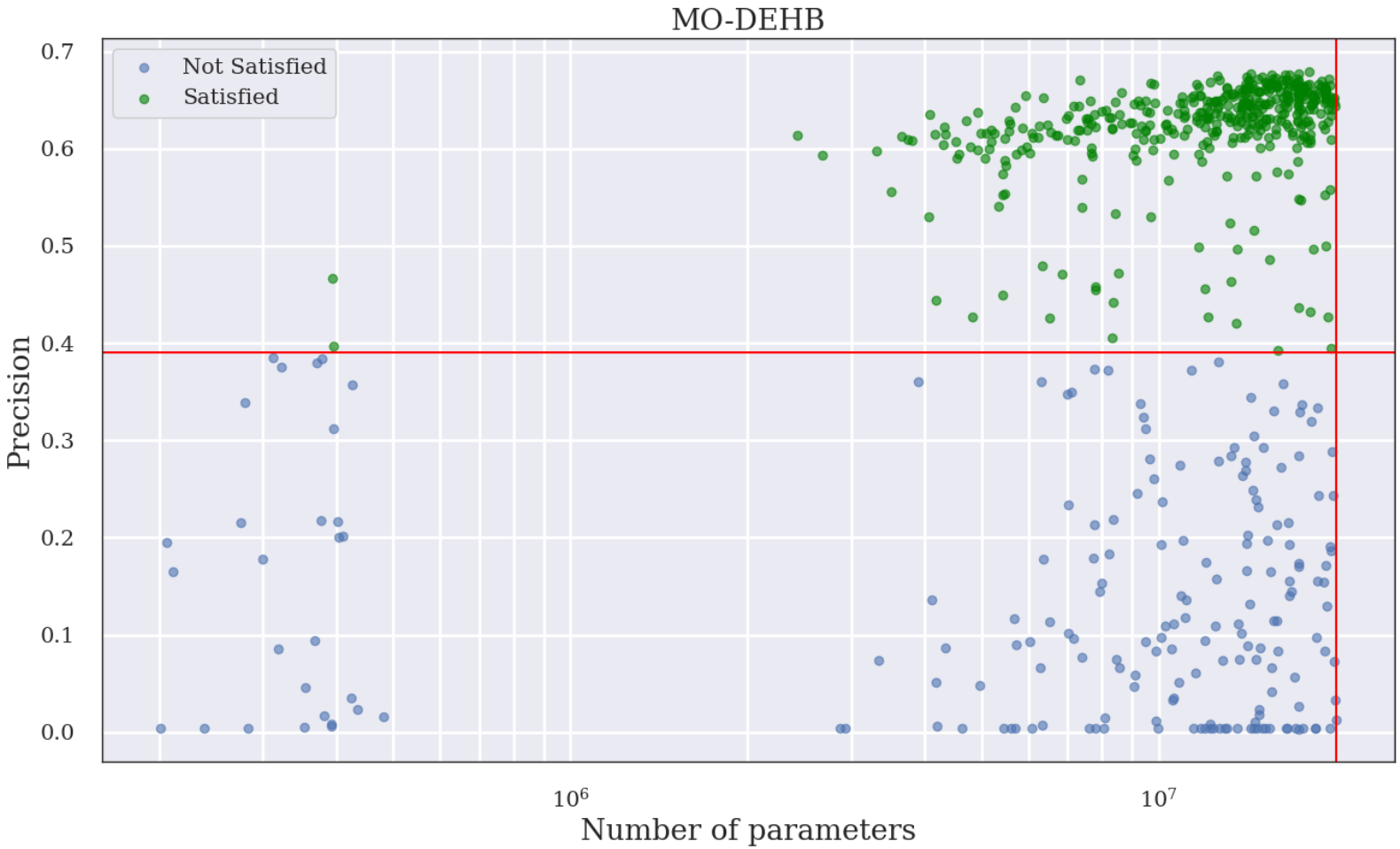
- DE is sensitive to its own hyperparameters
- Tune DE's hyperparameters (mutation factor, crossover rate, mutation strategy)
- Use data augmentation and add it in the configuration space to find the best ones

Thank You!

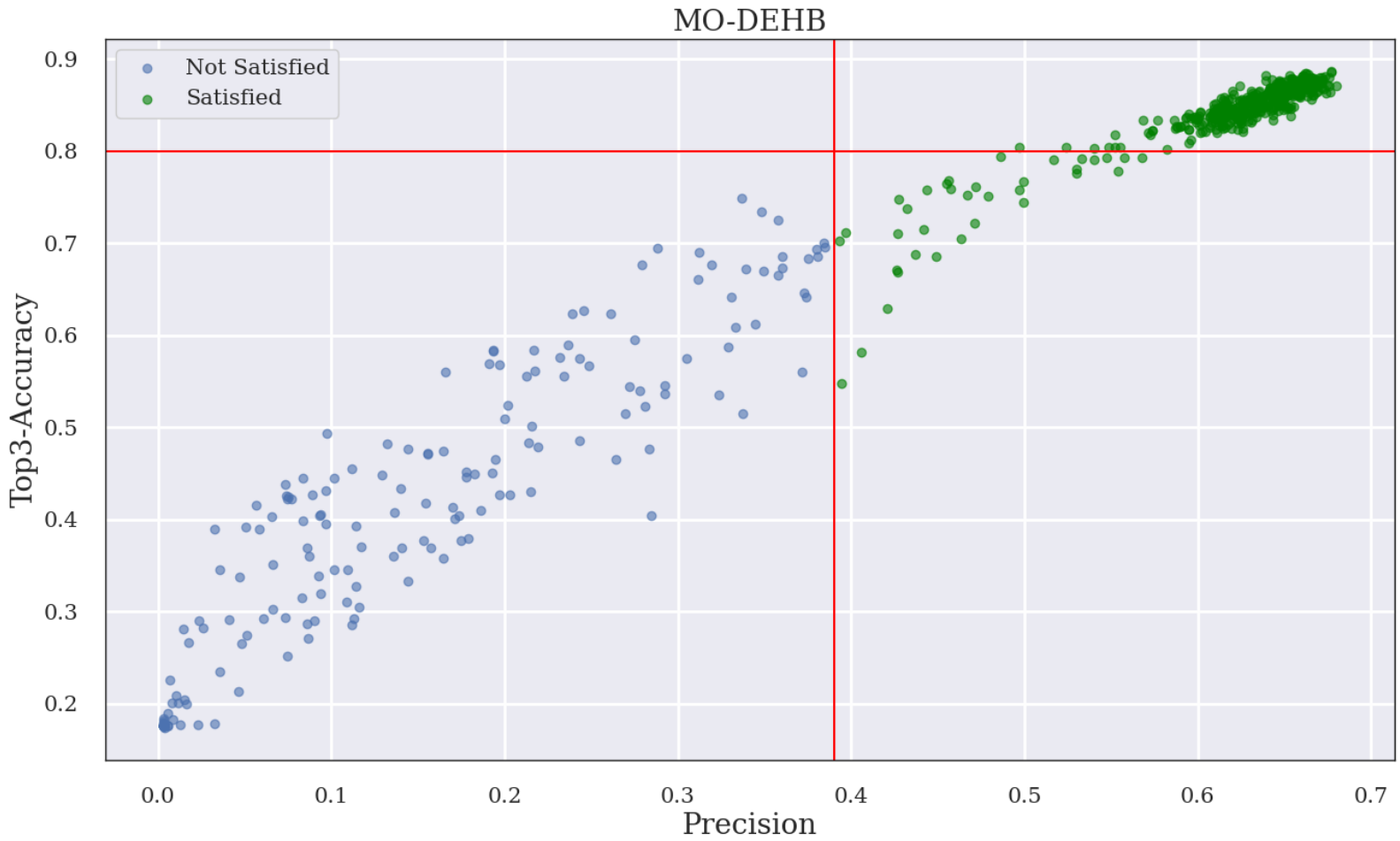
Results: Top3-Accuracy vs Model Size evaluated on full val set



Results: Precision vs Model Size evaluated on full val set

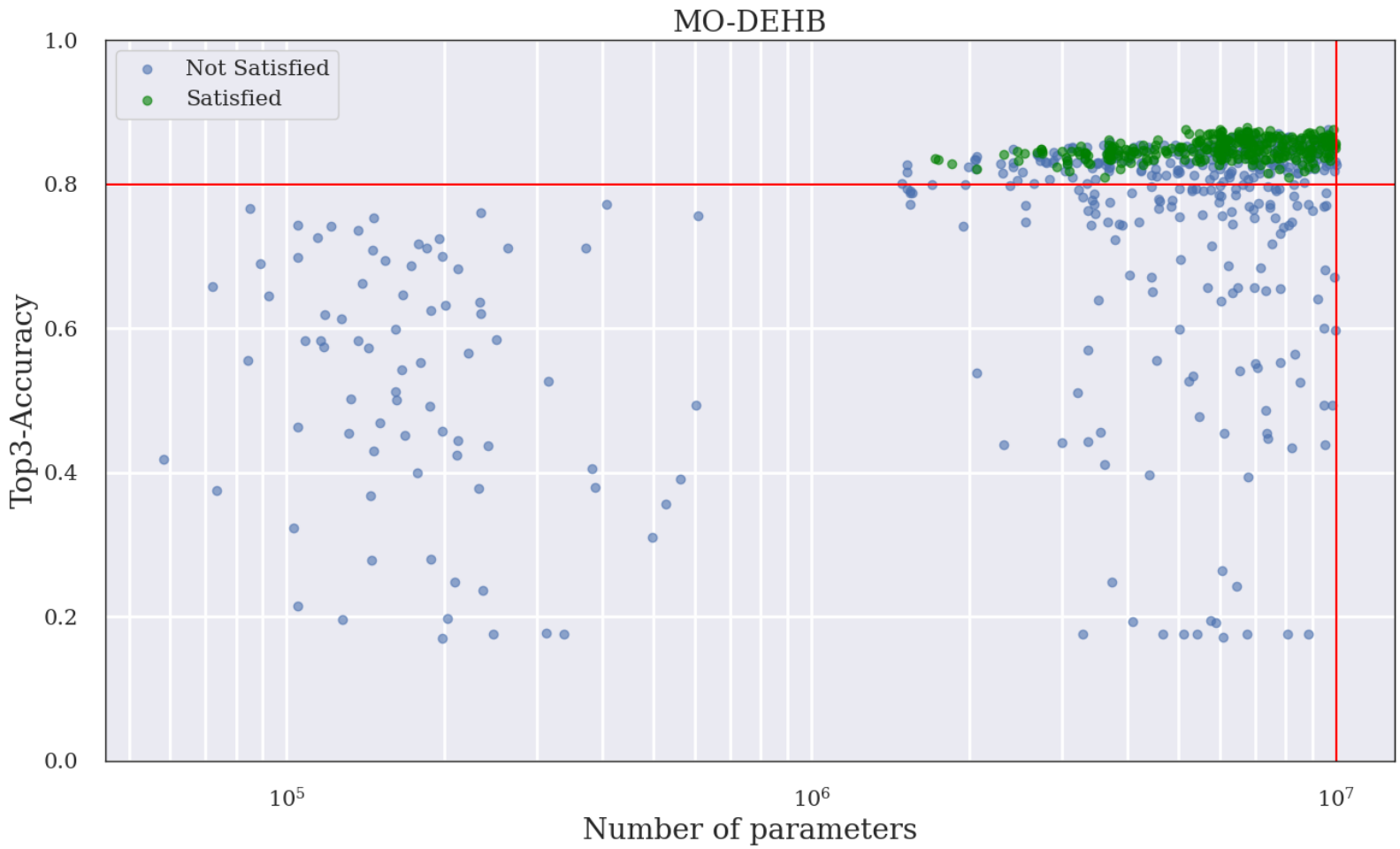


Results: Top3-Accuracy vs Precision evaluated on full val set



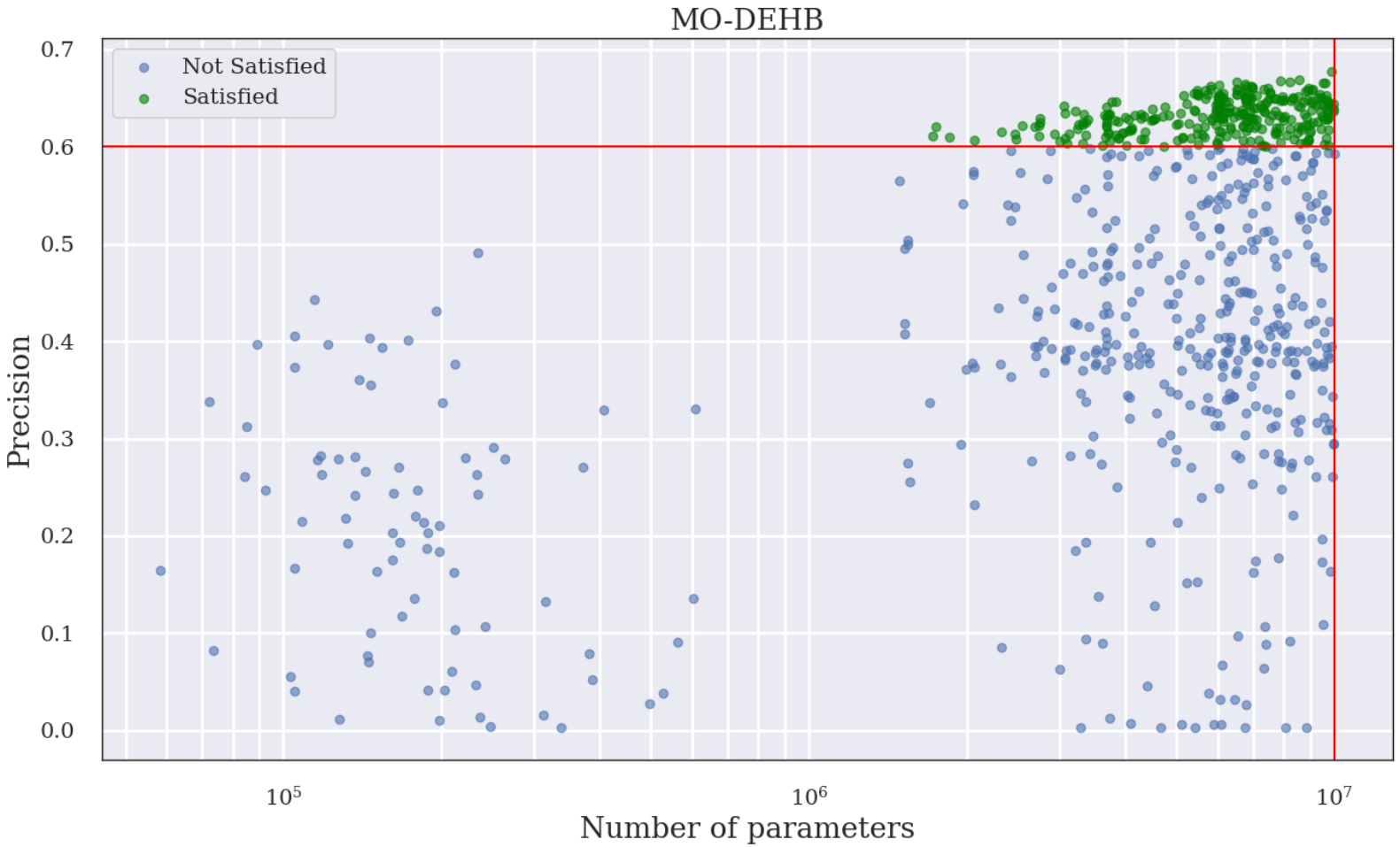
Results: Top3-Accuracy vs Model Size

Model size constraint: 10 million, Precision constraint: 0.6



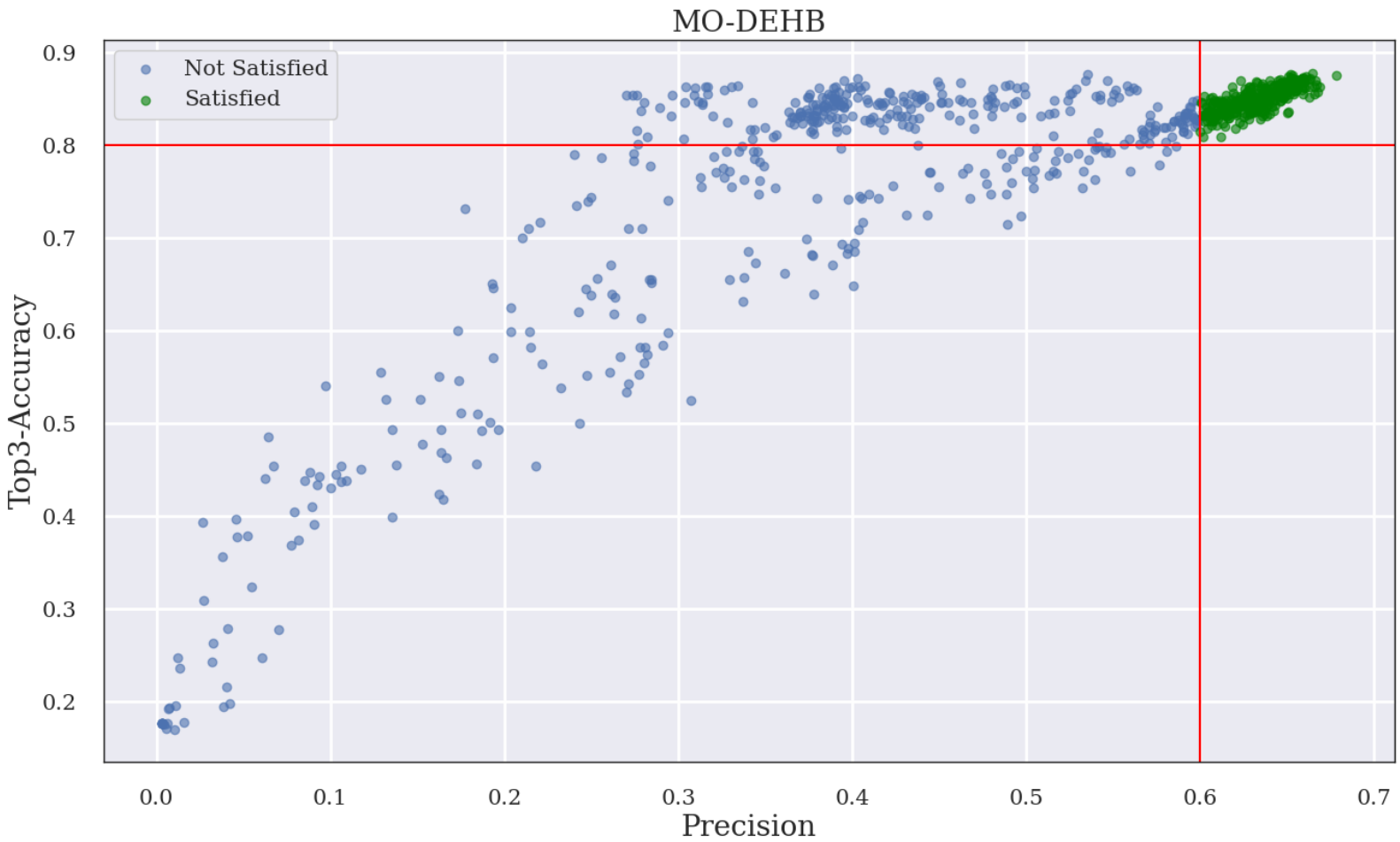
Results: Precision vs Model Size

Model size constraint: 10 million, Precision constraint: 0.6



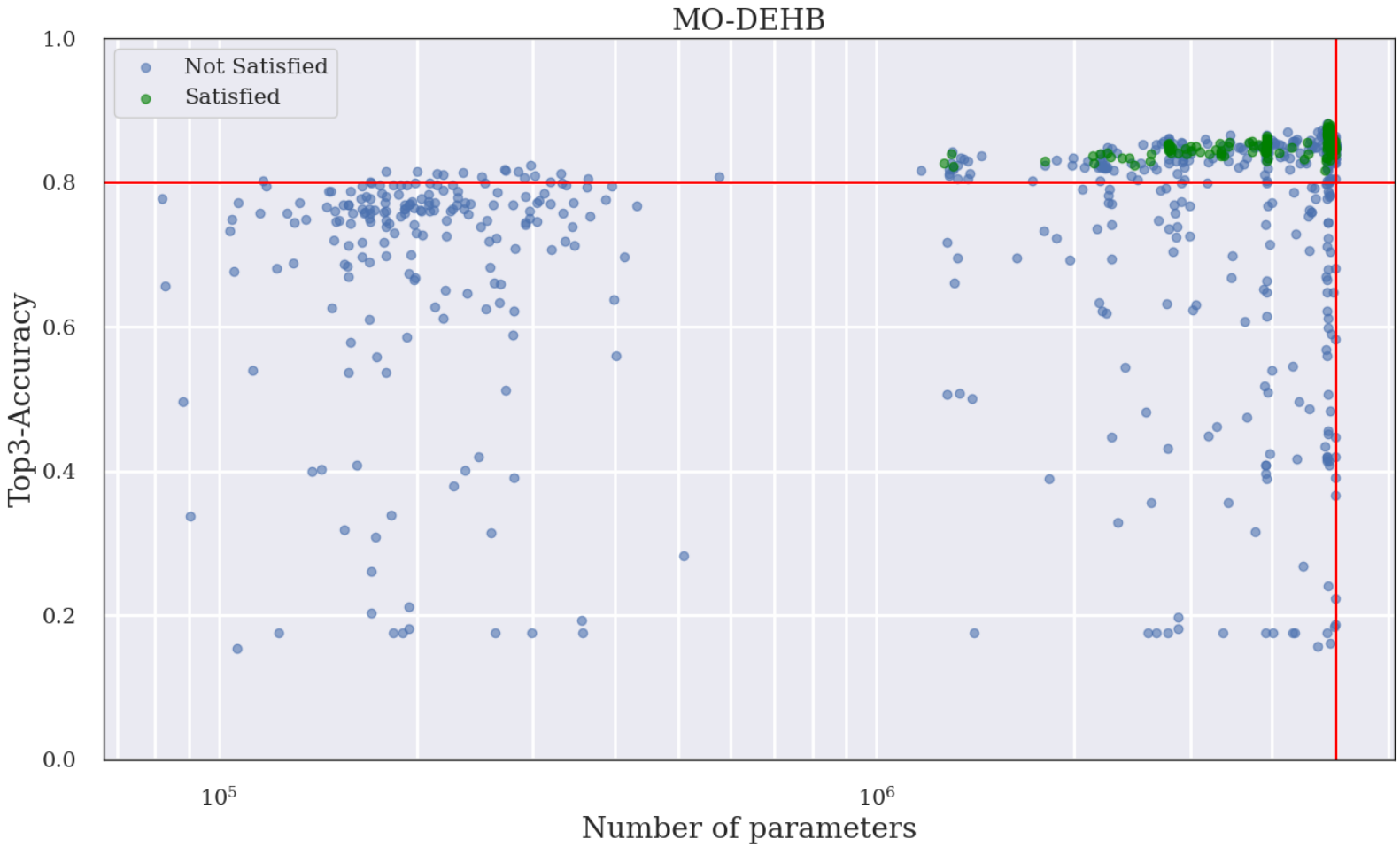
Results: Top3-Accuracy vs Precision

Model size constraint: 10 million, Precision constraint: 0.6



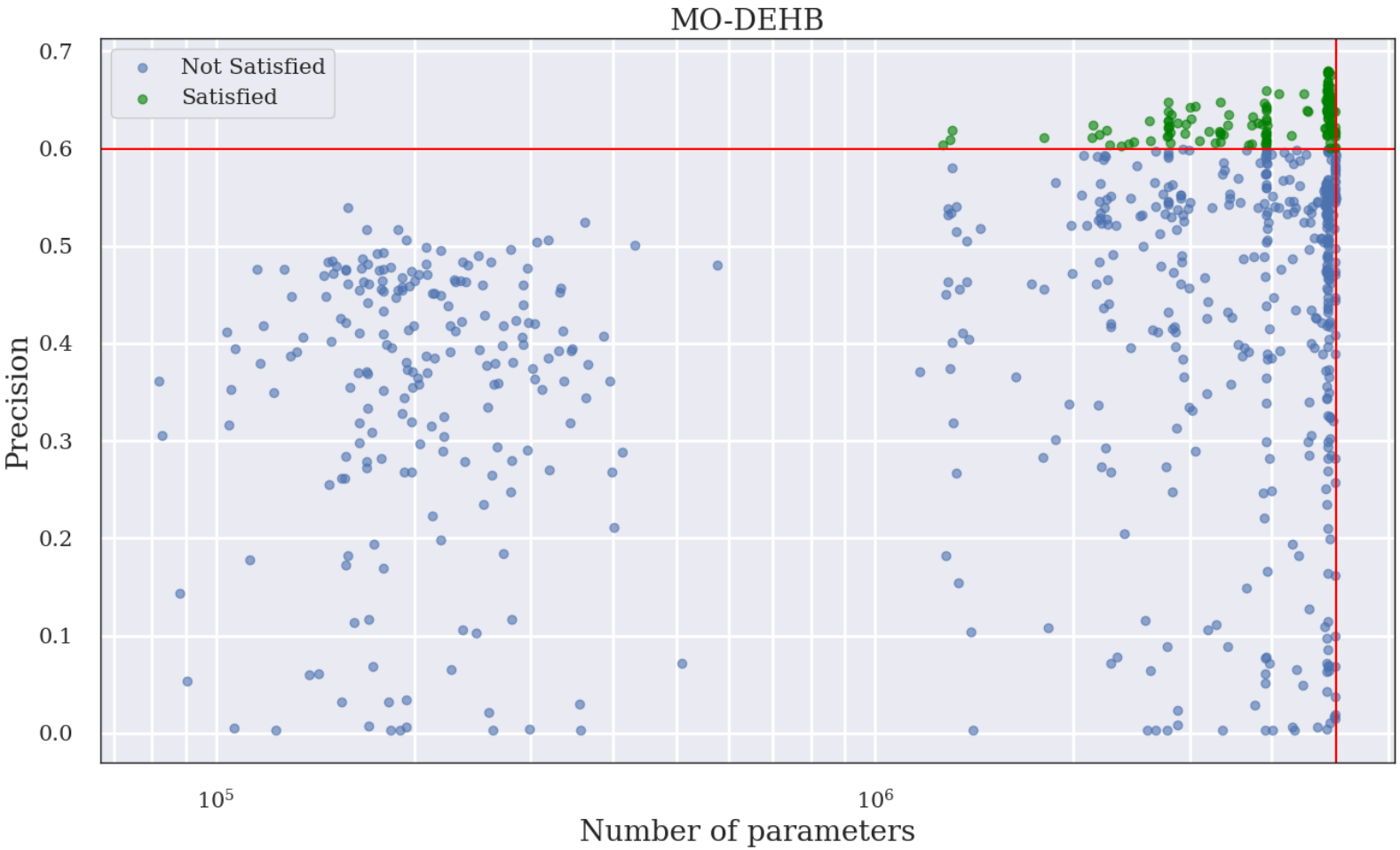
Results: Top3-Accuracy vs Model Size

Model size constraint: 5 million, Precision constraint: 0.6



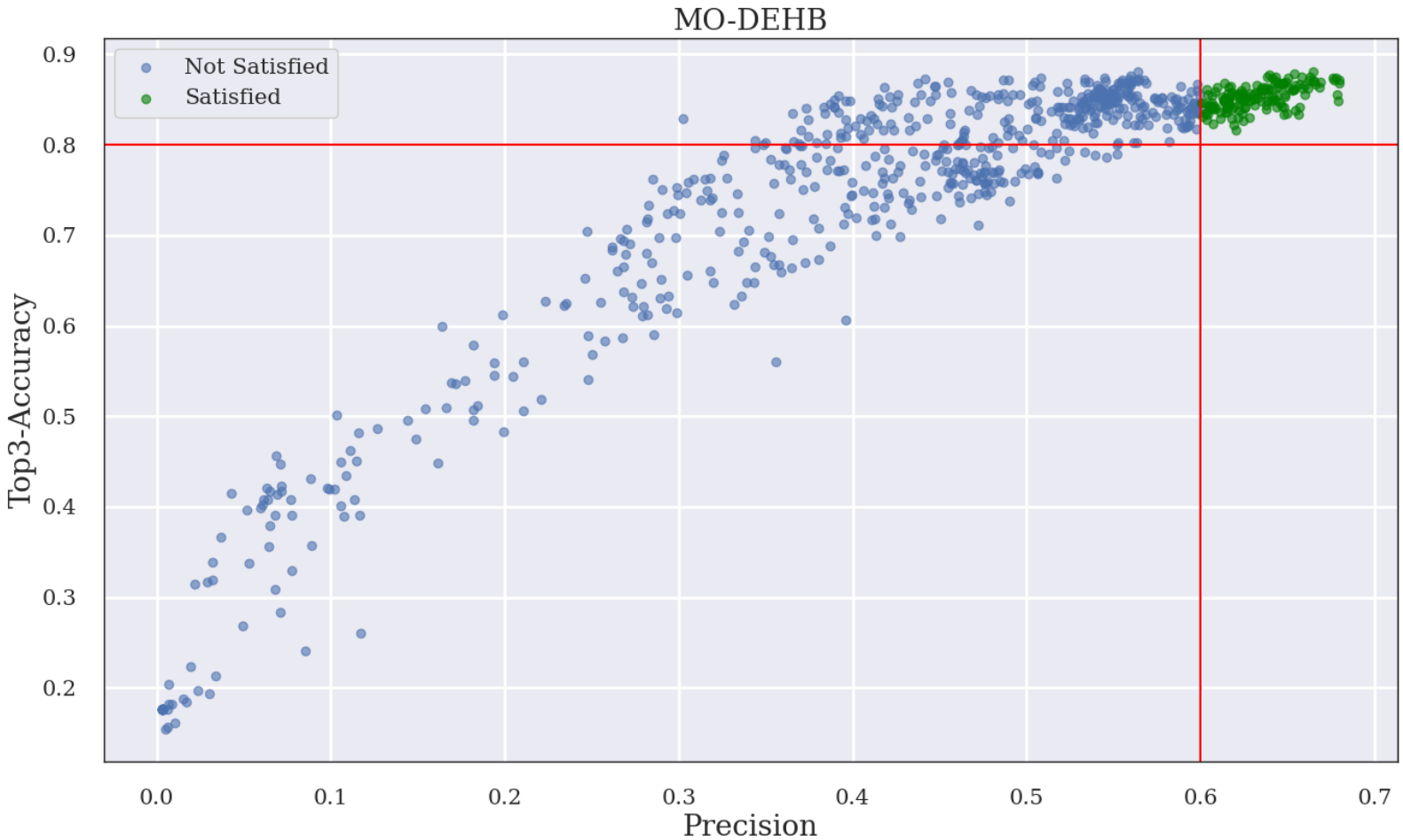
Results: Precision vs Model Size

Model size constraint: 5 million, Precision constraint: 0.6



Results: Top3-Accuracy vs Precision

Model size constraint: 5 million, Precision constraint: 0.6




Awad, N., Mallik, N., and Hutter, F. (2021).
Dehb: Evolutionary hyberband for scalable, robust and efficient hyperparameter optimization.
In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI'21)*. ijcai.org.


Wang, Y. and Cai, Z. (2012).
Combining multiobjective optimization with differential evolution to solve constrained optimization problems.
IEEE Transactions on Evolutionary Computation, 16:117–134.