Automatic Generation of Neural Network Architectures Using a Genetic Algorithm
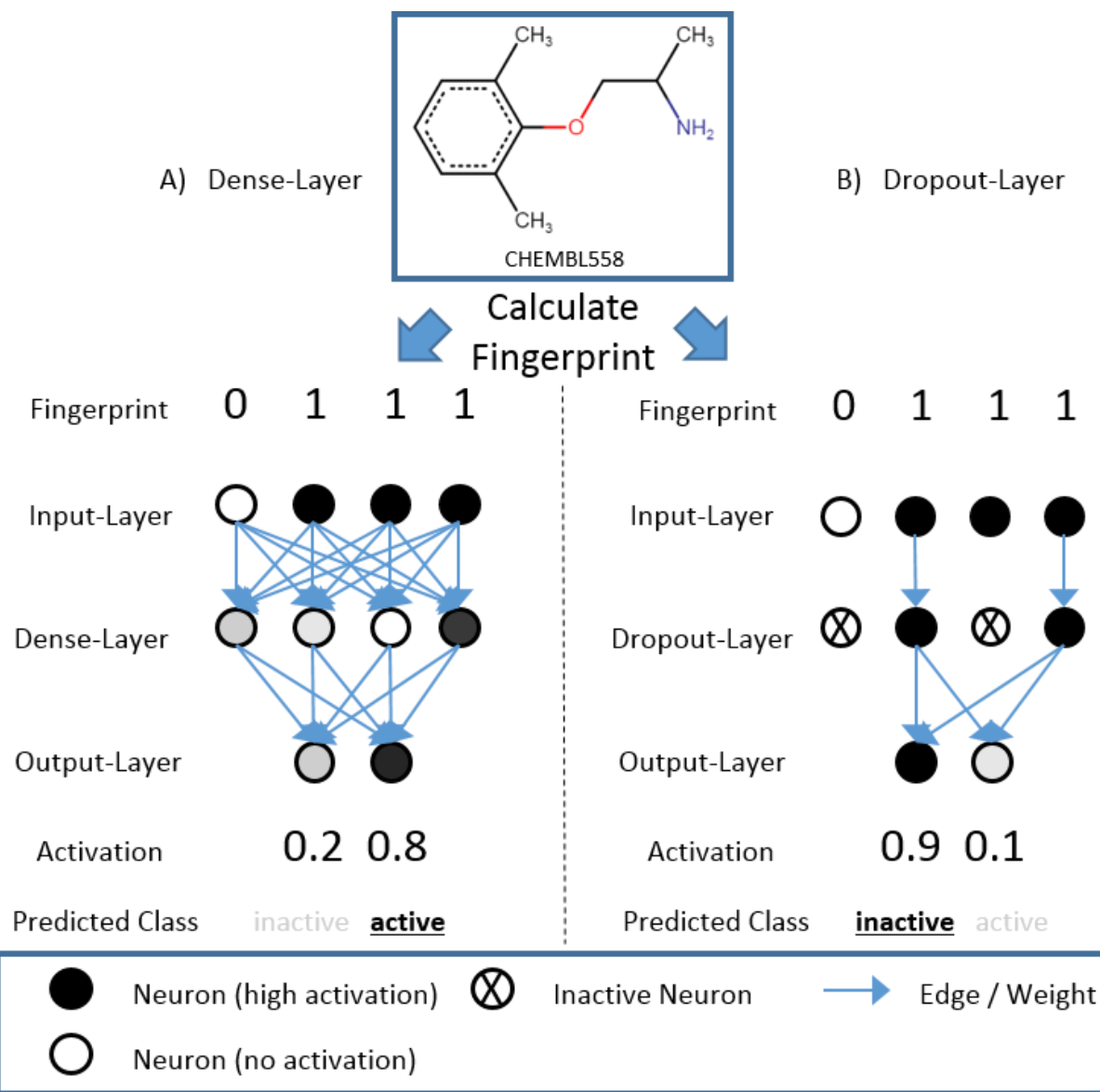
Wolf-Guido Bolick, Paul Czodrowski
RDKit UserGroupMeeting, Berlin, September 20th 2017

Merck

What is a DNN (deep neural network)?

- Morgan fingerprints as input
- Keras for learning deep neural networks
- scikit-learn for statistical measures & train/test set splits

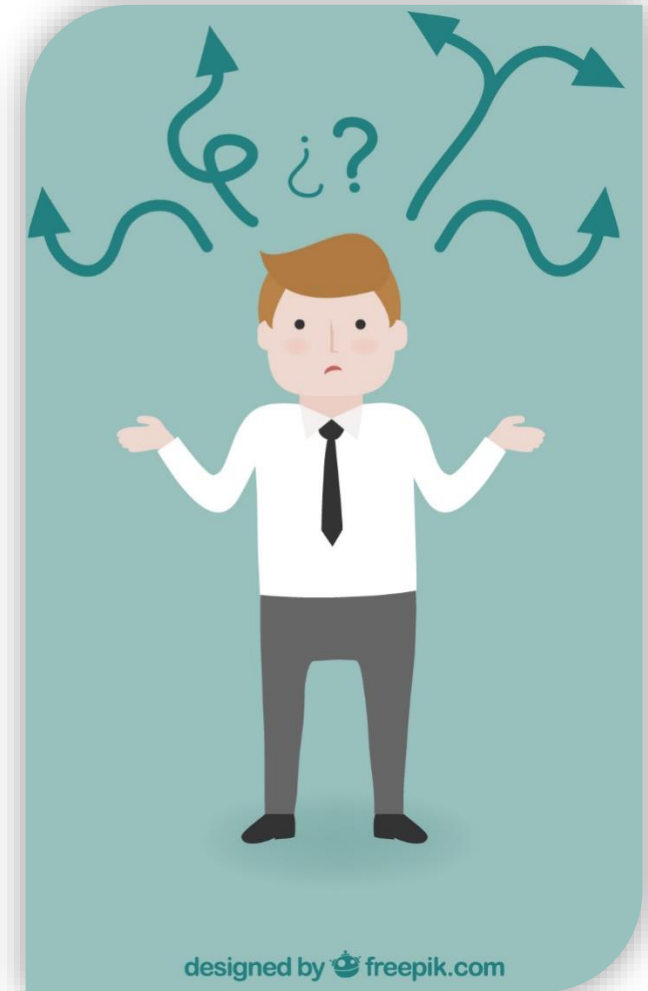# DNN Architectures & Hyperparameters

NN-Architecture
• Layer-Type • Layers • Neurons per Layer • Activation-Functions

Training-Parameters
• Optimizer • Learning-Rate • Weight-Decay • Batch-Size • Loss-Function • …

Hyperparameters

designed by freepik.com

Merck

Optimization of Hyperparameters

Expert

Lucky People

Everyone

- Hyperparameters derived from literature & **experience**
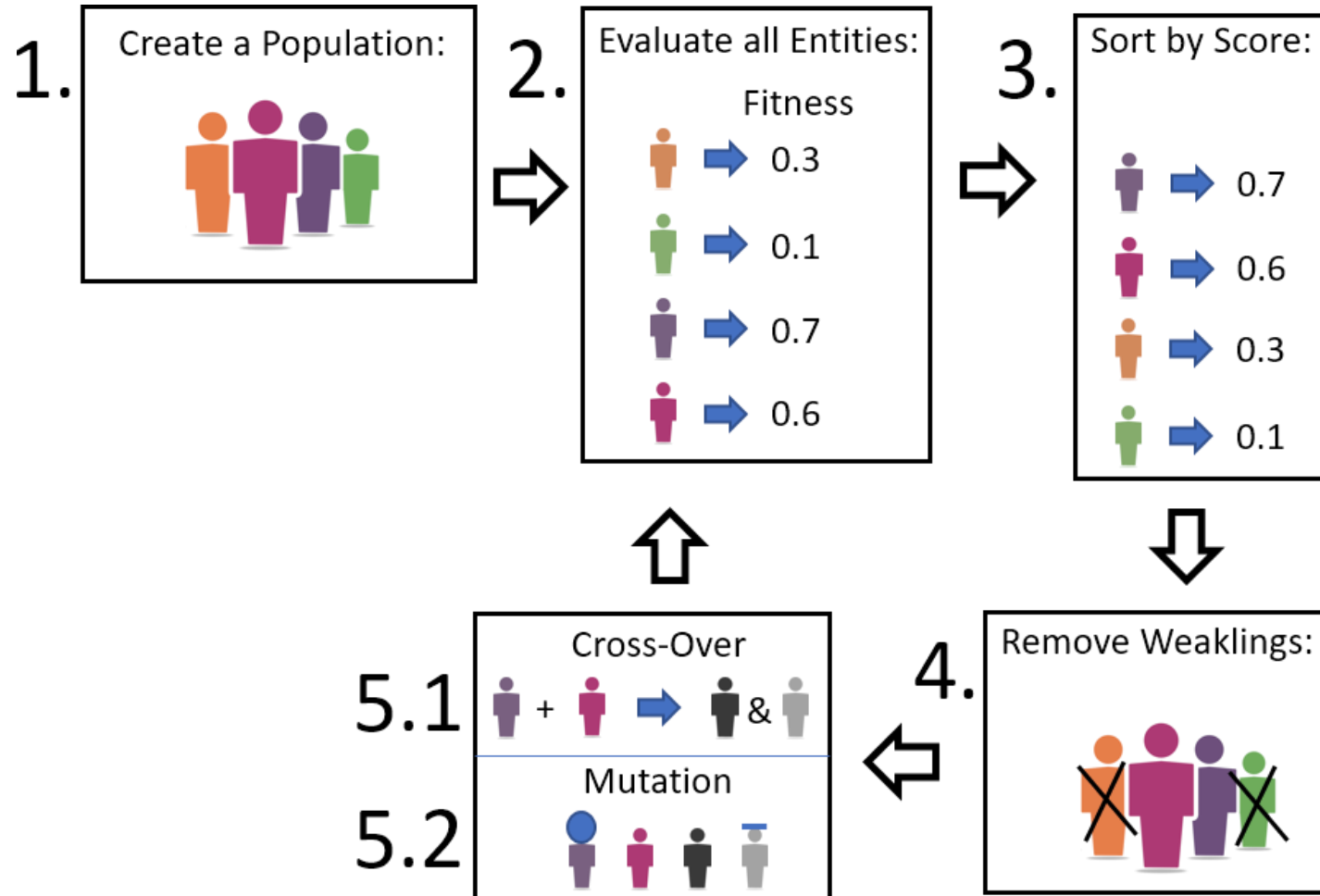- Hyperparameter search within promising parameter areas

- Random-Search

- Grid-Search
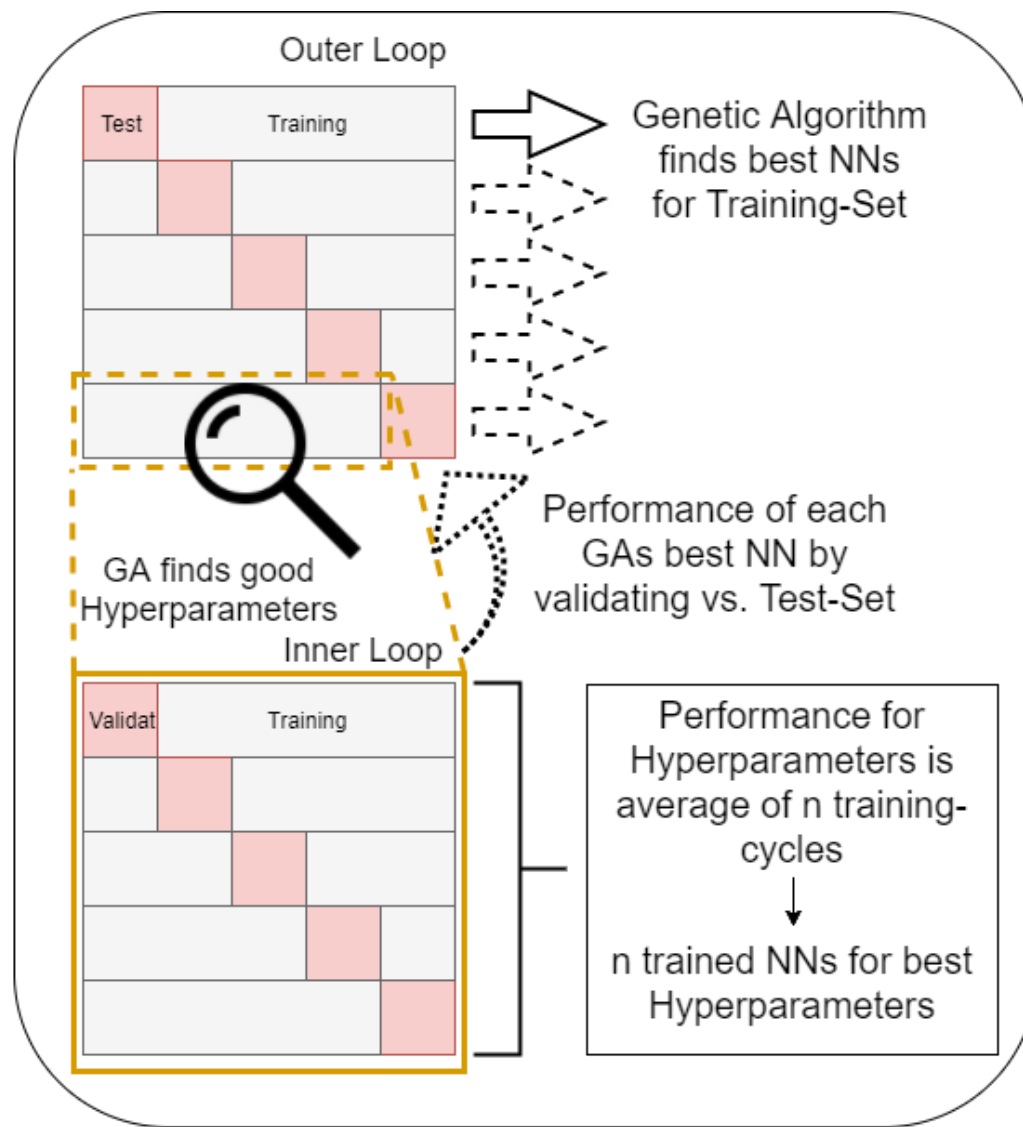- Probability based algorithms
- Directed Random-Search (*e.g.* **genetic algorithms**)

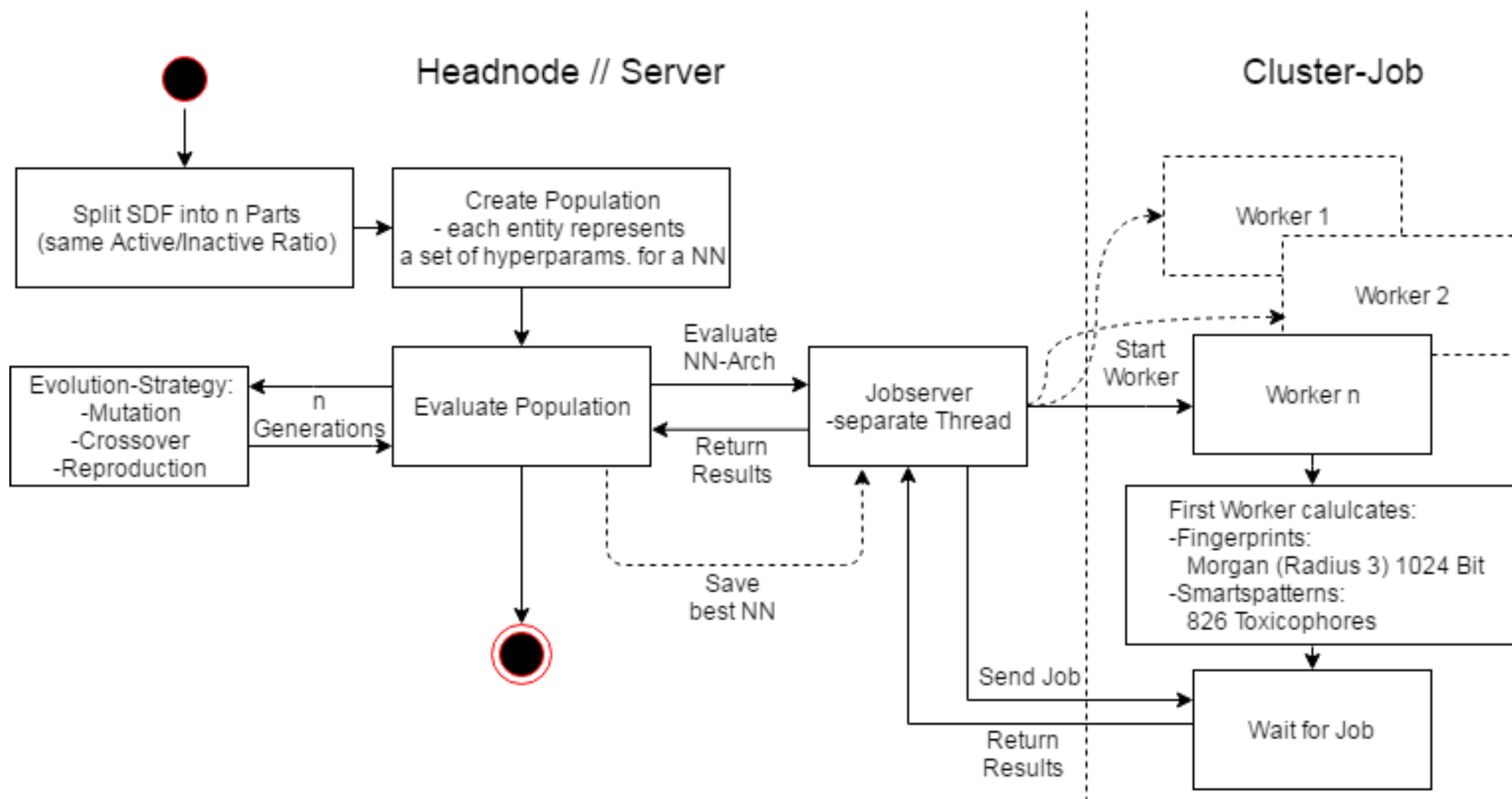**Merck**

What is a Genetic Algorithm (GA)?
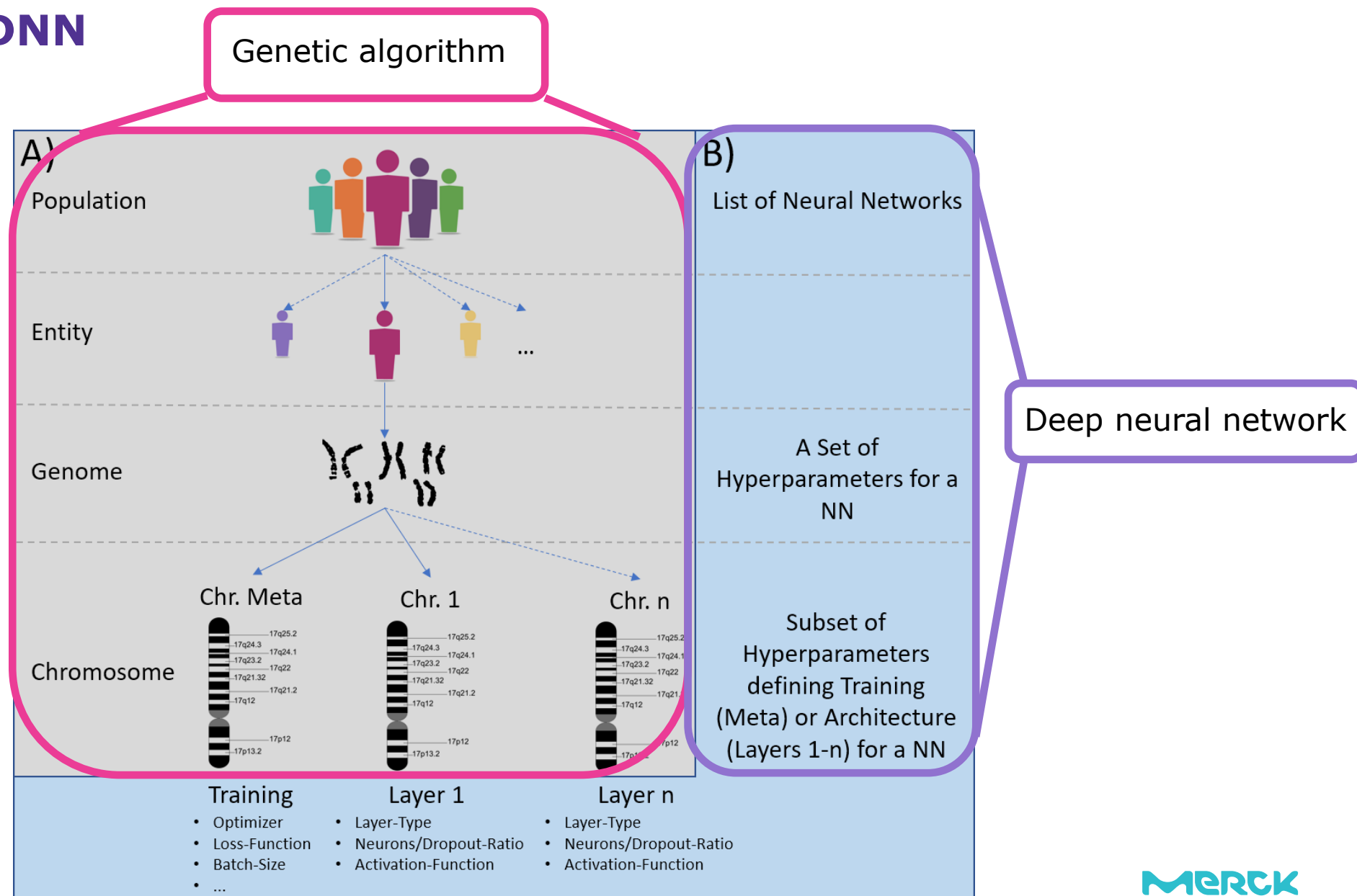
Validation strategy: nested cross validation

- **consensus Model of 25 individual Models**
- **Hyperparameter search inside inner loops & validated inside outer loops**
- **Every compound is represented in 16/25 individual Models**

MERCK

Implemented workflow

Analogy GA & DNN

GA settings

Parameter	Default value
Population size	100
Worker	10
Fingerprint size	1024
Evolution strategy	drop worst 50 %
SMARTS-Patterns	826

Parameter	Default value
Optimizer	sgd, rmsprop, adagrad, adadelta, adam, adamax, nadam
Loss-Function	mse, mae, msle
Learning-Rate	5E-2, 1E-1, 5E-1, 1.0
Weight-Decay	5E-7, 1E-7, 0.0
Momentum	0.0, 0.1, ..., 0.9
Nesterov	0, 1
Batch-Size	5%, 6%, ..., 20%

Parameter	Default value
Chromosomes	2 – 5
Layer types	Dense, Dropout
Neurons	32 – 512
Neurons stepsize	32
Dropout ratio	5, 10, ..., 90%
Activation functions	linear, sigmoid, hardsigmoid, softmax, relu, tanh

MerCK

Results for in house hERG data set

- **Outer kappa: almost 0.6 → moderate model**

- **Nested CV: computation time ca. 8 – 14 hours on GPU cluster**

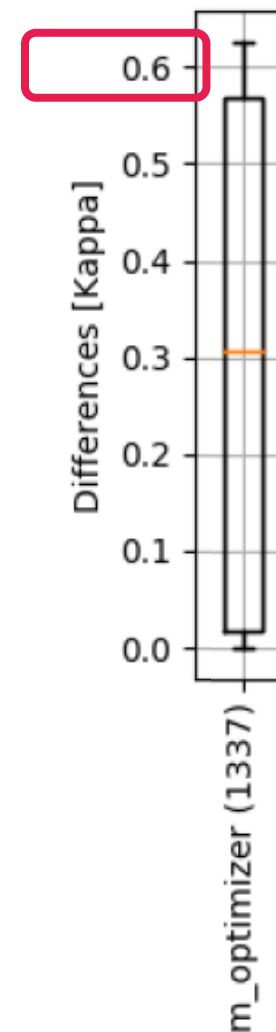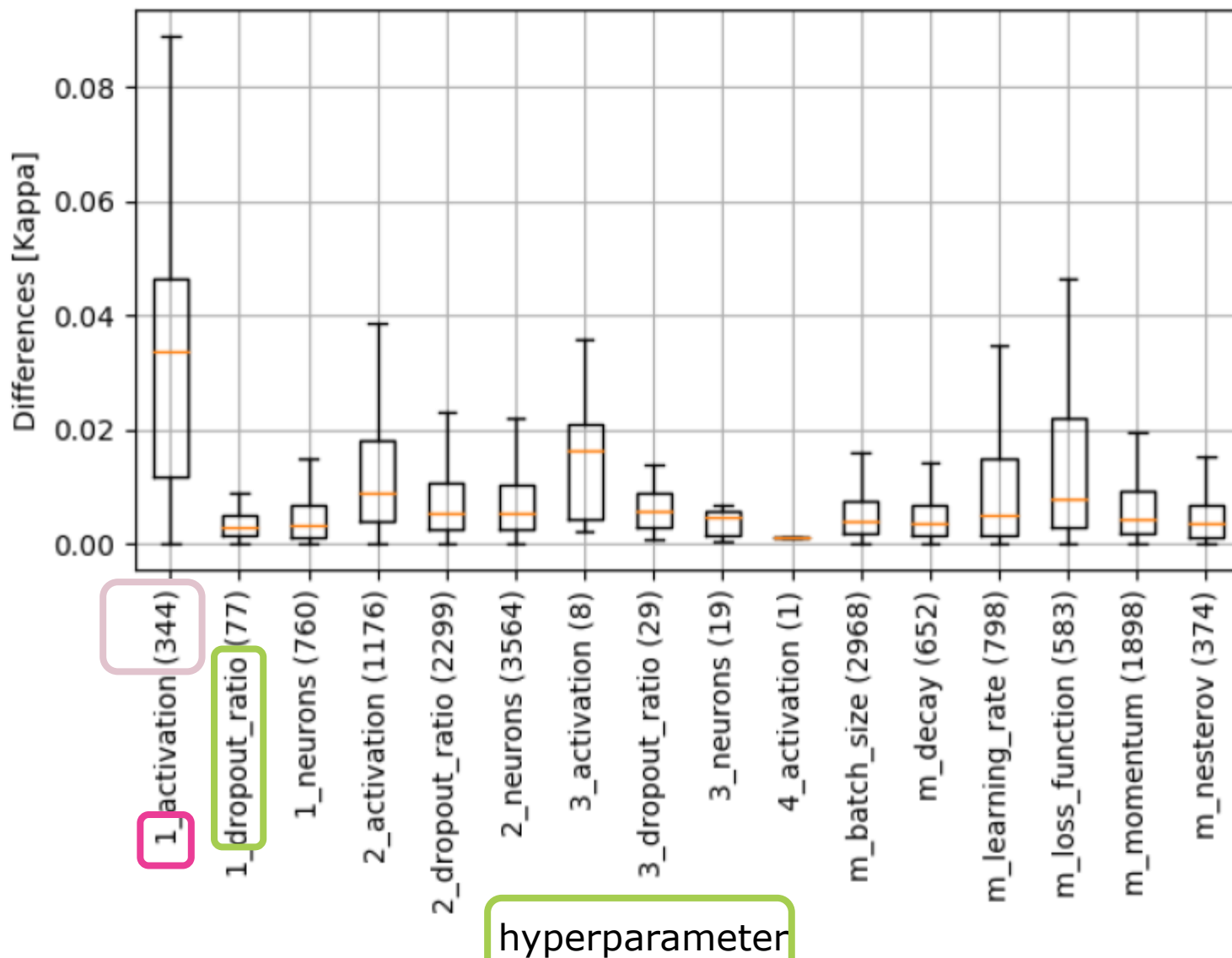| NN-Architecture	Split 1	Split 2	Split 3	Split 4	Split 5
Optimizer	sgd	adadelta	sgd	sgd	adadelta
Loss-Function	mae	msle	msle	mae	msle
Learning-Rate	1.0	0.5	1.0	1.0	1.0
Batch-Size	0.2	0.09	0.14	0.15	0.12
Momentum	0.1	0.8	0.6	0.4	0.3
Nesterov-Momentum	1.0	-	1.0	1.0	-
Weight-Decay	5e-07	5e-07	-	-	1e-07
Layer 1					
Layer-Type	dense	dense	dense	dense	dense
Activation-Function	relu	relu	tanh	relu	relu
Neurons	288	416	64	480	128
Dropout-Ratio	-	-	-	-	-
Layer 2					
Layer-Type	-	dense	dense	dropout	-
Activation-Function	-	relu	softmax	-	-
Neurons	-	64	384	-	-
Dropout-Ratio	-	-	-	0.15	-
Performance					
Inner Kappa	0.63 ± 0.02	0.62 ± 0.01	0.60 ± 0.03	0.63 ± 0.02	0.63 ± 0.02
Outer Kappa	0.57 ± 0.02	0.56 ± 0.04	0.56 ± 0.05	0.58 ± 0.02	0.59 ± 0.01
Generation	**20**	**40**	**79**	**19**	**49**
# Architectures	1318	1157	1146	1364	1403

MERCK

DNN performance over the GA generations
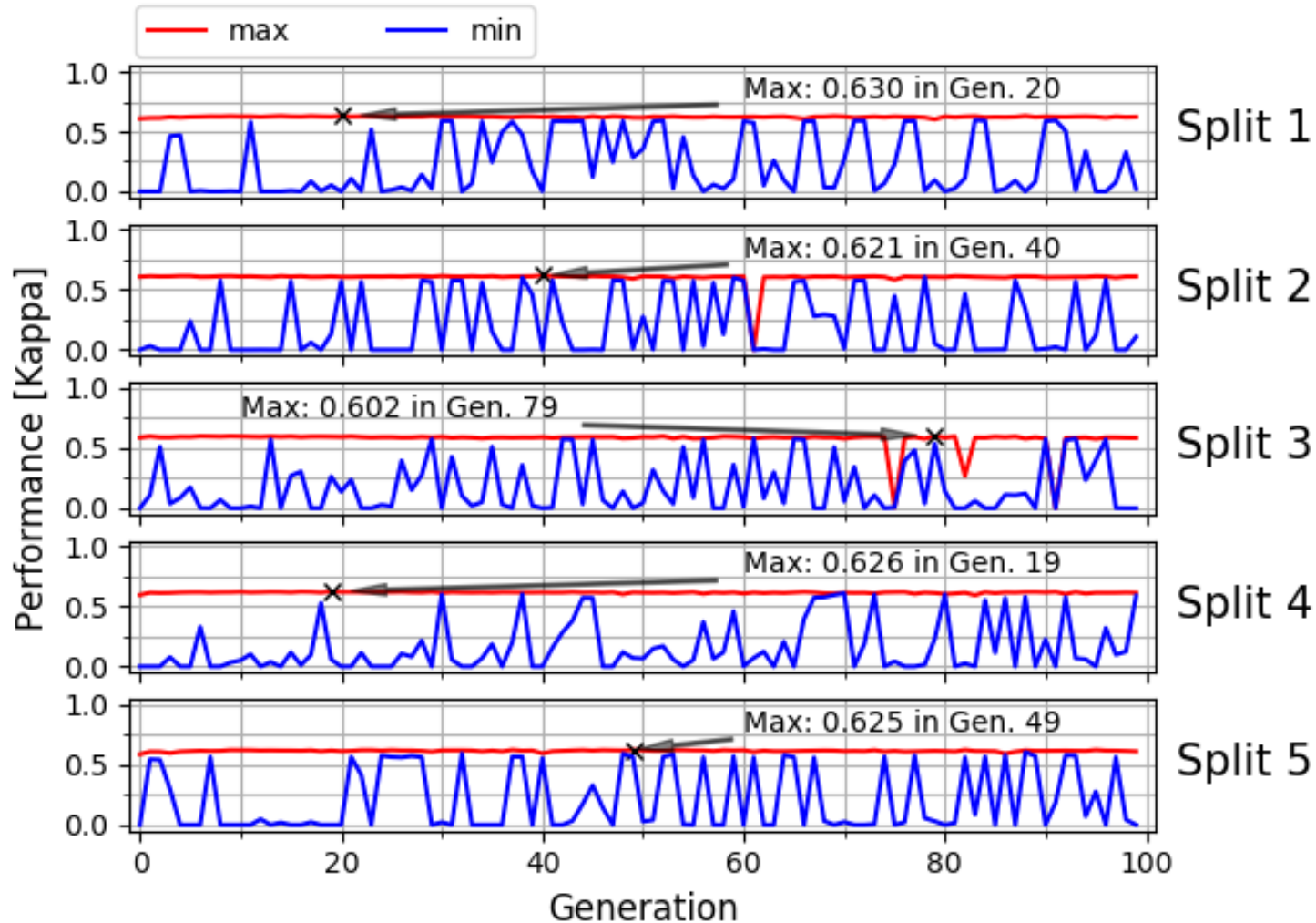
- **strong fluctuation of minimum performance of DNN**

- **in contrast, stable maximum performance**

Influence of individual parameters

Contributing pairs

layer

hyperparameter

[outliers not plotted]

12

Interface

- Interactive tracking of GA in the console
- Implemented in curses

Interface, ctd

chromosome 1
→ dropout-layer (35%)

chromosome 3
→ dense-layer (256 tanh)
→ activation-function: sigmoid

statistics:
Kappa: 0.09
Penal: 0.027 penalty-value
State: 100% → evaluation complete

```
MRL0D0XB2
35%-dropout
192-tanh
256-tanh
80%-dropout
kappa:    0.09
Penal:  0.027
State:100.0%
```

chromosome 0 with
hyperparameters
for training

chromosome 2
→ dense-layer (192 neurons)
→ activation-function: tanh

chromosome 4
→ dropout-layer (80%)

Detailed view of one particular architecture
Statistics given as well

Summary

- (Almost) stable implemention of a genetic algorithm the for hyperparameter search in Deep Neural Networks

  ➡ Follow-up studies on-going at Merck

- Comparison with other machine learning techniques

  ➡ Coming soon

- Roll-out of the code

  ➡ First step: publish a paper (not a single line written so far)!

MERCK

Acknowledgment

- Wolf-Guido Bolick

- Fabian Steinmetz

- Kristina Preuer, Günter Klambauer (Hochreiter group)

- Andreas Dominik (THM/JLU Gießen)