# Notes NOC

ALL

* Allele sharing
* Low-template DNA (low quality) -> drop-out
* Elevated stutter/drop-in

## nC-tool [1]

Derive NOC from

* TAC
* Reduce nr. of alleles by 5% from expected drop-out/alle sharing

## RFC19 model [2]

Derive NOC from

* Random forest ML model with 19 features

Percentage correct predictions for 2-5 NOC

* MAC = 69.2%
* nC-tool = 76.7%
* RFC19 model = 85%

Output = NOC + prob

## Decision Tree [3]

Derive NOC from

* Decision tree with

Tested various ML approached (RF / MLP / LDA), showing similar performance to the RF19 model. They obtained very high performance (96%) with a RFC 35 model.

Difference with [2] is “Benschop et al. used 1174 unique donors to construct 590 profiles [20], whereas the PROVEDIt dataset only had 26 unique donors within the 766 profiles used”

This means that the classifiers probably overfit to certain donors.

“In conclusion, the decision tree method for NoC assignment has been shown to be over 77% accurate, with increasing performance with improved stutter and artefact filters”

They used a decision tree to classify peaks as stutter or allele.

## Background STR mixture interpretation [4]

Information about how statistical analysis is done to determine the LR with the Hd and Hp. Showing that the LR is still the de-facto standard method.

“The peak height information is of benefit for analyzing mixed profiles.”

“The effect of incorrect estimation of the number of donors (caused by allele sharing) to the LR value was examined by Benschop (…) and was illustrated to exert a great effect on the LR” [5]

## Background about NFI-used software for LR calculation DNAStatistX

Shows the importance of correct NOC estimations: under-assigned number of contributors can cause the model the fail calculating the LR because the observed peaks cannot be well explained.

Also includes the NOC model + the generic RF11 model (with a lower accuracy of ~

Also includes the LoCIM method for inferring the major contributor.

# Notes metrics

* Accurate w.r.t. true effects of variables
* Speed?

## Counterfactuals [6]

“MOC returns a Pareto set of counterfactuals that represents different trade-offs between our proposed objectives, and which are constructed to be diverse in feature space.”

* Low number of feature changes (sparse explanations)
* Close to nearest observed data points (plausible explanations)

## ICE / PD plots /

ICE paper about PD plots: “Note that the approximation here is twofold: we estimate the true model with fˆ, the output of a statistical learning algorithm, and we estimate the integral over xC by averaging over the N xC values observed in the training set.”

“Visually, ICE plots disaggregate the output of classical PDPs. Rather than plot the target covariates’ average partial effect on the predicted response, we instead plot the N estimated conditional expectation curves: each reflects the predicted response as a function of covariate xS, conditional on an observed xC.”

1. Benschop, C., A. Backx, and T. Sijen, *Automated estimation of the number of contributors in autosomal STR profiles.* Forensic Science International: Genetics Supplement Series, 2019. **7**.

2. Benschop, C.C.G., et al., *Automated estimation of the number of contributors in autosomal short tandem repeat profiles using a machine learning approach.* Forensic Science International: Genetics, 2019. **43**: p. 102150.

3. Kruijver, M., et al., *Estimating the number of contributors to a DNA profile using decision trees.* Forensic Science International: Genetics.

4. Tao, R., et al., *Separation/extraction, detection, and interpretation of DNA mixtures in forensic science (review).* International Journal of Legal Medicine, 2018. **132**.

5. Benschop, C.C.G., et al., *The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures.* Forensic Science International: Genetics, 2015. **19**: p. 92-99.

6. Dandl, S., et al. *Multi-Objective Counterfactual Explanations*. in *Parallel Problem Solving from Nature – PPSN XVI*. 2020. Cham: Springer International Publishing.