Additional chapters

[DNA Mixture interpretation 2](#_Toc71045026)

[Short Tandem Repeat (STR) profiles 2](#_Toc71045027)

[Estimating the Number of Contributors (NOC) 2](#_Toc71045028)

[MAC method 4](#_Toc71045029)

[nC-tool 4](#_Toc71045030)

[Survey on mixture interpretation 6](#_Toc71045031)

[Set-up 6](#_Toc71045032)

[Question 1: workflow 6](#_Toc71045033)

[Analysis answers question 1 6](#_Toc71045034)

[Question 2: feature importance or counterfactual explanations 6](#_Toc71045035)

[Analysis answers question 2 7](#_Toc71045036)

[Data analysis 8](#_Toc71045037)

[Original 590 8](#_Toc71045038)

[Sampled 5000 8](#_Toc71045039)

[Benchmark 8](#_Toc71045040)

[(XAI techniques) <maybe not necessary since in paper> 9](#_Toc71045041)

[Experiments with various XAI techniques 9](#_Toc71045042)

[SHAP 9](#_Toc71045043)

[Anchors 9](#_Toc71045044)

[Counterfactuals 9](#_Toc71045045)

[WhatIf 9](#_Toc71045046)

[DiCE random 9](#_Toc71045047)

[DiCE genetic / GeCo 9](#_Toc71045048)

[Final user study 10](#_Toc71045049)

[References 11](#_Toc71045050)

# DNA Mixture interpretation

Experts can use DNA evidence to determine if certain people were involved in a crime by comparing the suspect DNA, victim DNA and other DNA samples to the evidence found at a crime scene. This interpretation become more difficult when the DNA profile consists of evidence from multiple people since information might overlap, or not every person contributed as much material. Even though software exists for analyzing this evidence, it is required that the expert inputs how many people contributed to the sample [1]. This chapter explains how to interpret a specific type of DNA profile, and highlights the different methods to determine the number of contributors.

## Short Tandem Repeat (STR) profiles

In forensic work, DNA evidence is often analyzed using *Short Tandem Repeat (STR)* profiles. These STR are specific tracks of repeated short DNA sequences of about two to six base pairs long that have been proven to show high variability between individuals in how many times the sequence repeats [2]. Most of these parts of the DNA or *loci* have been defined by CODIS, the United States national DNA database. We can capture the STR with a process called electrophoresis, which produces an electropherogram. In Figure 1, we see a simplified result that the electropherogram can produce for locus TH01. The y-axis shows the amount of information found in Relative Fluorescent Units (RFU), which is how the machine counts the quantity of DNA found. The x-axis shows the location of the locus on the DNA strand. Most importantly, we see two peaks, representing two alleles on this locus. These alleles are characterized by the number of repeats of the STR for locus TH01, which is [AATG]. On the right of Figure 1, we see the DNA sequence for six and eight repeats.

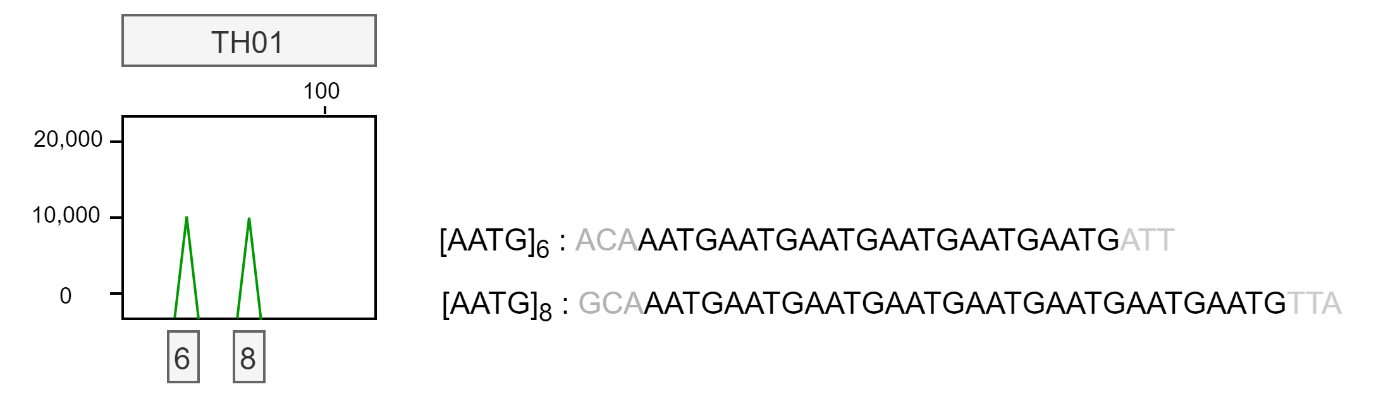


Figure 1: Simplified electropherogram result for locus TH01 showing two alleles with six and eight repeats each. The repeat sequence is shown on the right with arbitrary flanking regions.

One individual can have two different alleles for a single locus; one inherited from the mother, and one from the father. It is also possible that a person inherited the same allele from both of their parents, this means that they are homozygous at that locus. The peak will then be twice as large. We will now get into more detail of how to derive the number of contributors from an STR profile.

## Estimating the Number of Contributors (NOC)

The first step of DNA STR profile interpretation is to determine whether a sample has originated from a single source, or if the sample is a mixture [3]. This often easily discerned by checking whether or not there are loci with more than two alleles present. As we saw in Figure 1, a single person can contribute a maximum of two alleles per locus, so profiles with more alleles are considered a mixture. The next step is to determine the number of contributors. This step is necessary for DNA analysis software to calculate the weight of the evidence found [4]. When an incorrect NOC is used for further analysis involving the investigation of the DNA profiles, the results are unreliable [5]. It could make the difference between whether or not a person of interest is included in the evidence or not.

Determining the exact number of contributors is difficult. There are several obscuring factors that could make an expert underestimate the number of donors, especially when the number of donors increases [3, 6]:

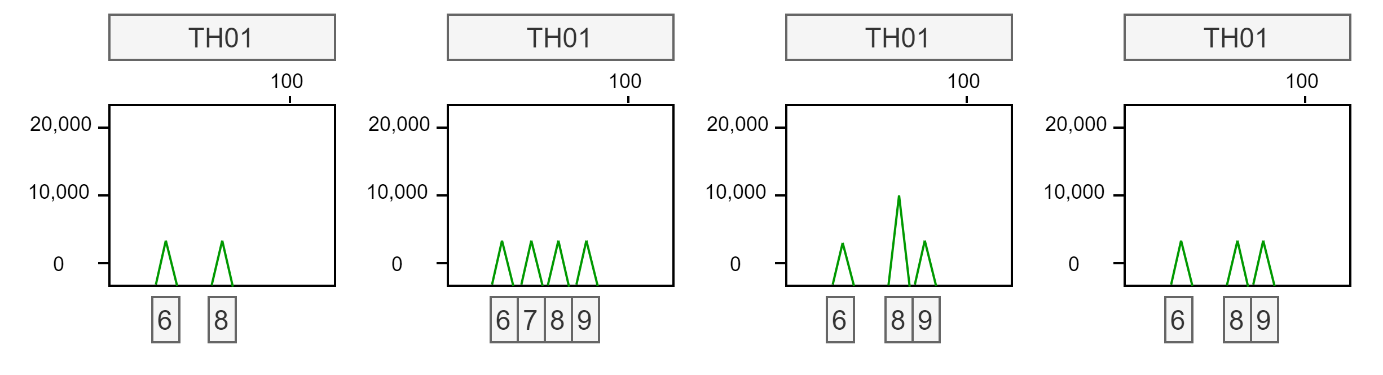


Figure 2: Four simplified electropherogram results for locus TH01. From left to right: Example of a single donor profile; Example of a 2-person mixture profile; Example of a 2-person mixture profile with allele sharing, peak 8 is twice as high compared to peak 6 and 9; Example of a 2-person mixture profile with drop-out, one peak has likely not been detected.

* Allele sharing: If two donors have the same allele at a locus, this is called allele sharing. It frequently occurs when donors are relatives, since siblings share a lot of DNA. It might be difficult to distinguish if an allele is shared between donors, or if a single donor simply is homozygous for this allele; in both cases, the peak height for that allele is higher. This can be seen in the third picture of Figure 2; allele 8 has twice as much information as alleles 6 and 9.
* Allele drop-out: If the DNA was degraded, for example due to sunlight, some parts of the DNA might not be present in the sample to measure. It is also possible that the amount of the DNA available is so small, that the alleles fall below a certain noise filter. Because of this low quality or quantity of DNA, some allele fragments might not show up in the profile at all, which is called drop-out.

These factors can decrease the number of alleles found in a certain profile, which could lead to an underestimation of the number of contributors. There are also factors that could lead to an overestimation of alleles present in a sample [3]:

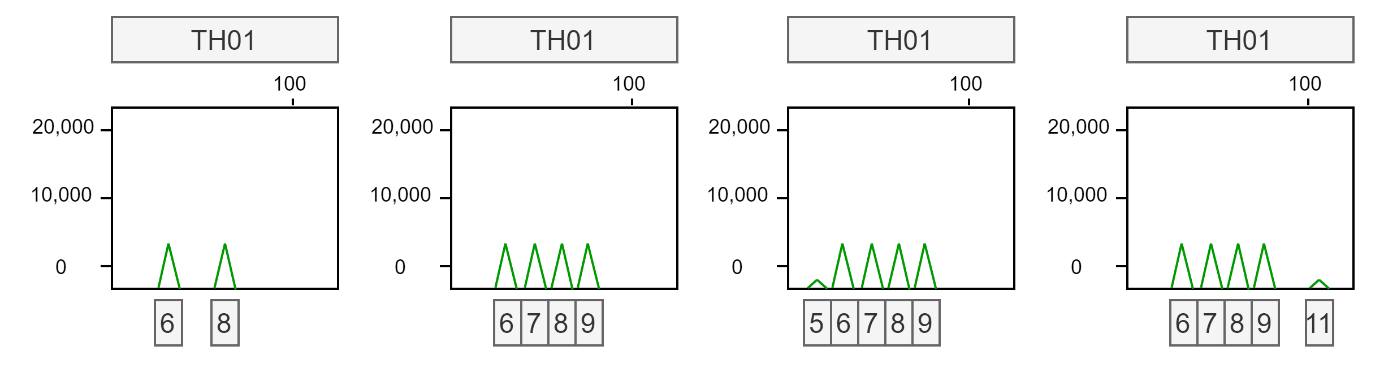


Figure 3: Four simplified electropherogram results for locus TH01. From left to right: Example of a single donor profile; Example of a 2-person mixture profile; Example of a 2-person mixture profile with a stutter peak at allele 5 caused by a STR with 6 repeats folding; Example of a 2-person mixture profile with a noise peak at location 11 caused by an error in reading.

* Stutter: During the process of measuring the STRs, a STR fragment can accidentally fold. This could cause the electropherogram to measure this strand to have one repeat fewer, since the folded part of the fragment is not correctly measured. In this way, a small stutter peak is found in the profile just before the valid peak. In Figure 3, this is
* Allele drop-in or other noise: The measuring process is not perfect, so some random noise might show up in the electropherogram, that does not contain any information about the DNA. In Figure 3, we can see that the rightmost image has a small peak at allele 11. Since it is not close to another allele, it is likely not a stutter peak.

Stutter peaks and noise are often filtered out using certain thresholds. As a result, some DNA information might also be lost due to a low-quantity donor.

### MAC method

The simplest method to get an estimate of the NOC is by using the Maximum Allele Count (MAC) [3, 7]. By taking the locus with the most alleles present, dividing that number by two, and rounding up, we can get an idea of the minimum NOC. Though this method is simple, it is unreliable due to the factors of allele sharing, drop-out, etc. Performance in general is quite poor, especially with 3 or more contributors, when there is a lot of allele sharing, or when the quality of the profile is low [8, 9]. On average, when assessing mixtures between 2-5 contributors, the MAC obtains correct predictions for about 60-70% of samples [9-11].

### nC-tool

Often experts use MAC in combination with the Total Allele Count (TAC), which measures the total number of alleles across all loci. However, this measure suffers from the same obscuring factors as the MAC. The nC-tool takes both these measures into account, as well as simulating various levels of dropout [11]. This achieves better results than using the MAC only, obtaining correct predictions for roughly 76% of 2-5 person mixtures [10, 11].

**< in progress >**

*Maximum likelihood approach [9],*

*Probabilistic mixture algorithm [12]*

*NOCit tool [13] which is a likelihood-based approach taking into account both qualitative and quantitative data.*

*1-3 person mixtures. Synthetically created from 58 known donors. Modelling the types of peaks, peak heights, drop out and stutter rates. posterior probability via a Monte Carlo-based approach. maximum probability method with NOCIt resulted in accuracies of approximately 83%. Computationally slow! Up to 9h for a 5 person mixture.*

*True allele up to ten contributors? probability model*

*nC-tool [11] uses both the MAC and Total Allele Count (TAC), as well as categories of drop-out to obtaining better results.*

*quantitative continuous model peak heights in the DNA profile and considers the*

*effect of artifacts and allelic drop-out. By using this software, the likelihoods of 1–4 persons’*

*contributions are calculated, and the most optimal number of contributors is automatically*

*determined; Kongoh was validated using 27 two-person mixtures, 27 three-person mixtures, and 18 fourperson mixtures. These mixtures were experimentally prepared using non-degraded DNA*

*from pristine blood samples.*

*Machine learning approach*

*PACE random forest [14]: 969 non-simulated DNA samples of 1 to 5 contributors generated from a combination of 120 individuals. They achieved about 90% accuracy on identifying 1, 2, 3, 4+ profiles.*

*Bayesian probability framework TrueAllele [15]. Closed-source. Minutes to longer for more complex*

**< in progress >**

## Survey on mixture interpretation

### Set-up

We wanted to gain insights into three topics:

1. What is the normal workflow of experts when estimating the NOC?
2. What type of explanation do experts prefer to help them make a decision (feature importance or counterfactual)?
3. What type of data do experts prefer to help them make a decision (summary statistics or individual peak information)?

### Question 1: workflow

This question described an average workflow as we have interpreted it. In summary:

* Inspect general information about the profile (peak heights, TAC, MAC, NOC tool prediction);
* Check the locus with the MAC to see if all peaks can be explained with the expected number of donors;
* Check for stutter peaks / extra peaks from another donor.

The users were asked to write any missing steps.

### Analysis answers question 1

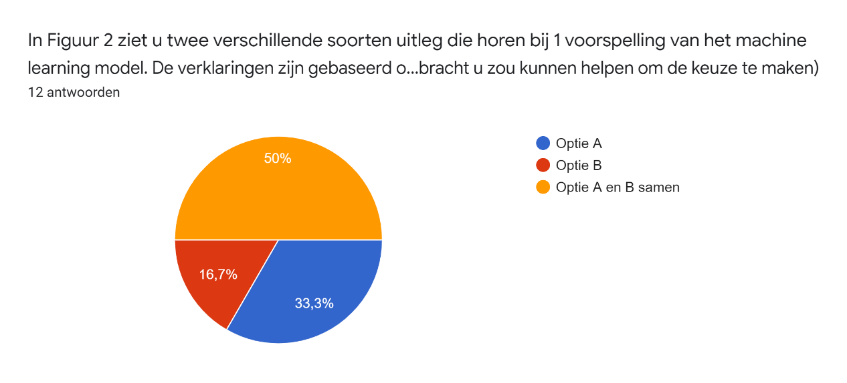
The following remarks were reported to be missing from the workflow:

* Check the number of peaks below the detection threshold (6x). This gives an indication of the DNA quality (1x) / the amount of dropout (3x).
* Experts can often not make a reliable choice between 4 or 5 donors based on the information (1x).
* None (3x).
* Locus SE33 (1x)

In summary, missing information concerns the number of peaks below the detection threshold which is not available to the machine learning model, so we could not incorporate this information. The remark about 4 or 5 donors demonstrates the difficulty of making decisions with more donors. Locus SE33 has the largest variety of alleles, which is why one user finds it more informative. **The remarks suggest that more and/or different features could be used for the machine learning model.**

### Question 2: feature importance or counterfactual explanations

This question described two explanations for the same prediction of a profile.



The users were asked to choose which one appealed to them more and why.

### Analysis answers question 2

Explanations for choosing A:

* You want to know why the model predicted its result (2x)
* Easier to understand (1x)
* Option B is also good, but a visual explanation is better (1x)
* Option B is also good, but can the model know if the expert is interested in 4 or 6? (2x)

Explanations for choosing B:

* Is more specific for comparing one to another (2x), which is relevant for criminal case investigations (1x), option A is more background information.
* Option A and B together is also a good option.

Explanations for choosing A and B:

* Option B can provide very specific information (1x) (e.g. if the allele count on one locus were lower to get a different NOC, and it could be explained by stutter).
* Option B is relevant when you came to a different NOC than the tool outputs (2x)
* Option A tells you why it came to its result in the first place (4x).
* Option B tells you where the threshold values lie (1x).
* Option B tells you if the predictions were close together (1x).
* More convincing (1x)
* Combination of information makes the decision complete (1x)

In summary, **most users liked the combination of explanations. People that picked one option, often also mention they liked the other as well. They enjoyed the general information of the feature attributions, and the specific values of the counterfactuals.**

Since option A had a nice visualization, as opposed to option B, it could have induced some presentation bias. However, this was not significant in the answers.

# Data analysis

## Original 590

## Sampled 5000

## Benchmark

# (XAI techniques) <maybe not necessary since in paper>

# Experiments with various XAI techniques

## SHAP

<How works for multi-class/regression>

* Interpretation by experts

It is an issue that feature importance methods split the impact on the model over correlated features. The result of this issue is that the importance value for correlated contributing features is underestimated, in contrast to if their importance was left undivided. However, since main goal of these values is to give an impression of the contributing factors to a prediction, the exact values are not a priority. This part of the explanation is to observe a general sense of which features contributed to the prediction in which direction. For this purpose, we deem SHAP adequate.

## Anchors

<Why this did not work out>

* + 1. Research into sampling
    2. Research into *anchored counterfactuals*
    3. Interpretation by experts

## Counterfactuals

<How best visualization was derived>

* Only show changed features
* Scale features
* Each feature has their own scale
* Interpretation by experts

### WhatIf

### DiCE random

### DiCE genetic / GeCo

# Final user study

# References

1. Taylor, D., J.-A. Bright, and J. Buckleton, *Interpreting forensic DNA profiling evidence without specifying the number of contributors.* Forensic Science International: Genetics, 2014. **13**: p. 269-280.

2. Boavida, A., et al., *PowerPlex® fusion 6C system: internal validation study.* Forensic sciences research, 2018. **3**(2): p. 130-137.

3. Intituut, N.F., *HBS: Interpretatie van autosomale STR DNA-profielen*. 2020. p. 16.

4. Bleka, Ø., G. Storvik, and P. Gill, *EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts.* Forensic Science International: Genetics, 2016. **21**: p. 35-44.

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. Coble, M.D., et al., *Uncertainty in the number of contributors in the proposed new CODIS set.* Forensic Science International: Genetics, 2015. **19**: p. 207-211.

7. Clayton, T.M., et al., *Analysis and interpretation of mixed forensic stains using DNA STR profiling.* Forensic Science International, 1998. **91**(1): p. 55-70.

8. Benschop, C., *PowerPlex Fusion 6C Profile analysis & interpretation.* 2020.

9. Haned, H., et al., *Estimating the Number of Contributors to Forensic DNA Mixtures: Does Maximum Likelihood Perform Better Than Maximum Allele Count?* Journal of Forensic Sciences, 2011. **56**(1): p. 23-28.

10. 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.

11. 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**.

12. Paoletti, D.R., et al., *Inferring the Number of Contributors to Mixed DNA Profiles.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2012. **9**(1): p. 113-122.

13. Swaminathan, H., et al., *NOCIt: A computational method to infer the number of contributors to DNA samples analyzed by STR genotyping.* Forensic Science International: Genetics, 2015. **16**: p. 172-180.

14. Marciano, M.A. and J.D. Adelman, *Developmental validation of PACE™: Automated artifact identification and contributor estimation for use with GlobalFiler™ and PowerPlex® fusion 6c generated data.* Forensic Science International: Genetics, 2019. **43**.

15. Bauer, D.W., et al., *Validating TrueAllele® Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors.* Journal of Forensic Sciences, 2020. **65**(2): p. 380-398.