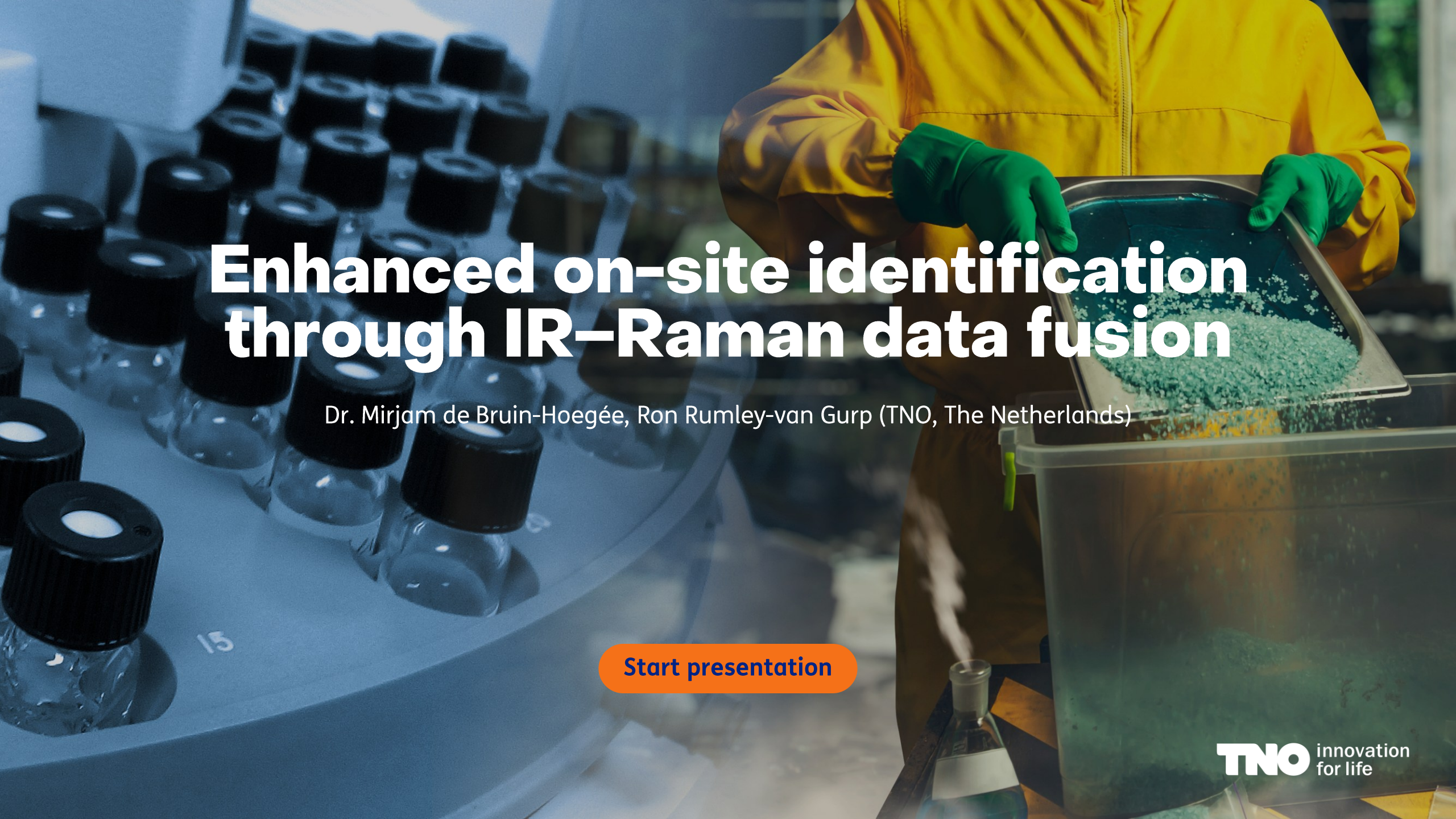




Enhanced on-site identification through IR–Raman data fusion

Dr. Mirjam de Bruin-Hoegée, Ron Rumley-van Gorp (TNO, The Netherlands)

[Start presentation](#)

Introduction

Forensic investigations

1. Collecting evidence on a crime scene
2. Laboratory analysis
3. Forensic report
4. Legal accountability



Challenges in sample collection

Sensitive analysis in complex samples using portable techniques

- During inspections, 2 portable detectors (LCD, Raman, FTIR, AP4C) need to identify a chemical before sampling
- Rapid degradation in the environment and metabolism in biological matrices
- Distinguishing between background concentrations
- Difficult to sensitively detect chemical attribution signatures at trace levels



Aim of the research

- Improving on-site identification of toxic chemicals through IR–Raman data fusion (proof-of-concept)
 - Complex matrices
 - Mixtures



Sample collection



Detection

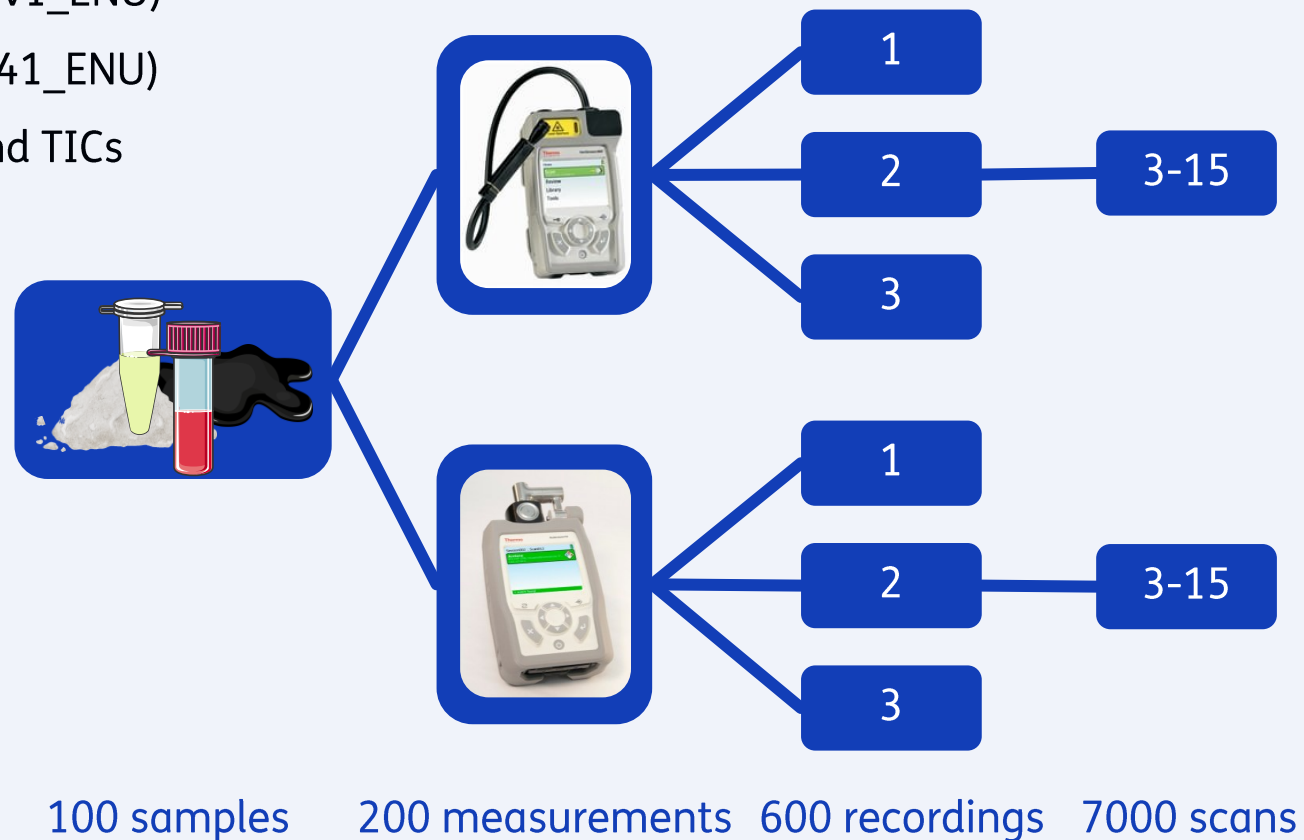


Data analysis

Method

- Data was collected by two portable sensors
 - TruDefender FTX (Software 1.5.1, library FT1.5v1_ENU)
 - FirstDefender RM (Software 4.1.0a, library RM41_ENU)
- Pure and 10-30% mixtures of max three CWAs and TICs
- 17 compounds

Code	Name
1PE	1-Phenylethanol
2CE	2-Chloroethanol
ACR	Acrolein
ALL	Allyl alcohol
BEN	Benzene
CS2	Carbon disulphide
GA	Tabun
GB	Sarin
GD	Soman
HD	Sulphur Mustard
HN1	Nitrogen mustard HN-1
IPC	Iron pentacarbonyl
L1	Lewisite (L1)
MCF	Methyl chloroformate
MHY	Methylhydrazine
PHE	Phenol
VX	VX

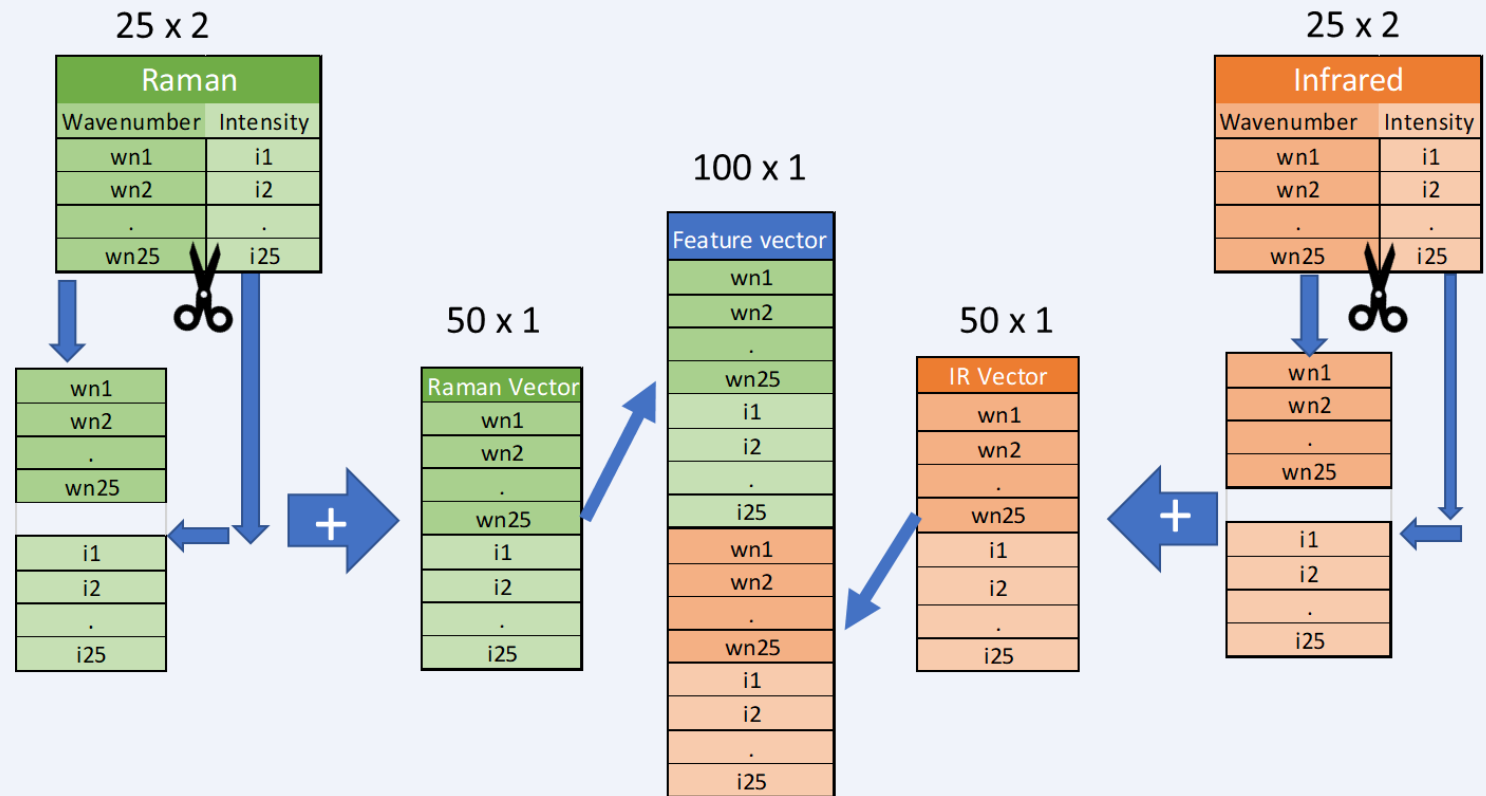


Method

- Selection of 25 highest peaks
- Concatenation of intensities
- AI models (python 3.14.3)
 - Decision Tree Classifier
 - Random Forest Classifier
 - Support Vector Machine
- *MultiOutputClassifier* function

Important considerations

- All pure samples in training set
- Train (70%), test (20%) and validation (10%) set
- No replicate recordings or scans in both train and test set



Results

- Identification with software provided on the sensor
 - No matches of mixtures with GD, VX, HD, HN1, and L1

Compound	Identification
1PE	match
2CE	match
ACR	match
ALL	match
BEN	match
CS2	match
GA	match
GB	match
GD	match
HD	match
HN1	match
IPC	match
L1	match
MCF	match
MHY	match
PHE	match
VX	match



Mixture	Raman	FTIR
GB-HN1-GA	GA-GB	no match
GD-GA-BEN	false positive	no match
GD-HD-L1	no match	no match
GD-MCF-CS2	MCF	no match
L1-ALL-BEN	BEN	no match
VX-HN1-CS2	CS2	no match
ACR-ALL-IPC	ACR-IPC	no match
BEN-2CE-MHY	BEN	no match
PHE-CS2-1PE	no match	no match
VX-L1-HD	no match	no match

Results

- Best performance by random forest classifier
 - Improvement of some mixed compounds and difficult matrices (diesel, sand, solvent CCl₄)
 - Accuracy alone is not informative, since 3 false negatives give an accuracy of 0.82 (14/17)

compound	TP	TN	FP	FN	TPR	TNR	FPR	FNR	accuracy	precision	recall	F1
HD	18	381	0	6	0.75	1	0	0.25	0.99	1	0.75	0.86
L1	6	375	0	24	0.2	1	0	0.8	0.94	1	0.2	0.33
VX	6	399	0	0	1	1	0	0	1	1	1	1



Discussion

- Careful implementation of AI models is required
 - Risk of data leaking if wrong choice train/test/validation set
 - How realistic is it if replicate measurements are in both train and test set (reproducibility)?
- Spectral preprocessing (first/second derivative)
- Feature scaling (wavenumbers vs intensity)
- Increasing number of samples



Conclusions and future work

- Data fusion of portable IR and Raman data is promising for improving identification of complex mixtures
- Improved field detection enables more accurate sampling, thereby supporting forensic verification in the lab

Future work

- Improve spectral processing, feature scaling, and mid level data fusion in close collaboration with data scientists
- Build models using synthetic data and use real data for testing and validation
- Apply data fusion to mass spectrometric data (GC-MS/LC-MS)
- Apply data fusion to data obtained by smaller and cheaper IR and Raman detectors





**THANK YOU FOR
YOUR ATTENTION**

ACKNOWLEDGEMENTS

TNO

MARCEL VAN DER SCHANS

JELLE DE KONING

OPCW

PETER HOTCHKISS

DUTCH MINISTRY OF DEFENCE