Detection of Gasoline in Fire Debris Using Artificial Intelligence: Image Transformation of Gas Chromatography and Mass Spectrometry Data and Deep Learning

Ting-Yu Huang, MS*, and Jorn Chi-Chung Yu, PhD, ABC-CC

Department of Forensic Science

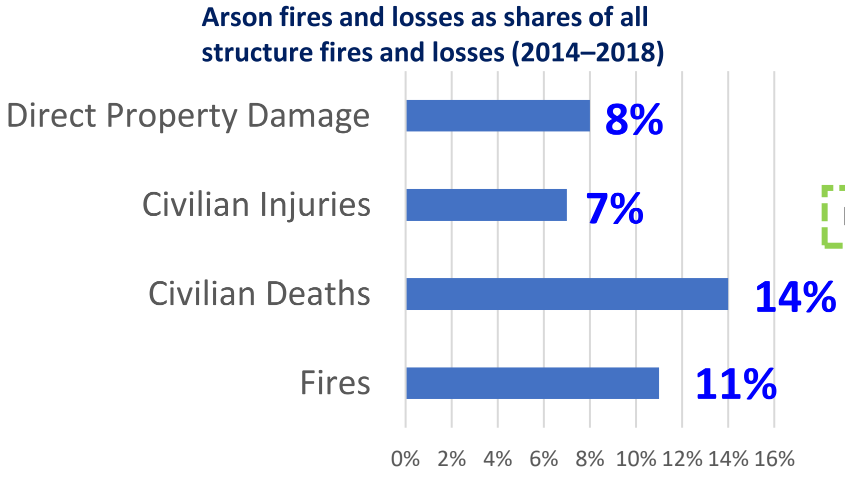
Sam Houston State University

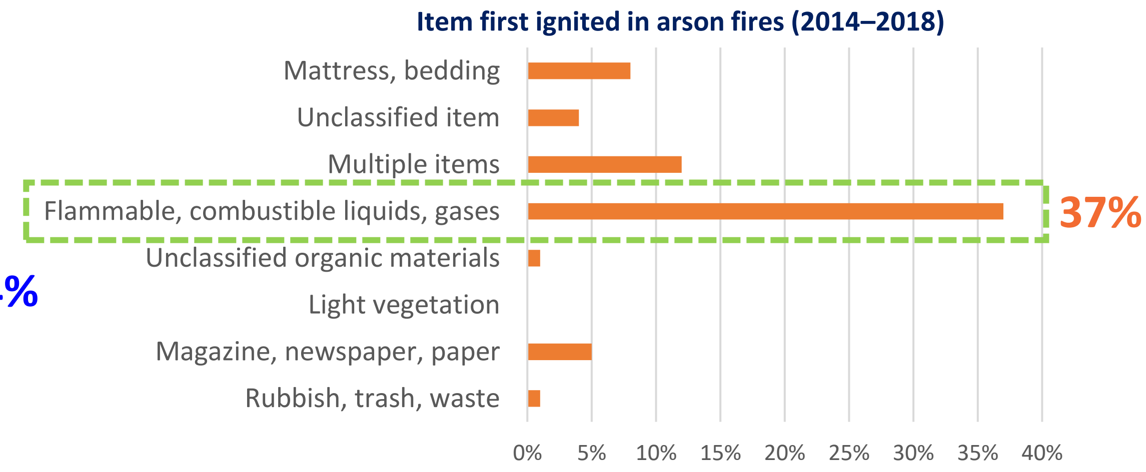
Huntsville, TX 77340

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

Gasoline in arson fires

- Annual average of 52,260 arson fires were reported (2014-2018)
- Caused 400 civilian deaths, 950 civilian injuries, and $815 million in direct property damage
- **Gasoline** is one of the most commonly-used ignitable liquids in arson fires
  - ➢ Easy to obtain and transport
- Standard test method for identifying gasoline is ASTM E1618-19
  - ➢ Visual comparison of chromatograms
  - ➢ Extracted ion profiling
  - ➢ Target compound analysis

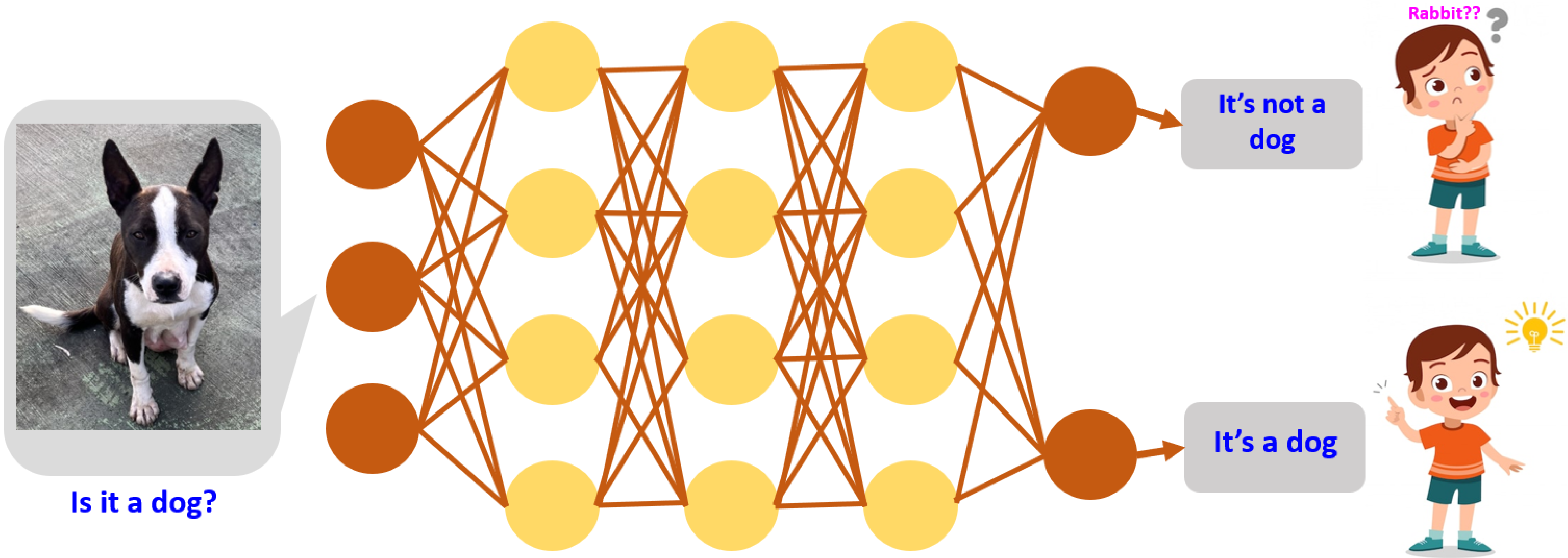




National Fire Protection Association (NFPA), Intentional Structure Fires. https://www.nfpa.org//-/media/Files/News-and-Research/Fire-statistics-and-reports/US-Fire-Problem/Fire-causes/osintentional.pdf, 2021.

Deep learning (1/2)

- Convolutional neural network (CNN)
  - A neural network with multiple layers
  - Automatically extract features
  - Unstructured data (images)

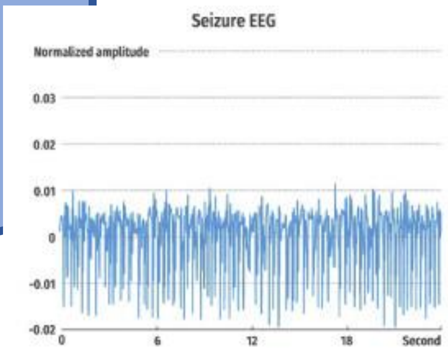


Sarigul, M., Ozyildirim, B. M., & Avci, M. (2019). Differential convolutional neural network. Neural Networks, 116, 279–287. https://doi.org/10.1016/j.neunet.2019.04.025

Deep learning (2/2)

Advantages

- Capable of extract complex features
- Higher classification performance
- Predictive modeling
- Superior capability to classify **images**
- Wide applications in diagnosing diseases



Disadvantages

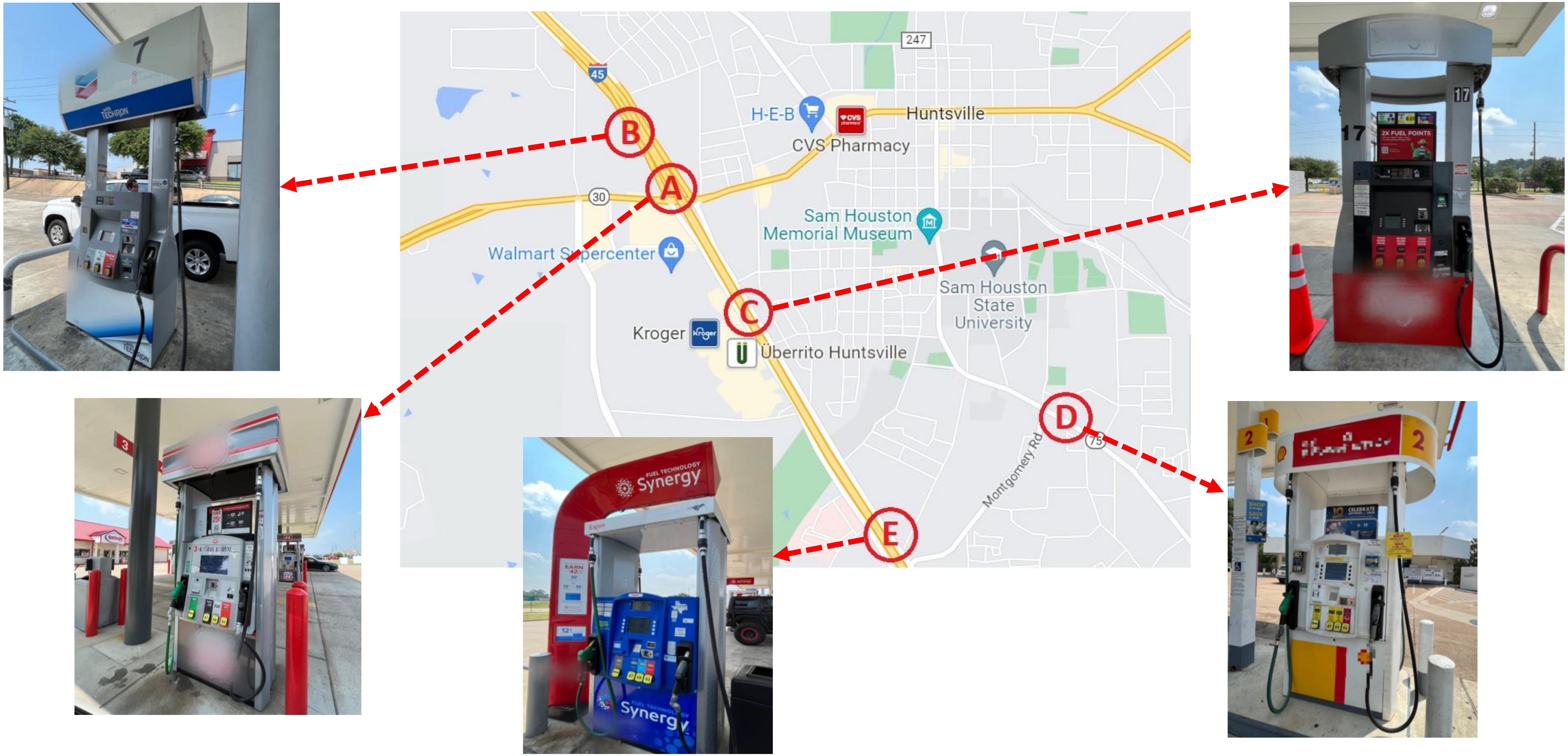
- Requires large-scale data collection
- High computational cost
- **Transfer learning**

- ***Can deep learning (transfer learning) be applied in gasoline detection if GC/MS data are presented in images?***

U. R. Acharya, O. Faust, Y. Hagiwara, J. H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Computers in Biology and Medicine. 100 (2017) 270–278. https://doi.org/10.1016/j.compbiomed.2017.09.017
S. S. Yadav, S. M. Jadhav, Deep convolutional neural network based medical image classification for disease diagnosis, Journal of Big Data. 6(2019) https://doi.org/10.1186/s40537-019-0276-2
M. A. Morid, A. Borjali, G. Del Fiol, A scoping review of transfer learning research on medical image analysis using ImageNet. Computers in Biology and Medicine. 128 (2020) 104115. https://doi.org/10.1016/j.compbiomed.2020.104115

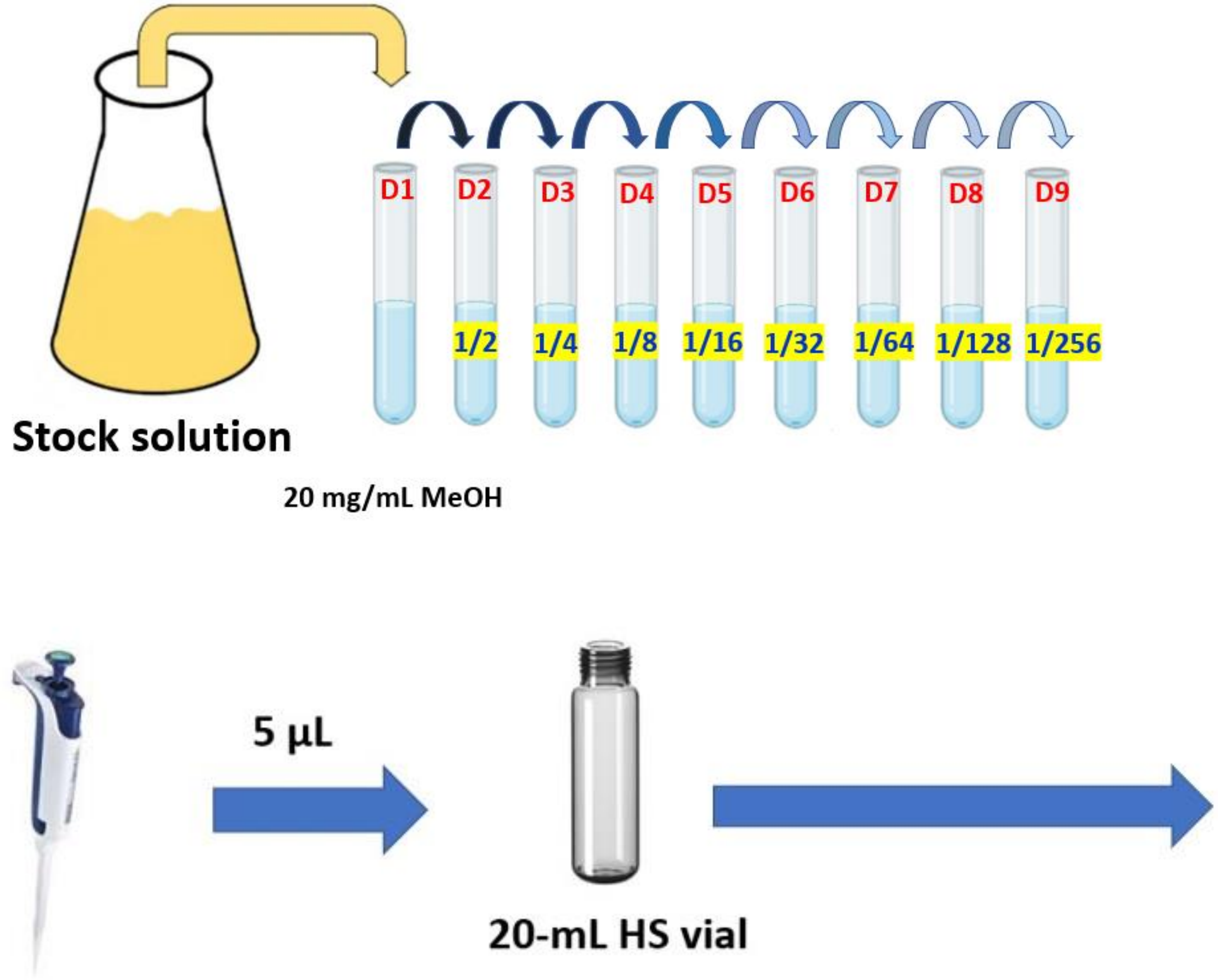
Sample collection and preparation (1/3)

- Five brands of gasoline were collected in Huntsville, Texas (Brand A, B, C, D, and E)



Sample collection and preparation (2/3)

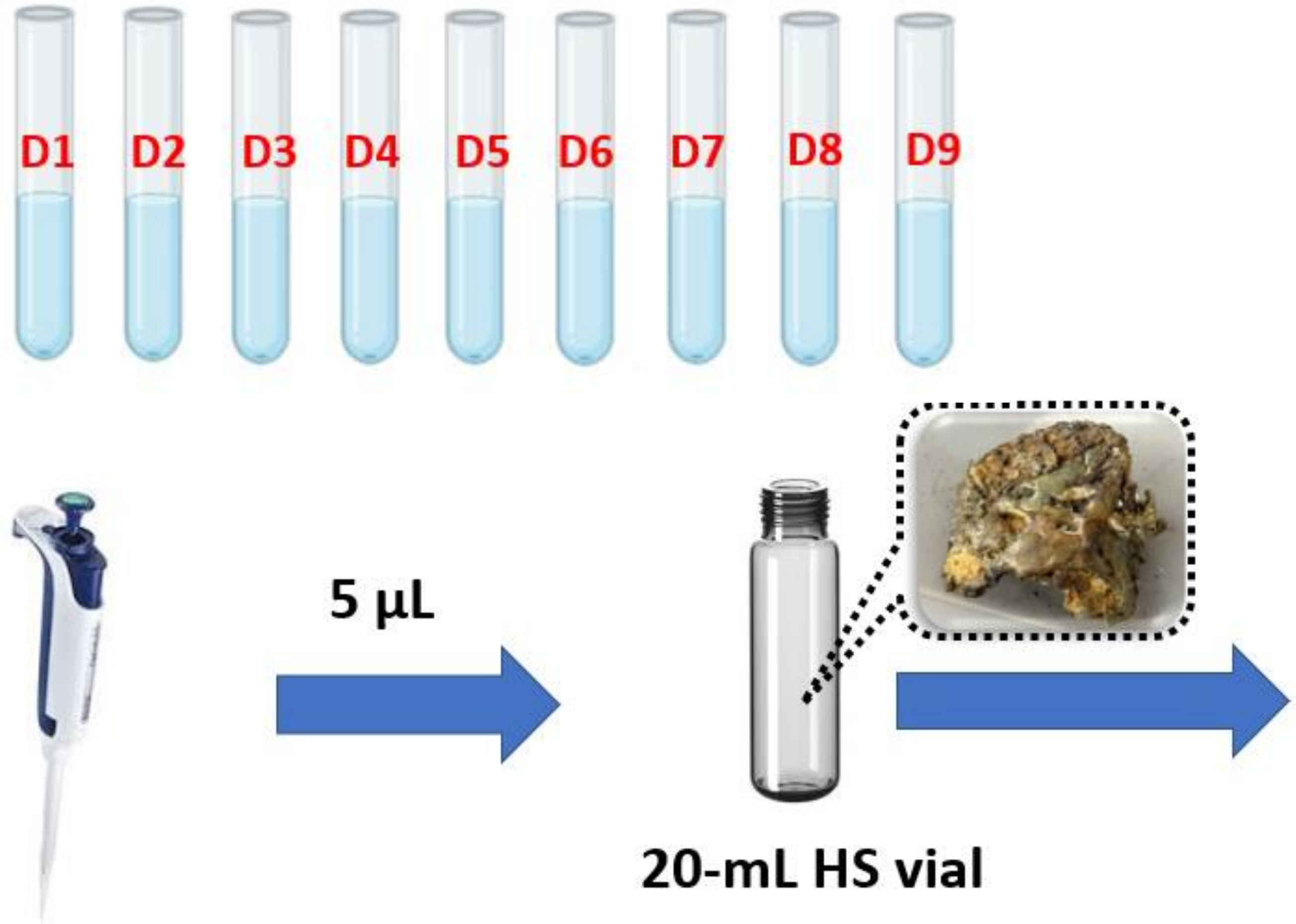
- Serial dilution of the gasoline samples (n=8 dilution series in triplicates)



	Analyte concentration
N1	100 µg gasoline/20-mL HS vial
N2	50 µg gasoline/20-mL HS vial
N3	25 µg gasoline/20-mL HS vial
N4	12.5 µg gasoline/20-mL HS vial
N5	6.3 µg gasoline/20-mL HS vial
N6	3.1 µg gasoline/20-mL HS vial
N7	1.6 µg gasoline/20-mL HS vial
N8	0.8 µg gasoline/20-mL HS vial
N9	0.4 µg gasoline/20-mL HS vial

Sample collection and preparation (3/3)

- Simulated fire debris samples
  - ➢ Nylon carpet
  - ➢ Direct heat method

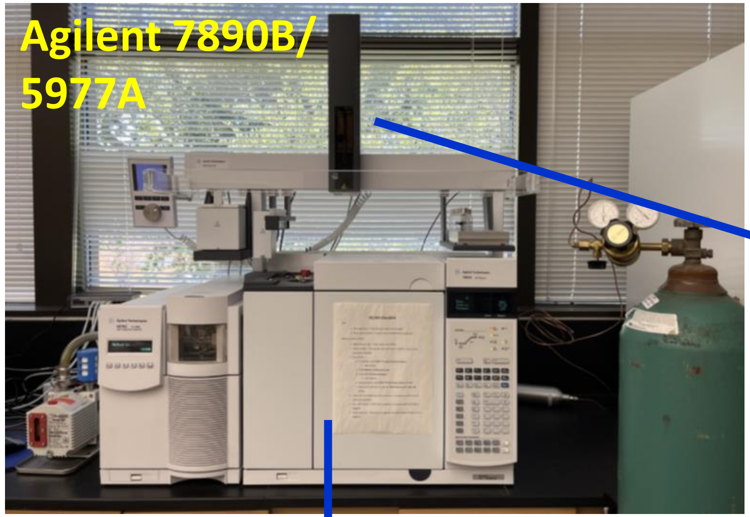


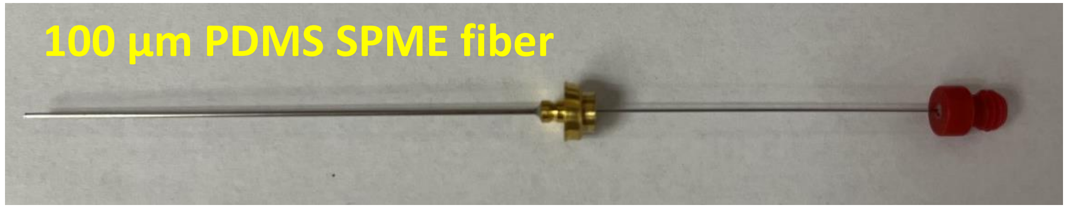
National Center for Forensic Science, https://ilrc.ucf.edu/substrate/criteria.php

	Analyte concentration
FD1	100 µg gasoline + 0.25 g substrate/ 20-mL HS vial
FD2	50 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD3	25 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD4	12.5 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD5	6.3 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD6	3.1 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD7	1.6 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD8	0.8 µg gasoline+ 0.25 g substrate/ 20-mL HS vial
FD9	0.4 µg gasoline+ 0.25 g substrate/ 20-mL HS vial

Instrumental analysis

- HS-SPME-GC/MS analysis





GC Oven Program Steps	Condition
GC oven initial temperature	40 °C
Hold time	5 min
Rate #1	10 °C/min
Oven temperature #1	150 °C
Rate #2	30 °C/min
Oven temperature #2	300 °C

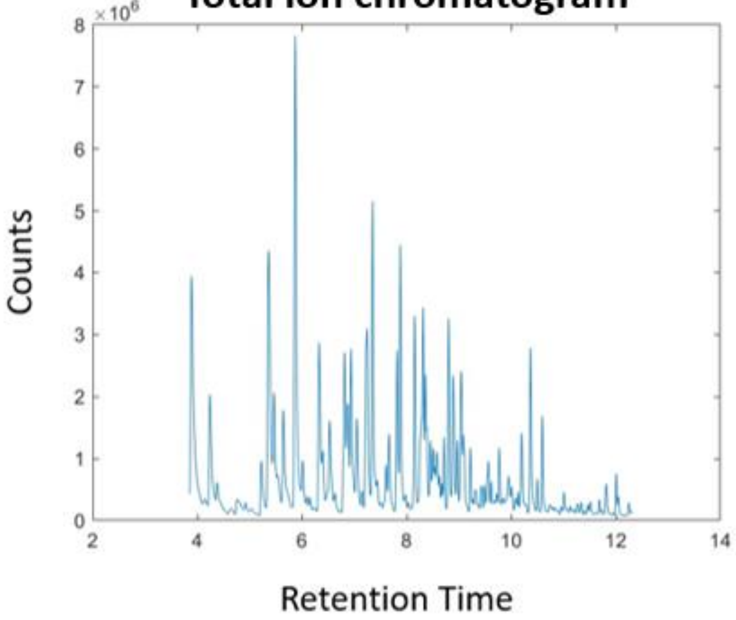
HHS-SPME Steps	Condition
Pre-Fiber Conditioning Temperature	250 °C
Pre-Fiber Conditioning Time	60 s
Pre-Incubation Time	300 s
Incubation Temperature	80 °C
Extraction Time	120 s
Desorb to	GC injection port
Desorption Time	120 s
Post-Fiber Conditioning Temperature	250 °C
Post-Fiber Conditioning Time	600 s
GC Runtime	1200 s

Image transformation

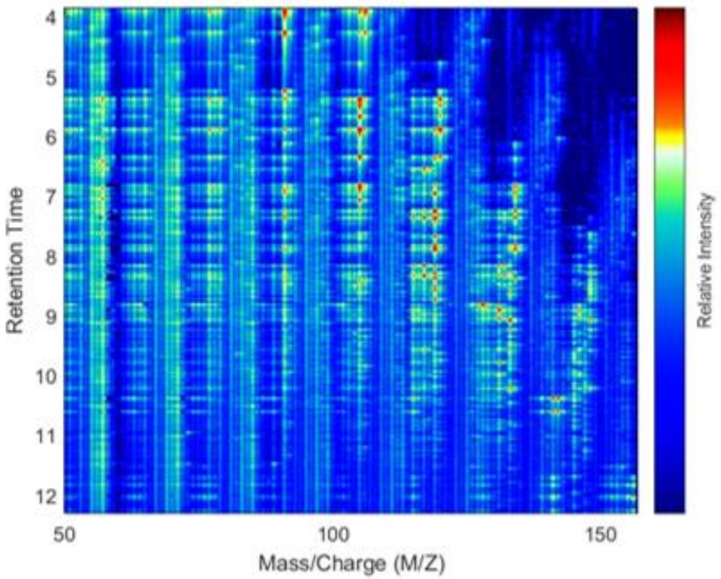
- GC/MS data were transformed into 3 types of images

12122022_PDMS_B_78_1.D **(Gasoline brand A, N1)**

Total ion chromatogram

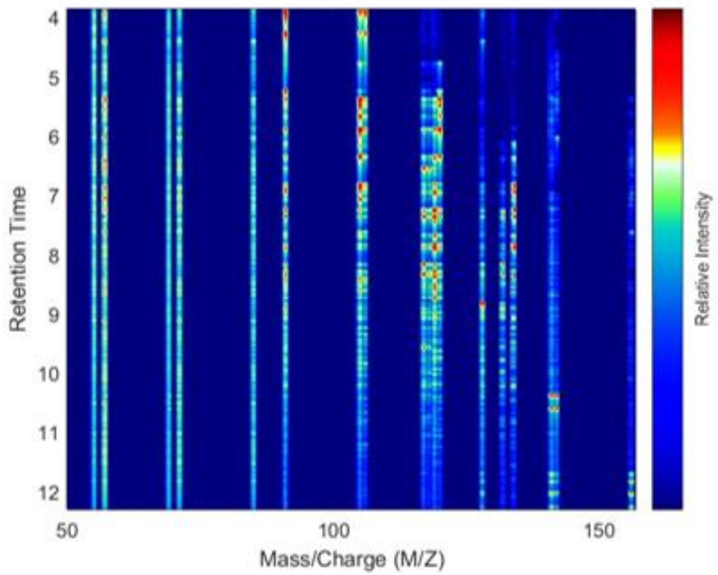


Heatmap



m/z **55-156**
Retention time 3-12 min

Extracted ion heatmap



Alkane ***m/z*** **57,71,85**
Cycloalkane and alkene ***m/z*** **55,69**
Alkylbenzenes ***m/z*** **91,105,106,119, 120,134**
Indanes ***m/z*** **117,118,132**
Alkylnaphthalenes ***m/z*** **128,142,156**

*Note: All axes and labels of the images were removed during transfer learning.

Transfer learning (1/2)

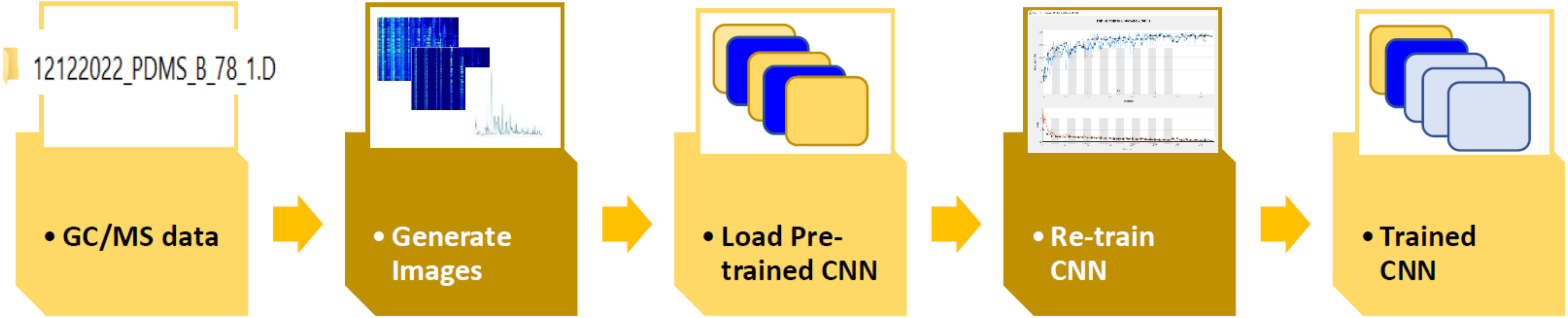
- Preparation of data sets

<table>
<tr><th></th><th colspan="4">Number of GC/MS data acquired</th><th colspan="2">Total number of transformed images</th></tr>
<tr><td rowspan="6">Training</td><td rowspan="5">Gasoline present</td><td rowspan="5">Neat samples</td><td>Brand A</td><td>63</td><td rowspan="5">315</td><td rowspan="6">390
• 80% for training
• 20% for validation</td></tr>
<tr><td>Brand B</td><td>63</td></tr>
<tr><td>Brand C</td><td>63</td></tr>
<tr><td>Brand D</td><td>63</td></tr>
<tr><td>Brand E</td><td>63</td></tr>
<tr><td>Gasoline absent</td><td colspan="2">Burned carpet</td><td>75</td><td>75</td></tr>
<tr><td rowspan="3">Verification</td><td rowspan="2">Gasoline present</td><td colspan="2">Neat samples</td><td>90</td><td rowspan="2">180</td><td rowspan="3">195</td></tr>
<tr><td colspan="2">Simulated fire debris samples</td><td>90</td></tr>
<tr><td>Gasoline absent</td><td colspan="2">Burned carpet</td><td>15</td><td>15</td></tr>
</table>

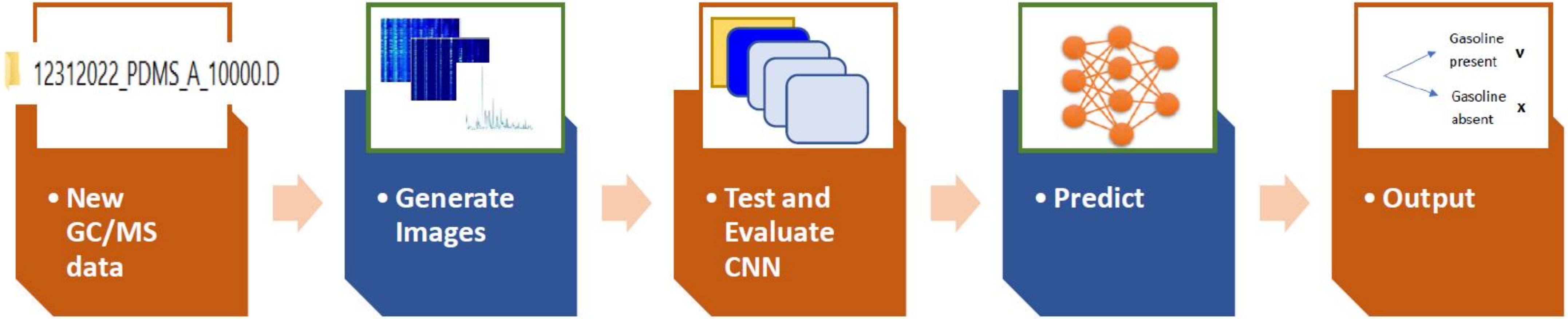
585

- Re-train GoogLeNet for gasoline detection

Training



Verification



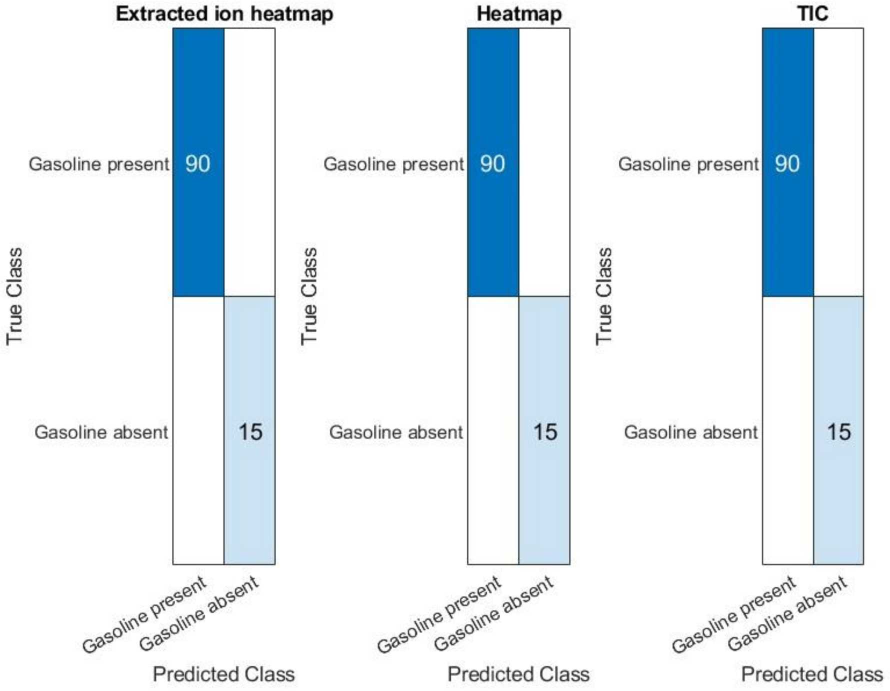
Predictions on the verification data set (1/5)

- Training: validation accuracy

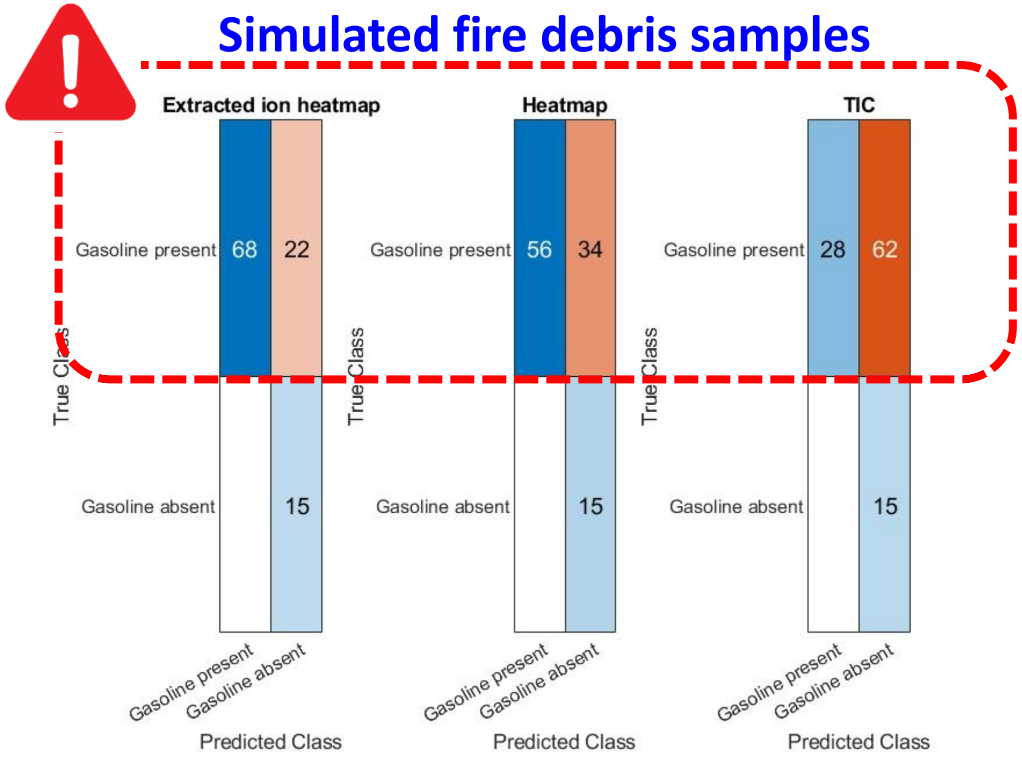
Extracted ion heatmap	Heatmap	TIC
100%	100%	100%

- Verification:

Neat samples

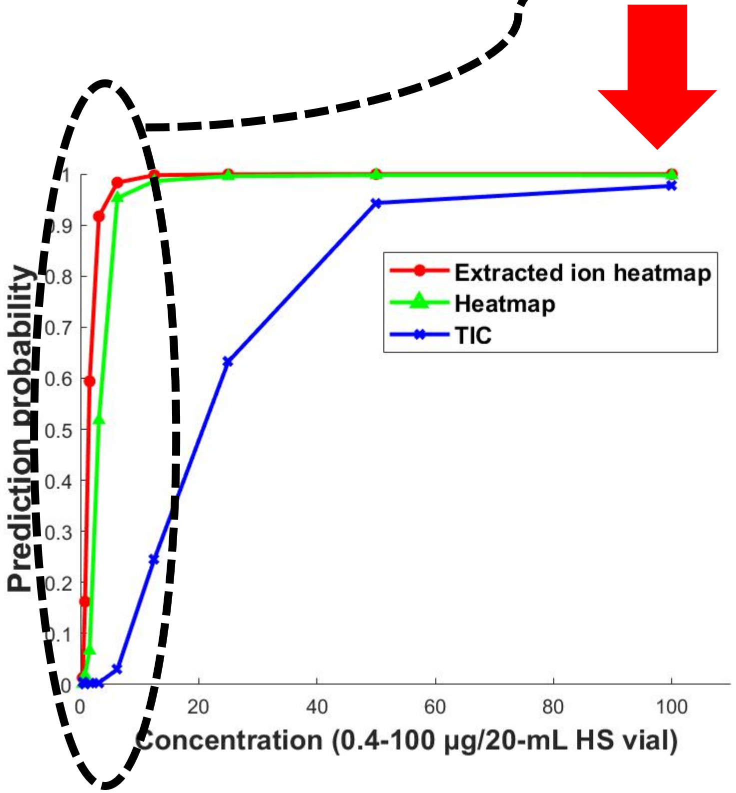


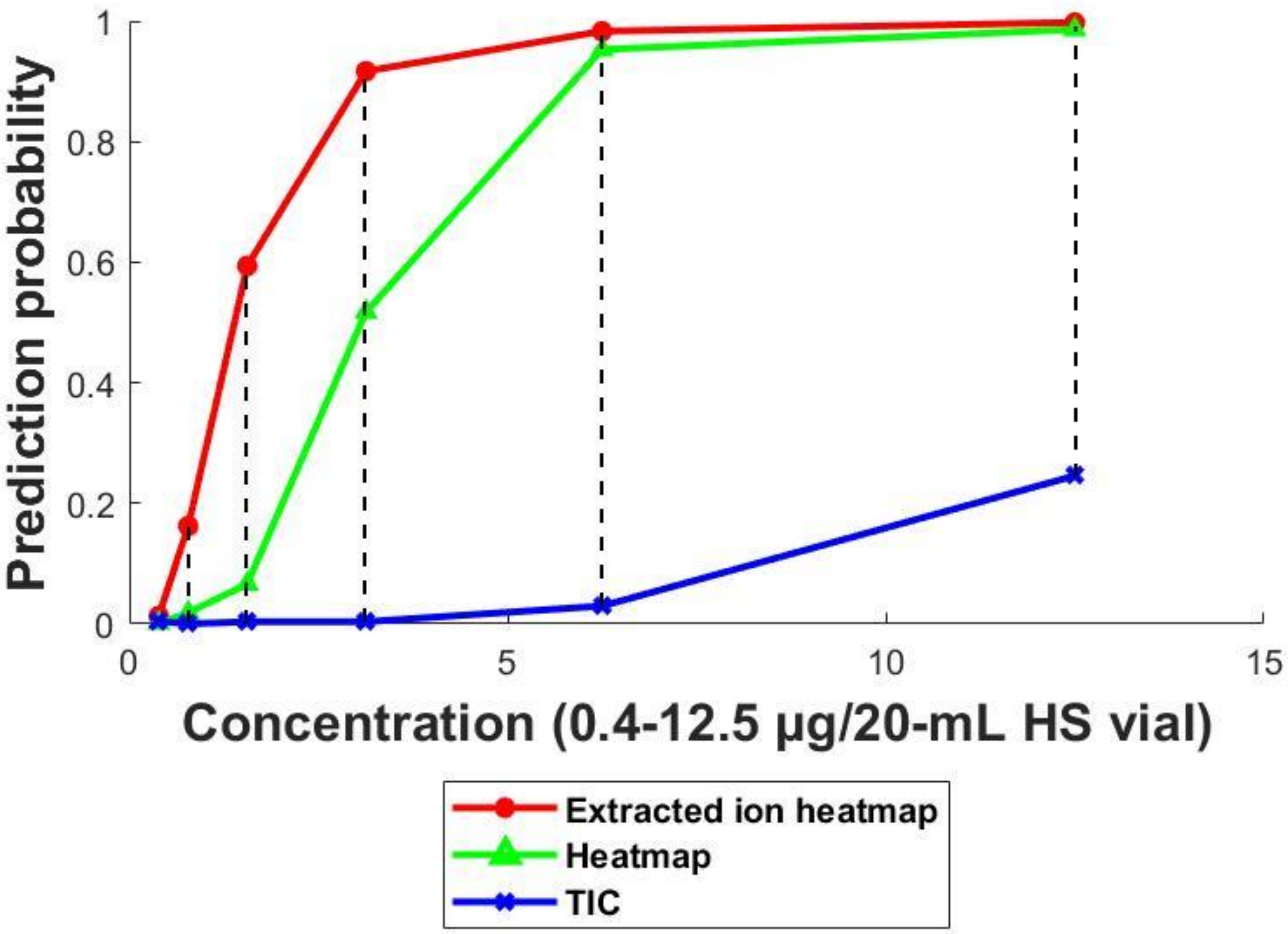
Simulated fire debris samples



Predictions on the verification data set (2/5)

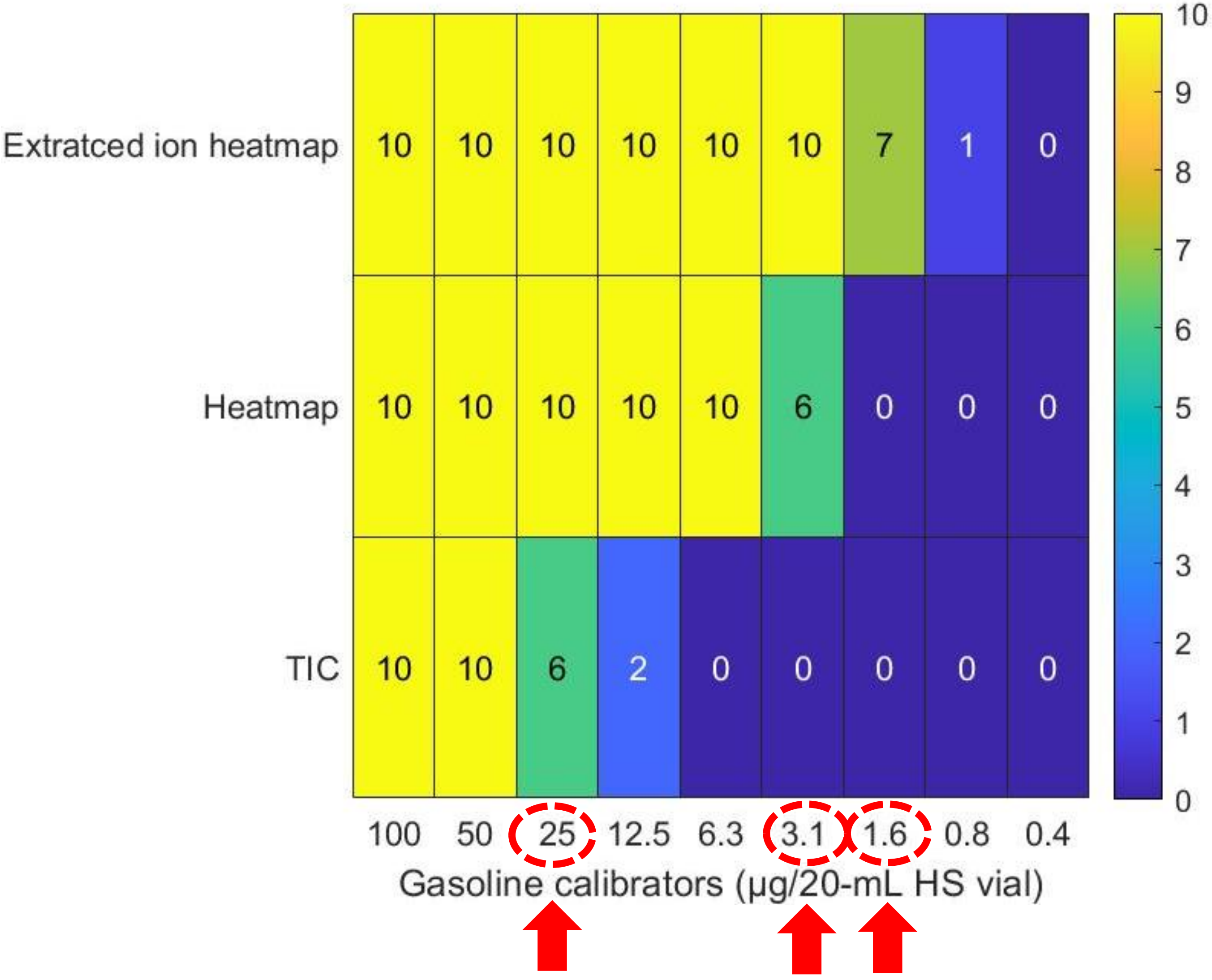
- Comparison of prediction probability for simulated fire debris samples





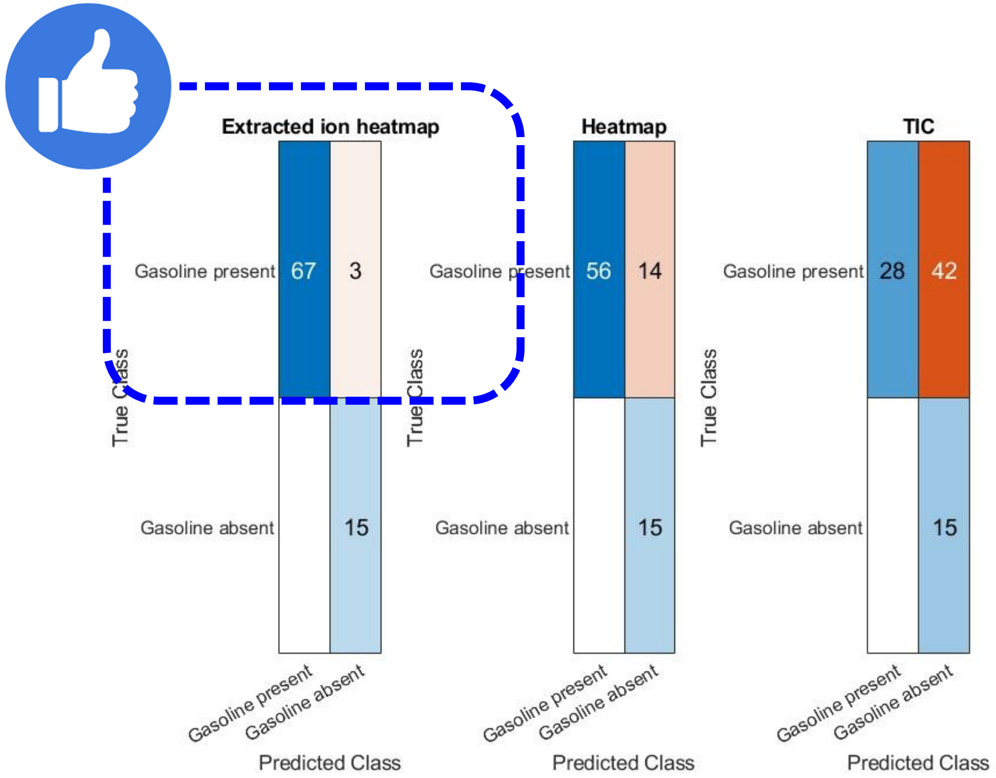
Predictions on the verification data set (3/5)

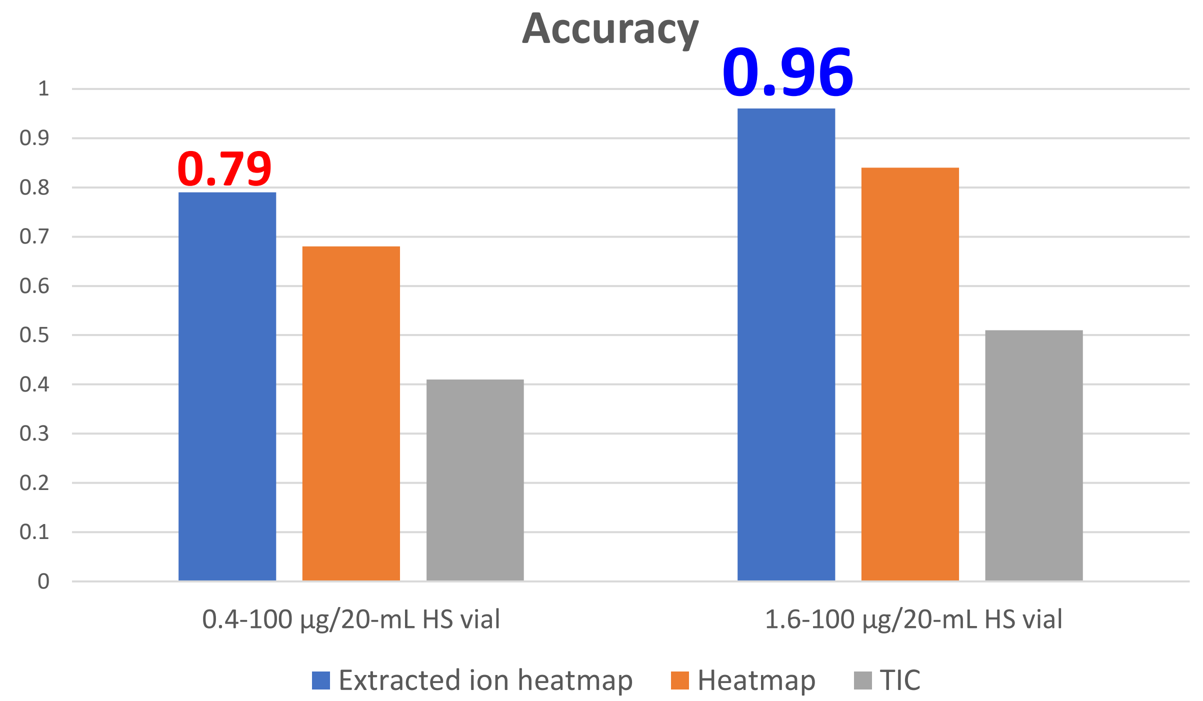
- Comparison of correct predictions for simulated fire debris samples



Predictions on the verification data set (4/5)

- Comparison of the predictions on the simulated fire debris samples at 1.6 - 100 µg gasoline sample/20-mL HS vial

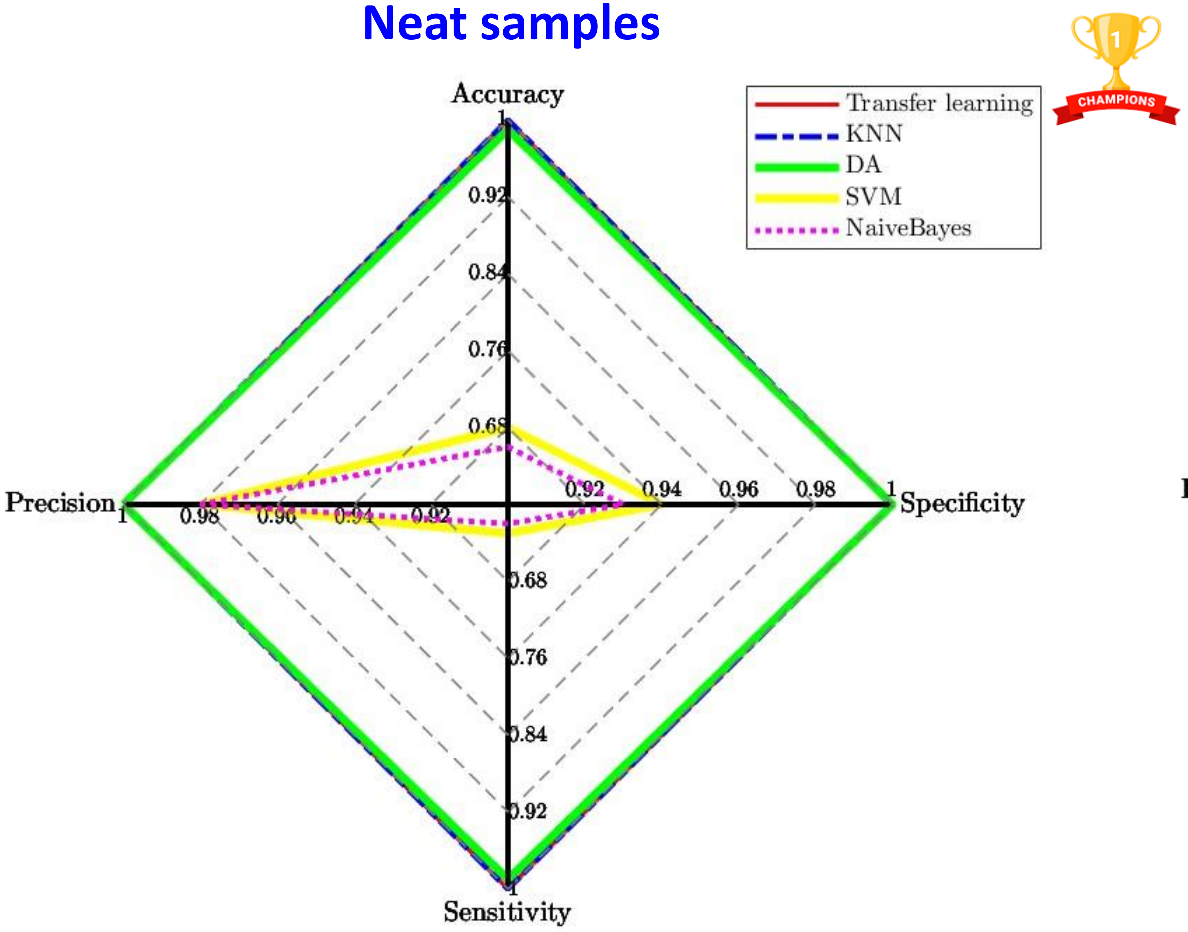




Predictions on the verification data set (5/5)

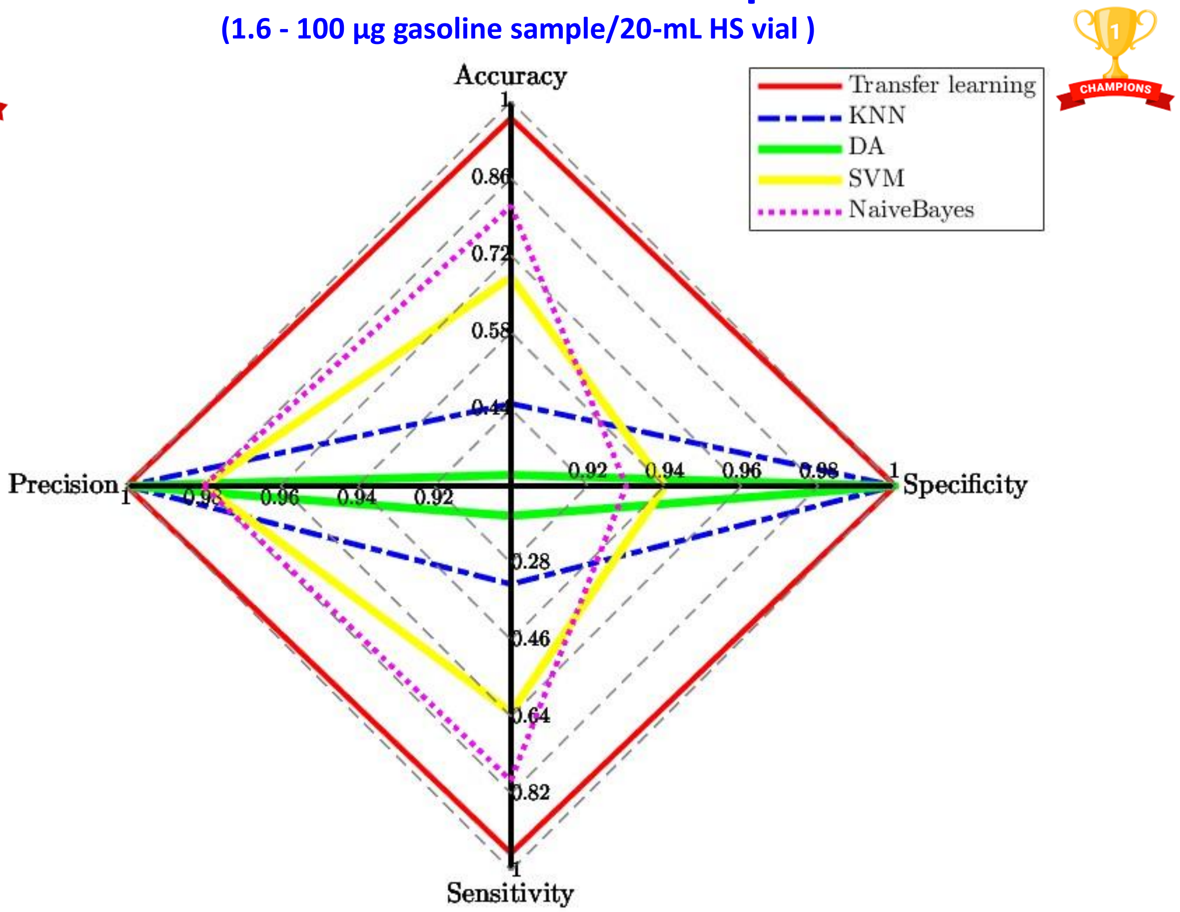
- Comparison of classification performance between the extracted ion heatmap and four ML models

Neat samples



Simulated fire debris samples

(1.6 - 100 µg gasoline sample/20-mL HS vial)



Conclusions

Experimental outcome

- TIC and heatmaps provided characteristic features of gasoline chemical profiles for transfer learning
- High performance for neat samples; limitation on fire debris samples
- Classification performance:
  - ➢ Heatmap > TIC
  - ➢ Extracted ion heatmap > all ion range heatmap
  - ➢ Extracted ion heatmap > ML models

Intelligent workflow

- No dependency on manual feature extraction
- Achieved high accuracy without large-scale data collection
- More capable of discriminating mixtures compared to other ML models

Acknowledgements

This work was partly funded by the 2022-2023 Forensic Sciences Foundation (FSF) Lucas Research Grants. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect those of the FSF. The authors also appreciate the funding support from the Ministry of Education, Taiwan.

Thank you

Ting-Yu Huang
txh038@shsu.edu