Machine learning in forensic fire debris analysis

Subjectivity and bias in existing Forensic Science methodologies can unfortunately distort the decision-making process in determining whether ignitable liquid residue is present or not in the debris from a fire. Investigating decision theory, statistical and machine learning (ML) methodologies in evaluating evidence from fire debris is the research focus of Professor Michael E. Sigman and Mary R. Williams, M.S., University of Central Florida. Their results demonstrate how forensic analysts could rely more on computers and ML to transition from making categorical to more nuanced statements about the value of evidence, possibly reducing flawed testimony, based only on human interpretation.

Destruction of land and property, not to mention loss of life, are all devastating effects of arson. Forensic investigators and scientists must collect and analyze evidence to help determine whether a fire arose from malicious intent (e.g., arson) or not. Arsonists often use ignitable liquids (IL) from commercial sources to start fires. The resulting IL residues are extracted from the fire debris and analyzed in a laboratory. The resulting data generates visual patterns based on the chemical compounds present in the sample.

FIRE REMNANTS LEAVE CLUES
The American Society for Testing Materials has a standard test method for IL (ASTM E1618-19). This method classifies them into one of seven classes (among them aromatic products, gasoline, and petroleum distillates) and a miscellaneous category for liquids with no class characteristics or multiple class characteristics. To determine whether a fire debris sample contains IL residue or not, the test makes use of the technique gas chromatography–mass spectrometry (GC-MS). This laboratory-based method separates out the chemical components in a sample (depending on their size, chemical and physical properties) and measures their relative concentrations.

A fire debris analyst makes visual comparisons of patterns contained in the sample readouts from the GC-MS data between a possible IL residue and a reference IL. The visual comparison is complicated by background contributions from partial burning of materials at the scene and changes in the IL pattern produced by the fire, resulting in obscured, complicated and variable data patterns. Historically, use of computer-based pattern recognition methods have met with resistance among the fire debris analyst community, partially due to the challenge to explain the methods in court. The important forensic decision in fire debris analysis is whether an IL residue is present or absent, but if only visual comparison is used the decision is susceptible to human error. Although the human mind is very good at pattern recognition, our performance is diminished by factors like fatigue and bias, making results prone to subjectivity. Machine learning (ML), a form of artificial intelligence, can help forensic analysts reach a decision. Computers use ML and large amounts of data to arrive at statistically reliable decision rules without human interference. In psychology and medicine, decisions which rely on statistical prediction rules, entirely or in combination with analyst experience, are reported to always be superior to human decisions that rely only on experience and training. A partnership between analysts and ML holds the promise of outperforming either one alone.

Arsonists often use ignitable liquids (IL) from commercial sources to start fires.

FINDING BIG DATA
Machine learning is a data hungry process. Depending on the ML method, the number of examples required to train the method can considerably exceed the amount of data that can reasonably be generated in a laboratory over a period of years. Dr Michael Sigman and Mary Williams at the University of Central Florida have combined records from databases of IL and burned substrates to produce computational (in-silico) fire debris data for training ML methods. Using this approach has allowed the research group to train ML methods on tens of thousands of examples.

The databases used to generate the in-silico examples have been developed and curated by Williams over the past 21 years. These databases include data on commercial ignitable liquids, products from partial burning of materials, and also fire debris. They are freely available online:
- http://ilrc.ucf.edu/
- http://ncfs.ucf.edu/internationaldb/index.php?design=ILRC1
- http://ilrc.ucf.edu/substrate/index.php
- http://ilrc.ucf.edu/firedebris/index.php

Testing ML models requires a set of samples that are experimentally prepared in such a way that they are known to contain mixtures of burned substrates with and without IL, in varying IL contributions. These are referred to as 'ground-truth' samples and they serve as the ultimate test for the accuracy of an ML model in predicting the presence or absence of IL residue. The research team has recently generated a database of nearly 1,000 such ground-truth samples.

Ground-truth samples are required to measure the performance of ML methods, nonetheless it is also important to test the methods on 'real-world' data that is representative

Destruction of land and property, not to mention loss of life, are all devastating effects of arson.

of casework samples. In a set of field experiments, the researchers built-out storage-containers to resemble two-room apartments, installed new furniture and flooring, and then burned the containers with the aid of carefully placed IL. After extinguishment, fire debris samples were collected from known locations where IL had been poured and locations away from the IL pour. The real-world samples were evaluated for the presence of IL residue by an 'informed analyst' who knew the GC-MS patterns for the IL used to start the fire and the sample locations relative to the pour locations. The informed analyst labelled each sample as positive or negative for IL residue.

NEURAL NETWORK ML MODELS
Recently, the researchers investigated the potential of using neural networks, biologically inspired ML models.

Selected ions from the GC-MS data were used as inputs into a neural network classification model. An optimal neural network model was selected from a subset of candidates trained on in-silico mixed fire debris samples. The University of Central Florida researchers found that sufficient information from the selected ions can allow models to be developed that discriminate between ground-truth samples containing an IL and those not containing an IL with a high accuracy. These results highlight the potential of neural network models to assist analysts in the evaluation of fire debris evidence.

MAKING BETTER DECISIONS
With its beginnings in radar equipment performance dating back to World War II, decision theory has found application in many walks of life from weather forecasting to predicting the finished quality of a Bordeaux wine. Its use in the forensic field of fire debris analysis, however, has largely been neglected until, in combination with ML techniques, it became the focus of research for Sigman and Williams. The researchers' results illustrate how ML can help analysts evaluate the evidentiary strength of fire debris samples. An unbiased estimate of evidentiary strength provides a basis for making better forensic decisions.

Fire debris samples either contain IL residue (designated positive) or they do not (designated negative). Decision theory relies on determining the rate of correct (True Positive Rate) and incorrect positive decisions (False Positive Rate) between the two alternatives as a function of the strength of the evidence. Accuracy of prediction rules can be evaluated objectively from 'Receiver Operating characteristic' (ROC) curves. Originally developed in World War II when the signal of enemy planes had to be discerned from random interference, ROC curves determine the percentage of true positive decisions ('hits') or false positive decisions (false alarms) as the strength of the evidence changes. At any point on the ROC curve, the slope of the curve is equal to the strength of the evidence. The slope of the ROC curve at any point is known as the likelihood ratio. Despite their ubiquity in the real world, ML and ROC curves were largely neglected in the practice of fire debris analysis until Sigman and Williams began to spearhead investigation into this compelling area of Forensic Science. Advances in their work signpost the way forward in reducing bias and subjectivity from result interpretation and assigning known error rates based on the strength of the evidence.

ARE DECISIONS NECESSARY?
Inequities in the judicial system can arise when analysts are required

Sigman and Williams investigate decision theory, statistical and machine learning (ML) methodologies in validating evidence from fire debris analysis.

to make decisions that result in categorical statements in court and reports, for example IL residue is present in the sample or it isn't, without disclosing the strength of the evidence that led to the decision. Each analyst has a personal threshold for declaring that IL is present, and variations among the individual thresholds lead to inequities and

presence or absence of IL residue patterns and promote transparency regarding evidential strength and decision thresholds.

The ability of jurors to understand and evaluate probabilistic statements of evidentiary value is an area for continued research in the social sciences.

The unbiased estimate of evidentiary strength provides a basis for making better forensic decisions.

The ROC curve is shown as a solid line. The slope of the dashed lines indicate the strength of evidence supporting the presence of IL residue. The strength of evidence A is stronger than evidence B. The red arrows point to the true positive rate and the false positive rate associated with a decision threshold corresponding to the evidential strength of sample A.
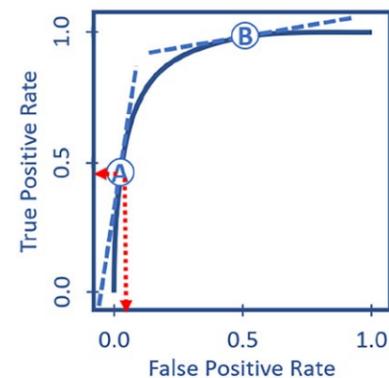
allows subjectivity into the legal system. The *European Network of Forensic Science Institutes* has emphasized replacing categorical decisions with a statistical approach to forensic reporting where written statements report the 'likelihood ratio', which is a measure of evidential strength. ROC curves are especially useful when considering likelihood ratios to obtain answers of forensic importance. This is an important step in making fire debris analysis a more reliable form of physical evidence. The use of ML and the likelihood ratio to guide important decisions is not a new concept but in the case of fire debris analysis their use would reduce reliance on visually recognizing the

THE PATH FORWARD
Forensic fire debris analysts and ML may form a promising partnership of the future. Analysts' current methods rely on visual pattern recognition and do not recognize differences in perceived level of confidence despite the strength of the evidence. On the other hand, ML methods must be properly trained and they may provide results of higher uncertainty when presented with new data from outside the distribution of training examples. A new area of artificial intelligence research known as evidential learning may help to solve these problems but until then, the partnership of analyst and ML seems a promising path forward.

Prof Michael E. Sigman

Mary R. Williams, M.S.

E: michael.sigman@ucf.edu   **T:** +1 407-823-6469   **W:** https://www.researchgate.net/profile/Michael-Sigman

Research Objectives

Professor Sigman and Ms Williams combine machine learning results with decision theory to validate evidence from fire debris analysis.

Detail

Address
University of Central Florida
P.O. Box 162367
Orlando, FL 32816-2367, USA

Bio
Michael E. Sigman is a Professor in the Chemistry Department and Director of the National Center for Forensic Science (NCFS) at the University of Central Florida. He has published numerous peer-reviewed articles, book chapters and reviewed reports, and

holds five patents. Sigman has served as Principal Investigator on research grants from multiple agencies.

Mary R. Williams holds an M.S. in Forensic Science and is currently a Research Specialist II in the NCFS at the University of Central Florida. She has published numerous peer-reviewed articles, book chapters and reviewed reports, and holds two patents. Williams is the curator for all fire debris databases at the NCFS,

co-principal investigator on multiple federal grants and assists in the direction of graduate students at NCFS.

Funding
National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

Collaborators
Fire debris analysts in forensic laboratories around the world.

References

- Thurn N.A., Wood T., Mary R. Williams M.R., Sigman M.E. (2021). Classification of ground-truth fire debris samples using artificial neural networks. *Forensic Chem.* 23, 100313. https://doi.org/10.1016/j.forc.2021.100313
- Allen, A., Williams M.R. & Sigman M.E. (2019). Application of likelihood ratios and optimal decision thresholds in fire debris analysis based on a partial least squares discriminant analysis (PLSDA) model. *Forensic Chem.* 16 (2019) 100188. https://doi.org/10.1016/j.forc.2019.100188
- Coulson R., Williams M.R., Allen A., Akmeemana A., Ni L., Sigman M.E. (2018). Model effects on likelihood ratios for fire debris analysis, *Forensic Chem.* 7, 38–46.
- Sigman M.E. & Williams M.R. (2016) Assessing evidentiary value in fire debris analysis by chemometric and likelihood ratio approaches. *Forensic Sci Int.* 264 113–121. https://doi.org/10.1016/j.forsciint.2016.03.051
- Waddell E.E., Song E.T., Rinke C.N., Williams M.R., Sigman M.E. (2014). Progress toward the determination of correct classification rates in fire debris analysis, *J. Forensic Sci.* 58, 887–896. doi:10.1111/1556-4029.12417
- Dawes, R. M. (2002). The ethics of using or not using statistical prediction rules in psychological practice and related consulting activities. *Philosophy of Science,* 69(S3), S178-S184. https://doi.org/10.1086/341844

Personal Response

How far away is laboratory-generated, procedure validated ground truth data? What needs to happen to remove barriers to entry?

❝ Ground truth data that has been laboratory-generated by a validated process is available in an open-access database (http://ilrc.ucf.edu/firedebris/index.php). ❞