We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

5,500
Open access books available

136,000
International authors and editors

170M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter

Explainable Artificial Intelligence for Digital Forensics: Opportunities, Challenges and a Drug Testing Case Study

Louise Kelly, Swati Sachan, Lei Ni, Fatima Almaghrabi, Richard Allmendinger and Yu-Wang Chen

Abstract

Forensic analysis is typically a complex and time-consuming process requiring forensic investigators to collect and analyse different pieces of evidence to arrive at a solid recommendation. Our interest lies in forensic drug testing, where evidence comprises a multitude of experimentally obtained data from samples (e.g. hair or nails), occasionally combined with questionnaire data, with a goal of quantifying the likelihood of drug use. The availability of intelligent data-driven technologies can support holistic decision-making in such scenarios, but this needs to be done in a transparent fashion (as opposed to using black-box models). To this end, this book chapter investigates the opportunities and challenges of developing interactive and eXplainable Artificial Intelligence (XAI) systems to support digital forensics and automate the decision-making process to enable fast and reliable generation of evidence for the court of law. Relevant XAI techniques and their applications in forensic testing, including feature section, missing data handling, XAI for multi-criteria and interactive learning, are discussed in detail. A case study on a forensic science company is used to demonstrate the real challenges of forensic reporting and potential for making use of forensic data to pave the way for future research towards XAI-driven digital forensics.

Keywords: digital forensics, drug testing, machine learning, explainable AI, decision-making, automation

1. Introduction

The primary focus of forensic analysis is the acquisition of accurate and reliable evidence through the utilisation of methodologies that have proven consistent and trustworthy across the domain [1]. The evidence is presented to the court of law and the prosecutor must be satisfied with its reliability, credibility and admissibility. Forensic evidence can be extremely sensitive and dangerous for law enforcement to handle and the use of incorrect or unreliable evidence threatens the safety of justice.
Digital forensics (DF) was introduced as a means of digitally making use of forensic data for both the discovery and interpretation of electronic evidence [2]. This area has become increasingly important with the surge in the volume, variety and velocity of forensic data. Currently, the major challenges faced by DF investigators are an increase in the number of cases and the complexity of cases [1]. The increase in cases could be due to a societal shift towards faith in DF techniques, with the common belief being that advanced tools are highly useful in skilfully extracting and using forensic information [2]. The increasing complexity of cases is simply a result of advances in technology, storage and applications [1]. Another challenge for DF investigators is the requirement for fast turnaround. Due to the nature of forensic inquiries, investigators wish to have faster, more advanced and more accurate tools, in order to prevent any setbacks that could adversely affect the case. Furthermore, it is expected that new challenges will arise for DF in the near future as pointed out by Mazurczyk et al. [3], p. 10: ‘modern digital forensics is a multidisciplinary effort that embraces several fields, including law, computer science, finance, networking, data mining and criminal justice’.

Artificial Intelligence (AI) is a technology that has been used for many decades, with growing importance in the modern day due to its uses for learning and reasoning. AI methods are extremely capable of learning and solving complex computational problems and have subsequently been considered crucial for future developments; from explaining the reasoning process of expert systems, to recognising patterns in artificial neural networks [4, 5]. Although AI models have been developed to support parts of the court cases, current judiciary systems may raise concerns over the reliability of decisions made by AI models. Moreover, these models can be useful but only when explained to judges and jurors, such as in a study by Vlek et al. [6] where they used scenario scheme idioms to construct Bayesian Networks (BN), in order to make the network easier to understand. This method attempted to explain why certain modelling choices were made as well as why the network arrived at the final output, given the choices made along the way. Another paper by Timmer et al. [7] used BNs to formalise the relationship between the hypothesis and the evidence presented in the network, and the authors derived a support graph to assist with interpretation of the BN, which could then be used for argument and evidence about the case.

In view of the importance of explainability, there emerges XAI, a collection of AI methods that focuses on producing outputs and recommendations that can be understood and interpreted by human experts. A focus of the AI community at the moment is to develop XAI methods that have a good balance between both transparency and explainability as well as power, performance and accuracy [8]. The application of XAI models to DF problems is scarce but would open up the possibility of using computer-based analysis for evidence in courts of law. It could become an extremely powerful tool for helping judges and jurors make decisions in the presence of many interconnected pieces of evidence.

This chapter investigates the opportunities and challenges of applying XAI to support DF. First, this chapter discusses DF and the applications of AI in the forensics domain. Second, it reviews existing literature on XAI, feature selection methods built on various types of variables such as images and electrodermal activity for drug and alcohol testing, missing data handling techniques and XAI for multi-criteria and interactive learning and their implementation in DF. Third, it discusses a current case study on drug testing that includes problem formulation, a description of the forensics data collected from questionnaires and analytical testing, and the high-level decision-making process for drug screening. Finally, the chapter presents a conclusion drawn from this study and further work.
2. Background

This section puts this chapter in context by reviewing the area of XAI and its application to DF, and discussing several data-related challenges one may need to address to make the most out of XAI methods, such as dealing with a large number of variables/features, missing data, multiple (conflicting) output criteria and interactions between the AI system and the practitioner.

2.1 XAI and its application in digital forensics

With ML being the core technology, AI systems have made remarkable achievements in solving increasingly complex computational tasks and making them critical aspects of the future development of human society [4]. However, in case of ML algorithmic models pursuing prediction accuracy and becoming increasingly opaque, the explainability becomes problematic for black-box techniques such as ensemble methods and deep neural networks [9].

To address the trade-off between interpretability and model performance, post-hoc interpretability techniques emerge, which approximate black-box models by techniques such as simplification, feature relevance estimation, or visualisation. Eventually, the opaque models are turned into glass-box, which achieve a good trade-off between interpretability and prediction accuracy. Examples of such techniques include local interpretable model-agnostic explanations (LIME) [10], which explain the predictions by approximating the opaque black-box model with simple models locally, and SHapley Additive exPlanations (SHAP) [11], which calculate the contribution of each feature to the prediction based on three desirable properties (i.e. local accuracy, missingness and consistency). These techniques are referred to as XAI, which propose creating a collection of ML techniques that generate more explainable, understandable and trustworthy models without losing out significantly in prediction accuracy [8]. XAI methods can be classified according to multiple criteria, including intrinsic or post hoc, model-specific or model-agnostic and local or global interpretability [12].

2.1.1 Intrinsic or post hoc?

This criterion distinguishes whether XAI is achieved intrinsically or post hoc. Intrinsic interpretability refers to ML models that are interpretable because of their simple structures (e.g. linear models, tree-based models). Post hoc interpretability refers to the use of methods like feature importance and partial dependency plots in explaining the black-box models (e.g. ensemble methods, neural network) after training.

2.1.2 Model-specific or model-agnostic?

For model-specific techniques, interpretability is incorporated within the internals (i.e. inherent structure and learning mechanisms) and is limited to specific models. In contrast, model-agnostic methods, as named, are irrelevant to the inner processing/structure of the model. They can be seamlessly used on any ML model and are applied after the model has been trained [12].

2.1.3 Local or global?

The scope of the interpretability, global to the model or local to the prediction, is another important criterion [10]. Global interpretability refers to the entire model
behaviour and answers ‘show does the trained model make predictions?’. Local interpretation methods explain a single prediction which influences a user’s confidence in the prediction and consequently, the user’s action.

DF, which requires the intelligent analysis of large amounts of complex data, is benefiting from AI. Mitchell [5] reviewed some of the basic AI techniques that have been applied to the DF arena. These include expert systems in explaining the reasoning process, Artificial Neural Network (ANN) in pattern recognition, and decision trees acting as learning the rules for pattern classification and expert system. Irons and Lallie [13] also identified the use of AI techniques to automate aspects like identification, gathering, preservation and analysis of evidence in DF process. In recent years, the importance and requirement of using explainable methods which achieve both the robustness of algorithms and transparency of reasoning have been increasingly acknowledged in DF. Interpretable ML classifiers like decision trees and rule-based models have been commonly applied to DF problem [14, 15]. To explain a legal case, the community has also applied the idea of BN [6, 7]. AfzaliSeresht et al. [16] presented an XAI model in which event-based rules are created to generate stories for detecting patterns in security event logs for assisting forensic investigators. Mahajan et al. [17] applied LIME towards toxic comment classification in cyber forensics and achieved both high accuracy and interpretability compared to various ML models. However, in terms of automated decision-making in DF, there are very few works that have been made to make it explainable. Figure 1 provides the classification of XAI techniques and their recent applications in DF.

2.2 Feature selection and dimensionality reduction

The increase in the availability of data due to a push in digitisation has led to high-dimensional data sets for training and testing AI algorithms. However, the
amount of available data is just as important as the quality of the data. To ensure high-quality data is being filtered out from redundant, irrelevant, or noisy data [18], one can apply feature selection. Selecting the most relevant features has been shown to increase prediction accuracy, since it simplifies the model [19] and removes redundant in features [20]. However, the situation of having too little data needs to be avoided where possible to reduce the risk of overfitting, which occurs when a function is too closely fit to a limited set of data points. It is worthwhile highlighting the difference between feature selection and dimensionality reduction: while both methods reduce the number of features in a dataset, feature selection is achieving this by simply selecting and excluding given features without changing them, dimensionality reduction transforms features into a lower dimension. Our focus is more on feature selection methods. However, commonly used dimensionality reduction methods include Principal Component Analysis (PCA), Random Projection, Partial Least Squares and Information Gain.

Feature selection methods are categorised in Figure 2 according to their process of ranking features into filter, wrapper and embedded techniques [21]. Filter methods are techniques that rank the relationship of features with an outcome without learning a model, such as Separability and Correlation Analysis (SEPCOR) [20]. Univariate filters calculate the ranking for each individual feature, while multivariate filters compute the ranking based on the correlation between the variables or between the variables and the outcome [22]. Wrapper techniques select features by comparing all the combinations of the included features before starting

![Figure 2. Classification of feature selection techniques.](image-url)
the prediction model, to find the most accurately predictive one [22]. Wrapper techniques are more computationally expensive than filters; however, they generally produce more accurate results. Finally, embedded methods are classifier-dependent selection methods, where the selection is built based on the classifiers’ chosen hypotheses [23].

Many comparative studies have been performed to find the best feature selection technique for high-dimensional data. For example, Hua et al. [24] compared a wide range of feature selection techniques for a variety of high-dimensional datasets. The authors followed a two-stage feature selection process to reduce computational time. In the first stage, feature selection methods that are independent from the classification process were applied. Following that, a further feature selection was implemented through classifier-specific feature selection techniques. The results show that wrapper methods have better performance in datasets with large samples, and filters have generally equal error trend. One of the main conclusions of their paper is that there is no feature selection technique performed best across all datasets. Another review of feature selection methods for high-dimensional datasets, which focused on filters, was conducted by Ferreira and Figueiredo [25]. The authors compared, amongst others, the following feature selection techniques for supervised learning: ReliefF, correlation-based filter selection, fast correlation-based filter, Fisher’s ratio and minimum redundancy maximum relevance. Other solutions to tackle high-dimensionality in feature selection are the choice of an adequate evaluation criteria, such as predictive measures designed for small sample datasets and ensemble feature selection methods, including combining multiple feature selection methods and boosting [26].

Table 1 provides an overview of different feature selection methods applied to forensic science applications. Shri and Sriraam [20] formulated a feature extraction and feature selection problem to detect the difference between alcoholics and control groups through measuring the impact of the use of alcohol in multichannel EEG signal regions. Feature subset selection was performed using separability and correlation analysis, which was proposed in the paper. The results illustrate that the introduced technique improved prediction accuracy, and further validation using

<table>
<thead>
<tr>
<th>Forensic application</th>
<th>Type of feature selection</th>
<th>Algorithm</th>
<th>Type of data</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol testing</td>
<td>Filter method</td>
<td>Separability and correlation analysis</td>
<td>EEG signals, eye blink artefact and motion artefact</td>
<td>[20]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feature ranking using area under the curve</td>
<td>Continuous data</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feature ranking using area under the curve</td>
<td>Categorical and continuous data</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear Discriminant Analysis (LDA)</td>
<td>Images</td>
<td>[29]</td>
</tr>
<tr>
<td>Screening substance use disorder</td>
<td></td>
<td>A discriminant function analysis</td>
<td>Categorical and continuous data</td>
<td>[30]</td>
</tr>
<tr>
<td>Drug testing</td>
<td>Wrapper method</td>
<td>Linear Discriminant Analysis (LDA)</td>
<td>Mass spectral data</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td>Exhaustive search method</td>
<td>Continuous and time domain features</td>
<td></td>
<td>[32]</td>
</tr>
</tbody>
</table>

Table 1. Selected applications of feature selection techniques in forensic research.
other classifiers and cross-validation is recommended. Another feature selection technique to enhance screening of alcohol use disorder was introduced by Mumtaz et al. [27]. The EEG features were recorded in 5-minutes eyes open and 5-minutes eyes closed segments. The implemented feature selection takes two steps. First, the relevance of each feature to the outcome is calculated using the ROC. Then, Markov blanket filtering combined with the ROC is used to remove redundant features. The second step has a high computational cost, which is one of the drawbacks of this method. The paper found that the inter-hemispheric coherence between the brain regions ranked the highest in classifying alcohol use disorder (AUD). Mumtaz et al. [28] designed a rank-based feature selection technique in response to the high-dimensionality in the dataset. Feature ranking was computed based on the area under the curve of that feature and represented the relevance of the feature to the outcome. The minimum number of features was chosen by adding the features to the model sequentially, starting from the highest-ranked features.

Another alcohol use detection method based on thermal infrared facial images was examined in [29]. The dimensionality reduction was carried out using PCA combined with Linear Discriminant Analysis (LDA) [33]. It was shown that LDA worked well if the data had no missing data [34]. In an application for feature selection [30], applied discriminant function analysis for substance use disorder detection. This disorder is usually related to P3 amplitude, addiction severity and impulsivity in predicting treatment completion. The research found that the P3 amplitude accounts for more variance compared to other variables.

Mahmud et al. [32] designed a method for quick detection of opioid intake using wrist-worn biosensor-generated data. The exhaustive search method was applied to seek a set of variables that achieved the highest accuracy. It helped to minimise the computational time and increased the prediction accuracy and sensitivity. Feature selection methods have also been applied to identify illegal drugs [31]. PCA followed by LDA was implemented for drug isomer differentiation. Three feature selection models that were tested included the full spectrum, exclusion of selected masses and the selected region, where ions are expected to contribute to the isomeric difference.

To summarise, feature selection methods have been implemented in forensic research and particularly for the detection of substance use. Their application covers various types of data, including images, EEG signals and time-series. Most of the reviewed methods were based on a filters approach. However, since most of these applications have selected the features for classification purposes, embedded techniques are designed to integrate the selection in the classification process. Therefore, it is important to investigate other embedded and wrapper feature selection methods.

2.3 Missing data

Forensic data contains a large number of features. A proportion of information in these features could be missing, which would reflect a different level of uncertainty because they are measured independently in laboratories [35]. High-dimensional forensic data presents challenges in establishing unbiased estimation and inference of ML models. Missing and uncertain forensic data must be treated in the data preprocessing stage, before the development of ML models. The deletion of incomplete instances and imputation of missing data is the most frequently used

\[1\] The P3 is a positive deflection of EEG that occurs when a low probability novel, target, or oddball stimulus is presented within a sequence of high probability non-targets or standards [30].
method of handling missing data, however the removal of the incomplete instances results in biased inference due to poor representation of complete samples [36, 37].

Statistical methods based on data imputation are largely utilised to handle missing data. The basic idea is to replace the missing values with the predicted values obtained from the observed data. There are three types of missing data—missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) [38]. The missingness mechanism by MCAR is independent of observed and unobserved data whereas, MAR is independent of unobserved data and dependent on the observed data. The missingness mechanism by MNAR is only dependent on unobserved data. The forensic datasets are usually MCAR type.

The missing forensics data can be imputed by methods such as Multivariate imputation by Chained Equations (MICE), Maximum likelihood estimation (MLE), Random Forest (RF), K-nearest neighbour (KNN) and MICE by Regularised regression. MICE run a series of regression models whereby each variable with missing data is modelled conditional upon the other variables in the data [39]. This implies that each variable can be modelled according to its distribution. The missing data can be imputed by MLE using the expectation-maximisation (EM) algorithm [40]. It iteratively solves complete data problems and then intuitively fills the missing data with the best guess under the current estimate of the unknown parameters in E-STEP, then re-estimates the parameters from the observed and filled-in missing data in M-STEP.

The method based on the RF called missForest was presented to impute missing continuous and categorical attributes [41]. It averages the multiple imputed unpruned classification or regression trees and estimates the imputation error by built-in out-of-bag error estimates of RF. A study showed that RF imputation method has less bias estimate and narrower confidence interval compared to MICE [42].

KNN imputes the closest instance in a multi-dimensional space by K-nearest neighbour imputation method. The similarity between two instances is measured by distance function such as Euclidean distance function. KNN imputation can handle instances with multiple missing variables without a need for the creation of a separate predictive model for each variable [43].

However, it suffers from the curse of dimensionality and could be computationally expensive as it searches for similar instances in the entire dataset.

A regularised regression model minimises the loss function by imposing some penalties. The superiority of regularised regression in terms of biases in imputed missing values in high-dimensional data is presented in [44]. In MICE by regularised regression the initial missing data are imputed by a simple method such as mean or frequency. The new parameters are estimated in the next iteration through the regression model and then missing values are replaced by predicted values. These steps are repeated for each variable with missing values. This procedure is conducted iteratively until convergence. After convergence, the final imputed data is utilised as input in a ML model.

2.4 XAI for multi-criteria problems

XAI techniques have shown promise in solving complex problems with multiple criteria. For example, decision trees, with tree-like structure in which internal nodes stand for tests on features and leaf nodes represent a class label [45], have been used as interpretable supervised classifiers in handling multi-criteria problems like medical diagnosis [46]. Vuong et al. [47] applied decision trees in forensic investigation to automatically produce detection rules used by the robotic vehicle in
cybersecurity based on both cyber criteria (network, CPU, disk data) and physical features (speed, vibration, power consumption).

While decision trees can be adopted for visual reasons to highlight the most influential features in a classification process [48], rules have a textual description and are also readily seen in multi-criteria decision aiding [49]. The most common rules are IF-THEN which discretise a high-dimensional, multivariate feature space into a series of simple and explainable decision statements [50]. Karabiyik and Aggarwal [51] proposed an automated disk forensic investigation tool that leverages a dynamic knowledge base created using rules in the form of IF-THEN statements. Belief-rule-base (BRB), an extension of the IF-THEN rule base, has also been used to address multi-criteria problems [52, 53]. The inference of BRB system is explained by using the evidential reasoning (ER) approach [54], which allows the representation of both qualitative and quantitative data by using belief distributions and the aggregation of belief-based information. In addition to interpretable models, model-agnostic XAI techniques such as using an extended Shapley Value [55] and augmentation-based surrogate model [56] have been adopted in the multi-criteria decision aiding models to further assist in explaining the result of these models to decision makers.

XAI techniques have also been used to solve decision problems with multiple objectives. For example, Pessach et al. [57] proposed a comprehensive analytical framework based on the Variable-Order Bayesian Network (VOBN) model to support HR recruiters in global recruitment scheme in balancing multiple organisational objectives. Other XAI techniques/systems developed to solve multi-objective problems include V2f-MORL (vector value function based multi-objective deep reinforcement learning) [58] and fuzzy rule-based systems with multi-objective evolutionary algorithms [59].

Indeed, the goal of XAI techniques is to have the simplest rules which are understandable for humans without sacrificing the performance, although simplicity and performance are often conflicting objectives [60]. To achieve both accuracy and comprehensibility, the two important but conflicting classifier properties, Piltaver et al. [61] proposed multi-objective learning of hybrid classifiers (MOLHC) algorithm in which the sub-trees in the initial classification decision tree are replaced with black-box classifiers so that the complete Pareto set of solutions (a set of solutions that do not dominate each other but are superior to the remaining solutions in the search space) is more likely to be found. Similarly, with objectives of maximising the model ability while minimising the complexity, Evans et al. [60] used multi-objective genetic programming, another tree-based construction method in which trees are evolved from a population of candidates rather than constructed greedily in a top-down manner, to construct model-agnostic representation of black-box estimators.

2.5 XAI in interactive learning

Interactive ML is an iterative process of learning that includes the interaction between humans and ML methods [62]. It has been applied for multiple purposes, such as visual analytics [63], interactive model analysis [64] and event sequence analysis [65]. Jiang et al. [62] reviewed recent research in interactive ML and its application to solve a variety of tasks, discussed research challenges and suggested future work in the area. One of the recommendations for future work is to combine XAI with interactive ML. For example, complex ML algorithms can be simplified by using easy to understand algorithms, which helps the process of model building and parameter tuning.
Previous research combining XAI with interactive learning was done, for example, by Spinner et al. [63]. This research used XAI to explain the output of a ML algorithm, searches for limitations within the models and optimises them. In addition, global monitoring and steering mechanisms were applied. A user study with nine participants was included to test the system, and the results indicated positive feedback from the users. Many other applications of XAI for interactive ML were applied in the form of visual analytics. A modular visual analytics framework was developed for topic modelling, which allows users to compare, evaluate and optimise topic models using a visual analytics dashboard [66]. The design of the framework is interpretable by users and adjusts to their optimisation goal, which is based on time-budget, analysis goal, expertise and the noisiness of the document collection.

A review of visual interaction, supporting dimensionality reduction systems and covering interpretable models, was conducted by Sacha et al. [67]. The paper constructed seven possible scenarios for the application of interactive ML in dimensionality reduction. These scenarios included: interactive feature selection, dimensionality reduction parameter tuning, defining constraints and dimensionality reduction type selection. The paper found that some of the previous studies investigated a combination of these scenarios and the maximum number of combined scenarios in a paper was four. The paper also observed that some of these scenarios were studied more in the literature, such as the feature selection, data selection and parameter tuning scenarios. The application of XAI for interactive learning in forensic science has not been explored yet but it is easy to see that this approach can be beneficial in this domain; for example, where collection of evidence can be controlled (e.g. if is obtained experimentally) but is expensive and/or time-consuming, then a suitable approach may be to use XAI in an interactive fashion with a user, who can decide to terminate evidence collection prematurely upon retrieval of sufficient evidence.

3. Case study

This case study describes the process of forensic investigation by experts from an existing forensic science company. It will explore the challenges faced by forensic experts in making decisions based on factual and heuristic knowledge gained through years of experience. It will discuss the opportunity to utilise the forensic data to develop an interpretable and trustworthy system for automation of the decision-making process [68–77].

3.1 Reporting challenges faced by forensic experts

Currently, a trained expert in this company makes a decision based on a combination of factors, including the analysis of the testing sample and other, external factors such as chemical treatments and more. The expert then produces a report explaining the reasoning behind their decision, outlining different standards and classifying their decisions into one of a plurality of outcomes surrounding likelihood of drug use and exposure to drugs.

The decision regarding likelihood of drug use or exposure is based on a multitude of considerations, including the level of drug detected, the specific metabolites, the client’s self-declarations and many more factors. When the decision process and report writing is conducted by individual experts, there can be some variability in the final decisions and reports that are produced. One of the main reasons for this is the high volume of features to be taken into consideration, which
may all have different levels of importance. Another is that with so many features, it is not possible to cover every potential case that may arise and therefore it is difficult to set specific guidelines for the experts to follow. There is also the potential for subjectivity of the expert when making the final decision—an issue which is difficult to eradicate when relying on human judgement. This can result in disagreement amongst individual experts, or uncertainty where experts may find it difficult to draw conclusions based on the evidence provided. Such differences in subjectivity could be due to personal experience, length of time in the role, previous encounters in different cases and many other potential effects.

When a metabolite is detected the machine generates information on the amount that was present in the sample or, in other words, the level. This is a continuous value which can be used by the experts to make decisions on whether the client was using a particular substance, whether they were exposed or if the client has not been in contact with a drug at all. The levels at which the expert defines use or exposure are up for debate. It can be difficult for them to pinpoint exact values where the judgement tips from likely exposure to likely use, and further problems arise when considering different levels within each category (e.g. highly likely, likely, etc.). Without set levels experts are using their own judgement to decide which category the client falls into, which again leaves room for disagreement across the board.

The most significant problem from a business-efficiency point of view is the length of time that it takes to write a report. A significant increase of new report instructions has resulted in the need for automation, as the current personnel are under high levels of pressure and demand for quick turnaround.

The need for automation is therefore not only to improve accuracy and reliability, but also to speed up delivery times and free up the time of the experts to allow them to undertake other key responsibilities such as research, training and dealing with abnormal cases. The current problem requires a system for automatic decision-making and report writing for the outcome of drug testing, to produce reports suitable for presentation in legal cases.

### 3.2 Forensic data

The features in the forensic data are collected through a combination of questionnaire data—which is completed by the client being tested—and the outcome of tests using forensic laboratory equipment. Each row represents an individual case and each column represents a feature. The forensic investigator collects the essential evidence such as hair and nails, as well as carrying out a structured questionnaire. The questionnaire consists of a number of sections, with a combination of multiple-choice options and Likert scale questions. The document collects information about medical history, drug and alcohol use, hair and nail care.

Hair and nail samples are an easy, non-invasive way of collecting the evidence required to detect the chemical and biological substances, which identify substance use or exposure. Depending on hair growth and the length of strands this can show up to 1 year of drug history, although typically only a maximum of 6 months is used during testing. Body hair is taken if there is less than 1 cm of hair available on the scalp. A nail sample is taken if scalp and body hair are both unavailable and can show up to 3–6 month of drug history. The evidence from hair and nail samples may fail the forensic test (false-negative results) if a suspect repeatedly cuts hair and nails, or uses certain chemical treatments. The forensic data from the questionnaire could gather missing features when some of the follow-up questions do not apply to a client. For example, follow-up questions for pregnancy would only apply only to females. The data could also be subject to inconsistencies due to inaccurate or false
Drug and alcohol intoxication can inhibit memory alone, making it difficult to obtain accurate information on both the quantity of the substance used/exposed to, and the number of days use/exposure, as the client is asked to recall over a period of up to 12 months. The analytical data collected through forensic laboratory tests could also be missing if the metabolites are not present in the client's body, as this would mean further testing is not required. The reason for this is that the testing equipment looks for every possible substance in the sample, rather than selecting those that have been instructed for analysis. The false-positive and false-negative test results affect the data quality. It could be due to external contamination in hair and nail samples, or having little to no body hair.

This type of forensic data can be used to develop decision support tools to fully automate the decision-making process and validation of the experts’ assessments against empirical data. The XAI model supports complex decision-making and can process large amount of data in minutes. The steps for the development of automated decision-making system in the forensic investigation are shown in Figure 3, where the relevant techniques are described in detail in Section 2 of this chapter.

3.3 Decision-making process for testing Drug X

The decision-making process for testing Drug X follows a hierarchical structure with binary outcomes, which has been simplified into a small decision tree shown in Figure 4. The specific metabolites have been anonymised, instead these have been renamed as ‘Metabolite 1’, ‘Metabolite 2’ and ‘Metabolite 3’. It is a snapshot of an interactive-decision-tree that allows visualisation and assessment of the entire decision-making process followed by an expert when drawing conclusions on whether or not a client has used or been exposed to Drug X.

First, based on the questionnaire data the expert will check to see whether the client has declared any use of Drug X in the last 12 months. If this is true then use is confirmed and no further testing is needed. If use has not been declared, based on the analytical data which has been extracted from the hair or nail sample, the expert will consider whether the data shows detection of the Metabolite 1 compound. If Metabolite 1 is detected, further testing is required to determine the levels of Metabolite 1 present in different sections of the hair as this will inform the expert whether the client has used or been exposed to the drug.

If Metabolite 1 is not detected, the expert checks for Metabolite 2. If Metabolite 2 is detected then it is concluded that the client has been exposed, but if it is not

---

2 Drug X has been used to anonymise the name of the specific drug compound being discussed.
detected then the final check is for Metabolite 3. If Metabolite 3 is detected then it is determined that there is no evidence of use or exposure, but if it is detected then the decision is either use or exposure. This is dependent on the levels of each metabolite detected.

4. Conclusion and future work

This chapter has discussed the application of XAI to digital forensics with a particular focus on forensic drug testing. We provided an overview of data-related challenges one may face when implementing an XAI solution including a large number of features (e.g. pieces of evidence), missing data, multiple conflicting decision criteria and the need for interactive learning. Different techniques for dealing with these challenges were reviewed and applications in digital forensics were highlighted. Finally, we outlined a case study on a forensic science company to demonstrate real challenges of forensic reporting and the potential for XAI to design a trustworthy automated system to present generated evidence in the court of law.

The chapter proposes important future directions for adopting XAI techniques to address challenges in digital forensics. These include, first and foremost, the validation of the manually derived decision trees. It would be interesting to derive decision trees automatically using the available data. These trees could differ from the manually derived trees and thus reveal alternative drivers and potential hidden biases. Another direction is the development of more advanced XAI methods.
including belief or fuzzy rule based models. To make these data-driven models more accurate, one can also investigate systematic ways of merging with knowledge base and rules provided by experts. Thus, updating the rules can be done in an interactive fashion, for example as and when new scientific insight from chemistry becomes available. Certainly, these directions of future research are relevant for forensics in drug testing but also for digital forensics in general.

Author details

Louise Kelly*, Swati Sachan†, Lei Ni†, Fatima Almaghrabi†, Richard Allmendinger and Yu-Wang Chen
University of Manchester, Manchester, UK

*Address all correspondence to: louise.kelly@manchester.ac.uk

† These authors are contributed equally.
References


programming approach for feature selection in highly dimensional skewed data. Neurocomputing. 2018;273:554-569


[38] Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-592


[51] Karabiyik U, Aggarwal S. Advanced automated disk investigation toolkit. In:


