https://doi.org/10.2352/EI.2022.34.4.MWSF-324


# Forensic Data Model for Artificial Intelligence based Media Forensics - Illustrated on the Example of DeepFake Detection

***Dennis Siegel*[1] *, Stefan Seidlitz*[1] *, Christian Kraetzer*[1] *, Jana Dittmann*[1]**

[1] ***Otto-von-Guericke University, Magdeburg, Germany***

## Abstract

*The recent development of AI systems and their frequent use for classification problems poses a challenge from a forensic perspective. In many application fields like DeepFake detection, black box approaches such as neural networks are commonly used. As a result, the underlying classification models usually lack explainability and interpretability.*
*In order to increase traceability of AI decisions and move a crucial step further towards precise & reproducible analysis descriptions and certifiable investigation procedures, in this paper a domain adapted forensic data model is introduced for media forensic investigations focusing on media forensic object manipulation detection, such as DeepFake detection.*

## Introduction

IT-forensics is a domain that, due to its novelty and the fast changes experienced in the threat landscape that has to be considered, still sees a lot of research activity. Many of the corresponding research initiatives unfortunately remain on a purely academic level, lacking the degree of maturity required for field application of analysis methods.
In this context the existence of standardized process models plays an important role on the path to mature solutions, because to achieve the ultimate benchmark for a forensic method (which would be its admissibility in court proceedings), it would require a standardization and certification of the tool(s) and procedures as well as training and certification of the practitioners / forensic experts. While much work exists on forensic process models (including crucial components such as data models) for older sub-disciplines of IT forensics, for the younger sub-discipline of media forensics domain adapted solutions are still amiss.

As main contribution of this paper, a domain adapted forensic data model is introduced for media forensic investigations focusing on media forensic object manipulation detection. The new data model is derived by domain transfer from established best practices. Furthermore, its applicability is demonstrated by using the new model to completely rework an analysis pipeline description form an earlier paper on DeepFake detection.
These results are considered important to move a crucial step further towards precise & reproducible analysis descriptions and certifiable investigation procedures. In addition they constitute an important step towards explainable artificial intelligence (XAI), fair AI and human oversight concepts who are major aspects of the upcoming EU Artificial Intelligence Act (AIA).

The paper is structured as follows: In section a short summary on the state of the art on forensic process models and corresponding data models is presented. In section a new domain adapted data model is derived from the existing state-of-the-art, which is then used in section to rework an existing investigation pipeline description for DeepFake detection to improve this description. At the end of the paper, section presents a short summary and presents starting points for potential future work.

## State-of-the-art on Forensic Process Models

Since the legislative and administrative process governing the usage of evidence in court (including expert testimony) is different for every country, it always has to be reflected in the light of the national regulations. In the German situation (which is relevant for the authors of this paper) one of the most important guidelines for IT forensics (and sub-disciplines) is the "Leitfaden IT-Forensik" [2] of the German Federal Office for Information Security (BSI; the national cyber security authority). It provides various means for modeling forensic processes, including the definition of a phase-driven investigation & reporting model, a basic data model and a classification of methods and tools. Since its last official update in 2011, it has been reflected upon and extended in many publications, such as [6] and [1].
What is currently amiss in this line of research is a domain specific adaptation to media forensics. This became apparent to the authors when analysis work performed in a previous publication (here: [12], where an analysis of video data with the aim of DeepFake detection is performed using three individual detection operators and alternative fusion operators) turned out to be hard (if not entirely impracticable) to project onto the pre-existing data models.

The following section elaborates more on this research gap while section briefly summarizes with the Data-Centric Examination Approach for Incident Response- and Forensics Process Modeling (DCEA) the latest extension to the BSI guidelines from [2], which is used here as starting point for the extension work.
The work in the following chapters is than focused primarily on extending the data model and secondarily on the impact to aspects of the investigation & reporting mode.

### *Media forensic processes*

Textbooks on media forensics such as [5] as well as relevant research work like [9] agree upon the fact that at the core of modern media forensics pipelines looking into questions of integrity one or more pattern recognition or anomaly detection mechanisms are to be found. After data collection and pre-processing operations either sequences or parallel networks of such operators (in

the latter case followed by fusion operators) are used to implement a set of analysis tasks. The output of the analyses will then have to be interpreted by an human expert, e.g., in form of an expert testimony in court.
While agreement exists in the community on the fundamental outline of analysis pipelines, the existing state-of-the-art lacks domain specific data models. Those are required to: a) facilitate efficient requirement engineering, design specification, implementation, certification and deployment of media forensic analysis pipelines, b) enable error, loss and uncertainty estimations in individual forensic analyses performed (see [6]) and c) ease processes aiming at the explainability and fairness in forensic investigations (novel factors that have to receive increased attention due to the current changes in legislation governing the application of AI, such as the upcoming EU Artificial Intelligence Act).
Due to the lack of such domain specific data models, this paper focuses on proposing such a model, suitable to the task at hand. This is done by performing a domain transfer on an established data model for digitized forensics (see section ).

### A Data-Centric Examination Approach for Incident Response- and Forensics Process Modeling

Forensic process models are an important cornerstone in the science and more importantly the practice of forensics. They guide investigations and make them comparable, reproducible as well as certifiable. Usually, the adherence to strict guidelines (i.e. process models) are regulated within any legal system (e.g. in the US by the fourth of the Daubert criteria ("*the existence and maintenance of standards and controls*" [3])). For mature forensic sciences, like for example dactyloscopy, internationally accepted standards (like the ACE-V process model for dactyloscopy) have been established over the last decades.
Due to the fact that IT forensics is a rather young discipline in this field (with media forensics being an even younger sub-discipline) it is hardly astonishing that here the forensic process models have not yet achieved the same degree of maturity as in other fields. Nevertheless, they would still be important to achieve universal court acceptability of methods. One well established forensic process model for IT forensics is the one proposed by the German Federal Office for Information Security (BSI). When it was originally published in 2011, its sole focus was on computer and network forensics but since then it has evolved to suite also to some extend the needs of other sub-disciplines such as digitized forensics. The latest major revision of this process model, which is used within this paper, can be found in [6] and is called the Data-Centric Examination Approach (DCEA). The core of DCEA consists of three main aspects: a model of the *phases* of a phase driven forensic process, a classification scheme for *forensic method classes* and *forensically relevant data types*.
The six DCEA *phases* are briefly summarized as: Strategic preparation (SP), Operational preparation (OP), Data gathering (DG), Data investigation (DI), Data analysis (DA) and Documentation (DO). While the first two (SP and OP) contain generic (SP) and case-specific (OP) preparation steps, the three phases represent the core of any forensic investigation. The phase DO is split in [6] into two aspects: case accompanying documentation (Chain-of-Custody, etc) as well as final documentation (e.g. the expert opinion statement presented in court). For details on the phase model the reader is referred, e.g. to [6] or [1].

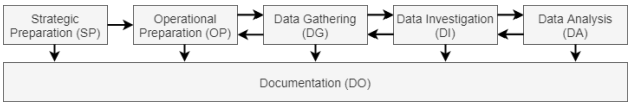


***Figure 1.*** *Phase model (based on [2])*

The second core aspect of DCEA is the definition of *forensic method classes* as presented in [6]. They consist of methods of: the Operating system (OS), the File system(s) (FS), IT applications (ITA), Explicit means of intrusion detection (EMID), Scaling of methods for evidence gathering (SMG) and Data processing and evaluation (DPE). Like the phases, this aspect is of limited relevance for this paper. For details on this classification scheme for investigation methods the reader is referred to [6].
The third (and in the context of this paper most relevant) aspect is the specification of *forensically relevant data types*. More recent publications, such as [1], have shown that this scheme needs to be extended accordingly if new investigation domains are considered.
The original set of data types, which was designed with digital IT forensics in mind, needs to be adapted towards every investigation domain. In [7] and [6] such an adaptation for the field of digitized forensics has been discussed for the field of dactyloscopy (forensic fingerprint analysis and comparison). This adaptation is summarized in Table 1. Because it is much closer to the requirements faced within this paper than the original data model, it is used as starting point for the modeling work performed here.

## Deriving a Forensic Data Model for Artificial Intelligence based Media Forensic Investigations focusing on Integrity

Performing abstract data modeling without precise knowledge about the context, in which the data type is supposed to be used, is a futile task. Therefore, first a generalized media forensic analysis process is briefly discussed in section . This is followed in section by an identification of the typical data streams within such a process. As the last step in the data modeling, the data streams are further differentiated into data types in section .

### Modeling a generalized media forensic analysis process

In general, each processing operation (or operator) is considered here as an atomar processing black box component with an identifier and (usually) a description of the processing performed in this operation. Each component has four well defined connectors: *input*, *output*, *parameters* and *log data*. To pay respects to the particularities of this field and make the following modeling task easier, a fifth connector is defined within this paper for a specific type of operator which requires a knowledge representation or a model for its processing operation. In that case, this fifth connector is labeled *model*. Depending on the nature of the operator this could be a rule set, signature set, statistical model, neural model, or any other form of knowledge representation.
Figure 2 shows the modeling for a small, exemplary selected processing sub-routine within a bigger media forensic investigation process (here the sub-routine of face segmentation as necessary step in DeepFake detection for videos). The first operator in this three step processing sub-routine is loading the video from its *in-*

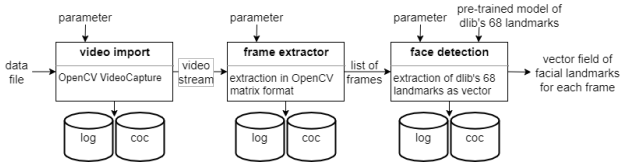


***Figure 2.*** *Exemplary modeling of the process for face detection.*

*put*. The *parameters* need to be chosen based on the video format and the *output* is stored as video stream. This video stream is then in the next operator split into single frames as necessary preprocessing for an image based face detection and segmentation algorithm. For the face detection and segmentation, a pre-trained model with 68 landmarks (here from [8]) is loaded at the third operators *model* connector. This is the only step in this small example where model data is used.
Each step provides corresponding process documentation in the form of logs and chain of custody (CoC) data at its *log data* connector.

### *Identifying typical data streams*

Based on the atomar operator description above and generalizing media forensic (i.e., passive) investigations focusing on analyzing the integrity of media objects, here five typical data streams are identified: The *process description* is proposed as a sourceable or instantiable template, which is generated before starting the investigation. It is supposed to be generated in the phase of Strategic preparation (SP) and contains general information (such as process layouts/graphs, interfaces and operators involved) independent from a specific investigation. Besides the actual process layout this stream inherits also information from DD7, DD9 and DD10 of the data types form digitized forensics (see table 1).
The second data stream *media data* contains all forms of media such as images, videos, audio and/or network streams used and created within the investigation process. Media data could be found both on input and output connectors of a component and would in case of an investigation in digitized forensics contain information from DD1, DD2 and DD8.
The non-media output of the individual examination steps is combined into the data stream *forensic process/pipeline internal data and reporting*. It contains actual (intermediate) investigation results and CoC data such as hashes and logs as well as error, loss and uncertainty indicators, meta data and traceability/explainability information (such as a risk and circumstantial evidence map (RCEM)). This output is gathered in the phases OP, DG, DI and DA and would in case of an digitized forensics investigation be described by DD2, DD3, DD8, DD9 and DD10.
Another important aspect is the combination of all settings used in the investigation, including all parameters and models used. This combination is defined as *process control data* and contains in digitized forensics DD3, DD4, DD7 and DD8.
The last data stream is *contextual data*, which contains all information regarding the context of a specific investigation. In general it contains information such as operator IDs, data source descriptors (e.g., camera types) and the results of a content analysis of the media objects required for plausibility and fairness evaluation. In case of an digitized forensics investigation contextual data would be found in DD3, DD8, DD9 and DD10.

This subdivision of the data associated with an investigation is a functional classification paying respect on one hand to the characteristics of data objects involved and on the other hand to operational and security requirements. The media data stream of an investigation might easily contain terabytes of video data which would require a access to a private cloud for efficient handling, while the reporting data would assumed be much smaller in data size but be more frequent and have other constraints like reliable time-stamping. From the operational and security perspective also different protection levels (and as a consequence security mechanisms) would be required depending on the nature of the objects in a stream and the risks associated.

### *Deriving the domain specific data model*

Taking the data streams identified above for media forensics into account, it is necessary to adapt the existing data models. As starting point, here the data types from digitized forensics are chosen because they require a less wide-ranging re-modeling. The objective of deriving a domain specific data model for integrity

| **Forensic data type** | **Description** (according to [6]) |
|---|---|
| DD1 Raw sensor data | Digital input data from the digitalization process (e.g. scans of test samples) |
| DD2 Processed signal data | Results of transformations to raw sensor data (e.g. visibility enhanced fingerprint pattern) |
| DD3 Contextual data | Contain environmental data (e.g. spatial information, spatial relation between traces, temperature, humidity) |
| DD4 Parameter data | Contain settings and other parameter used for acquisition, investigation and analysis |
| DD5 Trace characteristic feature data | Describe trace specific investigation results (e.g. level1/2/3 fingerprint features) |
| DD6 Substrate characteristic feature data | Describe trace carrier specific investigation results (e.g. surface type, individual surface characteristics) |
| DD7 Model data | Describe trained model data (e.g. surface specific scanner settings, reference data) |
| DD8 Classification result data | Describes classification results gained by applying machine learning and comparable approaches |
| DD9 Chain of custody data | Describe data used to ensure integrity and authenticity and process accompanying documentation (e.g. cryptographic hash sums, certificates, device identification, time stamps) |
| DD10 Report data | Describe data for the process accompanying documentation and for the final report |

**Forensic data types defined in [6] for an exemplary selected process in digitized forensics (here digital dactyloscopy) (updated from [7])**

focused media forensics is a specification and overlap-free representation of data types. As a result of the modeling performed here, eight media forensic data types (MFDT, see table 2) are defined, which are loosely derived from the ten data types of digitized forensics. *Digital input data* (MFDT1) is a re-definition based on DD1 and considers now any kind of media data as it is initially taken as input to the investigation. *Processed media data* (MFDT2) is derived from DD2 and contains all operator output which are media data. *Contextual data* (MFDT3) is derived from DD3 and includes case specific information regarding the investigation process and -objects. Contextual data can also be used to control targeted parametrization and thus allow case or objects specific parameter optimization. They also allow for plausibility and fairness evaluations as part of the assessment of an investigation performed. *Parameter data* (MFDT4) is similar to DD4 from digitized forensics and contains all configurations and parametrizations for operators in an investigation (except for model data, see MFDT6 below), including those who are used for training of classifiers and models before the actual investigation. *Examination data* (MFDT5) combines and extends the data types DD5, DD6 and DD8 from digitized forensics. It comprises all occurring non-media outputs (e.g., trace information, patterns and anomalies identified) of the investigation. *Model data* (MFDT6) corresponds to DD7 from digitized forensics. It includes trained models of machine learning algorithms like rule based approaches or decision trees as well as models of neural networks (incl. their network architecture). *Log data* (MFDT7) is an component of the documentation which is here newly added to the data model and is used for administration and maintenance (including Syslogs and information about the memory usage). Data in MFDT7 are not relevant for the specific case in the investigation, but are necessary for the administration of the system (e.g., to notice that the memory allocated for the task is not sufficient). *Chain of custody & report data* (MFDT8) is a combination of DD9 and DD10 from digitized forensics. They characterize the case relevant documentation for integrity and authenticity assurance as well as the accompanying documentation for the final report. For admissibility in court the final report would be required following the corresponding chain of custody guidelines.

Chain of custody & report data (MFDT8) also have to address the description of the deployed (process) modeling with regard to origin and provenance of decision (AI) models used. Especially in the context of neural networks a detailed specification of the network structure(s) (MFDT4, MFDT6) as well as the used parameters for training, (potential transfer-learning), testing and validation phases (MFDT4) would be required to allow for the necessary reproducibility of setups and corresponding error, loss and uncertainty as well as explainability considerations for explainable AI. But not only classifier designs and parameterizations have to be reported upon: Another aspect for the documentation refers to the data used in the process(es) of model generation, focusing on the training and validation sets taken from the content of data types MFDT1 and MFDT2. The decisive factors in this respect are origin, diversity and quantity of data (summarized within MFDT3). It is also significant for the documentation to characterize the differences between training and test/evaluation/validation phases of each mechanism. For example the consideration of disjoint data sets for training and testing yields a more generalizable and trustworthy result than a cross-validation would obtain. Furthermore, the documentation of initial control parameters (MFDT4: e.g., learning rate, optimizer, loss function) as well as information about the training process (MFDT7 & MFDT8: training duration, used hardware, etc.) are very important for traceability as well as interpretability.

Also important is the run-time of the detection process, which needs to be evaluated and documented in relation to the hardware used. Another documentation criteria refers to the type of result data (MFDT2 or MFDT5) calculated by methods such as neural network. In decision-based classification, the result is often represented by a classification/prediction label (MFDT5) and/or confidence estimate (MFDT5). In some cases it can also be an image or other media object (MFDT2) that represents relevant information such as a map of anomalies found, to be interpreted by a human investigator.

In field application, because of the typical black box usage of mostly Neural Networks, with an unknown internal behaviour in the hidden layers between in- and output, it might be possible that there exist no process data or feature vectors/data (MFDT5). But for a mature forensic method aiming for court admissibility such kind of black box behavior would not sufficient, because result data of forensic operators must be comprehensible. Because of that, methods focusing on explainability (e.g., LIME [11] or

| Data type | Derived from DD | Description |
|---|---|---|
| MFDT1 Digital input data | DD1 | The initial media data considered for the investigation. |
| MFDT2 Processed media data | DD2 | Results of transformations to media data (e.g. grayscale conversion, cropping) |
| MFDT3 Contextual data | DD3 | Case specific information (e.g. for fairness evaluation) |
| MFDT4 Parameter data | DD4 | Contain settings and other parameter used for acquisition, investigation and analysis |
| MFDT5 Examination data | DD5, DD6, DD8 | Including the traces, patterns, anomalies, etc that lead to an examination result |
| MFDT6 Model data | DD7 | Describe trained model data (e.g. face detection and model classification data) |
| MFDT7 Log data | newly defined | Data, which is relevant for the administration of the system (e.g. system logs) |
| MFDT8 Chain of custody & report data | DD9, DD10 | Describe data used to ensure integrity and authenticity (e.g. hashes and time stamps) as well as the accompanying documentation for the final report. |

**Media Forensic Data Types (MFDT) proposed in this work**

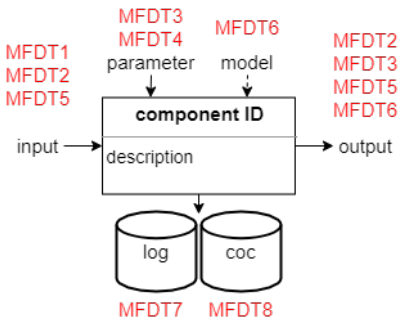


***Figure 3.*** *Template structure for a single component*

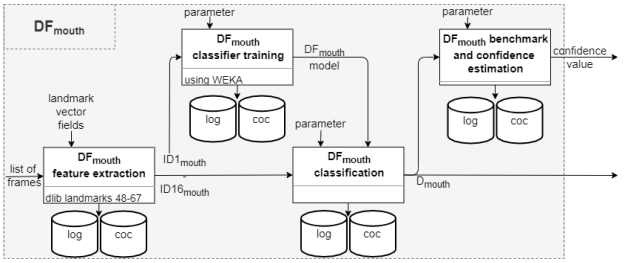


***Figure 4.*** *Illustration of the DeepFake detection based on mouth region* ***modeled as a template*** *in the proposed context model in the phase of* ***Strategic preparation (SP)***

LRP[10]) have to be included in the investigation. Moreover, the network structure could be expanded between hidden layers with more output layers to allow obtaining processed data (MFDT2) or feature vectors (MFDT5). As a necessary result, a neural network would become more transparent, interpretable and explainable.

Figure 3 shows the link between media forensic data types (MFDT) for the operator description presented above. As discussed in section , depending on whether a model is used in an operator or not, each component has four or five well defined connectors. The operator (i.e., process step itself; here shown as a box) has an unique identifier and a description of the process. This description should increase traceability as well as explainablity. The input of a component has a form of media data, the court exhibits itself (MFDT1) or after previously done preprocessing steps (MFDT2) or examination data (MFDT5). Depending of the processing step, the generated output could be media data (MFDT2), a derived information on the investigation context (MFDT3) or investigation results (MFDT5). It is also possible during the phase of Strategic preparation (SP) that a model is trained (MFDT6). The process control is done by parameters (MFDT4). Furthermore, the gathered contextual data (MFDT3) can be used for optimization of the parameters in the specific investigation. MFDT3 could for example be information about the recording device, resolution or lighting conditions, which might be useful to estimate decision uncertainty and thereby allowing to estimate the fairness of an investigation. The loading of a model (MFDT6) is limited to model-driven operators, which why it is shown by a dashed line. Process accompanying documentation will be divided and separately saved in log data (MFDT7) and chain of custody data (MFDT8) based on the modeled data types.

## Illustration of the practicability of applying the proposed new data model

As indicated in section , one motivation for this paper were apparent problems when projecting an exemplary selected media forensics processing pipeline designed for DeepFake from a previous paper (here [12]) of the authors onto existing data models. In this section it is shown, how the adapted data model from chapter  can be successfully used for the pipeline in that publication.
The modeling work is done in two separate steps. The first instantiation is focusing on training models for the operators in Strategical preparation (SP). This initialization is done using well established DeepFake reference data sets for the training. First, the original videos and corresponding DeepFakes are imported and pre-processed in a suitable format so that they can be further processed as a video stream. The video stream is then divided into individual frames (single images). The resulting list of images is used for both face detection and subsequent DeepFake detection (see figure 2). Assuming one face per frame, a pre-trained 68 landmark model is used for face detection. It locates the position of each of those facial landmarks and stores them in a vector field. The detection algorithm itself consists of the components feature extraction, classifier training, classification and benchmarking. In the feature extraction each frame is evaluated based on the corresponding landmarks relevant to the region it focuses on and generates a feature vector relevant to the classification. Exemplary for the detector $DF_{mouth}$, the classifier $DF_{mouth}$ model is created using the J48 classifier from Weka [4], testing different models and parameter settings. The optimal model then gets integrated into the classification, which then returns the decision (e.g. $D_{mouth}$). Afterwards a second instance of validation is done by benchmarking and confidence estimation. Based on the confidences the weights for a consecutive fusion step are determined. The same procedure is done for the algorithms $DF_{eye}$ and $DF_{foreground}$ with differences in the considered landmarks and generated features (for details see [12]). During the whole process each step gets documented and stored in the log and chain of custody databases respectively.
The second instantiation of the modeling corresponds to the actual investigation determining whether a DeepFake manipulation occurred in the presented videos. Considering the pipeline presented in figure 1, it covers all phases from OP to Documentation. The first processing steps are identical to those performed in the SP instantiation. This is to be expected, because both training and testing of an operator should be done under the same conditions (i.e., after identical pre-processing). Changes can be found in the application of the detection operators. Here the parts regarding model training are left out because the models pre-trained in SP are loaded instead, together with the used classifier parameters. Thus initialized the operators are applied to video material to determine traces of DeepFake manipulations. The respective individual decisions $D_{eye}$, $D_{mouth}$ and $D_{foreground}$ are then merged into the fusion module to determine a final decision $D_{fusion}$. The required fusion weights used for this purpose also come from the SP. A complete mapping of this process, including a labeling of the Media Forensic Data Types communicated at each connector, can be found in Figure 5.

## Conclusion and Future Work

In this paper a domain adapted forensic data model is introduced for media forensic investigations focusing on media

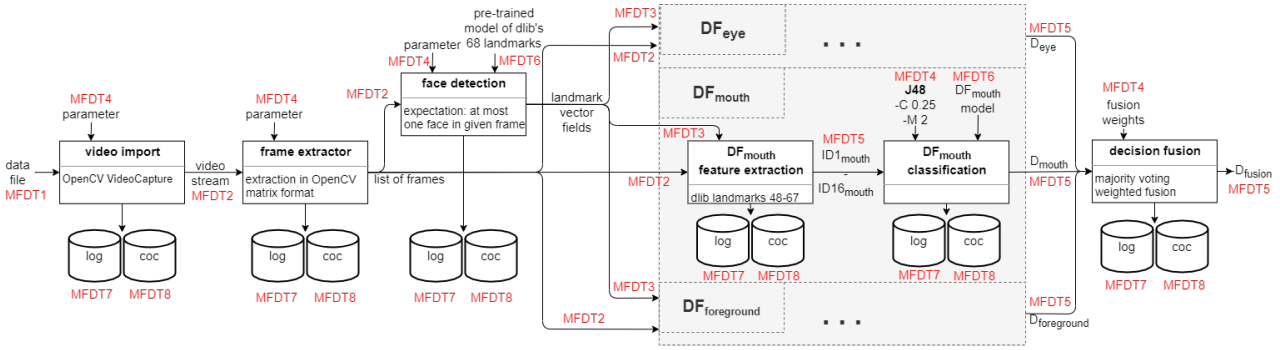


***Figure 5.*** *Illustration of the DeepFake detection pipeline* ***instantiated*** *in the forensic process model phase of* ***Operational preparation (OP)****, with the inclusion of occurring data types*

forensic object manipulation detection. The new data model is derived by domain transfer from established best practices. Furthermore, its applicability is demonstrated by using the new model to completely rework an analysis pipeline description form an earlier paper.

The work performed here motivates future work on the following aspects: First, on extending the considerations on templating and instantiation works in Strategic preparation (SP) and Operational preparation (OP) phases to move a further step towards precise and reproducible analysis descriptions and thereby towards certifiable investigation procedures.
Second, on expanding the modeling with regard to knowledge data generation and representation to be better able to include also more complex operations (e.g. modern training scenarios for neural network based detectors) as well as context dependent pipeline alternatives into forensic workflows.
Third, on extending the work on error, loss and uncertainty (on basis of [6]) as well as explainability and fairness in AI-driven forensics.

## Acknowledgements

The work in this paper is funded in part by the German Federal Ministry of Education and Research (BMBF) under grant number FKZ: 13N15736 (project “Fake-ID”). The authors wish to thank Dr. Stefan Kiltz for the discussions on the DCEA and its applicability as well as Dr. Robert Altschaffel and Dr. Mario Hildebrandt for discussions on domain adaptations required for forensic data models.
Author Contributions: Initial idea & methodology: Jana Dittmann (JD) and Christian Kraetzer (CK); Conceptualization: Dennis Siegel (DS), Stefan Seidlitz (StS), CK and JD; Modelling of the new data structure: CK, DS, StS; Modelling of the templating approach for media forensics in SP/OP: CK, DS, StS; Writing – original draft: DS; Writing – review & editing: CK, StS and JD.

## References

[1] Robert Altschaffel. *Computer forensics in cyber-physical systems : applying existing forensic knowledge and procedures from classical IT to automation and automotive*. PhD thesis, Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, 2020.

[2] BSI. *Leitfaden IT-Forensik*. German Federal Office for Information Security, 2011.

[3] Christophe Champod and Joëlle Vuille. Scientific evidence in europe - admissibility, evaluation and equality of arms. *International Commentary on Evidence*, 9(1), 2011.

[4] Mark A. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. *SIGKDD Explor.*, 11(1):10–18, 2009.

[5] Anthony T. S. Ho. *Handbook of digital forensics of multimedia data and devices / edited by Anthony T.S. Ho and Shujun Li, Department of Computing and Surrey Centre for Cyber Security (SCCS), University of Surrey, UK*. Wiley/IEEE Press, Hoboken, 2015.

[6] Stefan Kiltz. *Data-Centric Examination Approach (DCEA) for a qualitative determination of error, loss and uncertainty in digital and digitised forensics*. PhD thesis, Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, 2020.

[7] Stefan Kiltz, J. Dittmann, and C. Vielhauer. Supporting forensic design - a course profile to teach forensics. *2015 Ninth International Conference on IT Security Incident Management and IT Forensics*, pages 85–95, 2015.

[8] Davis E. King. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.*, 10:1755–1758, 2009.

[9] Christian Krätzer. *Statistical pattern recognition for audio-forensics*. PhD thesis, University of Magdeburg, 2013.

[10] Sebastian Lapuschkin, Alexander Binder, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek. The lrp toolbox for artificial neural networks. *Journal of Machine Learning Research*, 17(114):1–5, 2016.

[11] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ”why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144, 2016.

[12] Dennis Siegel, Christian Kraetzer, Stefan Seidlitz, and Jana Dittmann. Media forensics considerations on deepfake detection with hand-crafted features. *Journal of Imaging*, 7(7), 2021.

## Author Biography

***Jana Dittmann*** *is a Professor on multimedia and security at the University of Otto-von-Guericke University Magdeburg (OvGU). She is the leader of the Advanced Multimedia and Security Lab (AMSL) at OvGU, which is partner in national and international research projects and has a wide variety of well recognized publications in IT security.* ***Christian Kraetzer*** *is a post-doc researcher and* ***Dennis Siegel*** *as well as* ***Stefan Seidlitz*** *are PhD students at AMSL.*