Forensic Science International 276 (2017) 142–153

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation

Didier Meuwly [a,b,*], Daniel Ramos [c], Rudolf Haraksim [d]

[a] Netherlands Forensic Institute, Laan van Ypenburg 6, 2497GB The Hague, The Netherlands
[b] University of Twente, Drienerlolaan 5, 7522NB Enschede, The Netherlands
[c] ATVS – Biometric Recognition Group, Escuela Politecnica Superior, Universidad Autonoma de Madrid, C/Francisco Tomas y Valiente 11, 28049 Madrid, Spain
[d] LTS5 – Signal Processing Laboratory, École Polytechnique Fédérale de Lausanne, Faculty of Electrical Engineering, Station 11, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form 23 March 2016
Accepted 24 March 2016
Available online 26 April 2016

Keywords:
Method validation
Automatic interpretation method
Strength of evidence
Accreditation
Validation report

ABSTRACT

This Guideline proposes a protocol for the validation of forensic evaluation methods at the source level, using the Likelihood Ratio framework as defined within the Bayes' inference model. In the context of the inference of identity of source, the Likelihood Ratio is used to evaluate the strength of the evidence for a trace specimen, e.g. a fingermark, and a reference specimen, e.g. a fingerprint, to originate from common or different sources.

Some theoretical aspects of probabilities necessary for this Guideline were discussed prior to its elaboration, which started after a workshop of forensic researchers and practitioners involved in this topic. In the workshop, the following questions were addressed: *"which aspects of a forensic evaluation scenario need to be validated?"*, *"what is the role of the LR as part of a decision process?"* and *"how to deal with uncertainty in the LR calculation?"*. The questions: *"what to validate?"* focuses on the validation methods and criteria and *"how to validate?"* deals with the implementation of the validation protocol.

Answers to these questions were deemed necessary with several objectives. First, concepts typical for validation standards [1], such as *performance characteristics*, *performance metrics* and *validation criteria*, will be adapted or applied by analogy to the LR framework. Second, a validation strategy will be defined. Third, validation methods will be described. Finally, a validation protocol and an example of validation report will be proposed, which can be applied to the forensic fields developing and validating LR methods for the evaluation of the strength of evidence at source level under the following propositions:

H_1/Hss: The trace and reference originate from the same source.
H_2/Hds: The trace and reference originate from different sources.

1. Introduction

1.1. Preliminary considerations

This Guideline aims at providing assistance to the forensic practitioners in determining the scope of validity[1] and applicability of the LR methods developed and to validate the LR's produced as forensic evidence in practice. Even though the empirical examples given (taken over from forensic fingerprints) are shaped around the LRs computed from scores of a biometric system (namely *score-based LRs*), the Guideline proposed is general and can be applied to any forensic method producing LR values, whether it is biometric or not, and whether it is score-based or feature-based.

It is worth noting, that there is an on-going discussion in the forensic community regarding issues related to the concepts of probability and of the Likelihood ratio (LR). Especially concerning is the concept of uncertainty of computed LRs, which leads to different methods for the measurement of performance of LRs methods, which may not necessarily be compatible. This has direct consequences on the definition of the criteria for the validation of computer-assisted LR methods developed for forensic evaluation. Therefore, the points of view regarding the concepts of probability and of the LR will be discussed prior to the introduction of the performance characteristics and criteria.

* Corresponding author at: Netherlands Forensic Institute, Laan van Ypenburg 6, 2497GB The Hague, The Netherlands. Tel.: +31 708886344.
E-mail addresses: d.meuwly@nfi.minvenj.nl, d.meuwly@utwente.nl (D. Meuwly).

[1] The scope of validity is to be understood as the range of conditions for which the method has been tested.

http://dx.doi.org/10.1016/j.forsciint.2016.03.048
0379-0738/

1.2. Definitions

In the context of the interpretation of the evidence by LR values we understand validation as the process followed in order to determine the scope of validity of a method used to compute LR values. The latter means that we allow the method to be used in forensic casework in the future.

Here, we define important concepts that are typical in validation strategies in other contexts. Later our definitions adapt to the LR methodology.

- A *performance characteristic* is a characteristic of a LR method that is thought to have an influence in the validation of a given method. For instance, LR values should be discriminating in order to be valid, provide clear distinction between comparisons under different hypotheses. In this case, discriminating power is a performance characteristic.
- A *performance metric* is a variable whose numerical or categorical value measures a performance characteristic. For instance, the minimum log-likelihood ratio cost ($minC_{llr}$) can be interpreted as a measure of discriminating power, and therefore it can be used as a performance metric of the discriminating power.
- A *validation criterion* presents a condition related to the performance characteristics that has to be met as a necessary condition for the LR method to be deemed as valid. For instance, a validation criterion can be formulated as follows: *only methods producing rates of misleading evidence smaller than 1% can be considered as valid*. Note that a single validation criterion is not sufficient in general, and therefore several validation criteria might be necessary in order to determine the validity of the method.

2. Computation of likelihood ratios for forensic evaluation

Many different methods have been described in the literature to compute LR values [2–7], feature-based [3,4] and score-based [5,7–11]. This Guideline considers both classes of LR methods, score and feature-based, and an example of comparison of these methods can be found in [5].

In a score-based method illustrated in Fig. 1, the LR values are calculated from the comparison scores [7,10], which are typically the result of a comparison performed by pattern-recognition algorithms. These extract and compare the features of trace (T) and reference (R) specimens. The score (E) resulting from this comparison is used to compute a likelihood ratio with the LR method (Bayes' inference model), using a dataset of trace specimens (DB Traces) and a dataset of reference specimens (DB References). Score-based approaches are traditionally used in forensic biometrics and a typical example can be found in [2].

Feature-based LR methods illustrated in Fig. 2 exploit directly the features of the specimens in comparison and produce a LR value without the previous computation of a comparison score. Several examples of feature-based LR methods are described in [3–5]. These methods involve statistical modeling at the level of the features, using for example probability density functions for either of the propositions to produce the LR values. Feature-based approaches are traditionally used in forensic chemistry and examples can be found in [5,12,13].

2.1. The LR as part of the forensic evaluation process

Forensic research makes progress in the field of evaluation of forensic evidence. Currently, a uniform and logical inference model is used for evaluating and reporting forensic evidence [14]. It uses a likelihood ratio (LR) approach based on the Bayes inference model (Theorem of conditional probabilities). Standards and Guidelines have been proposed for the formulation of evaluative forensic science expert opinion first in UK by the Association of Forensic Science Providers (AFSP) [15] and then in Europe, within the European Network of Forensic Science Institutes (ENFSI) [16].

The LR methods are extensively used, for example, for the interpretation of DNA profiles. Some recommendations on the interpretation of the DNA mixtures have been issued in 2006 [17]:

R1: "LR is the preferred approach to (DNA) mixture interpretation".
R2: "Even if the legal system does not implicitly appear to support the use of the likelihood ratio, it is recommended that the scientist is trained in the methodology and routinely uses it in case notes".

Even though this Guideline does not use examples from the DNA, we endorse and follow these recommendations, because the logic of the inference model remains, independently of the type of traces considered [18].

Computer-assisted methods have been developed to compute LRs, assisting forensic practitioners in their role of forensic evaluators to perform inferences at source level [19]. Very early principles for using the LR approach in forensic evaluation can be found in the analysis of glass microtraces [3]. It has also been used in forensic fields focusing on human individualization, such as fingermark [20,21], earmark [22], speaker recognition [7,23] and hair [24]; or object individualization such as toolmarks [25], fibre [26] and glass microtraces [3,6,12,13,27] (which represents a very early practical example of the use of the LR approach). But the LR approach has been firstly implemented in a casework process as a standard for the evaluation of DNA profiles [14] and several computer-assisted methods are being developed and validated to assess the value of DNA mixture profiles [28–32].

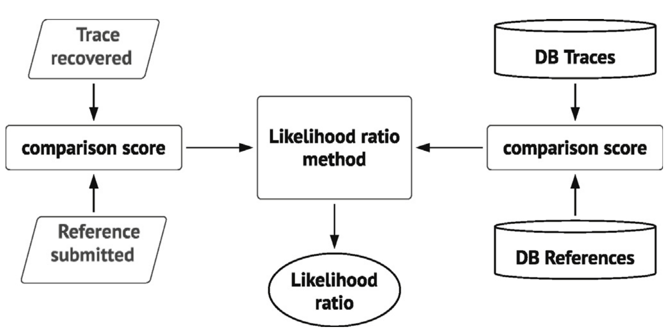

Fig. 1. Score-based LR computation.

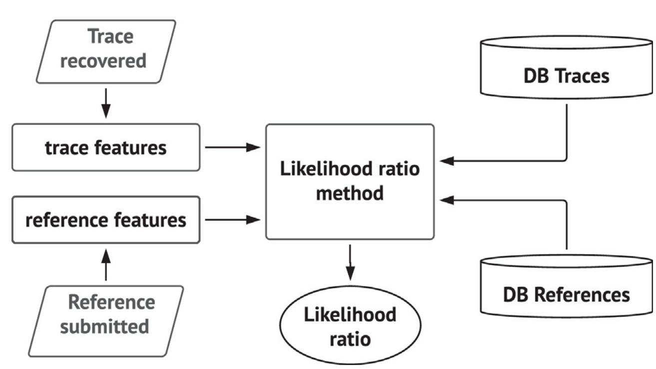

Fig. 2. Feature-based LR computation.