



Explainable offline automatic signature verifier to support forensic handwriting examiners

Moises Diaz¹ · Miguel A. Ferrer¹ · Gennaro Vessio²

Received: 26 May 2023 / Accepted: 20 October 2023 / Published online: 20 November 2023
© The Author(s) 2023

Abstract

Signature verification is a critical task in many applications, including forensic science, legal judgments, and financial markets. However, current signature verification systems are often difficult to explain, which can limit their acceptance in these applications. In this paper, we propose a novel explainable offline automatic signature verifier (ASV) to support forensic handwriting examiners. Our ASV is based on a universal background model (UBM) constructed from offline signature images. It allows us to assign a questioned signature to the UBM and to a reference set of known signatures using simple distance measures. This makes it possible to explain the verifier's decision in a way that is understandable to non-experts. We evaluated our ASV on publicly available databases and found that it achieves competitive performance with state-of-the-art ASVs, even when challenging 1 versus 1 comparisons are considered. Our results demonstrate that it is possible to develop an explainable ASV that is also competitive in terms of performance. We believe that our ASV has the potential to improve the acceptance of signature verification in critical applications such as forensic science and legal judgments.

Keywords Handwriting · Signature verification · Explainability · Transparency · Forensic · Biometrics · Universal background model · Synthesis

Abbreviations

ASV	Automatic signature verification
FHEs	Forensic handwriting examiners
UBM	Universal background model
LR_q	Likelihood ratio
$P(U)$	Probability of membership in the UBM
$P(R)$	Probability of membership in the reference set
DL	Deep learning
FCN	Fully convolutional network
DET	Detection error trade-off

FAR	False acceptance rate
FRR	False rejection rate
RF	Random forgery
SF	Skilled forgery
EER	Equal error rate

1 Introduction

Forensic Handwriting Examiners (FHEs) are tasked with comparing a questioned signature to one or more genuine specimens. This practice is based on the assumption that it is highly unlikely for two different writers to produce identical handwriting. However, all writers exhibit inherent variability in their writing or signature. Despite FHEs relying on established best practices and accumulating experience over time, their traditional examination methodology often faces criticism. This criticism stems from concerns about the reproducibility and validity of their visually-based assessments, which inherently possess subjective elements. In other words, two experts can arrive at different conclusions. Therefore, there is a pressing need

✉ Moises Diaz
moises.diaz@ulpgc.es

Miguel A. Ferrer
miguelangel.ferrer@ulpgc.es

Gennaro Vessio
gennaro.vessio@uniba.it

¹ Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, Las Palmas de Gran Canaria, Spain

² Department of Computer Science, University of Bari Aldo Moro, Bari, Italy

for an objective and quantifiable methodology in the field of forensic signature analysis, one that seeks to establish new standards and introduce scientific rigor into the process.

The pattern recognition and computer vision community have developed several automatic signature verifiers (ASVs). As a result, outstanding performances have been achieved on several benchmark datasets. In particular, exceptional results have been possible thanks to sophisticated and effective deep learning models that comprise hundreds of layers and millions of parameters [16, 19]. Nevertheless, performance is not the only attribute for signature verification systems. In addition to the seven features identified by Jain for biometric systems [33] and other pertinent aspects like security, fairness, and privacy [34], explainability has emerged as a crucial factor for the practical implementation of these systems [5]. The lack of explanation in the functioning of these systems prevents their use in some practical contexts, such as finance, healthcare, government, commercial transactions, and security. An easy-to-explain and straightforward computer system with principles close to the practice of forensic examination is preferable for a jury [44].

The trade-off between the explainability and security of ASV systems presents a multifaceted consideration. While transparent and interpretable features and classifiers enhance comprehensibility, they also introduce potential vulnerabilities that adversaries could exploit in forging signatures, as emphasized by [28]. Striking the right balance necessitates meticulous evaluation, possibly entailing the amalgamation of transparent features for typical scenarios and the incorporation of additional, less transparent security layers for safeguarding sensitive applications. The imperative for regular updates and continuous research cannot be overstated in mitigating evolving threats. Ultimately, the decision regarding the trade-off among explainability, security, and performance hinges upon the specific use case and the risk tolerance of the deploying organization.

Greater transparency should be considered to justify the use of an ASV system in a courtroom: The automatic extraction of parameters from a signature should be directly linked to its physical shape. So a matching technique that is easy to interpret and explain in an objective and understandable way is needed. Moreover, there is a requirement for a measure to weigh the evidence of a conclusion drawn with simple, transparent, meaningful, and understandable scores. In summary, there is a practical use for a signature system capable of being explained in a human-to-human interaction [52]. Such a system would provide the forensic expert not only with a powerful tool to make the findings less questionable, but also a way to add absolute transparency in communicating the value of

evidence, which would also ensure reproducibility and convince authorities to use biometric technology to explain the decision-making processes.

To this end, we propose an offline ASV inspired by the new technological paradigm of explainable artificial intelligence [52], which shows desirable features for FHEs. The characteristics of our system and its analysis are detailed below:

- It is designed in a modular way to improve its explicability [52]. Thus, at each step in the pipeline (i.e., for each module), the proposed techniques can be easily changed according to their accuracy and/or transparency.
- The ASV uses explainable features, which have a physical meaning close to the structure of the signature. Furthermore, our experiments consider exploring the performance with non-explicable deep learning, DL, features for performance comparison purposes.
- Understandable matching distances are studied to match signatures. We select the most straightforward ones to make the meaning clear across all available distances.
- Forensic analysis requires a transparent model of the “world.” In our case, we use a universal background model (UBM) constructed with a third-party set of signatures. Synthetic signatures are also studied in the UBM to improve the privacy of the system.
- The system is also quite robust in supporting 1 versus 1 signature comparisons, representing the worst working scenario for an FHE.
- The results are provided in terms of evidence, i.e., likelihood ratio, LR, and probability of belonging to the UBM and to the reference population, which are criteria familiar to forensic experts.

The results validate the use of explainable and DL features in several situations. Also, we assess the use of a real versus synthetic UBM. We selected several publicly available offline signature databases for these purposes, whose acquisition process is also transparent. As for the synthetic UBM, the generation of synthetic signatures is based on motor control procedures, thus assuring its transparency. This also increases the data privacy property of the proposed explainable system. Finally, it is worth noting that while we are witnessing a paradigm shift in which signatures are increasingly acquired through online acquisition systems such as digitizing tablets and smartphones, FHEs are still required to verify signatures acquired primarily in *offline* mode. For this reason, we focus here on the latter case.

1.1 Subsequent work and main contribution

In the initial version of this work [18], an offline automatic signature verifier was proposed for forensic applications and validated on two public databases. We also analyzed individual handcrafted features to assess their impact on the verifier's performance.

However, subsequent reviews and publications have emphasized the need for quality assurance and the reliability of systems in legal contexts [2, 14, 39], highlighting a general lack of forensic handwriting examiners [14].

Taking these factors into consideration, we believe that developing explainable systems that can provide objective and reliable decisions in signature verification would be a valuable contribution to forensic handwriting analysis.

To this end, we have extended our previous work and proposed an explainable offline automatic signature verifier for forensic applications based on the verifier presented at the conference. Our experimentation now includes more databases and a more comprehensive comparative analysis between explainable and non-explainable signature verification systems. Our detailed analysis leads us to present this system as an explainable ASV.

The rest of this paper is organized as follows. First, we review related work in Sect. 2. Then, Sect. 3 introduces the proposed automatic system for forensics, which is set up in Sect. 4. Experimental evaluation and results are reported in Sect. 5, while the article is concluded in Sect. 6.

2 Related work

The demand for interpretable systems has primarily been tackled within the field of biometrics to elucidate the outcomes of deep neural network-based pattern recognition. One notable illustration is Grad-CAM, a widely recognized post-processing explanation technique [62], which functions as a visual aid. Its use has been demonstrated for ear recognition systems [3] as it can identify the most distinctive patterns in images, thus helping to provide textual explanations and even detecting the most discriminating neurons in the network. Similar instances of interpretable techniques are evident in iris recognition [10], fingerprint segmentation [36], and face recognition [73], all of which leverage the physical attributes of the subject [60].

All of these biometric recognizers are based on DL technology, which has improved explainability with visualization patterns. In offline signature verification, the rise of DL has also achieved impressive results on different benchmark datasets, e.g. [25, 26, 74]. For instance, Grad-CAM can identify regions of interest within an image of a

signature that the model deems important to its decision, thereby assisting forensic experts in comprehending which parts of a signature played a crucial role in determining its authenticity. Nevertheless, our objective is not to elucidate why a neural network-based model classified a signature as genuine or forged, but rather to create a system that is inherently easy to explain. Although Grad-CAM can offer insights into the model's decision-making process, neural networks are still perceived as complex models, and conveying their understanding and explanation to a court necessitates a background unfamiliar to most forensic examiners. Furthermore, to effectively utilize neural networks and, by extension, Grad-CAM, access to a substantial amount of labeled data is essential, which is not always the case in a typical work scenario where only a few signatures are accessible. To the best of our knowledge, a fully explainable biometric system in the context of offline ASV, up to the development of our system, has not been available.

Traditionally, to examine signatures, FHEs use various illumination and magnification tools, such as stereo-microscopes, light panels, specialized grids printed on transparent films, and so on. The comparison results are then typically provided on a five or nine-point scale, based on some hypotheses about the genuineness of the questioned signature [68].

The literature is sparse on systems specifically oriented to supporting forensic investigations. Some automated tools have been developed in recent years to support the daily activity of FHEs, such as FLASH ID, iFOX, and D-Scribe [16]. However, FHEs have traditionally made very limited use of such systems. Additionally, some tools have some drawbacks that affect the possibility of their use in real-world forensic cases. For example, in a courtroom, a careful understanding and explanation of the decision process are crucial [69]. Ultimately, none of these systems is so widespread today that its usefulness can be fully established.

As an illustration, the use of signatures is prevalent in bank checks [23]. The verification of authorship can be open to interpretation by an FHE and may often require justification in a court of law. The solution presented in this article offers evidence of the authorship of signatures through the use of likelihood ratios as metrics. Furthermore, the transparency of our offline ASV system opens the door to the acceptance of the system by judges and other professionals with limited technological knowledge.

To partially overcome this problem, and to provide empirical evidence on the reliability of decisions made by computer systems, competitions were organized by biometric researchers, in collaboration with forensic examiners, to evaluate the performance of automated tools in forensic cases. Popular competitions include the

4NsigComp, SigComp, SigWiComp series [7, 41, 42, 48, 49], in which sub-corpora closest to the real forensic cases were used, and organizers presented the results in terms of LR. The ultimate objective in these competitions was performance. These competitions did not relate to the explainability of the ASVs nor was the use considered of one signature as a reference.

To try to bridge the gap between biometrics and forensics, M. I. Malik's thesis [47] proposed new signature verification systems based on local features and on the assumption that local information contains essential clues for clear explanation rather than taking an holistic approach. Similarly, Marcelli et al. [50] proposed a method to link the features FHEs are most familiar with and measurements on the digital image of a signature on a paper document. The method was intended to provide a quantitative estimate of the variability of features when examined in different contexts. More recently, Okawa [55] has proposed an approach to mimic the cognitive processes that FHEs use to reach their decision, based on a “bag-of-visual words” and a vector of locally aggregated descriptors. The main idea is to focus on the salient local regions of the signatures in order to conduct the comparisons. Inspired by these works, we also use local features to represent the signatures.

By contrasting them with relevant published work on offline automatic signature verification systems, Table 1 provides several techniques and algorithms. It is important to note that the explainability of various systems varies. However, it is outside the purview of this study to categorize these systems in terms of their explainability. Instead, we concentrated on developing a method that could be useful to forensic handwriting experts.

In this paper, we propose a different approach. In addition to comparing a questioned signature with a reference set, we compare such a signature with a UBM

developed with external signatures. This approach is similar to a police procedure, in which the police have their own UBM to compare the cases they receive. The use of a UBM is not new in signature verification. We identify a few previous works that used a trained UBM in online signature verifiers. For example, a Gaussian Mixture Model was introduced with features of all signatures in a UBM in [51, 75]. Instead, an Ergodic-HMM was preferred to train a UBM in [4]. The signature system used the trained UBM to support the final decision in these cases. However, these trained UBMs fail in their explicability since they are too graphically opaque for human users to understand [28].

Our explainable ASV is designed with several modules of features and classifiers. These modules can be customized to meet the specific needs of different applications, as some applications require a higher level of explainability than others. Additionally, our work explicitly considers the security and forensic requirements of ASV, such as the need for robustness to attacks and the need to generate explainable predictions. Finally, our ASV has been evaluated on large and diverse signature databases, demonstrating its effectiveness.

3 Explainable ASV architecture

The architecture of our system aims to verify a questioned offline signature automatically and is composed of several modules, as illustrated in Fig. 1. Each module is a step in a pipeline that can be implemented with different techniques to analyze their impact on the final performance. Additionally, each module can be easily replaced depending on the desired level of explainability in an end application or for a user.

Our system requires a reference set composed of enrolled signatures to the ASV, i.e., real reference

Table 1 Offline automatic signature verification techniques for forensic handwriting examination

Algorithm/technique	Description	References
Graphology analysis	Examines psychological aspects of handwriting to assess authenticity	[56]
Machine learning	Utilizes supervised learning models for signature verification based on training data	[23]
Local features	Identify keypoints or interest points in signatures for robust feature matching.	[47]
Stroke sequence analysis	Studies the order and sequence of strokes in a signature	[50]
Local binary patterns	Encodes texture information for signature representation and comparison	[53]
Handwriting feature analysis	Analyzes individual handwriting features like slant, size, and pressure	[68]
Graph-based approaches	Represent signatures as graphs and analyze structural information for authentication	[45]
Forensic document analysis	Investigates paper, ink, and other physical aspects of the document	[55, 69]
Signature comparison	Compares questioned signatures to known reference signatures	[17]
Biometric authentication	Uses biometric data like pen pressure and speed for verification	[19]
Neural networks	Employs deep learning models for complex signature analysis	[43, 74]

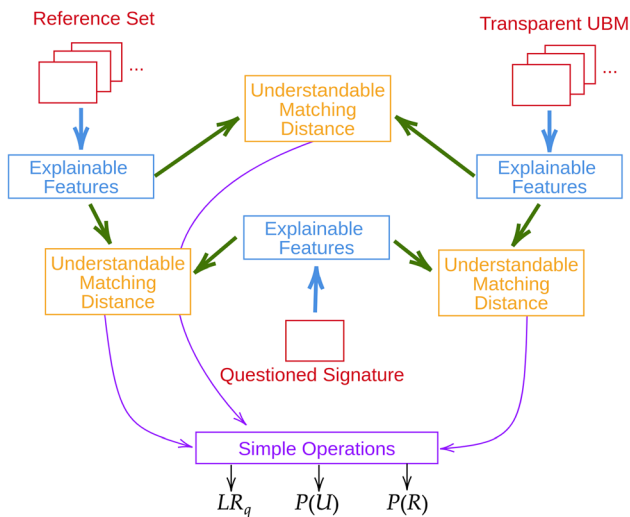


Fig. 1 Block diagram illustrating our explainable offline ASV system. The inputs consist of the reference set containing known signatures, a transparent Universal Background Model (UBM) constructed using a third-party set of signatures, and questioned signatures that have never been encountered by the ASV. The outputs are presented in terms of evidence, specifically the likelihood ratio (LR_q), as well as the probabilities of membership in the UBM and the reference set ($P(U)$, $P(R)$)

signatures of individuals. Additionally, a transparent UBM, which ought to be understood by itself [40], is used as input. The UBM is a gathering of representative sample from the population. Toward an explainable system, we propose to use a pool of signatures where the features are extracted and compared directly with the questioned and reference set. It is worth highlighting that the UBM is challenging to build, as it should represent all signature populations except for the reference set. A practical solution is to build up this set with signatures signed by other people. To this end, we used publicly offline signatures databases. Ideally, this UBM should be independent of the written script.

The output of the system numerically estimates whether the questioned signature is genuine or not through the probabilities of belonging to the reference set and the UBM, and the LR. Rather than a simple numerical score, the LR estimates from the evidence whether a questioned specimen is closer to the reference signatures or to the UBM [11]. Furthermore, the design of the verifier uses an explainable framework with features and matching mechanisms that are more understandable to human beings. In summary, the hypothesis behind this system is that a genuine signature must be closer to the reference set than the UBM set; and a fake specimen must be farther from the reference set than the UBM.

Formally speaking, given an image-based signature in the form of a grayscale matrix, \mathbf{X} , our system initially converts it into a specific feature vector of m elements,

$\mathbf{x} = [x_1, x_2, \dots, x_m]$. Thus, our system proceeds with specimens interpreted as features. Accordingly, feature extraction is directly carried out with all signatures in the UBM, the reference signatures, and the questioned one. To achieve interpretability, the features used should be easily understood by a human being and related to physical or geometrical characteristics of the signatures. To this aim, explainable handcrafted features are used for pursuing model explainability. For the sake of transparency, we select some of the well-known explainable features such as geometrical relations, curvature properties and slants.

Let $r \in R$ be a reference specimen and R the reference set. Also, let U be the total number of signatures included in the UBM and u a single signature of the UBM. Finally, we define q as a questioned signature to be verified automatically. Next, the explainable ASV calculates an LR and two types of probabilities.

3.1 Likelihood ratio

We compute the likelihood ratio, LR_q , for the questioned signature. It evidences how many times q is closer to the reference signatures than to the UBM. To compute LR_q , we calculate two distances. A first distance, $\delta_1(q)$, represents the alternative hypothesis, H_1 . This distance quantifies how close a signature is to other signers, $u \in U$. Mathematically, it is denoted as follows:

$$\delta_1(q) = \min_{\forall u \in U} d(q, u) \tag{1}$$

where $d(\cdot)$ denotes a generic distance matching between the questioned signature and a signature belonging to the UBM, $u \in U$.

We express the null hypothesis, H_0 , by measuring the distance between the questioned signature and the available reference signatures, $r \in R$, belonging to the claimed signer. The following relation is performed to compute $\delta_2(q)$:

$$\delta_2(q) = \min_{\forall r \in R} d(q, r) \tag{2}$$

Finally, the evidence is computed in terms of $LR(q)$ as:

$$LR_q = -2 \log \left(\frac{\delta_2(q)}{\delta_1(q)} \right) \tag{3}$$

which means how many times the questioned signature is closer to the reference signatures than to any other signature. New scores in terms of probabilities are worked out to improve the interpretability of such a ratio.

3.2 Probabilities

In addition to LR_q , the system calculates two probabilities, $[P(U), P(R)]$, which can provide more transparency to the

verification. $P(U)$ denotes the probability of belonging to the UBM. Alternatively, $P(R)$ gives the probability of belonging to the reference set. While the first probability can always be computed, the second requires more than one signature in the reference set in order to be estimated. It is, therefore, expected that the more reference signatures are available, the better the estimation of $P(R)$. As such, our offline ASV cannot compute $P(R)$ in the case of 1 versus 1 signature verification.

3.2.1 Probability of belonging to the UBM

The metric $P(U)$ represents the probability that the questioned LR belongs to the UBM. It can be computed as follows:

$$P(U) = 1 - F_{LR_{u_i}}(LR_q | \mu_{LR_{u_i}}, \sigma_{LR_{u_i}}) \tag{4}$$

Here, $F_{LR_{u_i}}$ is the normal cumulative distribution function (CDF) with a mean and standard deviation $(\mu_{LR_{u_i}}, \sigma_{LR_{u_i}})$, evaluated at the questioned LR_q . The graphical representation of $P(U)$ is the area under the normal probability density function (PDF), $f_{LR_{u_i}}$ in the interval $[LR_q, +\infty]$. To obtain this, we need to compute the LR of each specimen in the UBM, LR_{u_i} :

$$LR_{u_i} = -2 \log \left(\frac{\delta_{U,2}(u_i)}{\delta_{U,1}(u_i)} \right) \tag{5}$$

where the signature i in the UBM is denoted by u_i and $(\delta_{U,1}, \delta_{U,2})$ are two distances that depend on the UBM and the reference set, respectively. These two distances can be estimated as follows:

$$\delta_{U,1}(u_i) = \min_{\forall u \in U} d_{u \setminus u_i}(u_i, u) \tag{6}$$

$$\delta_{U,2}(u_i) = \min_{\forall r \in R} d(u_i, r) \tag{7}$$

$d(\cdot)$ being a matching distance of two signatures in the form of feature vectors.

3.2.2 Probability of belonging to the reference set

This probability, $P(R)$, assesses how probable it is that the questioned signature may belong to the reference specimens. The following expression can be used to calculate it:

$$P(R) = F_{LR_{r_i}}(LR_q | \mu_{LR_{r_i}}, \sigma_{LR_{r_i}}) \tag{8}$$

In this case, $P(R)$ represents the area under the $f_{LR_{r_i}}$ curve where LR_q is likely to fall within the interval $[-\infty, LR_q]$. Thus, the LR_q is evaluated in the CDF, $F_{LR_{r_i}}$, with a mean

and standard deviation $(\mu_{LR_{r_i}}, \sigma_{LR_{r_i}})$. Similarly, we can determine the LR of each reference signature as follows:

$$LR_{r_i} = -2 \log \left(\frac{\delta_{R,1}(r_i)}{\delta_{R,2}(r_i)} \right) \tag{9}$$

where $(\delta_{R,1}(r_i), \delta_{R,2}(r_i))$ are two distances that depend on references and UBM specimens:

$$\delta_{R,1}(r_i) = \min_{\forall r \in R} d_{r \setminus r_i}(r_i, r) \tag{10}$$

$$\delta_{R,2}(r_i) = \min_{\forall u \in U} d(r_i, u) \tag{11}$$

The above procedure aims to describe a transparent system which is generic, not only on the feature space but also on distance matching. A visual easy-to-understand example of outputs generated by our system can be viewed in Fig. 2.

4 Setting up the proposed system

This section details the explainable features and understandable matching distances used in the modules of the ASV. We also identify the databases used and the experimental protocol for evaluation.

4.1 Databases

We used two different databases to develop the UBM. We also used three other databases for the experiments to avoid bias in the results. This strategy contributes significantly to the ASV explainability [5]. All used signatures are also transparent in terms of the acquisition protocol and composition.

Two UBMs were developed, one with real signatures and another with synthetic ones. The computational model to generate the synthetic signatures was based on well-known control motor processes, and the procedure is completely transparent [22]. It provides complete privacy to the explainability of the UBM because none of the signatures could be identified. The former was created with the first n genuine signatures of GPDS960 [72], whereas the latter with the Synthetic10000 [22]. The motivation is to study our system in these two situations, the actual use of which would depend on the real application. The experiments were conducted with offline specimens available in the MCYT-75 [57], BiosecuID [24], Thai [12], and CEDAR [37] databases. Please note that all these databases are publicly available and allow experiments using two scripts.

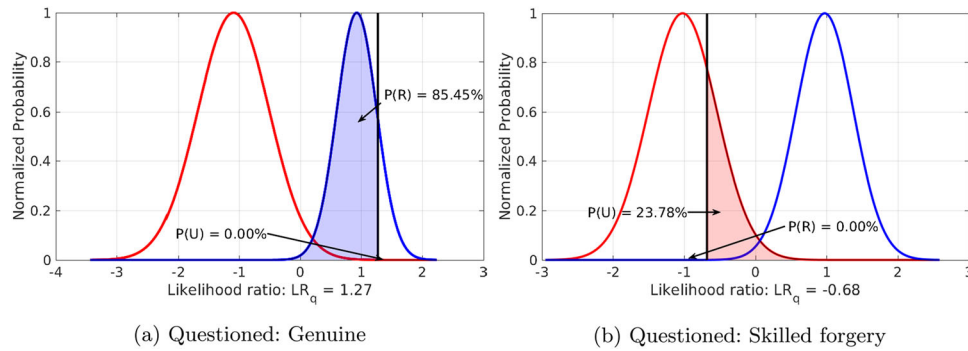


Fig. 2 An interpretable visualization of the output results of the proposed ASV system is presented, showcasing scenarios where questioned signatures are either a genuine (on the left) or a skilled forgery (on the right). The normalized probabilities are depicted as functions of the likelihood ratios. In both figures, the solid red and blue curves represent the normalized probability density functions of

4.2 Features

We used four types of handcrafted features in our proposal as explainable features. DL-based features were also considered to quantify the performance for each feature. Explainable features could be accepted in a court if they were close to the physical features made by ink deposition on paper, as non-expert humans may then understand how they were obtained. Moreover, the value of the parameters is related to the physical phenomenon of ink deposition of the signature. On the other hand, DL features can perform better [28]. The features utilized in this paper can be shared upon request.

As for explainable features, the first type is based on several geometrical descriptions of the shape of the signature. We work out these measures in Cartesian and polar coordinates based on their contours [20]. All geometrical features were concatenated into a vector, resulting in a dimension of 445. The second type consisted in dividing the signatures recursively into a quadtree at two levels. Then we computed the gradient on each level to explain the texture properties of the signatures [63]. The dimension was 200. The third type was based on counting the run-lengths of the binary images. The feature vector dimension was 400, and it explains the width of the strokes in the four main directions: vertical, horizontal, diagonal 45° , and diagonal 135° [9]. The last feature type consisted in determining the physical textural properties of offline signatures [21]. Specifically, we elaborated the local binary patterns and local derivative patterns of the images. The dimension of these vectors is 765 and 255, respectively. Moreover, we applied the discrete cosine transform to these vectors, obtaining dimensions of 168 and 167, respectively. As a result, four feature vectors associated with textural

the UBM and the reference sets, respectively, denoted as $f_{LR_{U_i}}$ and $f_{LR_{R_i}}$. As expected, when the questioned signature is genuine, the probability of belonging to the UBM ($P(U)$) is lower, while the probability of belonging to the reference set ($P(R)$) is higher

features were obtained. These features are summarized in Table 2.

There are two main advantages in using these features. The first is that they have been successfully applied to offline ASVs in previous work and some forensic-based signature competitions such as 4Nsig-Comp2010 [7] and SigWlcomp2015 [49]. The second is that the LBP, LDP, and gradient features, organized in quadtree structures, quantify how the ink was deposited on the paper from different points of view. Run-length and geometrical features work on binary images. Therefore, they mainly reflect aspects related to the shape of the signature from different perspectives. Furthermore, as the extraction process is entirely different in each case, we take advantage of the complementarity of their information content.

As for the DL features, three deep neural networks were involved. The first two models consisted of popular convolutional neural networks, namely VGG19 [66] and ResNet_v2 [31], with weights pre-trained on ImageNet. To perform feature extraction, we followed the common practice of removing the top-level classifier, adding a global average pooling on top of the convolutional base, and performing *transfer learning*. Additionally, we conducted transfer learning not from ImageNet but from a knowledge base more similar to our context. To this end, we also used an ad hoc, fully convolutional network (FCN) with weights we have pre-trained on the very popular MNIST dataset of handwritten digits. Having no fully connected layers, this model can accept inputs of any size, so it is suitable for processing small MNIST digits and higher resolution signature images. We applied commonly suggested “generic” (VGG19-g) pre-processing for all three models, which consists of scaling the signatures to 224×224 pixels and normalizing their values between 0 and 1. In this way, the global average pooling provides 512

Table 2 Feature vectors used in the explainable ASV

Explainable features			Deep learning features		
Label	Features	Dimension	Label	Features	Dimension
g	Geometrical	445	d1	VGG19-g.	512
qt	Quadtree	200	d2	VGG19-s.	512
rl	Run-length	400	d3	ResNet_v2	512
t1-t4	Textural	(765, 255, 168, 167)	d4	FCN	64

features from VGG, 2048 for ResNet_v2, further reduced to 512 with a simple auto-encoder, and a smaller 64-dimensional feature vector from the simpler FCN. Also, to experiment with a more “specific” (VGG19-s) pre-processing for signature images (as done for example in [15]), we resized the images to 155×220 pixels (to preserve the aspect ratio), inverted the pixel values so that the background pixels have value 0 and finally normalizing these values between 0 and 1. For simplicity, we have applied this specific pre-processing only with VGG19.

It is worth noting that our goal was to assess whether the explainable features are competitive enough, when compared with automatically learned features, without much loss of performance. For this reason, we used the common practice of feature extraction with transfer learning without training much more specialized DL-based signature verification systems. Indeed, developing such a system was not the aim of our research. At the end of the experimental section, a comparison with more advanced methods is reported.

4.3 Matching distance

Many powerful offline ASVs use machine learning classifiers. However, we propose to use understandable matching distances for function-based features as they offer a more straightforward explanation for our ASV. In the literature, there exist matching distances for function-based features of different or equal length, such as the ℓ^2 norm. In our case, our matching will be based on features of the same size. We experimented with three distances for their extensive use in signature verification [16].

We decided to use a simple version of DTW. This matching distance builds a dissimilarity matrix with the Euclidean distances between all the members of two feature sequences to calculate the optimal distance between the elements of the two feature vectors. Excellent results have been achieved in offline and online signature verification when this distance was used, e.g. [64]. Manhattan distance or ℓ^1 norm has also been used because of its good results in signature verification [61]. Finally, cosine distance has also been chosen for our experiments since it is based on the simple Euclidean dot product formula. It has

proven its effectiveness via DL features over others, such as the Euclidean distance [58].

Motivated by their simplicity and effectiveness in signature verification, we integrated these methods into the proposed ASV. In addition, their use would help our aim to develop an explainable automatic signature verifier for offline signatures.

4.4 Evaluation

As usual in biometric systems and automatic signature verification, performance was evaluated through detection error trade-off (DET) curves. We computed the false acceptance rate (FAR) and false rejection rate (FRR) curves for this purpose. First, we used the signatures in the reference set and the questioned signatures to develop the FAR curve. To allow a fair comparison, the reference signatures were the initial signatures enrolled per user. Subsequently, the remaining genuine signatures were employed as questioned signatures. Then, two FAR curves were built for random and skilled forgeries. The random forgery (RF) set was worked out with a random signature out of 74 random users, whereas all fake signatures were used for the skilled forgery (SF) set. To quantify the performance in the verification task, the equal error rate (EER) was used in all cases.

It is noteworthy that various metrics have been proposed for the assessment of automatic signature verification systems. These metrics encompass the FAR, FRR, and Half Total Error Rate (HTER), as detailed in [6]. Furthermore, a recent standard ISO Central Secretary [32] promotes the adoption of the Attack Presentation Classification Error Rate (APCER) at fixed Bona Fide Presentation Classification Error Rate (BPCER) values. In our research, we have chosen to uphold the universally accepted metrics [16, 59] to prevent potential confusion in the long-term development of the automatic signature verification field. It is important to note that these same metrics were employed in the most recent international ASV competition to evaluate system performance [70].

To evaluate our system, we used only one signature as a reference. Our motivation was to set up the system to cope with a very challenging case in ASV. Hence, it was expected that the more signatures are used in the reference

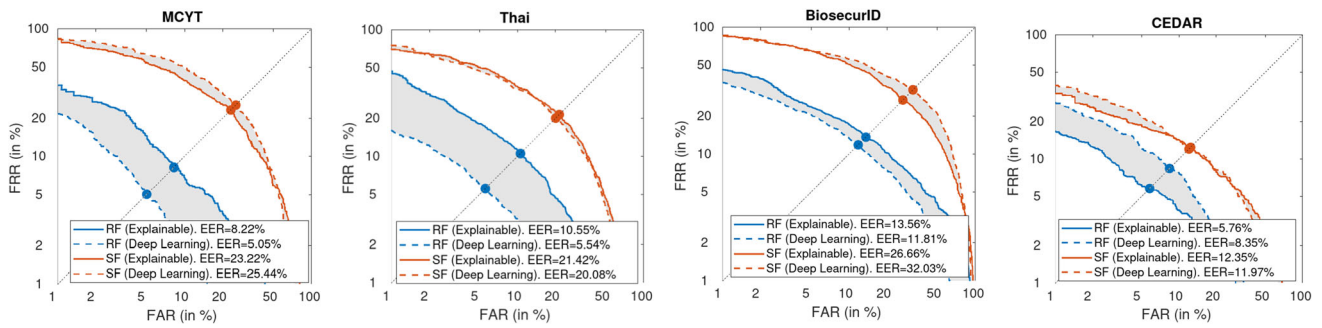


Fig. 3 Detection error trade-off curves for the proposed ASV system, utilizing explainable features derived from both geometric and textural properties of the signature, as well as deep learning features extracted from VGG19 and ResNet_v2

Table 3 Equal error rates (%) using three understandable matching distances in the ASV

Database	Matching distance	Explainable features	Deep learn features	
MCYT	DTW	8.22	5.05	Random forgeries
	ℓ^1 norm	6.29	4.29	
	Cosine	8.94	4.09	
Thai	DTW	10.55	5.54	
	ℓ^1 norm	7.81	4.76	
	Cosine	8.86	4.77	
BiosecurID	DTW	13.56	11.81	
	ℓ^1 norm	8.16	9.85	
	Cosine	11.59	10.15	
CEDAR	DTW	5.76	8.35	
	ℓ^1 norm	5.19	8.01	
	Cosine	7.17	6.50	
MCYT	DTW	23.22	25.44	Skilled forgeries
	ℓ^1 norm	21.17	23.93	
	Cosine	21.89	24.47	
Thai	DTW	21.42	20.08	
	ℓ^1 norm	20.58	20.25	
	Cosine	23.83	20.00	
BiosecurID	DTW	26.66	32.03	
	ℓ^1 norm	26.03	31.08	
	Cosine	27.92	31.33	
CEDAR	DTW	12.35	11.97	
	ℓ^1 norm	12.12	8.41	
	Cosine	13.18	11.14	

population, the better the performance. This evaluation allows us to study our proposed system for forensics in the context of other computational ASVs.

5 Experimental results

This section demonstrates that the proposed offline ASV is adequate for performance and explainability. We also studied the 1 versus 1 verification case to adjust the ASV.

Finally, the best configuration is evaluated by increasing the number of signatures in the reference set.

5.1 Studying explainable and deep learning features

Feature extraction is a crucial step in an offline ASV. Indeed, many researchers have proposed several feature extractors to improve the performance of the systems [16]. However, in the previous decade, DL strategies seemed to be the optimal technique for representing patterns in terms

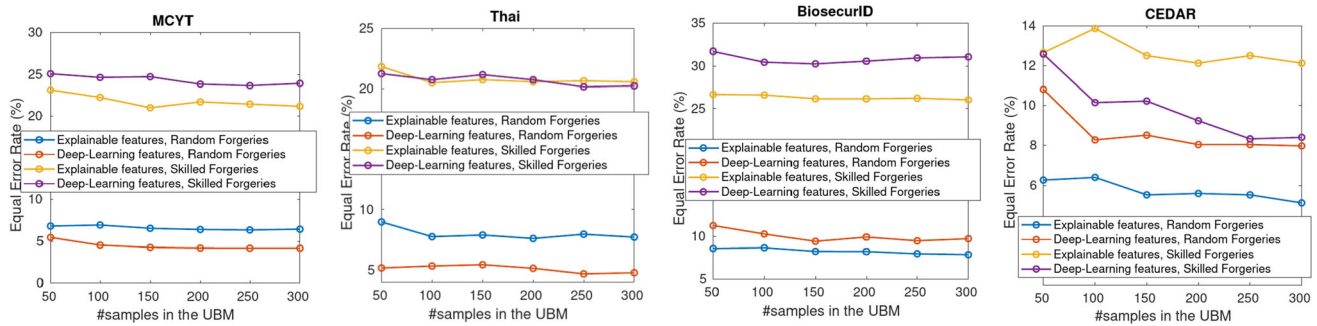


Fig. 4 Exploring the optimal number of genuine signatures in the UBM across three databases. The experiments encompassed random and skilled forgeries, employing both explainable and deep learning features in each scenario

of performance [29]. In this section, we quantify the performance of our ASV system when explainable or DL features are used.

We set up the system using DTW as matching distance, 300 real signatures in the UBM, and a single reference signature. Besides using this setup in our previous work [18], training with one signature represents the most challenging case in signature verification [9]. To compute a global performance, we fused LRs and probabilities as if they were considered scores in a biometric-based signature verifier. We then applied a weighted sum for the explainable features (see Table 2) as follows: $s_{fused}^h = \omega_1(LR_g + P_g) + \omega_2(LR_{qt} + P_{qt}) + \omega_3(LR_{r1} + P_{r1}) + \omega_4(LR_{t1} + P_{t1})$ where the weights were proportionally adjusted, thus regarding the performance of each feature as:

Table 4 Equal error rates (%) using real and synthetic genuine signatures in the universal background model using ℓ^1 norm as matching distance

	Database	UBM	RF	SF
Explainable features	MCYT	Real	6.29	21.17
		Synt.	7.24	23.04
	Thai	Real	7.81	20.58
		Synt.	12.00	24.25
	BiosecurID	Real	8.16	26.03
		Synt.	8.33	25.27
	CEDAR	Real	5.19	12.12
		Synt.	7.00	11.82
Deep learning features	MCYT	Real	4.29	23.93
		Synt.	4.99	22.42
	Thai	Real	4.76	20.25
		Synt.	6.81	22.92
	BiosecurID	Real	9.85	31.08
		Synt.	8.04	23.37
	CEDAR	Real	8.01	8.41
		Synt.	4.58	5.15

$\Omega^h = (0.1, 0.75, 0.05, 0.1)$. For those who are interested, it may be helpful to consult the individual performance of handcrafted features as presented in [18]. The same fusion strategy was applied to the DL features: $s_{fused}^{dl} = \omega_1(LR_{d1} + P_{d1}) + \omega_2(LR_{d2} + P_{d2}) + \omega_3(LR_{d3} + P_{d3}) + \omega_4(LR_{d4} + P_{d4})$, the weights being $\Omega^{dl} = (0.5, 0.25, 0.15, 0.1)$.

Figure 3 shows the DET plots when explainable or DL features are used for each dataset and random, RF, and skilled forgeries, SF. For the MCYT corpus, a similar performance was obtained for SF with both features. The area between DET plots was 1.90. In the case of RF, slightly better performance was achieved with DL features with 3.02 of the difference between DET plots. These results are also consistent with the other two databases. We obtained 2.48 and 0.94 in DET curves for RF and SF in Thai, respectively. Similar differences were found in BiosecurID, with 2.30 and 4.36 for RF and SF. Thus, for RF and SF, the differences in the CEDAR case were 0.22 and 0.74. In the case of EER, Fig. 3 shows the performance obtained in each case. In general, using this initial configuration of our system, explainable features offer similar results as DL ones, especially in SF, which represents the most challenging signature verification experiment.

5.2 Studying different understandable matching distances

In this subsection, our goal is to analyze the performance of our system with DTW, ℓ^1 norm and cosine, as understandable matching distances. Therefore, we kept the other setting options of our system unchanged, such as the use of real signatures in the UBM, one signature as a reference, and the results when the features are fused.

Table 3 summarizes the results we obtained for both random and skilled forgeries. We can compare the performance in terms of EER when explainable and DL features are used with each matching distance. It is observed that the lower loss of performance with explainable

Table 5 Equal error rates (%) increasing the number of reference signatures

Database	UBM	Feature	Reference signatures						
			1	3	5	7	10		
MCYT	Real	Explain.	6.29	4.32	3.86	2.72	1.80	Random forgeries	
		Deep L.	4.29	2.52	1.73	1.03	0.74		
	Synt.	Explain.	7.24	4.43	3.57	2.88	1.82		
		Deep L.	4.99	2.67	1.77	1.30	0.90		
	Thai	Real	Explain.	7.81	2.92	2.35	2.34		2.00
		Deep L.	4.76	1.47	1.00	0.80	0.81		
BiosecurID	Real	Explain.	8.16	5.28	5.09	4.95	4.58		
		Deep L.	9.85	6.11	5.58	5.55	4.16		
	Synt.	Explain.	8.33	5.70	5.16	5.58	4.41		
		Deep L.	8.04	5.07	4.54	4.56	3.31		
CEDAR	Real	Explain.	5.19	3.64	1.99	1.85	1.25		
		Deep L.	8.01	4.92	4.21	3.64	3.54		
	Synt.	Explain.	7.00	4.38	3.13	2.49	2.39		
		Deep L.	4.58	1.58	1.31	0.94	1.01		
MCYT	Real	Explain.	21.17	18.24	15.93	13.79	10.85	Skilled forgeries	
		Deep L.	23.93	16.99	14.86	12.46	9.79		
	Synt.	Explain.	23.04	18.77	16.19	14.95	12.37		
		Deep L.	22.42	16.28	12.63	10.59	8.72		
Thai	Real	Explain.	20.58	12.42	9.58	9.33	8.50		
		Deep L.	20.25	10.25	7.75	8.08	7.75		
	Synt.	Explain.	24.25	14.83	12.42	12.17	11.08		
		Deep L.	22.92	11.50	9.08	8.83	8.50		
BiosecurID	Real	Explain.	26.03	20.28	19.20	18.64	17.62		
		Deep L.	31.08	22.30	19.96	20.59	18.19		
	Synt.	Explain.	25.27	19.52	18.89	18.38	17.31		
		Deep L.	23.37	17.94	14.91	15.22	13.33		
CEDAR	Real	Explain.	12.12	7.20	4.70	4.24	3.71		
		Deep L.	8.41	6.67	5.53	4.77	4.70		
	Synt.	Explain.	11.82	6.59	3.71	3.48	3.18		
		Deep L.	5.15	2.95	2.27	1.52	1.59		

features is obtained with the ℓ^1 norm. In the column “Explainable features,” we see that this distance outperforms the other two. Additionally, we can compare the performance loss row by row with explainable and DL features. In terms of explainability, all distances support the concept of explainability in automatic signature verification. Among them, ℓ^1 norm distance can be seen as the most useful in critical applications and the easiest to understand since it is based on the sum of the difference of individual elements in a vector. Here we will continue to explore the best system configuration with this matching distance.

5.3 Studying the universal background model

In this subsection, we explored several aspects of the UBM. Once again, we use one signature as a reference, feature fusion at score level and the ℓ^1 norm as understandable matching distance.

We collected the performance for both random and skilled forgeries by gradually increasing the number of genuine signatures in the UBM from 50 to 300. Figure 4 shows three subplots, one per dataset considered. Each subplot shows four solid lines for random and skilled forgery experiments, with explainable and DL features. Relatively stable performance is observed in all cases for more than 100 different genuine signatures. Therefore, we cannot decide on a precise n_{opt} number. In the following

Table 6 State-of-the-art results in offline ASV

Method	Ref.	Performance (EER %)			
		th_{global}		th_{user}	
		RF	SF	RF	SF
<i>Database: GPDS</i>					
CNN-BiLSTM [43]	FCV*	–	10.16	–	–
HOG, DRT + DMML [67]	5	2.15	20.94	–	–
SigNet + SVM [29]	12	–	–	–	1.69
Curvelet trans. + OC-SVM [27]	8	–	15.95	–	–
Tangent Angle+ SVM [8]	12	–	14.82	–	–
Poset-oriented grid + SVM [77]	5	–	–	–	9.87
<i>Database: CEDAR</i>					
CNN-BiLSTM [43]	FCV*	–	0	–	–
on-2-off, SigCNN [35]	5	–	6.41	–	4.50
on-2-off, SigCNN [35]	10	–	5.27	–	3.48
CNN+Con. Loss Layer [71]	5	–	2.50	–	–
CNN+Con. Loss Layer [71]	10	–	1.66	–	–
Point-to-Set + CNN [76]	5	–	9.29	–	5.22
Triplet Nets-Graph [46]	10	–	12.27	–	5.91
Archetypal analysis [78]	5	–	–	–	2.07
SigNet + SVM [29]	8	–	–	–	4.77
Poset-oriented grid + SVM [77]	5	–	–	–	4.12
Curvelet trans. + OC-SVM [27]	8	–	7.83	–	–
DCNN + PDSN [38]	5	–	7.40	–	4.37
Waveforms+GP [65]	8	–	8.71	–	–
Meta-learning [30]	8	–	10.21	–	7.07
<i>This work: explainable ASV</i>	1	5.19	12.12	–	–
<i>This work: explainable ASV</i>	5	1.99	4.70	–	–
<i>This work: explainable ASV</i>	10	1.25	3.71	–	–
<i>Database: MCYT</i>					
On-2-off, SigCNN [35]	5	–	5.82	–	3.42
On-2-off, SigCNN [35]	10	–	4.55	–	2.01
CNN+Con. Loss Layer Tsourounis et al. [71]	5	–	2.61	–	–
CNN+Con. Loss Layer Tsourounis et al. [71]	10	–	1.62	–	–
HOG, DRT+DMML [67]	5	1.73	13.44	–	–
Poset-oriented grid + SVM [77]	5	–	–	–	6.02
Archetypal analysis [78]	5	–	–	–	3.97
SigNet + SVM [29]	10	–	–	–	2.87
VLAD + KAZE [54]	10	–	5.60	–	–
DCNN + PDSN [38]	5	–	7.12	–	3.78
Point-to-Set + CNN [76]	5	–	9.21	–	4.86
Triplet Nets-Graph [46]	10	–	9.16	–	3.91
Waveforms+GP [65]	10	–	7.55	–	–
Meta-learning [30]	5	–	15.37	–	12.77
Handcrafted + DTW [18]	1	6.79	19.71	–	–
<i>This work: Explainable ASV</i>	1	6.29	21.17	–	–
<i>This work: Explainable ASV</i>	5	3.86	15.93	–	–
<i>This work: Explainable ASV</i>	10	1.80	10.85	–	–
<i>Database: Thai</i>					
SCUT + CNN [13]	5	0.19	7.10	–	–

Table 6 (continued)

Method	Ref.	Performance (EER %)			
		th_{global}		th_{user}	
		RF	SF	RF	SF
LTP + oBIFs [13]	5	1.09	10.91	–	–
ERL [13]	5	3.02	17.80	–	–
Textural + HMM [12]	5	2.01	11.08	–	–
<i>This work: Explainable ASV</i>	1	7.81	20.58	–	–
<i>This work: Explainable ASV</i>	5	2.35	9.58	–	–
<i>Database: BiosecurID</i>					
LBP + SVM + Attributes [53]	4	1.66	15.55	–	–
LBP + SVM [24]	4	4.81	20.28	–	–
Handcrafted + DTW [18]	1	10.65	25.91	–	–
<i>This work: explainable ASV</i>	1	8.16	26.03	–	–
<i>This work: explainable ASV</i>	5	5.09	19.20	–	–

*FCV Fold cross validation process

experiments, we set $n_{opt} = 300$, as we did in our previous work [18]. Additionally, we see that the curves for random and skilled forgeries are slightly closer together. In these cases, we also see that DL features do not always guarantee better performance, especially for skilled forgeries.

It is worth noting that the size of the UBM can be increased, which may be advantageous for computer vision applications. However, the objective of our explainable offline ASV is to serve as an automated tool for FHE who are accustomed to designing manual UBM. Therefore, a size of 300 signatures could be deemed large for FHE, but it would be considered a standard UBM size for computer vision purposes.

Also, we included fake signatures in the UBM used for matching the genuine one. As such, the UBM was enlarged to $2 \cdot n_{opt}$ offline specimens. In general, we observed that adding forgeries in the UBM barely improves performance. In MCYT or Thai with explainable features, the results were worsened by less than 1. No relevant improvements were seen in RF with BiosecurID or Thai when DL features were used. As for the SF, BiosecurID constantly improves, whereas no effect is observed in Thai, and MCYT is slightly worsened. As for an easy-to-explain system in a court, we avoid complicating the UBM by introducing forgeries. Hence, this analysis suggests using only genuine signatures in the UBM.

Data privacy helps in the design of a confidentiality system, which is another goal considered in the explicability of the algorithms [5]. Accordingly, we designed a UBM with artificial specimens as a further option for our system. To this end, we randomly chose 300 different signatures from the first identities in a synthetic offline database [22]. All experiments were repeated using this

synthetic UBM. To analyze the use of this UBM, we report the results obtained in Table 4, when a real UBM was used under the same conditions. Overall, we can observe that the performance does not change significantly due to the effect of the real versus synthetic UBM. On the contrary, it sometimes improves, as in the case of BiosecurID, or is slightly worse (see results with Thai).

In some applications in biometrics, synthetic signatures alleviate conflicts with data protection regulations and cope with insufficient training data. However, even though privacy is one of the requirements in explainable systems [5], some applications prefer real signatures in the UBM. According to this dichotomy, we analyze below the proposed explainable ASV, thus offering results with a real and a synthetic UBM.

5.4 Effect of using multiple reference signatures

We further compared the performance with explainable and DL features when different reference signatures were used. Commonly, the more knowledge a system has, the less the error will be. For this experiment, we used the best configurations found: fused features and ℓ^1 norm. In addition to the LR and probabilities of belonging to the UBM, we computed the probability of belonging to the reference set when more than one signature is used. These latter probabilities are fused at the score level by using the same weights already found.

We can see in Table 5 a consistent behavior of the proposed system since the more reference signatures there are, the better the performance. This means that we obtain a better representation of $P(R)$. As expected, the best results were obtained using DL features. Aside from

BiosecurID, real signatures in the UBM seem to be a better option for performance. It should be considered therefore that interpretability was successful, despite the performance loss, since it leads to explainable ASVs. This experiment allows us to compare our results with the state-of-the-art.

5.5 Comparative analysis with the state of the art

The proposed explainable ASV is now compared to previous work. We aim to determine whether the achieved performance using our explainable system is adequate for practical use.

Indeed, one of the most challenging tasks is to evaluate fairly the state-of-the-art results in offline signature verification. Despite recent efforts with the Thai dataset [13], the main reason is the lack of standard benchmarks or competitions fixing experimental protocols and metrics for ASV evaluation.

Despite this difficulty, Table 6 tries to overview results in ASV using several publicly available popular databases. Other complications to reasonably analyzing prior literature are using global versus custom thresholds in the system or experimenting with random or skilled forgeries. Moreover, evaluation metrics like accuracy, FAR, and FRR are commonly used throughout the literature. For simplicity and tradition in the field [16], Table shows some works that use the EER and have a performance which is competitive.

More importantly, in this work, the contributions in Table 6 are based on machine learning (ML) and DL techniques. According to [1, 28], DL is at the extreme of unexplainable systems, and some ML techniques could be considered easier to explain, such as random forests (RFs), which are not commonly used in ASV because of their poor performance. Consequently, the use of the challenging systems by the FHE in a courtroom cannot be guaranteed because their results cannot be easily explained [44].

To this end, using a fully explainable system for automatically verifying signatures would lead to a loss of performance. This work can quantify such a performance loss by comparing our results with the state-of-the-art in Table 6. Our most explainable configuration is based on explainable features, ℓ^1 norm as matching distance, real signatures in the UBM, and using a global threshold for evaluation. We estimate a weak performance of 2.16 and 2.48 perceptual points for RF and SF, respectively, for the most challenging system in the Thai database. Nevertheless, it is worth pointing out that our system outperformed the results given in [12]. In the case of BiosecurID, we lost

about 3.62 and 4.73 of performance for RF and SF. This is compared with [53], which used four signatures as references and was the best performing system in Table. Regarding the work proposed in [67], we quantify a performance loss of our system of 2.13 and 2.49 for RF and SF, respectively, with the MCYT corpus. In summary, even if we lose some performance, the advantage of offering explicable ASVs validates the use of our system in specific applications, where this characteristic is critical, such as in forensic applications.

6 Conclusion

In this paper, we proposed a novel explainable offline signature verifier to support FHEs. We introduced a universal background model of signatures from third-party signers to improve the accuracy of our system. To preserve privacy, we added synthetic signatures to the UBM. We also considered explainable features and understandable distance matching.

Our ASV can provide objective evidence of whether a signature is genuine or not in terms of likelihood ratios and probabilities of belonging to the UBM and the reference set. This makes it suitable for use in forensic settings, where it is important to be able to explain the decisions made by the ASV.

Our experiments demonstrated that our ASV can achieve competitive performance in an explainable setting. We also showed that explainable features and an understandable distance matching based on the ℓ^1 norm can be used to maintain a state-of-the-art performance level. This suggests that handcrafted features should still be considered for ASV applications where system explainability is crucial. We believe that this research will help to narrow the gap between the forensic and pattern recognition communities by providing a novel and explainable offline signature verifier.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research is partially supported by the Spanish MINECO (PID2019-109099RB-C41 project), the European Union (FEDER program), and the Italian Ministry of University and Research through the PON AIM 1852414 project.

Availability of data and materials The full databases are freely available.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate The authors are grateful to Daniela Mazzolini and Patrizia Pavan, forensic handwriting examiners enrolled as technical consultants at Civil and Criminal Courts in Italy, for fruitful discussions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6:52138–52160
- Adeyinka OA, Adesesan BA (2021) The reproducibility and repeatability of modified likelihood ratio for forensics handwriting examination. *Int J Comput Inf Eng* 15(5):322–328
- Alshazly H, Linse C, Barth E et al (2021) Towards explainable ear recognition systems using deep residual networks. *IEEE Access* 9:122254–122273
- Argones E, Pérez-Piñar D, Alba JL (2009) Ergodic HMM-UBM system for on-line signature verification. European workshop on biometrics and identity management. Springer, Berlin, pp 340–347
- Arrieta AB, Díaz-Rodríguez N, Del Ser J et al (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible ai. *Inf Fusion* 58:82–115
- Bengio S, Marcel C, Marcel S et al (2002) Confidence measures for multimodal identity verification. *Inf Fusion* 3(4):267–276
- Blumenstein M, Ferrer MA, Vargas J (2010) The 4NSigComp2010 off-line signature verification competition: Scenario 2. In: 12th international conference on frontiers in handwriting recognition, IEEE, pp 721–726
- Bonde AS, Narwade P, Bonde SV (2022) Offline signature verification using gaussian weighting based tangent angle. In: 8th international conference on signal processing and communication (ICSC), pp 458–462
- Bouamra W, Djeddi C, Nini B et al (2018) Towards the design of an offline signature verifier based on a small number of genuine samples for training. *Expert Syst Appl* 107:182–195
- Chen C, Ross A (2021) An explainable attention-guided iris presentation attack detector. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 97–106
- Chen Xh, Champod C, Yang X et al (2018) Assessment of signature handwriting evidence via score-based likelihood ratio based on comparative measurement of relevant dynamic features. *Forensic Sci Int* 282:101–110
- Das A, Suwanwiwat H, Ferrer MA et al (2018) Thai automatic signature verification system employing textural features. *IET Biom* 7(6):615–627
- Das A, Suwanwiwat H, Pal U, et al (2020) ICFHR 2020 competition on short answer assessment and thai student signature and name components recognition and verification (SASIGCOM 2020). In: International conference on frontiers in handwriting recognition, pp 222–227
- Deviterne-Lapeyre M, Ibrahim S (2023) Interpol questioned documents review 2019–2022. *Forensic Sci Int Synergy* 6:100300
- Dey S, Dutta A, Toledo JI, et al (2017) Signet: convolutional siamese network for writer independent offline signature verification. arXiv preprint [arXiv:1707.02131](https://arxiv.org/abs/1707.02131)
- Diaz M, Ferrer MA, Impedovo D et al (2019) A perspective analysis of handwritten signature technology. *ACM Comput Surv (Csur)* 51(6):1–39
- Diaz M, Ferrer MA, Ramalingam S et al (2019) Investigating the common authorship of signatures by off-line automatic signature verification without the use of reference signatures. *IEEE Trans Inf Forensics Secur* 15:487–499
- Diaz M, Ferrer M, Alonso JB, et al (2021) One vs. one offline signature verification: a forensic handwriting examiners perspective. In: International Carnahan conference on security technology, pp 1–6
- Faundez-Zanuy M, Fierrez J, Ferrer MA et al (2020) Handwriting biometrics: applications and future trends in e-security and e-health. *Cogn Comput* 12:940–953
- Ferrer MA, Alonso JB, Travieso CM (2005) Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *IEEE Trans Pattern Anal Mach Intell* 27(6):993–997
- Ferrer MA, Vargas JF, Morales A et al (2012) Robustness of offline signature verification based on gray level features. *IEEE Trans Inf Forensics Secur* 7(3):966–977
- Ferrer MA, Diaz M, Carmona-Duarte C et al (2016) A behavioral handwriting model for static and dynamic signature synthesis. *IEEE Trans Pattern Anal Mach Intell* 39(6):1041–1053
- Foroozandeh A, Akbari Y, Jalili MJ et al (2012) A novel and practical system for verifying signatures on persian handwritten bank checks. *Int J Pattern Recognit Artif Intell* 26(06):1256014
- Galbally J, Diaz-Cabrera M, Ferrer MA et al (2015) On-line signature recognition through the combination of real dynamic data and synthetically generated static data. *Pattern Recogn* 48(9):2921–2934
- Ghosh R (2021) A recurrent neural network based deep learning model for offline signature verification and recognition system. *Expert Syst Appl* 168:114249
- Ghosh S, Ghosh S, Kumar P et al (2021) A novel spatio-temporal siamese network for 3d signature recognition. *Pattern Recogn Lett* 144:13–20
- Guerbai Y, Chibani Y, Hadjadj B (2015) The effective use of the one-class svm classifier for handwritten signature verification based on writer-independent parameters. *Pattern Recogn* 48(1):103–113
- Gunning D, Stefik M, Choi J et al (2019) XAI - explainable artificial intelligence. *Sci Robot* 4(37):eaay7120
- Hafemann LG, Sabourin R, Oliveira LS (2017) Learning features for offline handwritten signature verification using deep convolutional neural networks. *Pattern Recogn* 70:163–176
- Hafemann LG, Sabourin R, Oliveira LS (2019) Meta-learning for fast classifier adaptation to new users of signature verification systems. *IEEE Trans Inf Forensics Secur* 15:1735–1745
- He K, Zhang X, Ren S, et al (2016) Identity mappings in deep residual networks. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14, Springer, pp 630–645
- ISO Central Secretary (2020) Information technology - Biometric presentation attack detection - Part 4: Profile for testing of mobile devices. Standard ISO/IEC 30107-4:2020, International

- Organization for Standardization, Geneva, CH, <https://www.iso.org/standard/75301.html>
33. Jain A, Bolle R, Pankanti S (1999) Biometrics: personal identification in networked society, vol 479. Springer Science & Business Media, Berlin
 34. Jain AK, Deb D, Engelsma JJ (2021) Biometrics: trust, but verify. *IEEE Trans Biom Behav Identity Sci* 4(3):303–323
 35. Jiang J, Lai S, Jin L et al (2022) Forgery-free signature verification with stroke-aware cycle-consistent generative adversarial network. *Neurocomputing* 507:345–357
 36. Joshi I, Kothari R, Utkarsh A, et al (2021) Explainable fingerprint roi segmentation using monte carlo dropout. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 60–69
 37. Kalera MK, Srihari S, Xu A (2004) Offline signature verification and identification using distance statistics. *Int J Pattern Recognit Artif Intell* 18(07):1339–1360
 38. Lai S, Jin L (2018) Learning discriminative feature hierarchies for off-line signature verification. In: International conference on frontiers in handwriting recognition, pp 175–180
 39. Li B, Li N (2019) Handwriting expertise reliability: a review. *J Forensic Sci Med* 5(4):181–186
 40. Lipton ZC (2018) The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue* 16(3):31–57
 41. Liwicki M, Malik MI, Van Den Heuvel CE, et al (2011) Signature verification competition for online and offline skilled forgeries (SigComp2011). In: International conference on document analysis and recognition, pp 1480–1484
 42. Liwicki M, Malik MI, Alewijnse L, et al (2012) ICFHR 2012 competition on automatic forensic signature verification (4NSigComp 2012). In: International conference on frontiers in handwriting recognition, pp 823–828
 43. Longjam T, Kisku DR, Gupta P (2023) Writer independent handwritten signature verification on multi-scripted signatures using hybrid cnn-bilstm: a novel approach. *Expert Syst Appl* 214:119111
 44. Lund SP, Iyer HK (2017) Likelihood ratio as weight of forensic evidence: a closer look. *Journal of Research (NIST JRES)*, National Institute of Standards and Technology, pp 1–32
 45. Maergner P, Riesen K, Ingold R, et al (2017) A structural approach to offline signature verification using graph edit distance. In: 14th IAPR international conference on document analysis and recognition (ICDAR), IEEE, pp 1216–1222
 46. Maergner P, Pondenkandath V, Alberti M et al (2019) Combining graph edit distance and triplet networks for offline signature verification. *Pattern Recognit Lett* 125:527–533
 47. Malik MI (2015) Automatic signature verification: Bridging the gap between existing pattern recognition methods and forensic science. PhD thesis, Department of Computer Science Technische Universitat Kaiserslautern
 48. Malik MI, Liwicki M, et al (2013) ICDAR 2013 competitions on signature verification and writer identification for on-and offline skilled forgeries (SigWiComp 2013). In: 12th International conference on document analysis and recognition, pp 1477–1483
 49. Malik MI, Ahmed S, Marcelli A, et al (2015) ICDAR2015 competition on signature verification and writer identification for on-and off-line skilled forgeries (SigWiComp2015). In: International conference on document analysis and recognition, pp 1186–1190
 50. Marcelli A, Parziale A, De Stefano C (2015) Quantitative evaluation of features for forensic handwriting examination. In: 13th International conference on Document Analysis and Recognition, pp 1266–1271
 51. Martinez-Diaz M, Fierrez J, Ortega-Garcia J (2007) Universal background models for dynamic signature verification. In: International conference on biometrics: theory, applications, and systems, pp 1–6
 52. Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. *Artif Intell* 267:1–38
 53. Morales A, Morocho D, Fierrez J et al (2017) Signature authentication based on human intervention: performance and complementarity with automatic systems. *IET Biom* 6(4):307–315
 54. Okawa M (2017) Offline signature verification with vlad using fused kaze features from foreground and background signature images. In: International conference on document analysis and recognition, pp 1198–1203
 55. Okawa M (2018) From BoVW to VLAD with KAZE features: offline signature verification considering cognitive processes of forensic experts. *Pattern Recognit Lett* 113:75–82
 56. Oliveira LS, Justino E, Freitas C, et al (2005) The graphology applied to signature verification. In: 12th conference of the international graphonomics society, pp 286–290
 57. Ortega-Garcia J et al (2003) MCYT baseline corpus: a bimodal biometric database. *IEE Proc Vis Image Signal Process* 150(6):395–401
 58. Pan C, Huang J, Hao J et al (2020) Towards zero-shot learning generalization via a cosine distance loss. *Neurocomputing* 381:167–176
 59. Plamondon R, Lorette G (1989) Automatic signature verification and writer identification-the state of the art. *Pattern Recogn* 22(2):107–131
 60. RichardWebster B, et al (2018) Visual psychophysics for making face recognition algorithms more explainable. In: European conference on computer vision, pp 252–270
 61. Sae-Bae N, Memon N (2014) Online signature verification on mobile devices. *IEEE Trans Inf Forensics Secur* 9(6):933–947
 62. Selvaraju RR, Cogswell M, Das A, et al (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: International conference on computer vision, pp 618–626
 63. Serdouk Y, Nemmour H, Chibani Y (2017) Handwritten signature verification using the quad-tree histogram of templates and a support vector-based artificial immune classification. *Image Vis Comput* 66:26–35
 64. Shanker AP, Rajagopalan A (2007) Off-line signature verification using dtw. *Pattern Recognit Lett* 28(12):1407–1414
 65. Shariatmadari S, Emadi S, Akbari Y (2020) Nonlinear dynamics tools for offline signature verification using one-class gaussian process. *Int J Pattern Recognit Artif Intell* 34(01):2053001
 66. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. [arXiv preprint arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
 67. Soleimani A, Araabi BN, Fouladi K (2016) Deep multitask metric learning for offline signature verification. *Pattern Recognit Lett* 80:84–90
 68. Stewart L (2017) The process of forensic handwriting examinations. *Forensic Res Criminol Int J* 4(5):139–141
 69. Taroni F, Marquis R, Schmittbuhl M et al (2012) The use of the likelihood ratio for evaluative and investigative purposes in comparative forensic handwriting examination. *Forensic Sci Int* 214(1–3):189–194
 70. Tolosana R, Vera-Rodriguez R, Gonzalez-Garcia C, et al (2021) ICDAR 2021 competition on on-line signature verification. In: Document analysis and recognition–ICDAR 2021: 16th international conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV 16, Springer, pp 723–737
 71. Tsourounis D, Theodorakopoulos I, Zois EN et al (2022) From text to signatures: Knowledge transfer for efficient deep feature learning in offline signature verification. *Expert Syst Appl* 189:116136

72. Vargas F, Ferrer M, Travieso C, et al (2007) Off-line handwritten signature GPDS-960 corpus. In: 9th International conference on document analysis and recognition, pp 764–768
73. Williford JR, May BB, Byrne J (2020) Explainable face recognition. European conference on computer vision. Springer, Berlin, pp 248–263
74. Yapıcı MM, Tekerek A, Topaloğlu N (2021) Deep learning-based data augmentation method and signature verification system for offline handwritten signature. *Pattern Anal Appl* 24(1):165–179
75. Zeinali H, BabaAli B, Hadian H (2018) Online signature verification using i-vector representation. *IET Biom* 7(5):405–414
76. Zhu Y, Lai S, Li Z, et al (2020) Point-to-set similarity based deep metric learning for offline signature verification. In: *Int. Conf. on Frontiers in Handwriting Recognition*, pp 282–287
77. Zois EN, Alewijnse L, Economou G (2016) Offline signature verification and quality characterization using poset-oriented grid features. *Pattern Recogn* 54:162–177
78. Zois EN, Theodorakopoulos I, Economou G (2017) Offline handwritten signature modeling and verification based on archetypal analysis. In: *Proceedings of the IEEE international conference on computer vision*, pp 5514–5523

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.