IET Biometrics

ORIGINAL RESEARCH PAPER

Improving biometric recognition by means of score ratio, the likelihood ratio for non-probabilistic classifiers. A benchmarking study

Carlos Vivaracho-Pascual [ID] | Arancha Simon-Hurtado | Esperanza Manso-Martinez

Computer Science Department, University of Valladolid, Valladolid, Spain

Correspondence

Carlos Vivaracho-Pascual, Computer Science Department, University of Valladolid, Valladolid, Spain.
Email: cevp@infor.uva.es

Abstract

One of the ever present goals in biometrics research is to improve system performance. Herein, an alternative method is proposed that is independent of the biometric characteristic and the system, as this proposal, Score Ratio, is applied to the output (comparison score) of the classifier. The Likelihood Ratio is widely used with probabilistic classifiers because it performs well in these circumstances. However, when the classifiers are non-probabilistic, then this ratio is not used. This is our proposal: with non-probabilistic classifier based systems, the decision is taken solely through the score, supposing that the biometric feature, X, belongs to the Claimant (H_0 hypothesis), here, it is also proposed to make use of the score considering that X does not belong to the Claimant (H_1 hypothesis); more specifically, using the ratio between these two scores: the Score Ratio. For more objective results, benchmarking and reproducibility are used in the experiments, applying our proposal with third-party (benchmarking) experimental protocols, databases, classifiers and performance measures for fingerprint, iris and finger vein recognition. Statistically significant improvements have been obtained when the Score Ratio is used with regard to not using it in all cases tested.

1 | INTRODUCTION

Biometric recognition encompasses biometric verification and identification. Here, the focus is on verification, but the proposed technique could also be used for identification, as it is applied at comparison score (score in short from now on) level, to improve the discriminative capacity.

If we take a biometric feature (feature vector) X, the verification problem can be written in the following form (hypothesis test):

H_0: X is from Claimant H_1: X is not from Claimant.

The chosen hypothesis can be carried out as in equation (1), using Claimant (C) [1] information only, or with the likelihood ratio test shown in equation (2), in which 'Non-Claimant' (NC) information is included, with NC being any user other than the Claimant. $p(X/H_0)$ and $p(X/H_1)$ are, respectively, the probability density functions for hypotheses H_0 and H_1 at input instance X, while θ is the decision threshold.

$$p(X/H_0) \begin{cases} \geq \theta & Accept & H_0 \\ < \theta & Reject & H_0 \end{cases} \quad (1)$$

$$\frac{p(X/H_0)}{p(X/H_1)} \begin{cases} \geq \theta & Accept & H_0 \\ < \theta & Reject & H_0 \end{cases} \quad (2)$$

The problem is approached using pattern recognition. Under this approach, each Claimant C is represented by a biometric model or template λ_C [1], its output (score) $s(X/\lambda_C)$ being used to estimate (approximate) $p(X/H_0)$. With regard to the likelihood denominator ($p(X/H_1)$), the NC class is also represented by a model or template, $\lambda_{\bar{C}}$, using the score $s(X/\lambda_{\bar{C}})$ to estimate $p(X/H_1)$; in Section 3, the different alternatives found in the bibliography to model $\lambda_{\bar{C}}$, which is not an easy task, are shown.

From this pattern recognition approach, equation (1) transforms into equation (3), while equation (2) transforms into the **Score Ratio** that can be seen in equation (4).

$$s(X/\lambda_C) \left\{ \begin{array}{lll} \geq\theta & Accept & H_0 \\ <\theta & Reject & H_0 \end{array} \right) \qquad (3)$$

$$\frac{s(X/\lambda_C)}{s(X/\lambda_C^-)} \left\{ \begin{array}{lll} \geq\theta & Accept & H_0 \\ <\theta & Reject & H_0 \end{array} \right) \qquad (4)$$

With probability-based classifiers, such as Gaussian Mixture Models or Hidden Markov Models, the system output $s(X/\lambda_C)$ is a probability, $p(X/\lambda_C)$. With these classifiers, the decision is generally carried out using the likelihood ratio $\frac{p(X/\lambda_C)}{p(X/\lambda_C^-)}$ test [2–4], as a better performance is obtained. Nevertheless, this Score Ratio (equation (4)) has not been used with non-probabilistic classifiers, to the best of our knowledge. Based on this idea, in a previous work [5], we proposed also extending the ratio to this type of classifiers.

That work, [5], focused principally on describing the new proposal, Score Ratio, and carrying out a series of exploratory experiments, showing the promising advantages of using our Score Ratio in terms of performance improvements. Here, we approach the application of the proposal in a broad and detailed study, proving the advantages of using the Score Ratio with respect to not using it, that is, with regard to using, as is usual, only $s(X/\lambda_C)$ to take the decision (equation (3)). The goal of this work is to show that the Score Ratio is a real alternative to improve the performance of biometric systems.

In Figure 1(b), our Score Ratio proposal is shown graphically, and compared with a habitual system (Figure 1(a)) where $s(X/\lambda_C)$ is solely used to take the decision.

Several biometric characteristics have been tested to broaden the study; fingerprint and iris, because they are two of the most mature biometrics, and finger veins, which is an emergent and very promising biometric technology. Since a behavioural trait, that is, signature, was approached in Ref. [5], we decided to focus here on the other type of biometric characteristics, the static or biological ones. To test the Score Ratio performance in the most objective way,
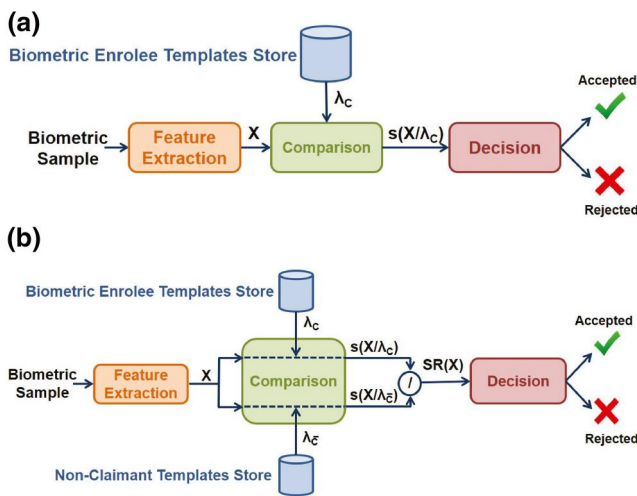
benchmark and public databases and systems are used, following reference experimental and performance evaluation protocols [6]. Our own systems have only been used with the finger vein biometric characteristic to test under different conditions (Section 8). All of the main scripts and configuration files used in our experiments are publicly available (the download links are shown in the corresponding sections) to guarantee the reproducibility of the experiments.

Focusing on the classifiers, the study is extended to non-probabilistic ones in general, including but not limited to the distance-based ones. The difference is the score (classifier output) interpretation. In distance-based classifiers, $s(X/\lambda_{C_i}) < s(X/\lambda_{C_j})$ means that the input X is closer to the C_i Claimant than the C_j Claimant. Nevertheless, there are non-probabilistic classifiers, as will be shown, where $s(X/\lambda_{C_i}) < s(X/\lambda_{C_j})$ means just the opposite, that is, the input X is closer to the C_j Claimant than the C_i Claimant. We refer to the latter as *probability-like classifiers*. For these, the decision is taken as in equations (3) and (4). However, for distance-based ones, the decision must be changed, as shown in equations (5) and (6).

$$s(X/\lambda_C) \left\{ \begin{array}{lll} \leq\theta & Accept & H_0 \\ >\theta & Reject & H_0 \end{array} \right) \qquad (5)$$

$$\frac{s(X/\lambda_C)}{s(X/\lambda_C^-)} \left\{ \begin{array}{lll} \leq\theta & Accept & H_0 \\ >\theta & Reject & H_0 \end{array} \right) \qquad (6)$$

Distance-based classifiers are tested with iris and finger vein. Probability-like classifiers are tested with finger vein and fingerprint.

Another point of variability tested is the way to calculate the Equal Error Rate (EER), used as the performance measure (Section 5.3). This measure can be calculated with a different threshold for each biometric data subject or with the same for all; in the latter case, the use of score normalisation (Section 5.1) usually improves the results. Following the reference performance evaluations, the general or common threshold has been used with fingerprint and iris; while the individual threshold is the one tested with the finger vein. With the common threshold, the performance of Score Ratio has been tested both with and without the use of score normalisation (Section 5.1.2).

Finally, the statistical significance of the improvements has been measured. This is very important to confirm that the improvement is due to Score Ratio and not simply to chance.

The content is organised as follows. Section 2 provides the notation so as to facilitate the reading of the work. Section 3 shows the theoretical background of our Score Ratio proposal and the related works, followed by its description in Section 4. Following the description of the general experimental setup in Section 5, a more specific description of each biometric system, the experimental protocol, the results and their analysis can be seen in Sections 6–8. The conclusions are shown in Section 9.

FIGURE 1 Main elements in a biometric system, without Score Ratio (a) and with Score Ratio (b)

2 | NOTATION

First of all, we shall fix the notation and terminology to be used, herein, to make it easier to read. With regard to biometrics, the standard vocabulary has been followed (ISO/IEC 2382-37:2017).

- **C** is used to refer to the Claimant in general, that is, to any biometric data subject (subject in short from now).
- $\mathbf{C_i}$ is a specific Claimant, that is, a specific subject.
- **Cohort Set, ChS.** Individuals contributing biometric data, who are not subjects, and who are utilised to obtain $s(X/\lambda_{\bar{C}})$, that is, the model or template for the Non-Claimant class.
- $\mathbf{Ch_i}$ is each element of the prior set: $ChS = \{Ch_1, Ch_2, …, Ch_H\}$, where **H** is the size of the set.
- **N**: the number of components of the cohort set ($N \le H$) that are used to estimate $s(X/\lambda_{\bar{C}})$.
- **M**: the number of a priori selected components ($M << H$) from the ChS to obtain $s(X/\lambda_{\bar{C}})$, with the goal of reducing the computing load of the Score Ratio usage (Sec. 4.2).
- **Normalisation Set, NrS.** The set of individuals contributing biometric data, who are not subjects, used for score normalisation (Sec. 5.1).
- $\mathbf{Nr_i}$ is each of the elements of the normalisation set: $NrS = \{Nr_1, Nr_2, …, Nr_R\}$, where **R** is the number of elements of this set.
- **E**. Any element of the biometric database. The Claimant C, the ChS elements and the NrS elements are all biometric data records. All come from the same database, but the task or role for which they are used is different. We then need a general identifier: E (Element).
- λ_E is the biometric model or biometric template for E. λ_C, λ_{C_i}, λ_{Ch_i} and λ_{Nr_i} are particular examples of λ_E.
- **X** is used to refer to the biometric feature. It is a feature vector $X = \{x_1, x_2, …, x_Q\}$, extracted from the biometric sample (Figure 1), x_k being the k component (which can also be a vector) and Q being the number of them. This vector is used for comparison.

3 | THEORETICAL BACKGROUND AND RELATED WORKS

Likelihood rate is a concept closely linked to the statistical test, where the problem to solve could be stated as shown in the following.

Suppose that $p(X/H_0)$ and $p(X/H_1)$ are the two possible distributions that X can have. This defines two Hypotheses:

H_0: *X has the probability density function* $p(X/H_0)$
H_1: *X has the probability density function* $p(X/H_1)$

From a statistical approach, the test can be built on the basis of a straightforward idea: if $X = X_o$ is observed, where X_o is an X observation, $p(X_o/H_1) > p(X_o/H_0)$ is evidence supporting option H_1, otherwise, the evidence is in favour of H_0.

The likelihood ratio function $L(X)$ is determined, from the previous distributions, as can be seen in equation (7). The statistical test can now be constructed since: if $L(X_o)$ has a small value, then this is evidence supporting H_1, while if $L(X_o)$ has high values, it is evidence supporting H_0. So, it is reasonable to utilise $L(X)$ to take the decision about which Hypothesis to select, using a threshold value θ, such that the hypothesis H_0 will be rejected if and only if $L(X_o) < \theta$. θ is estimated with the significance level $\alpha = p(L(X) < \theta/H_0)$, that is, fixing the value of false-negative probability.

$$L(X) = \frac{p(X/H_0)}{p(X/H_1)} \qquad (7)$$

Having determined the value of θ, $p(L > \theta/H_1)$, which is the false-positive probability, can be calculated. Neyman–Pearson [7,8] showed that the use of the likelihood ratio minimises the false-positive probability, which makes the above test more powerful.

There may sometimes be a set of distributions, $\{p_k(X/H_0)\}$ and/or $\{p_i(X/H_1)\}$, for Hypotheses H_0 and/or H_1, respectively, instead of a single one, as in the previous. This happens in our problem. Neyman–Pearson extended the likelihood ratio function, as can be seen in equation (8).

$$L(X) = \frac{\max_k(p_k(X/H_0))}{\max_i(p_i(X/H_1))} \qquad (8)$$

Focusing on Biometrics, the input value X must be classified as coming or not from the Target Class or Claimant C; so, the initial Hypotheses become:

H_0: *X is from C* H_1: *X is not from C*.

In practice, $p(X/H_0)$ is estimated through the output, $p(X/\lambda_C)$, of a statistical model, λ_C, (such as GMM or HMM) of C; while $p(X/H_1)$ is estimated through a model of the Non-Target Class (Non-Claimant, NC, or biometric impostor in biometrics terms), $p(X/\lambda_{\bar{C}})$. The problem is how to calculate or estimate this last probability function.

Two methods to obtain this likelihood can be seen in the literature: with a cohort set or representative set of the NC class [2,4,9], or using a model to capture the behaviour of the NC class [3,9].

When the second approximation is utilised to obtain $p(X/\lambda_{\bar{C}})$, the model is achieved by means of numerous examples of NCs, that is, data from many individuals different from the Claimant C. One example is the Universal Background Model, UBM, [3,9]. In this case, the Score Ratio is carried out using the log likelihood: $\log \frac{p(X/\lambda_C)}{p(X/UBM)}$, with λ_C being attained by means of the adaptation of the UBM with the biometric enrolment data records of the Claimant.

With the first approach, the cohort set, $p(X/\lambda_{\bar{C}})$, is estimated using a composite hypothesis, that is, using a set of probability functions $\{p(X/\lambda_{Ch_i})\}$. The Neyman–Pearson approach gives the solution (Eq. (8)). Nevertheless, in practice, instead of using the maximum one only, in general, the N

maximum probabilities are used [2,4,9], as a better performance is thus achieved. The reason is that to use the $N > 1$ ChS elements nearest to the claimant can approximate, in general, λc better than if only the closest one, $N = 1$, is used; equation (9) shows the likelihood ratio function [4,10], where $1 \le N < H$ and $\{p(X/\lambda_{Ch_k})\}= \{N\max_i(p(X/\lambda_{Ch_i}))\}\ 1 \le i \le H$.

$$L(X) = \frac{p(X/\lambda_C)}{p(X/\lambda_{\bar{C}})} = \frac{p(X/\lambda_C)}{\frac{\sum_{k=1}^{N} p(X/\lambda_{Ch_k})}{N}} \tag{9}$$

4 | SCORE RATIO

4.1 | General approach

When non-probabilistic classifiers are used, the problem with Score Ratio (Figure 1 (b)) is to estimate $s(X/\lambda_{\bar{C}})$ in equations (4) or (6). From the approaches seen in the previous section to obtain $p(X/\lambda_{\bar{C}})$, only the cohort set one can generally be applied, since many of the systems that use non-probabilistic classifiers are based on biometric templates, and in this case, it is not possible to obtain a single impostor model.

So, following equation (9), our Score Ratio proposal can be seen in equation (10). Here, the identical strategy of utilising the N ChS components nearest to the Claimant is used, but with scores. Given a biometric probe X, the selection of these components to apply the score ratio in equation (10) is:

- For *distance-based classifiers*: since low scores imply high similarity, those ChS elements with the minimal scores are selected ($\{s(X/\lambda_{Ch_k})\} = \{N\min_i(s(X/\lambda_{Ch_i}))\ 1 \le i \le H$).
- For *probability-like classifiers*: since high scores imply high similarity, those ChS elements with the maximal scores are selected ($\{s(X/\lambda_{Ch_k})\}= \{N\max_i\ (s(X/\lambda_{Ch_i}))\}\ 1 \le i \le H$).

We call this proposal *Score Ratio Basic Approach, SRBA*.

$$SRBA(X) = \frac{s(X/\lambda_C)}{s(X/\lambda_{\bar{C}})} = \frac{s(X/\lambda_C)}{\frac{\sum_{k=1}^{N} s(X/\lambda_{Ch_k})}{N}} \tag{10}$$

In the experiments, different values of H (the size of the ChS), and N are tested. The exact values are specified in each case.

4.2 | Reduced calculations: A Priori Cohort Selection

Given a feature vector X to classify, the application of Score Ratio (equation (10)) supposes that H (the cohort set size) additional scores must be calculated ($\{s(X/\lambda_{Ch_k})\}\ 1 \le k \le H$) with regard to using only $s(X/\lambda_C)$ (equation (10)

numerator). In addition, these scores must be ordered to select the N closest ones, though the computing load of this operation is in fact negligible because of the ChS sizes tested.

An analysis of the cost in time for those extra calculations was carried out in Ref. [5] to study whether this prevents a real-time system response; the conclusion is that this depends on the system. However, as improving the response time is interesting in all cases, an alternative to decrease the computing load to estimate $s(X/\lambda_{\bar{C}})$ is proposed and tested.

The method proposed is based on reducing the number of elements of the cohort set from which the N closest ones are selected, to calculate equation (10) denominator. To achieve this, $M << H$ elements of the ChS closest to the Claimant are first selected, and then, the score ratio is performed using this subset instead of the entire cohort set.

The similarity between a certain Claimant C and each element of the ChS Ch_i, $s(Ch_i, C)$, is accomplished by means of their respective models or templates, $s(Ch_i, C) =s(\lambda_{Ch_i}, \lambda_C)$. Using these scores, the M elements of the cohort set, $\{Ch_v\}$, closest to C are selected as follows:

- For distance-based classifiers: $\{Ch_v\} = \{M\min_i(s(Ch_i, C))\}\ 1 \le i \le H$.
- For probability-like classifiers: $\{Ch_v\} = \{M\max_i(s(Ch_i, C))\}\ 1 \le i \le H$.

The Score Ratio is carried out, as shown in equation (10), yet now, only the preselected subset of elements from the cohort set $\{Ch_v\}\ 1 \le v \le M$ is used.

The $s(Ch_i, C)$ calculation is system dependent. We can use an example to provide a better understanding of this. Let us suppose that we use a distance-based classifier (the generalisation for systems based on probability-like classifiers is immediate), and each subject of the biometric database, E, is modelled using a T size template ($\lambda_E = \{X_E^1, X_E^2, ..., X_E^T\}$), where X_E^j is the feature vector extracted from the j biometric enrolment data record of E. Then, for a Claimant C, his/her template will be $\lambda_C = \{X_C^1, X_C^2, ..., X_C^T\}$, and given a biometric feature X, $s(X/\lambda_C) =f(d(X/X_C^j))\ with\ 1 \le j \le T$, with $d(X/X_C^j)$ being the distance between X and $X_C^j f()$ will be a mixture function of the T distances calculated, such as, for example, min, max, $mean$, $\sum$, etc. Therefore, $s(Ch_i, C) = s(\lambda_{Ch_i}, \lambda_C) = f(s(X_{Ch_i}^k/\lambda_C))\ with\ 1 \le k \le T$.

We call this proposal *Score Ratio a Priori Cohort Selection, SRaPCS*.

4.3 | Score Ratio in operation

We believe it is interesting, even though the results are set out later, to show graphically how Score Ratio operates over the Match and Non-Match distributions to achieve improvements in the system performance. This can give a better understanding of the proposal.

For this study, we use a representative and statistically significant example of the tests performed. In particular, the example was extracted from the tests performed in fingerprint recognition (Section 6). The classifier is probability-like type.

Match and Non-Match distributions of the example, with and without Score Ratio, are shown in Figure 2(a). As can be seen, the Score Ratio produces a narrowing in the score distributions (Figure 2 (b) and (c)). In addition, the Non-Match distribution with Score Ratio is moved to the left with respect to the distribution without Score Ratio (Figure 2(c)). This displacement, though small in the figure, is statistically highly significant (the Mann–Whitney–Wilcoxon test between Non-Match distributions with and without Score Ratio gives a p-value $< 2.2 \times 10^{-16}$) and the system improvement achieved is, as will be seen, also important. As the distributions are not Gaussian, the Mann–Whitney–Wilcoxon test (using R software) has been used to measure whether the difference between distributions is statistically significant; if the p-value > 0.05, the null hypothesis (H_0: the population distributions are identical) is not rejected, and rejected otherwise.

In the end, Score Ratio decreases the overlap between distributions (Figure 2 (d)), thus improving the results. This can be seen in Figure 3, where the system performance with and without Score Ratio is shown by means of a Detection error trade-off (DET) curve [11] for a more complete comparison. The Score Ratio improves the system whatever the chosen threshold, the improvement being of at least 15%.

5 | EXPERIMENTAL ENVIRONMENT

5.1 | Score normalisation

It is usual in biometrics that the Match and Non-Match distributions of a system vary from one subject to another (Figure 4a). To avoid this, if necessary, score normalisation must be performed, since with this technique, the scores of the Claimant matchers are transformed into a common domain (Figure 4b).

Several score normalisation techniques [5] have been proposed, but here, due to the number of samples used to build
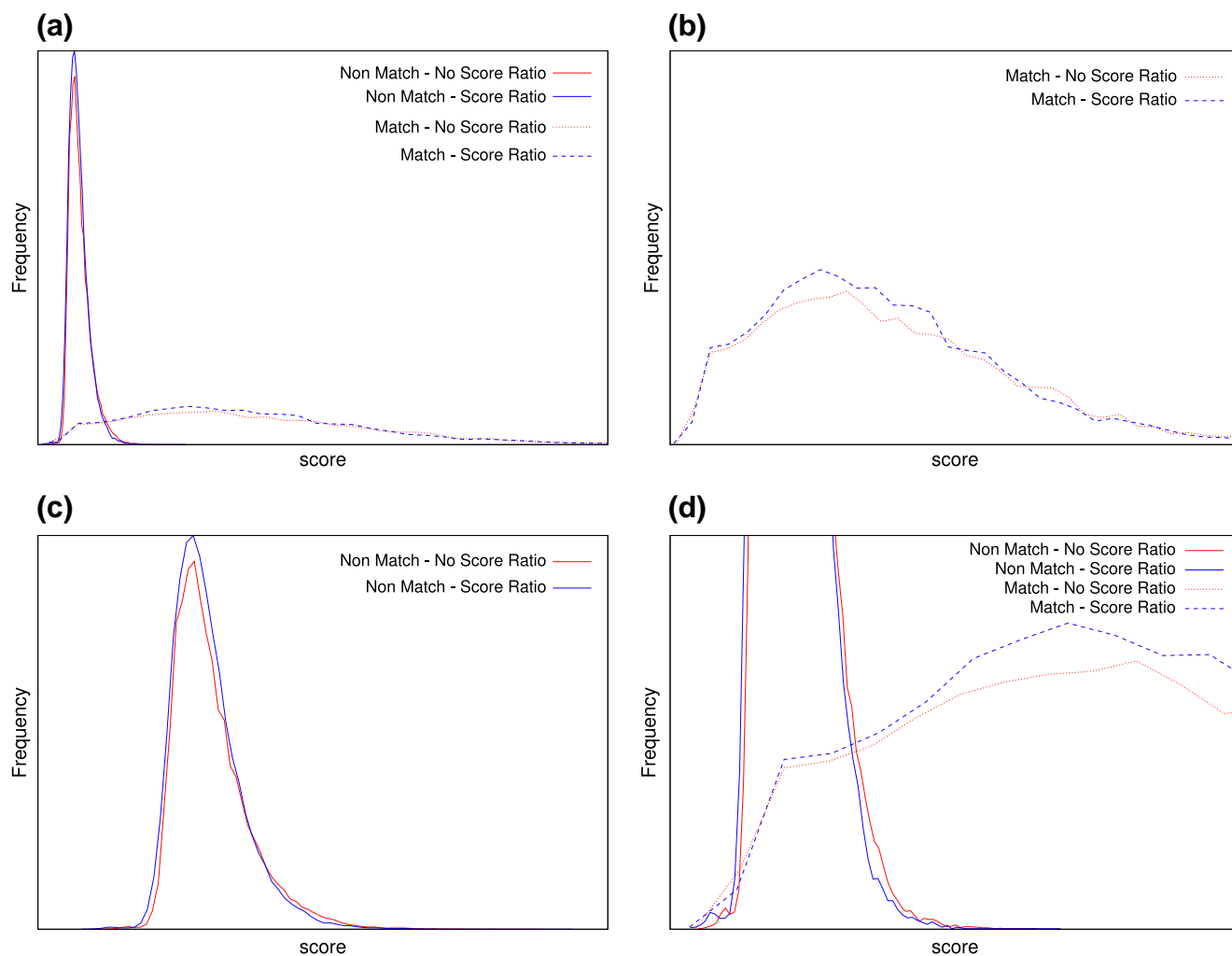
FIGURE 2 Example of Match and Non-Match distributions with and without Score Ratio (a). Match and Non-Match distributions are only shown in (b) and (c), respectively. The intersection zone is zoomed in on figure (d)

the Claimant template (one in iris and fingerprint), only Impostor-Centric techniques can be used. From these, ZNorm is one of the most commonly used in the literature: $s_{\{norm\}} = (s - \widehat{\mu}_C^{\{Non\}})/\widehat{\mu}_C^{\{Non\}}$, where $\widehat{\mu}_C^{\{Non\}}$ and $\widehat{\mu}_C^{\{Non\}}$ are the mean and the standard deviation of the Non-Match distribution for the Claimant C classifier, estimated using the *Normalisation Set*, NrS (Section 5.2), as shown below. This normalisation technique is the one tested here.

5.1.1 | Score normalisation without Score Ratio

s in the ZNorm equation is: $s = s(X/\lambda_C)$, if Score Ratio is not applied.

The Non-Match distribution is estimated as follows. From each subject Nr_i of the NrS, a biometric sample (sample in short from now on) is randomly selected, extracting from it the
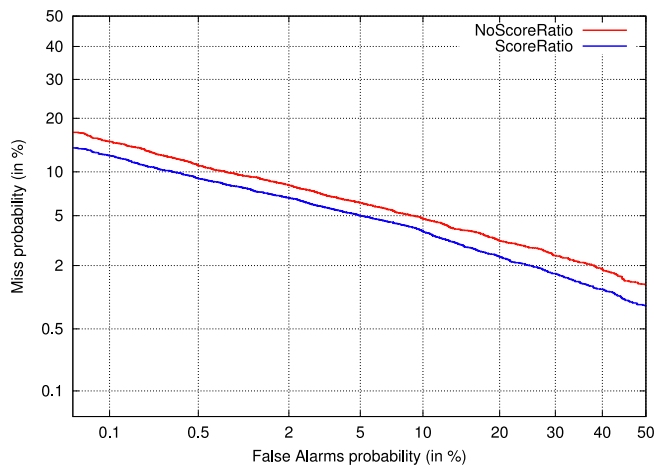
feature vector X_{Nr_i}, thus forming the so-called *Normalisation Gallery*, $NrG = \{X_{Nr_i}\}\ 1 \le i \le R$, where R is, let us remember, the Normalisation Set size. For a Claimant C, the score for each element in NrG is obtained, achieving the Non-Match Score Set, $NMSS = \{s(X_{Nr_i}/\lambda_C)\}\ 1 \le i \le R$. This set is an a priori estimation of the Non-Match distribution for Claimant C, that is, the impostor score distribution estimation for this Claimant. Then, the mean and standard deviation of NMSS are used to approximate $\widehat{\mu}_C^{\{Non\}}$ and $\widehat{\mu}_C^{\{Non\}}$.

5.1.2 | Score normalisation with Score Ratio

Now, $s = \frac{s(X/\lambda_C)}{s(X/\lambda_C^-)}$ in the ZNorm equation. Then, it is necessary to apply the Score Ratio to the sets utilised to estimate the statistics.

So, NMSS becomes the NMSS_Score Ratio, $NMSS\_SR = \{\frac{s(X_{Nr_i}/\lambda_C)}{s(X_{Nr_i}/\lambda_C^-)}\}\ 1 \le i \le R$. From $NMSS\_SR$, $\widehat{\mu}_C^{\{Non\}}$ and $\widehat{\mu}_C^{\{Non\}}$ are estimated.

5.2 | Experimental sets

Each biometric database was divided into the following subsets:

- **Normalisation Set, NrS.** When score normalisation must be accomplished (for fingerprint and iris, as will be seen), this set is randomly selected from the corpus.
- **Cohort Set, ChS.** The components of this set are randomly selected from the users in the database not utilised for the NrS (fingerprint and iris) or from the complete corpus (finger vein). The effects of the size of this set on the Score Ratio performance were studied, testing different sizes.

FIGURE 3 Example of system performance with and without Score Ratio by means of a DET curve
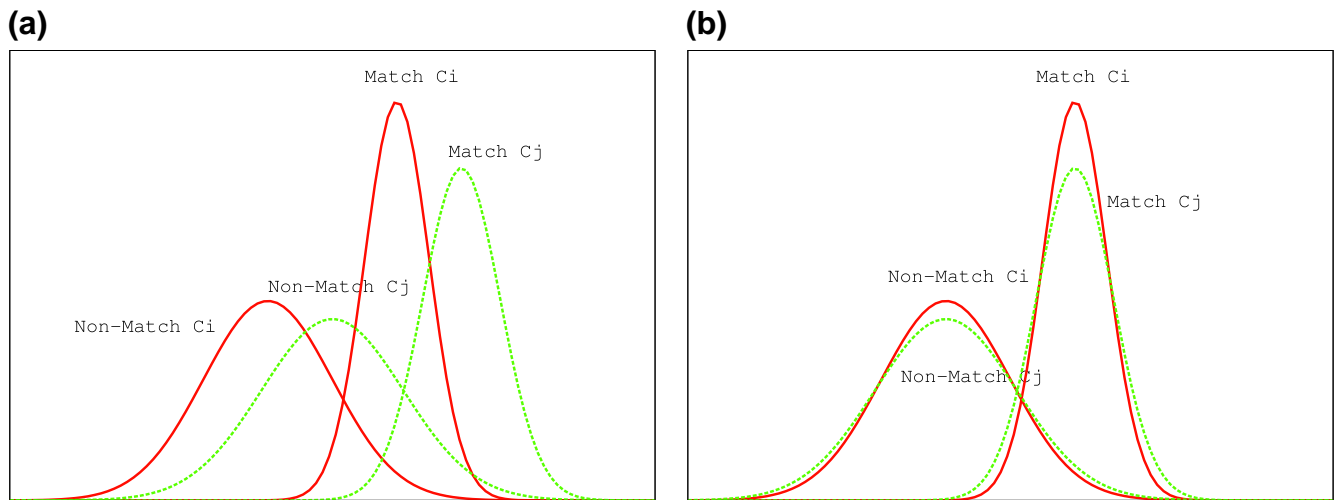

FIGURE 4 Visualisation of the score normalisation effect. The Match and Non-Match distributions of two subjects, initially in different domains (a), are moved to the same with its application (b)

- **Test Set (TS)**, used for testing, consists of the users not included in the other sets. For more objective results, the same set was used in all the tests carried out.

The size or sizes of the previous sets are specified in the next sections, where each biometric system and the results are shown.

5.3 | Performance measure

Performance measure can be accomplished by means of a graphical representation, such as a DET plot [11] or a ROC (Receiver Operator Characteristic) curve, or by means of a measure based on a single number. This last is easier to handle and simpler to understand when the amount of comparison is high, which is the case here; so, this is the measure selected in the experiments. More specifically, the EER, one of the most commonly used in the biometric bibliography, is the one used here.

To achieve the final EER of the test, two approaches can be found in the literature:

- An **individual EER** is calculated for each Claimant in the TS, the final EER being the mean of these individual EERs. Under this approach, also called *with individual threshold*, score normalisation (Section 5.1) is not necessary.
- A **global EER** is calculated using a Claimant scores set and another with the impostors' scores, created by joining all the genuine and impostor test results, respectively. Under this approach, also called *with global threshold*, score normalisation (Section 5.1) is usually necessary.

Although the second approach is the most usual, here, both are addressed.

5.4 | Statistical significance measure

We consider it is important to show the statistical significance of the results achieved. More specifically, if the improvements or the worsening in the system performance (in our case, the difference between the result with and without Score Ratio) are significant or not, that is, if they are 'real' or simply due to chance.

The statistical confidence of the performance estimation is not a straightforward problem in biometrics [12]. Several approximations have been proposed; for example, we can find two rules that approach the relation between the confidence bounds and the test size: the *Rule of 3* [12] and the *Rule of 30* [13]. We can also find the estimation of the confidence bounds on the observed error rates [12]. However, all of these proposals are based on one or more of the following approximations and are not always true in biometrics:

- Independent trials. This is not true if multiple samples per person are used in the tests, which is the usual.

- Error equally distributed among classes. This is not true due to the so-called *biometric menagerie*, described first by Doddington in speaker recognition [14], and noticed in other biometrics [15].
- The observed error rates follow a Gaussian (or normal) distribution.

To avoid these problems, the use of the Bootstrap non-parametric technique is proposed in Ref. [12] and incorporated in the ISO/IEC 19795-1:2006 standard about *Biometric performance testing and reporting, Part 1: Principles and framework*. The advantage of the bootstrap estimation is that it reduces the need to make assumptions about the underlying distribution of the observed error rates and the dependencies between attempts.

Following this technique, a bootstrap test set is created by sampling with replacement from the original test set. The original test set, as will be shown in the next sections, is composed of S subjects, each having G biometric mated comparison trials (historically referred to as 'genuine trials') and I biometric non-mated comparison trials ('impostor trials'). Each bootstrap test set is constructed from the original one in a such way that it replicates the structure and dependencies of this set:

1. S subjects, $\{Cb_i\}$, are sampled with replacement. Sampling with replacement means the list is likely to contain more than one occurrence of the same item.
2. For each Cb_i, G genuine trials are sampled with replacement.
3. For each Cb_i, I impostor trials are sampled with replacement.

Many bootstrap test sets are generated and the EER is calculated for each. The distribution of the bootstrap EER values is used to approximate that of the observed EER. Following Ref. [12], 1000 bootstrap test sets have been created in each experiment to get 95% confidence in the statistical calculations.

The statistical significance of the difference between the system performance with and without Score Ratio is evaluated as follows, in each of the experiments shown in the next sections:

1. By means of the original test set, the EER of the experiment is calculated with and without Score Ratio.
2. 1000 different bootstrap test sets are obtained and the EER without Score Ratio is calculated for each.
3. Another 1000 different bootstrap test sets are obtained and the EER with Score Ratio is calculated for each.
4. Figure 5 shows a typical distribution of the EERs calculated in the previous two steps. It can be seen that the distribution fits a Gaussian one. The mean of those distributions are, very approximately, the corresponding EERs achieved with the original test sets. For example, the EERs achieved with and without Score Ratio with the original test set of Figure 5 are 2.66% and 3.13%, respectively; while the corresponding means of the bootstrap sets are 2.65% and 3.11%.
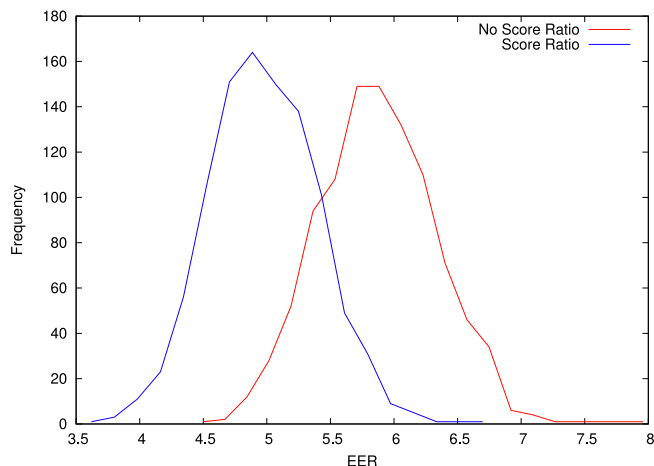
FIGURE 5 Bootstrap test sets EER distribution with and without Score Ratio

Then, under the conditions shown, the t-test can be applied to the EER distributions achieved by means of the bootstrap test sets to determine whether the difference between the results with and without Score Ratio is statistically significant or not. R software has been used for this test. If the p-value of the t-test is greater than 0.05 with a 95% confidence, we can say that the difference in the performance is not significant (H_0 is not rejected), being significant otherwise. The lower the p-value, the more significant that difference will be.

6 | FINGERPRINT

Fingerprint is the biometric characteristic that has been used for the longest: modern fingerprint identification methods were provided at the end of nineteenth century. It is one of the most well-known biometrics and is by far the most important technology in the biometrics market. As a very mature mode, well established reference databases, experimental protocols and systems can be found [6], and which are used in our work, as can be seen in the next sections. The main scripts and configuration files of fingerprint experiments are available in http://www.infor.uva.es/cevp/Download/Fingerprint.zip.

6.1 | Experimental setup

The MCYT biometric database [16] has been used. This corpus is very popular in fingerprint recognition and can be considered a benchmark [6]. Two types of acquisition devices are used: CMOS-based capacitive and optical. For each individual in the database, 12 different samples of each fingerprint were acquired. For the tests, each finger is considered a different subject [6] that is a different Claimant. Then, we have 12 impressions (fingerprint images) for each subject per sensor.

The acquisition control is accomplished in three levels:

- Three samples with low level of control: the individual puts his/her finger on the screen sensor without any position restrictions.
- Three more samples with medium level of control: in this stage, the individual him/herself must observe the fingerprint in a computer screen while the finger is located on the sensor.
- Six more samples with high level of control: the acquisition is accomplished as in the above stage, but with more control in the finger position.

The description of each subset in Section 5.2 and the test performed is:

- **Normalisation Set, NrS.** Ten individuals were randomly selected from the database. From each individual, one sample of each fingerprint was selected (the same that was used as the Claimant template). The size of this set is, then: 10 individuals × 10 print/individual × 1 sample/print = 100 samples, for each device.
- **Cohort Set, ChS.** Sizes of 5, 10, 15 and 20 individuals were tested. From each individual his/her 10 fingerprints were used. Then, in fact, the sizes of this set tested are 50, 100, 150 and 200 subjects, respectively.

Here, besides the size of the ChS, the acquisition control level was also tested. So, we have three ChSs for each ChS size ($H = \{50, 100, 150, 200\}$) tested: one consisted of fingerprints with low control and the other two with middle- and high-control-level fingerprints, respectively.

- **Test Set (TS).** Eighty-three individuals were used. Thus, the size of this set is: 83 individuals × 10 fingers/individual = 830 fingers (subjects or Claimants). Following the reference protocol [6], one impression per finger with low control during acquisition is used as template, this being the first image acquired. The rest of the fingerprint samples are used for biometric mated comparison trials. For non-mated ones, one impression per finger (that used for the template) of the rest of the individuals in the TS different from the Claimant are used. That is, we have 9130 ($83 \times 10 \times 11$) genuine trials and 68,060 ($83 \times 82 \times 10$) impostor trials.

6.2 | Recognition system

Both here and in the next biometric characteristics, we shall not give a detailed description of the biometric systems. Our interest focuses on the Score Ratio usage, so only the important parts of the system are described for a better understanding of the work performed. A more indepth description of the systems can be found in the bibliography.
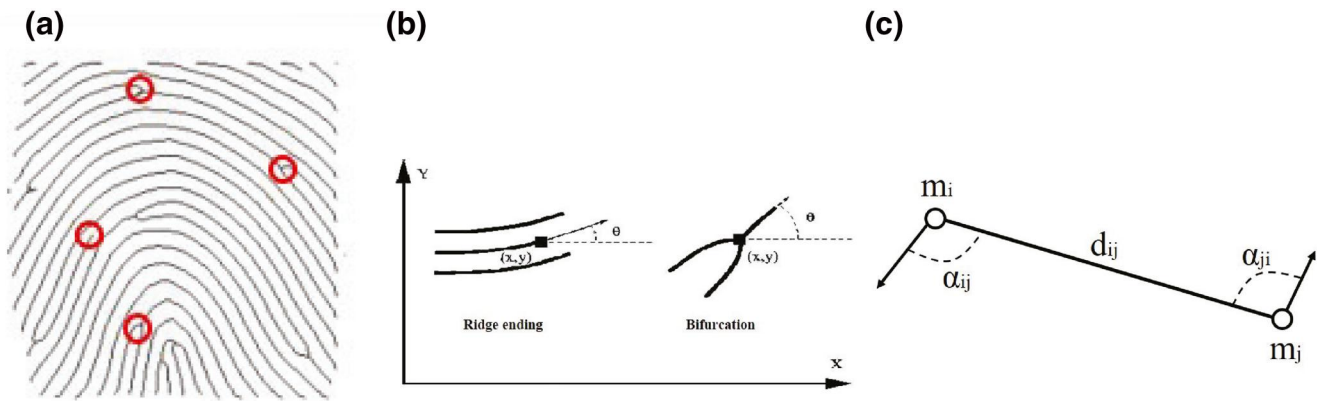
FIGURE 6 Example of fingerprint minutiae (a), the features extracted from each (b) and the invariant measures extracted from two minutiae (c)

TABLE 1 Fingerprint (optical device) recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio *Basic Approach*, and with and without score normalisation. In 'ChS-*H*', *H* means the cohort set size. This size does not affect the results without Score Ratio (ChS is only used to apply Score Ratio, $N \neq 0$ rows), so, for a better comparison, the EER of the $N = 0$ row has been repeated in all *ChS-H* columns. The ChS fingerprints control level appears in the *CL* column: low (*L*), middle (*M*) and high (*H*) (as was shown in the benchmarking protocol, when Score Ratio is not used, $N = 0$ row, CL is not applicable). The cell colour code used is: light grey for no significant differences between results with and without Score Ratio, dark grey for no improvements with Score Ratio, and normal (white) when Score Ratio significantly improves the reference system (EER in $N = 0$ row). Best results are in boldface

		No Score Normalisation				Score Normalisation			
N	CL	ChS-50	ChS-100	ChS-150	ChS-200	ChS-50	ChS-100	ChS-150	ChS-200
0	-	3.57%	3.57%	3.57%	3.57%	3.13%	3.13%	3.13%	3.13%
1	L	3.21%	3.10%	3.09%	3.09%	2.99%	2.99%	2.93%	2.92%
	M	3.22%	3.10%	3.06%	3.08%	3.10%	3.10%	2.85%	2.84%
	H	3.28%	3.13%	3.04%	3.21%	2.92%	2.92%	2.87%	2.95%
3	L	2.97%	2.96%	2.94%	3.01%	2.85%	2.85%	2.85%	2.80%
	M	3.00%	2.90%	2.86%	3.03%	2.89%	2.89%	2.70%	2.73%
	H	3.15%	2.97%	2.92%	3.00%	2.71%	2.71%	2.65%	2.73%
5	L	2.91%	2.90%	2.95%	2.96%	2.79%	2.79%	2.83%	2.75%
	M	2.96%	2.87%	2.88%	2.94%	2.84%	2.84%	2.67%	2.68%
	H	2.99%	2.97%	2.86%	2.93%	2.68%	2.68%	2.69%	2.65%
10	L	2.93%	2.92%	2.89%	2.91%	2.72%	2.72%	2.77%	2.76%
	M	2.86%	2.87%	**2.85%**	2.89%	2.72%	2.72%	2.65%	2.65%
	H	2.90%	2.88%	2.86%	2.86%	2.63%	2.63%	2.65%	**2.61%**

The public reference system NBIS[1] (NIST Biometric Image Software), release 5.0.0, has been used for fingerprint recognition. The main system characteristics are:

- Feature extraction stage based on minutiae (ridge ending and bifurcations, Figure 6) detection using the MINDTCT package.
- From the minutiae extracted, the similarity between two fingerprints is measured by means of the very well-known matching algorithm BOZORTH3 [17]. This algorithm uses invariant measurements, such as the distance between two minutiae or the angle between each minutia's orientation and the intervening line between both minutiae (Figure 6). The score achieved is the type probability-like that is the higher the score the more similar the fingerprints are.

Following the benchmark protocol [6], the *global EER* approach (Section 5.3) is used here to measure system performance.

6.3 | Results with optical device

Here, the results (EER) achieved with the optical device are shown. In Table 1, those with the *Score Ratio Basic Approach* are shown, both when score normalisation is applied and when it is not. The results with the *Score Ratio a Priori Cohort Selection* can be seen in Table 2.

6.4 | Results with capacitive device

The results with the capacitive device using the *Score Ratio Basic Approach* are shown in Table 3, and those with the *Score Ratio a Priori Cohort Selection* can be seen in Table 4. The

TABLE 2 Fingerprint (optical device) recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio with *a Priori Cohort Selection*. ChS-200 is used to select the a priori M elements closest to the Claimant. If $N > M$, then the denominator in the Score Ratio equation (equation (10)) cannot be calculated, so these rows are empty. The remaining columns and rows, as well as cell colour code, are the same as in the previous table

N	CL	No Score Normalisation					Score Normalisation				
		$M = 5$	$M = 10$	$M = 15$	$M = 20$	$M = 25$	$M = 5$	$M = 10$	$M = 15$	$M = 20$	$M = 25$
0	-	3.57%	3.57%	3.57%	3.57%	3.57%	3.13%	3.13%	3.13%	3.13%	3.13%
1	L	3.74%	3.54%	3.29%	3.23%	3.18%	3.66%	3.22%	3.13%	3.06%	2.94%
	M	3.88%	3.44%	3.34%	3.33%	3.15%	3.56%	3.24%	3.10%	3.03%	2.94%
	H	3.90%	3.65%	3.40%	3.32%	3.21%	3.62%	3.25%	3.10%	3.00%	2.91%
3	L	3.41%	3.18%	3.04%	3.05%	2.96%	3.17%	2.87%	2.76%	2.78%	2.69%
	M	3.54%	3.16%	3.01%	3.03%	2.97%	3.11%	2.90%	2.83%	2.82%	2.71%
	H	3.43%	3.30%	3.11%	3.00%	2.98%	3.04%	2.84%	2.81%	2.72%	2.66%
5	L	3.36%	3.03%	2.99%	2.95%	2.92%	3.03%	2.81%	2.69%	2.68%	2.58%
	M	3.43%	3.07%	2.91%	2.94%	2.86%	3.01%	2.80%	2.71%	2.73%	2.66%
	H	3.34%	3.11%	2.93%	2.99%	2.88%	2.93%	2.65%	2.65%	2.62%	2.58%
10	L		2.95%	2.85%	2.91%	2.78%		2.66%	2.60%	2.63%	2.52%
	M		2.94%	2.86%	2.77%	2.77%		2.69%	2.69%	2.61%	2.58%
	H		2.97%	2.75%	2.72%	**2.71%**		2.63%	2.56%	2.50%	**2.46%**

N	CL	No Score Normalisation				Score Normalisation			
		ChS-50	ChS-100	ChS-150	ChS-200	ChS-50	ChS-100	ChS-150	ChS-200
0	-	6.84%	6.84%	6.84%	6.84%	5.91%	5.91%	5.91%	5.91%
1	L	6.39%	6.12%	6.10%	6.11%	5.58%	5.35%	5.46%	5.41%
	M	6.60%	6.36%	6.38%	6.27%	5.77%	5.56%	5.57%	5.40%
	H	6.54%	6.45%	6.37%	6.21%	5.95%	5.66%	5.54%	5.42%
3	L	6.22%	5.98%	5.97%	5.93%	5.37%	5.25%	**5.21%**	5.29%
	M	6.36%	6.27%	6.19%	6.14%	5.47%	5.39%	5.36%	5.30%
	H	6.28%	6.20%	6.26%	6.12%	5.59%	5.47%	5.39%	5.35%
5	L	6.14%	5.96%	5.96%	5.98%	5.30%	5.23%	**5.21%**	5.24%
	M	6.30%	6.19%	6.21%	6.14%	5.34%	5.28%	5.34%	5.29%
	H	6.26%	6.21%	6.25%	6.13%	5.49%	5.36%	5.40%	5.30%
10	L	6.22%	6.00%	**5.90%**	5.95%	5.32%	5.24%	**5.21%**	5.24%
	M	6.24%	6.18%	6.14%	6.13%	5.29%	**5.21%**	5.26%	5.25%
	H	6.28%	6.17%	6.21%	6.10%	5.41%	5.32%	5.32%	5.22%

TABLE 3 Fingerprint (capacitive device) recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio *Basic Approach*. Columns and rows, as well as cell colour code, are the same as in previous tables

results of this section are interesting, as compared with the previous one, to test the Score Ratio with the worst system performance, as the images with the capacitive device are worse than those with the optical one.

6.5 | Results analysis

The first important aspect to note is that significant improvements have been achieved with Score Ratio, both with the *Basic Approach* and with the *a Priori Cohort Selection*. These improvements have been achieved with both devices, independently of the control level of the samples used in the ChS and with and without score normalisation, showing the consistency of the proposal with regard to the data.

Focusing on the Score Ratio parameters tested.

- With regard to N (let us remember that N is the number of cohort set elements used to calculate the Score Ratio, equation (10), denominator), it is advisable to use values bigger

TABLE 4 Fingerprint (capacitive device) recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio with *a Priori Cohort Selection*. ChS-200 is used to select a priori M elements closest to the Claimant. If $N > M$, then the denominator in the Score Ratio equation (equation (10)) cannot be calculated, so these rows are empty. The remaining columns and rows, as well as cell colour code, are the same as in previous tables

N	CL	No Score Normalisation					Score Normalisation				
		M = 5	M = 10	M = 15	M = 20	M = 25	M = 5	M = 10	M = 15	M = 20	M = 25
0	-	6.84%	6.84%	6.84%	6.84%	6.84%	5.91%	5.91%	5.91%	5.91%	5.91%
1	L	6.82%	6.73%	6.39%	6.28%	6.30%	6.17%	6.10%	5.83%	5.70%	5.62%
	M	6.98%	6.83%	6.47%	6.36%	6.33%	6.40%	6.10%	5.85%	5.81%	5.67%
	H	7.12%	6.60%	6.58%	6.39%	6.33%	6.37%	5.93%	5.84%	5.63%	5.64%
3	L	6.41%	6.27%	6.01%	5.91%	5.96%	5.68%	5.64%	5.43%	5.26%	5.27%
	M	6.50%	6.49%	6.36%	6.22%	6.18%	5.91%	5.660%	5.58%	5.46%	5.38%
	H	6.61%	6.21%	6.14%	6.02%	6.04%	5.85%	5.46%	5.56%	5.39%	5.40%
5	L	6.26%	6.09%	5.89%	5.85%	5.86%	5.56%	5.44%	5.24%	5.22%	5.20%
	M	6.44%	6.36%	6.31%	6.20%	6.14%	5.78%	5.58%	5.49%	5.43%	5.29%
	H	6.46%	6.11%	6.02%	5.93%	6.02%	5.61%	5.48%	5.48%	5.32%	5.34%
10	L		6.01%	5.80%	**5.76%**	5.79%		5.30%	5.10%	5.12%	**5.07%**
	M		6.29%	6.20%	6.07%	6.12%		5.49%	5.35%	5.29%	5.28%
	H		6.00%	6.01%	5.94%	5.89%		5.18%	5.32%	5.16%	5.20%

than 1. Although the performance is good for the rest of the values, in general, the best results have been achieved with $N = 10$. This allows us to conclude that using a large enough N value is recommended. In addition, this value does not affect the computer load of the proposal. In the next modes, bigger values have been tested, showing that from a particular value onwards the results do not improve, which is why we previously said *large enough*.

- As for the cohort set size in the *Basic Approach*, improvements have been achieved with all of them. However, although with no very large differences, the best ones have been achieved with the larger sizes: 150 and 200. This parameter does affect the computer load of the Score Ratio proposal. So, if this were important, we can use small ChS sizes with a slightly worse Score Ratio performance, or we can use the *a Priori Cohort Selection* approach, which, also considerably reduces the computer load, as well as improving the performance, as we will analyse next.

To show some concrete figures, the application of the *Score Ratio Basic Approach* has achieved an improvement in the system performance of 20%[2] (from 3.57% to 2.85%) without score normalisation and 17% (from 3.13% to 2.61%) with score normalisation, using the optical device data; while, with the capacitive ones, the best improvements have been 14% (from 6.84% to 5.90%) without score normalisation and 12% (from 5.91% to 5.21%) with score normalisation. With the *a Priori Cohort Selection* approach, the best improvements have been 24% (from 3.57% to 2.71%), 21% (from 3.13% to 2.46%), 16%

(from 6.84% to 5.76%) and 14% (from 5.91% to 5.07%). As can be seen, first, the better the reference system is, the bigger the improvements that have been achieved with Score Ratio and, second, the *a Priori Cohort Selection* approach not only reduces the computational load, but has also improved the *basic Approach* performance, as already pointed out.

Finally, focusing on the *a Priori Cohort Selection* approach, the Score Ratio improves the reference system in all of the tests except for $N = 1$ and small values of M (5 and 10). In general, the results improve as M and N increase. However, good results can be found for small values of M, showing that the Score Ratio can significantly improve the reference system with a small computer load increase (remember that the value of N does not affect the processing time).

7 | IRIS

Together with fingerprint, iris is the biometric characteristic with the oldest biometrics solution for authentication on computerised systems. Very important in the current biometric market, it is also a very mature mode. As with fingerprint, the benchmark database, and the experimental protocol and system have been used in the experiments. Scripts and configuration files of iris experiments are available in http://www.infor.uva.es/cevp/Download/Iris.zip.

7.1 | Experimental setup

The BIOSECURID multimodal biometric database [18] has been used. This database includes eight unimodal biometric

[2]The improvement percentage is calculated as: $((EEE_{NoSR} - EER_{SR})/EER_{NoSR}) * 100$

traits, namely: speech, iris, face, handwritten signature and handwritten text, fingerprints, hand and keystroking. It is a database acquired under realistic conditions and with balanced gender and population distributions. Here, the iris part is used. This part comprises 400 subjects: two eyes of 200 individuals (as in fingerprint, each eye is considered a different subject). Four samples of each eye were acquired in four different sessions that is we have a total of 16 samples per subject. This implies that the database includes 6400 iris images, being one of the biggest public ones. As it is used in many works, this database represents a good benchmark.

The iris database was split into the following subsets (Section 5.2):

- **Normalisation Set, NrS.** Twenty-five individuals were randomly selected from the database. From each individual, one sample of each eye was selected (the one that was used as template). The size of this set is, then: 25 individuals $\times$ 2 eyes/individual$\times$ 1 sample/eye = 50 samples.
- **Cohort Set, ChS.** Sizes of 30, 60 and 90 (15, 30 and 45 individuals, two eyes per individual) were tested. Bigger sizes were tested at the beginning of the experiments, but, with no improvements, we decided not to try larger sizes and thus increase the TS.
- **Test Set (TS).** Two hundred and sixty subjects (130 individuals, two eyes each one) were used. Following the reference experimental protocol [6], one iris image is used as template, this being the first image acquired. The remaining images were used for biometric mated comparison trials. For impostor tests (biometric non-mated comparison trials), one image per eye (that used for the template) of the rest of the individuals in the TS, different from the Claimant, were used. That is, we had 3900 (260 $\times$ 15) genuine trials and 67,340 (260 $\times$ 259) impostor trials.

7.2 | Recognition system

The OSIRIS open-source reference system [6] was used. This system is inspired in Doughman's approach [19], which is the main benchmark in iris recognition. Briefly, this approach consists of:

- First, the iris is isolated from the image captured (Figure 7 left).
- The iris image is normalised into a fixed rectangular size and enhanced (Figure 7 middle).
- Using two-dimensional Gabor filters, the iris is finally transformed into a binary pattern called *iris code* (Figure 7 right).
- Comparison between irises is made using the Hamming distance, in which *iris codes* are compared using the XOR technique. So, the iris system uses a distance-based classifier.

As in fingerprint, the *global EER* approach (Section 5.3) is used in the benchmark protocol [6] to measure system performance and this approach is followed here.

7.3 | Results

In Tables 5 and 6, the results without Score Ratio compared with those with Score Ratio *basic Approach* and *a Priori Cohort Selection*, respectively, are shown.

7.4 | Results analysis

As with fingerprint, the application of Score Ratio has outperformed the reference system with both the *Basic Approach* (in all tests) and with the *a Priori Cohort Selection* (in all tests except for $N = 1$). Although the improvements are smaller here, they are still statistically significant. The best figures are 6% (from 5.56% to 5.23%) without score normalisation and 7% (from 5.44% to 5.05%) with score normalisation when the *Basic Approach* is applied; the same values are achieved with the *a Priori Cohort Selection* approach.

Bigger values of N with regard to those tested in the fingerprint mode have been proved here. However, values over 10 have only shown improvements with the *a Priori Cohort Selection* approach. In the same way, a minimal value of N ($N \geq 3$) is necessary to achieve good results, as with fingerprint.

With regard to the ChS size, the conclusion in the *Basic Approach* is similar here to that with fingerprint: the best performance has been obtained, mainly, with the larger values (60 and 90), but the reference system is also outperformed with small values.

The good performance of the *a Priori Cohort Selection* has also been demonstrated here: similar results to those with the *Basic Approach* have been achieved. Moreover, the results are, in general, better with this approach than with the general one, even with small values of M ($M = 10, 15$).

8 | FINGER VEIN

Finger vein is an emerging biometric characteristic based on the vascular patterns that exist in the finger. Although used in commercial systems, it has not been until very recently that the scientific community has paid attention to it. Unlike the previous mature modes addressed in this work, we can say that there are no reference databases or systems in vein recognition. It is not easy to find public databases either. One of the largest and most complete is the University of Twente Finger Vascular Pattern (UTFVP) database [20] (available in: https://scs.ewi.utwente.nl/downloads/show,FingerThis is what we have used here. Also, the experimental protocol in the publications related to the UTFVP database [20,21] isfollowed.

The interest in testing this biometric characteristic is to prove Score Ratio under different conditions with respect to those in fingerprint and iris. The most important are: i) first, following the reference protocol in Ref. [20], *individual EER* (Section 5.3) is used, with the intention of showing that Score Ratio can also improve system performance at an individual level; ii) second, and following the reference
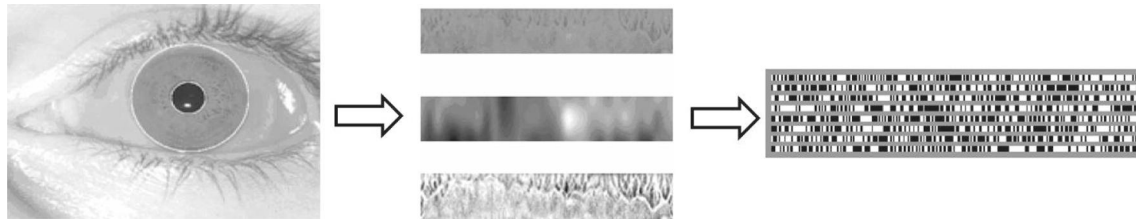
FIGURE 7 Main modules (stages) of the iris recognition system

TABLE 5 Iris recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio *Basic Approach*. Columns and rows, as well as cell colour code, are the same as in previous tables

N	No Score Normalisation			Score Normalisation		
	ChS-30	ChS-60	ChS-90	ChS-30	ChS-60	ChS-90
0	5.56%	5.56%	5.56%	5.44%	5.44%	5.44%
1	5.53%	5.33%	5.41%	5.28%	5.28%	5.30%
3	5.30%	**5.23%**	5.35%	5.23%	5.12%	5.27%
5	5.30%	5.25%	5.38%	5.18%	**5.05%**	5.10%
10	5.38%	5.30%	5.38%	5.15%	5.23%	**5.05%**
15	5.41%	5.33%	5.35%	5.18%	5.26%	5.13%
20	5.35%	5.33%	5.35%	5.23%	5.23%	5.21%
25	5.35%	5.35%	5.35%	5.33%	5.23%	5.18%

experimental protocol, the subject template is made up of more than one sample, more specifically, two samples are used. Scripts and configuration files of finger vein experiments are available in http://www.infor.uva.es/cevp/Download/Vein.zip.

8.1 | Experimental setup

The UTFVP database is made up of images of 60 individuals, captured with a custom designed device. From each individual, four images in two different sessions (two per session) where acquired of six fingers that is 1440 images in total.

The database was split into the following subsets (Section 5.2).

- **Normalisation Set, NrS.** This is not considered here, since *individual EER* is used and the normalisation of each Claimant Match and Non-Match score distributions (Section 5.3) is not necessary.
- **Cohort Set, ChS.** To maximise the TS, only sizes of 36 (6 individuals × 6 fingers) and 54 (9 individual × 6 fingers) were tested.
- **Test Set (TS).** 306 subjects (51 individuals, six fingers each one) were used. Following the reference protocol [20,21], two finger vein images, in this case, the first image acquired in each session, were used as the template. The rest of the images are used as genuine tests, two per

subject. For impostor tests, in order not to have very unbalanced test sets, one image (the first one acquired in the first session) of a single finger (randomly selected) for the rest of the individuals in the TS different from the Claimant are used; that is we have 50 impostor tests per subject.

8.2 | Recognition systems

Following the bibliography, the original infrared image is pre-processed prior to the feature extraction, isolating the finger from the background, extracting the ROI (Region of Interest), enhancing the vascular pattern and normalizing the size of the final image (Figure 8).

Here, since the goal is to test Score Ratio, our own very simple systems were used, looking for a variety of approaches:

- **Discrete Fourier Transformation (DCT)-based approach.** Feature extraction is accomplished by means of DCT coefficients extraction. The feature vector is made up of the 75 low-frequency components. Feature vectors are compared using Euclidean **distance.**
- **Binarisation-based approach.** The pre-processed image is binarised, so that the feature vector is now a binary pattern. Comparison between images is accomplished by means of two different approaches:
  ○ Applying the *AND* binary operation and counting the number of ones in the result. Under this approach, the score achieved is the type **probability-like** that is the higher the score, the more similar the finger vein patterns are.
  ○ Applying the *Hamming distance.* This approach is similar to applying the XOR operation between the binary images and counting the number of ones in the result. Thus, we identify this approach as *XOR* in the results section below.

8.3 | Results

The performance of the finger vein systems with and without Score Ratio is shown in Tables 7 and 8, when the *Score Ratio Basic Approach* and the *a Priori Cohort Selection* are used, respectively. Remember that here, the system EER is the mean of the individual EERs per Claimant.

	No Score Normalisation					Score Normalisation				
N	M = 5	M = 10	M = 15	M = 20	M = 25	M = 5	M = 10	M = 15	M = 20	M = 25
0	5.56%	5.56%	5.56%	5.56%	5.56%	5.44%	5.44%	5.44%	5.44%	5.44%
1	5.56%	5.65%	5.67%	5.58%	5.62%	5.67%	5.66%	5.55%	5.49%%	5.44%
3	5.40%	5.48%	5.51%	5.48%	5.56%	5.41%	5.31%	5.30%	5.26%	5.27%
5	5.40%	5.41%	5.48%	5.42%	5.47%	5.31%	5.21%	5.22%	5.21%	5.26%
10		**5.24%**	5.38%	5.35%	5.42%		5.10%	5.15%	5.12%	5.17%
15			5.25%	5.29%	5.36%			5.15%	5.13%	5.11%
20				5.31%	5.34%				**5.07%**	5.11%
25					5.30%					5.08%

TABLE 6 Iris recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio with *a Priori Cohort Selection*. ChS-90 is used to select the a priori M elements closest to the Claimant. If $N > M$, then the denominator in the Score Ratio equation (equation (10)) cannot be calculated, so these rows are empty. The remaining columns and rows, as well as cell colour code, are the same as in previous tables

FIGURE 8　Main modules (stages) of the finger vein system

8.4 | Results analysis

This third biometric mode once more proves that Score Ratio can improve the reference systems, but now under different experimental conditions with regard to those in the previous tests. In spite of this, the same main conclusions as in the previous modes can be extracted:

- Both *Basic Approach* and *a Priori Cohort Selection* outperform the reference system in all of the tests, except for $N = 1$ in some cases.
- The second approach can achieve even better results than those with the general one, besides reducing the computer load.
- Good results have been achieved with small values of M ($M = 10, 15$) in the *a Priori Cohort Selection*, which implies that the Score Ratio can improve the reference systems with very small computer load increases.

The biggest improvements achieved here are: 13% (from 7.09% to 6.17%) in the DCT-based system, 13% (from 7.17% to 6.21%) in the XOR-based system and 43% (from 14.35% to 8.11%) in the AND-based system with the *Basic Approach* and 25% (from 7.09% to 5.3%), 21% (from 7.17% to 5.7%) and 41% (from 14.35% to 8.5%), respectively, with the *a Priori Cohort Selection* approach.

9 | CONCLUSIONS

Herein, a proposal, named Score Ratio, for using Non-Claimant Class information in the biometric verification problem with non-probabilistic classifiers, has been shown and widely tested.

TABLE 7　Finger vein recognition performance (EER in %) using and not using ($N = 0$ row) Score Ratio *Basic Approach*, with all of the system tested: **DCT**-based and binarisation-based applying **XOR** and **AND** operations. Columns and rows, as well as cell colour code, are the same as in previous tables

	DCT		XOR		And	
N	ChS-36	ChS-54	ChS-36	ChS-54	ChS-36	ChS-54
0	7.09%	7.09%	7.17%	7.17%	14.35%	14.35%
1	7.67%	7.30%	7.76%	7.79%	9.78%	8.96%
3	7.03%	6.73%	7.01%	7.00%	9.08%	8.49%
5	7.01%	6.73%	6.84%	6.79%	9.03%	8.42%
10	6.61%	6.51%	6.48%	6.46%	8.62%	8.28%
15	6.48%	6.25%	6.39%	6.31%	8.59%	**8.11%**
20	6.42%	6.21%	6.29%	6.28%	8.56%	8.12%
25	6.33%	**6.17%**	**6.21%**	6.24%	8.50%	8.16%

The proposal has been tested on two very mature biometric characteristics (fingerprint and iris) as well as an emerging one (finger vein). Benchmark databases, experimental protocols and/or systems have been used in the tests for more objective results. Furthermore, their statistical significance has been measured.

Two different approaches of Score Ratio have been tested, the *Basic Approach* and the *a Priori Cohort Selection* approach, with the aim of decreasing the computational load of the Score Ratio, if necessary, with the second. The results show that using Score Ratio improves the performance of the reference systems in the great majority of the experiments carried out. Furthermore, these improvements

TABLE 8 Finger vein recognition performance (EER in %) using and not using (N = 0 row) Score Ratio with *a Priori Cohort Selection*, with all of the system tested. ChS-54 is used to select the a priori M (M = 5/10/15/ 20/25) elements closest to the Claimant. If N > M, then the denominator in the Score Ratio equation (equation (10)) cannot be calculated, so these rows are empty. The remaining columns and rows, as well as cell colour code, are the same as in previous tables

N	DCT	XOR	AND
0	7.1%	7.2%	14.4%
1	8.3%/8.1%/7.9%/7.7%/7.7%	9.9%/9.1%/9.3%/8.8%/8.6%	9.5%/9.1%/9.1%/9.0%/9.1%
3	6.5%/6.8%/6.9%/6.8%/6.9%	7.6%/7.5%/7.4%/7.2%/7.0%	9.5%/8.8%/8.6%/8.6%/8.6%
5	5.8%/6.3%/6.5%/6.6%/6.6%	6.1%/6.7%/6.9%/6.8%/6.7%	10.1%/8.9%/8.7%/8.7%/8.6%
10	-/5.4%/5.8%/6.1%/6.1%	-/5.9%/6.3%/6.3%/6.3%	-/9.6%/8.9%/8.6%/**8.5%**
15	-/-/**5.3%**/5.8%/5.9%	-/-/5.9%/5.9%/6.0%	-/-/9.7%/9.0%/8.6%
20	-/-/-/5.6%/5.7%	-/-/-/**5.7%**/5.8%	-/-/-/9.5%/8.8%
25	-/-/-/-/5.6%	-/-/-/-/**5.7%**	-/-/-/-/9.1%

are statistically significant, thus strengthening the confidence in the results.

Following all of the above, it can be concluded that Score Ratio is an interesting alternative to improve non-probabilistic biometric based-systems, in the same way that the likelihood ratio is for the probabilistic based ones.

ORCID
Carlos Vivaracho-Pascual https://orcid.org/0000-0001-7732-6170

REFERENCES
1. Wayman, J., et al.: Vocabulary harmonisation for biometrics: the development of iso/iec 2382 part 37. IET Biom. 3(1), 1–8 (2014)
2. Matsui, T., Furui, S.: Likelihood normalization for speaker verification using a phoneme and speaker independent model. Speech Commun. 17(1), 109–116 (1995)
3. Argones Rua, E., Alba Castro, J.L.: Online signature verification based on generative models. IEEE Trans. Syst. Man Cybern. B Cybern. 42(4), 1231–1242 (2012)
4. Nanni, L., Lumini, A.: A supervised method to discriminate between impostors and genuine in biometry. Expert. Syst. Appl. 36(7), 10401–10407 (2009)
5. Vivaracho Pascual, C., Simon Hurtado, A., Manso Martinez, E.: Using the score ratio with distance-based classifiers: a theoretical and practical study in biometric signature recognition. Neurocomputing, 248, 57–66 (2017). Neural Networks: Learning Algorithms and Classification Systems
6. Petrovska Delacrétaz, D., Chollet, G., Dorizzi, B.: Guide to biometric reference systems and performance evaluation. Springer, London (2009)
7. Lehmann, E.L.: Testing statistical hypotheses, 2nd ed. Chapman and Hall, New York (1986)
8. Ferguson, T.S.: Mathematical statistics: a decision theoretic approach. Academic Press, New York (1967)
9. Reynolds, D.A., ISCA EURASIP: Speaker identification and verification using Gaussian mixture models. Speech Commun. 17(1-2), 91–108 (1995)
10. Matsui, T., Furui, S.: Similarity normalization method for speaker verification based on a posteriori probability. Proceedings of the workshop on automatic speaker recognition identification and verification, pp. 59–62. Martigny, Switzerland (1994)
11. Martin, A., et al.: The DET curve assessment of detection task performance. Proc. EuroSpeech. 4, 1895–1898 (1997)
12. Mansfield, A.J., Wayman, J.L.: Best practices in testing and reporting performance of biometric devices. Centre for Mathematics and Scientific Computing, National Physical Laboratory, Teddington (2002)
13. Doddington, G.R., et al.: The nist speaker recognition evaluation: overview, methodology, systems, results, perspective. Speech Commun. 31(2), 225–254 (2000)
14. Doddington, G., et al.: Sheep, goats, lambs and wolves. a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. Proc. ICSLP (1998)
15. Yager, N., Dunstone, T.: The biometric menagerie. IEEE Trans. Pattern Anal. Mach. Intell. 32(2), 220–229 (2010)
16. Ortega-Garcia, J., et al.: MCYT baseline corpus: a bimodal biometric database. IEE Proc. Visual Image Signal Process. 150(6), 395–401 (2003)
17. Jain, A.K., et al.: An identity-authentication system using fingerprints. Proc. IEEE. 9(85), 1365–1388 (1997)
18. Fierrez, J., et al.: Biosecurid: a multimodal biometric database. Pattern Anal. Appl. 13(2), 235–246 (2010)
19. Daugman, J.: How iris recognition works. IEEE Trans. Circ. Syst. Video Technol. 14(1), 21–30 (2004)
20. Ton, B.T., Veldhuis, R.N.J.: A high quality finger vascular pattern dataset collected using a custom designed capturing device. In: Proc. 2013 International Conference on Biometrics (ICB), pp. 1–5. Madrid (2013)
21. Nibbelke, V.: Vascular pattern recognition for finger veins using biometric graph matching (2014). https://www.semanticscholar.org/paper/Vascular-pattern-recognition-for-finger-veins-using-Nibbelke/f8c22cdafe6eb8f1d192d8ac0bfb07e573a51fd6. Accessed September 2020

How to cite this article: Vivaracho-Pascual C, Simon-Hurtado A, Manso-Martinez E. Improving biometric recognition by means of score ratio, the likelihood ratio for non-probabilistic classifiers. A benchmarking study. *IET Biome*. 2021;10:127−141. https://doi.org/10.1049/bme2.12011