

Curvewise DET confidence regions and pointwise EER confidence intervals using radial sweep methodology

Michael E. Schuckers¹, Yordan Minev¹ and Andy Adler²

¹ Department of Mathematics, Computer Science and Statistics
St. Lawrence University, Canton, NY USA

² Department of Systems and Computer Engineering
Carleton University, Ottawa, ON Canada

Abstract. One methodology for evaluating the matching performance of biometric authentication systems is the detection error tradeoff (DET) curve. The DET curve graphically illustrates the relationship between false rejects and false accepts when varying a threshold across a genuine and an imposter match score distributions. This paper makes two contributions to the literature on the matching performance evaluation of biometric identification or bioauthentication systems. First, we create curvewise DET confidence regions using radial sweep methods. Second we use this same methodology to create pointwise confidence intervals for the equal error rate (EER). The EER is the rate at which the false accept rate and the false reject rate are identical. We utilize resampling or bootstrap methods to estimate the variability in both the DET and the EER. Our radial sweep is based on converting the false reject and false accept errors to polar coordinates. Application is made of these methods to data from three different biometric modalities and we discuss the results of these applications.

1 Introduction

Biometric identification is the process of matching an individual's physical or behavioural traits to a stored version of those same traits. Measuring how well a bioauthentication process matches these traits is an important facet of the overall performance of the system in which it is embedded. The focus of this article will be the statistical estimation of the matching performance of these bioauthentication systems. In this context genuine users are generally those that the system should accept and imposters are those that the system should reject. An effective biometrics authentication system evaluates a given match score as a genuine or an imposter match score with a minimal error. There are two possible error rates that can occur. The proportion of imposter match scores which are accepted from the system is called the False Accept Rate (*FAR*). The proportion of genuine match scores which are rejected by the system is called the False Reject Rate (*FRR*). Some authors prefer the use of false match and false non-match rate. For generality here we will use FAR and FRR. See Mansfield and Wayman

[1] for a discussion of the differences. Our interest focuses on the representation of the tradeoff between these two curves known as the DET. The DET curve evaluates the matching performance of a biometric system. One factor that is often overlooked in the use of DET's is that the curves are sample estimates of a population DET and, thus, there is variability in the estimated curves. This paper quantifies this variability and builds confidence regions for the population DET curve based upon the observed sample.

There is a good deal of literature on DET's or their equivalent the receiver operating characteristic curve (ROC) beginning with the work of Green and Swets [2]. In this paper we will concentrate on the DET; however, the methods described here apply directly to the ROC. An ROC is a plot of the FAR against the true accept rate (TAR) — $TAR = 1 - FRR$ — while we use a DET that is a plot of the FAR against the FRR. (Some authors, e.g. Poh *et al* [3] transform the FAR and FRR to achieve a DET.) A good deal has been written about the ROC in a variety of contexts. See, e.g. Zhou *et al* [4] or Hernández-Orallo *et al* [5]. Macskassy *et al* [6] compared three methods that have been proposed for ROC confidence bands. These include: a simultaneous joint confidence region and a fixed width approach both due to Campbell [7], a Working-Hotelling band technique due to Ma and Hall [8]. The simultaneous joint confidence approach converts confidence rectangles at various points along the ROC to create a joint region. The fixed width approach creates a region of single width in a direction determined by number of observations used to calculate each FAR and TAR. Working-Hotelling bands are based upon confidence regions for a regression line. Based upon its performance in their evaluation, Macskassy *et al* recommend the fixed width approach. The drawback of this methodology for bioauthentication applications is that it produces large regions near the ends of the ROC or DET which is often the area of interest for such applications. Recall that larger intervals imply less knowledge about an estimate. Recently, Poh *et al* [9] suggested a new DET confidence methodology based upon the radial sweep methodology of Adler and Schuckers [10]. However, this methodology is a pointwise approach to confidence unlike the curvewise approach described by Macskassy *et al* [6]. The difference between them is that a curvewise approach aims to capture the entire DET curve while a pointwise approach only provides confidence about a single point and thus neglects the correlation between points along the curve. This is often referred to as the multiple comparison problem. We will utilize the pointwise approach to make a confidence interval for the EER since it is appropriate for a single point; however, we take the curvewise approach for DET estimation. Finally, we mention the work by Dass *et al* [11] to create a confidence region for FRR's given a range of values for FAR. This approach is essentially univariate rather than the bivariate approach taken here.

The approach we propose in this paper is a new curvewise method for creating a confidence region for the DET as a function of both the FAR and FRR based on a radial sweep methodology. By curvewise in this context we mean that our confidence will be based upon the entire curve rather than on each individual point. This yields a wider confidence interval – due to the correlation between

points on the curve – but one that is more accurate for variability of the entire curve. This approach, described below, transforms each curve from the (FAR, FRR) space to polar coordinates, (radius, angle). By using a polar representation of the error rate data we can create regions that are appropriate for inference about both the FAR and FRR simultaneously. Adler and Schuckers [10] originally used this approach to average across several DET’s to create a composite DET. We extend their work to create a curvewise confidence region. In addition, this methodology has variable width which allows for smaller intervals as the DET curves approach the axes. This is a distinct advantage over the fixed width approach advocated by Macskassy *et al* [6]. These regions — where the FAR and FRR approach unity — are often the focus of biometric authentication applications and threshold settings. Having created such a confidence region, one gains a better understanding of the variability in the system and the location of the population DET based upon the sample selected. Further, it is possible to derive using a pointwise confidence interval for the equal error rate (EER) as a carryover of the work done to create the DET region. The rest of this article is organized in the following way. Section 2 discusses the calculation of a DET and the conversion to polar coordinates. Our approach for making a confidence region for a DET is given in Section 3. This section also describes the methodology for creating an EER confidence interval. We apply our inferential estimation approach to data from three biometric modalities: face, fingerprint and hand geometry in Section 4. These data comes from the paper by Ross and Jain [12]. Section 5 discusses these results and their possible extensions.

2 DET curve estimation

We begin by outlining the methodology for estimating a DET curve. For the purposes here, we assume that we have a population of *genuine* and *imposter* matching scores from a biometric system. Further we assume that these scores represent a sample of an ongoing process which we would like to evaluate. The distributions of match scores, t , are denoted by $g(t)$ and $f(t)$ for the genuine and imposters distributions respectively. From these distributions, a DET, R , is plotted as the false accept rate (FAR) on the x -axis against the false reject rate (FRR) on the y -axis, by varying a threshold τ , and calculating

$$FAR(\tau) = \int_{\tau}^{\infty} f(x)dx \quad \text{and} \quad FRR(\tau) = \int_{-\infty}^{\tau} g(y)dy. \quad (1)$$

If $f(t)$ and $g(t)$ are known then R can be calculated based upon Equation (1) by varying τ . In general this is rarely the case and these distributions must be estimated. In that case we can calculate, \hat{R} based upon an approximation of $f(t)$ and $g(t)$.

$$\hat{FAR}(\tau) = \int_{\tau}^{\infty} \hat{f}(x)dx = \frac{\sum_x I(x > \tau)}{\sum_x I(x \in \mathbb{R})} \quad \text{and} \quad (2)$$

$$\hat{FRR}(\tau) = \int_{-\infty}^{\tau} \hat{g}(y)dy = \frac{\sum_y I(y \leq \tau)}{\sum_y I(y \in \mathbb{R})} \quad (3)$$

where $I()$ represents an indicator function and x and y are observed scores from imposter and genuine distributions, respectively. As with R , \hat{R} is computed by calculating FAR and FRR for various values of τ . In this way the sample DET curve summarizes the authentication performance of the biometric system on the sample on which it is calculated. The FAR and FRR calculations above are based on a criterion of accept if a match score is greater than τ . An alternative formulation can easily be made for other decisions. Thus, we will treat this DET, \hat{R} as an estimate of the population DET, R .

Having calculated a DET, either sample or population, We next convert the coordinates of the curve from rectangular to polar coordinates. Then our DET curve over a range of values Θ is $R = \{(r_\theta, \theta) : \theta \in \Theta = (\theta_L, \theta_U)\}$ and θ_L and θ_U represent lower and upper values for the range of the radial angles. In practice we sweep across a range of T θ 's using a discrete list of values $\Theta^* = (\theta_L, \theta_L + \delta, \theta_L + 2\delta, \dots, \theta_L + (T-2)\delta, \theta_U)$ where $\Theta = (\theta_L, \theta_U)$ is given. Then we can write $\hat{R} = \{(\hat{r}_\theta, \theta) : \theta \in \Theta^*\}$. One important issue in radial sweeping is the location of the radial center, (c_x, c_y) . Because we would like the confidence region to not depend on which error rate is on which axis and in order to have a confidence interval for the EER as a derived outcome for our data, we limited possible center points $c_x = c_y = c$ for some specified constant c . (See Figure 1 from [10] for a graphical representation of the conversion to polar coordinates.) It is immediately clear that choosing a center along the $FAR = FRR$ line results in an average curve that is independent of the selection of axes. Once we have determined the origin of the polar coordinate system, we can convert the coordinates of every point on the DET curve, R , to polar coordinates. We next turn our attention to deriving a confidence region for the DET, R .

3 A curvewise DET confidence region

In this section, we develop our methodology for creating a curvewise DET confidence region. To derive the confidence regions, we need to estimate the variance of \hat{r}_θ for each θ . Several methods are possible. We use bootstrap methods to approximate this radial variance. Bootstrap procedures, such as those proposed by Poh *et al* [9] or Bolle *et al* [13], resample the observed data to estimate in a non-parametric way the distributions which generated this data. Here, we resample both the genuine and imposter distributions. The general outline for our confidence region technique is as follows:

1. Calculate estimated DET, \hat{R} .
2. Bootstrap both genuine and imposter distributions.
3. Calculate bootstrapped DET, R^\dagger .
4. Repeat previous two steps M times and store each generated DET.
5. Determine curvewise bounds for confidence region, \mathfrak{R} .

Let $R^\dagger = \{(r_{m\theta}^\dagger, \theta) : \theta \in \Theta^*\}$ be the m^{th} bootstrapped DET $m = 1, \dots, M$ for some large number M . $r_{m\theta}^\dagger$ represents the radial length for the m^{th} bootstrapped DET at the angle θ . Next we find the adjusted standard deviation of $r_{m\theta}$'s at

each θ in Θ^* by taking the square root of the variance plus a slight adjustment, ϵ . We will denote this by $s_{\hat{r}_\theta} = \sqrt{\text{var}(r_\theta^\dagger) + \epsilon}$. This adjustment is made to ensure that some variability is present in the confidence region when no radial variability is present in the bootstrapped replications.

We define the DET confidence region \mathfrak{R} as the region bounded by a spline of $r_{L,\theta}$'s, the lower limits for the region at angle θ , and a spline of $r_{U,\theta}$'s, the upper limits for the region at angle θ for each $\theta \in \Theta^*$. We want that region to have the property that $P(R \in \mathfrak{R}) = 1 - \alpha$ where $1 - \alpha$ is the confidence level. We say the population DET, R , is captured by the $(1 - \alpha)100\%$ ROC confidence region and denote it by $R \in \mathfrak{R}$ if $r_{L,\theta} \leq r_\theta \leq r_{U,\theta}$ for all $\theta \in \Theta^*$. In order to determine the DET confidence region, \mathfrak{R} , we need to create a region that captures the entire curve R with probability $(1 - \alpha)$. The pointwise approach of Poh *et al* [9], is to take the $\alpha/2 \times 100\text{th}$ and $(1 - \alpha/2) \times 100\text{th}$ percentiles of the r^\dagger 's at each θ . As mentioned above, this ignores the correlation between radii at different values of θ , particularly neighboring values of θ in Θ^* . Consequently it fails to yield a confidence region with $P(R \in \mathfrak{R}) = 1 - \alpha$.

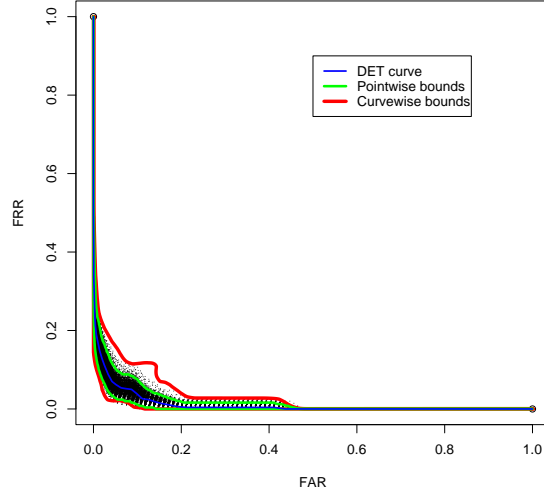


Fig. 1. 95% confidence region for DET of Facial Recognition System

Several options exist for finding an DET confidence region, \mathfrak{R} , that meets our criteria. We prefer an approach that uses a different width at each θ to match the variability in the data. To achieve this systematically we build our DET region by connecting the endpoints $(\hat{r}_\theta + \eta_L s_{r_\theta}, \hat{r}_\theta + \eta_U s_{r_\theta})$ at each $\theta \in \Theta^*$ where η_L and η_U represent lower and upper constants, respectively, that yield the desired

$(1 - \alpha) \times 100\%$ confidence level. Below we describe the process for determining η_L and η_U . First, we calculate a standardized residual, $e_{m\theta}$ for each $r_{m\theta}$,

$$e_{m\theta} = \frac{r_{m\theta} - \hat{r}_\theta}{s_{\hat{r}_\theta}}. \quad (4)$$

This residual tells us how far each bootstrapped radius, \hat{r}_θ , is from the estimated DET, \hat{R} as measured by the variability, s_{r_θ} at each θ . Next, the maximum absolute standardized residual, ω_m , for the m^{th} DET curve is calculated. This value, ω_m , tells us the furthest amount that each bootstrapped curve is from the estimated DET, \hat{R} . We do this in order to determine the largest standardized valued needed to have $(1 - \alpha)100\%$ of the bootstrapped curves within the DET region, \mathfrak{R} . We define ω_m to be:

$$\omega_m = \begin{cases} \min_\theta(e_{m\theta}) & \text{if } |\min_\theta(e_{m\theta})| < |\max_\theta(e_{m\theta})| \\ \max_\theta(e_{m\theta}) & \text{if } |\min_\theta(e_{m\theta})| \geq |\max_\theta(e_{m\theta})| \end{cases} \quad (5)$$

Once we have all M ω_m 's we find the $\alpha/2^{th}$ and the $1 - \alpha/2^{th}$ percentiles of the ω_m 's and they become η_L and η_U respectively. Note that the ω_m 's can take both positive and negative values. In fact, in our experience it is likely that η_L will be negative. Having found η_L and η_U we then build our confidence region \mathfrak{R} , as the region bounded by the splines created by connecting the upper endpoints, $\hat{r}_\theta + \eta_U s_{\hat{r}_\theta}$, and the splines created by connecting the lower endpoints, $\hat{r}_\theta + \eta_L s_{\hat{r}_\theta}$.

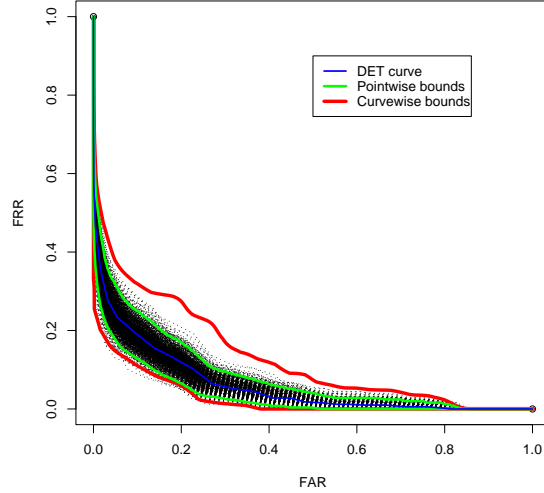


Fig. 2. 95% confidence region for DET of a Hand Geometry System

4 Application and results

In this section we apply the methodology proposed above to three different biometric modalities. These are facial recognition, fingerprints and a hand geometry. These data is described in Ross and Jain [12]. We briefly summarize them here. Biometric captures were taken from 50 individuals. Ten genuine comparisons were made for each individual and five imposter comparisons were made for each cross-comparison pair. All cross-comparison pairs were run. Thus, there were $10 \times 50 = 500$ genuine comparisons and $50 \times 49 \times 5 = 12250$ imposter comparisons. For these regions we followed Adler and Schuckers [10] in establishing the radial center at $(c = 1, c = 1)$. This value was chosen because the radial angles from this center point seem match the typical curvature of a DET curve and, hence, are more likely to be perpendicular to such curves. Note that it is possible to use this methodology to create regions strictly for the FAR or FRR like those proposed by Dass *et al* [11] by allowing the coordinate of the polar center for the error rate of interest to become very small i.e. $(0, -C)$ for some larger number C . We selected T , the number of angles in Θ^* , to be 1000 so that there were 1000 different angles, θ 's, used to create our confidence region. Due to our choice of $c = 1$, θ_L was π and θ_U was $3\pi/2$. We adjusted the radial variance by $\epsilon = 2\bar{n}^{-2}$ where \bar{n} is the average number of comparisons between the genuine and imposter samples. For the data here that is $\bar{n} = 6375$. The use of this ϵ is meant to approximate the use of the $+2/+4$ adjustment suggested by Agresti and Coull [14]. Finally it is worth noting that, in practice, we truncate the final DET confidence region by limiting it to values in the range of $(0, 1) \times (0, 1)$. In some cases the confidence region included negative values because of the width of \mathfrak{R} due to the calculation of η_L and η_U . Note also that since η_L and η_U are global values in the sense that they have to hold for all θ that they do not necessarily reflect the variability at a particular θ . We use 95% for the confidence level in these applications. Figures 1, 2, 3 contain the DET curves and the derived 95% confidence regions for each of these curves. In these figures the center line is the estimated DET, \hat{R} , the next pair of lines moving outward are the pointwise confidence bounds and the last pair of lines are the curvewise confidence bounds which define the confidence region, \mathfrak{R} . The lighter dashed lines represent the $M = 1000$ bootstrapped DET's. Table 1 gives a summary of the percentage of ROC curves that are inside each of these bounds. The confidence region for the

Modality	Pointwise Region	Curvewise Region
Face	64.5%	95.0%
Hand Geometry	62.5%	94.9%
Fingerprint	81.2%	95.0%

Table 1. Percent of bootstrapped ROC curves inside the bounds of 95% confidence regions

facial recognition system and the hand geometry classifier tend to both be symmetric around the EER while the fingerprint region tends to diminish quickly along the FRR axis and very slowly along the FAR axis. The latter effect is due to the larger variability in bootstrapped DET's in that region. This can be seen by the dashed lines, each being a bootstrapped DET, in Figure 3. Thus the DET confidence region accurately reflects the variability in the sample data.

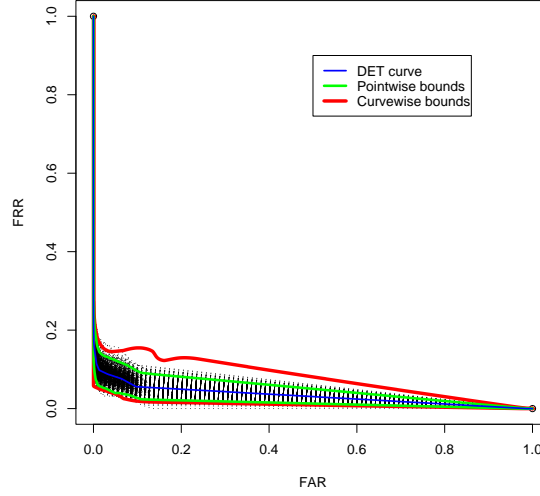


Fig. 3. 95% confidence region for the DET of a Fingerprint System

One byproduct of this methodology is that we can derive a confidence interval for the EER. As mentioned above this quantity is a one that is often reported as an overall measure of performance for a biometric classifier. As with other sample measures, there is a need to quantify the variability in such estimates. To that end, it is possible to create an approximate $(1 - \alpha)100\%$ confidence interval for the EER from the bootstrapped DET's. Since the EER is found at a single θ , there is no need to worry about the multiple comparisons problem outlined above. Consequently, we can simply find the $\alpha/2^{th}$ percentile and the $1 - \alpha/2^{th}$ percentile for the appropriate radius. (In the case when T , the number of angles, is an even number, an approximation can be calculated by taking the average of the two appropriate endpoints.) Table 2 contains 95% confidence intervals for the EER's of the three classifiers described above. Note that the EER for the facial recognition system here is the lowest; however, there is a great deal of variability in that estimate. The width of the error rate is approximately 4.75%. The hand geometry system has the highest estimated EER and the upper bound

on that EER is approximately 18%. Given that we are dealing with non-Gaussian distributions it is not a surprise that the confidence regions are asymmetric. It is perhaps surprising that the facial recognition system outperforms the fingerprint system in this particular application; however, the larger overlap in these two distributions suggests that their EER's are not significantly different.

Modality	Lower Bound	Equal Error Rate	Upper Bound
Face	0.0384	0.0575	0.0847
Hand Geometry	0.1168	0.1494	0.1810
Fingerprint	0.0430	0.0715	0.0972

Table 2. Pointwise 95% EER confidence intervals for three modalities

5 Discussion

This paper proposes a methodology for a curvewise DET confidence region. The region and the methods that created it are flexible and adhere to the variability in the sample data rather than being based upon a fixed width. We achieve this confidence region by converting DET curve data to polar coordinates and calculating the variability in the polar radii at a larger number of angles based upon bootstrapping of the original data. Additionally we have shown how to produce an EER confidence interval using a pointwise approach. It is clear from the results here that a pointwise approach to DET estimation yields regions that are too small. This work also has extensions beyond the confidence regions described here. The region defined here is a two-sided region. A one-sided region which would yield a single upper confidence bound for a DET would be straightforward to create. (One would create that bound by using $\hat{r} + \eta s_{r_\theta}$ where η would be based upon the $(1 - \alpha) \times 100^{th}$ percentile). Another extension would be to compare the confidence region based upon a sample to a hypothesized DET curve. This would allow the acceptance or rejection of that hypothesized curve based upon whether or not the curve fell within the confidence region. Dass *et al* [11] tests a hypothesized curve with their univariate error rate approach. Likewise a corresponding approach could be taken for testing of an EER. Lastly, we note that all of this work has been done on axes defined by the FAR and FRR. There are equivalent confidence regions for any one-to-one transformations of the axes and these regions will have the same properties as the confidence regions described here.

Acknowledgements This work is supported by NSF grants CNS-0325640 and CNS-0520990 (Dr. Schuckers) which is cooperatively funded by the National Science Foundation and the United States Department of Homeland Security and by NSERC Canada (Dr. Adler).

References

1. Tony Mansfield and James L. Wayman, “Best practices in testing and reporting performance of biometric devices,” on the web at www.cesg.gov.uk/site/ast/biometrics/media/BestPractice.pdf, 2002.
2. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, John Wiley & Sons, 1966.
3. Norman Poh and Samy Bengio, “Database, protocol and tools for evaluating score-level fusion algorithms in biometric authentication,” *Pattern Recognition Journal*, 2005.
4. X.-H. Zhou, D. K. McClish, and Nancy A. Obuchowski, *Statistical Methods in Diagnostic Medicine*, John W. Wiley & Sons, 2002.
5. J. Hernández-Orallo, C. Ferri, N. Lachiche, and P. A. Flach, Eds., *ROC Analysis in Artificial Intelligence, 1st Int. Workshop, ROCAI-2004*, Valencia, Spain, 2004.
6. Sofus A. Macskassy, Foster J. Provost, and Saharon Rosset, “Roc confidence bands: an empirical evaluation,” in *ICML*, Luc De Raedt and Stefan Wrobel, Eds. 2005, pp. 537–544, ACM.
7. G. Campbell, “Advance in statistical methodology for the evaluation of diagnostic and laboratory tests,” *Statistics in Medicine*, vol. 13, pp. 499–508, 1994.
8. G. Ma and W. J. Hall, “Confidence bands for receiver operating characteristics curves,” *Medical Decision Making*, vol. 13, pp. 191–197, 1993.
9. Norman Poh, Alvin Martin, and Sammy Bengio, “Performance generalization in biometric authentication using joint user-specific and sample bootstraps,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear.
10. Andy Adler and Michael E. Schuckers, “Calculation of a composite DET curve,” in *Proceedings of Audio- and Video-based Biometric Person Authentication 2005 (AVBPA2005)*, Takeo Kanade, Anil Jain, and Nalini K. Ratha, Eds. 2005, vol. 3546 of *Lecture Notes in Computer Science*, pp. 279–288, Springer.
11. Sarat C. Dass, Y. Zhu, and Anil K. Jain, “Validating a biometric authentication system: Sample size requirements,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 19–30, 2007.
12. Arun Ross and Anil K. Jain, “Information fusion in biometrics,” *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, September 2003.
13. Ruud M. Bolle, Nalini K. Ratha, and Sharath Pankanti, “Error analysis of pattern recognition systems – the subsets bootstrap,” *Computer Vision and Image Understanding*, vol. 93, pp. 1–33, 2004.
14. Alan Agresti and Brent A. Coull, “Approximate is better than “exact” for interval estimation of binomial proportions,” *The American Statistician*, vol. 52, no. 2, pp. 119–126, 1998.