# Biometric features and privacy
**Condemned, based upon your finger print**

Jan-Willem Bullée & Raymond Veldhuis

University of Twente
Faculty of EEMCS
P.O. Box 217, 7500 AE Enschede, The Netherlands

j.h.bullee@student.utwente.nl

**Abstract**

What information is available in biometric features besides that needed for the biometric recognition process? What if a biometric feature contains Personally Identifiable Information? Will the whole biometric system become a threat to privacy? This paper is an attempt to quantifiy the link between biometrics and privacy. For a number of biometric-personal(ity) combinations, the availability, detectability and retrievability of Personally Identifiable Information from biometric features is calculated. This paper should make the reader more aware of the possibilities in this area and inspire further research. By the use of a meta-analysis, the possible risks of the fingerprint biometric and three personal(ity) traits are inventoried. Based on $d'$-values the retrievability of Personally Identifiable Information is determined. The results show that in all three possible biometric-personal(ity) combinations Personally Identifiable Information can be retrieved better than 50/50 guessing. This implies storing biometric data in the clear can be a threat to privacy.

# 1 Introduction

Biometrics authentication has the advantage, over tokens and keys, that biometrics cannot be forgotten, and disclosure to a third party is difficult. In the case of a lost token, this can easily be replaced by providing a new password or changing the lock. If someone steals your biometric information- by taking your finger print from a shiny surface or a downloaded database - this is not replaceable, and you have a problem. Your biometric information is yours, uniquely for you, irreplaceable, almost unchangeable and provided only once.

All biometric information is classified as Personally Identifiable Information (PII). PII is all information that can be used to distinguish, trace, or be linked to an individual. This includes a wide range of information, for example; name, social security number or information that can be linked to race, geographical indicators and educational information. Included in PII are also personal characteristics, facial pictures, fingerprints and other biometric information or stored template data [7]. The treatment of PII is described in the Federal Law. This law is at least applicable for Canada, the United States of America, the member states of the European Union, the United Kingdom and Ireland.

The importance of protecting biometric information lies in its sensitivity and traceability to an individual. Jain describes eight possible places to attack within a generic biometric system [6]. One of the places of interest is just after the sensor level, during the communication to the next module. The sensor is the earliest place in the system that comes in touch with PII. This implies that the whole biometric system becomes privacy sensitive and needs to be handled with special care.

The patterns in biometrics are unique and thus can be used to determine our identity. What if a biometric contains more information than needed for the identification process? For example, information that also can be classified as PII. If this is the case, threats to the privacy of a user or a whole group can occur. To eliminate this threat, in an ideal situation, all PII needs to be removed as early as possible, preferably at the sensor level.

Imagine a fingerprint authentication system with knowledge of biometrics and gender differences. If this is a corrupt system, it can discriminate on gender, and only allow females to pass. This may seem like an innocent example. Imagine however the biometric system would know if you used drugs the last two days or reveals something about your sexual orientation based on the retinal veins.

In this research, a meta-analysis is used to gather information about the biometric and personal(ity) topics. Publications from different domains are used in the analysis. The analysis is based on the descriptive statistics and the statistical conclusions of correlations between biometric features and the personal(ity) traits. This research aims to provide a helicopter view of the field where the biometrics correlate with the personal(ity) traits from a privacy perspective. We use the main research question: "*Can a biometric feature reveal Personally Identiable Information about that person?*".

A personality trait can become visual in certain parts of biometric features. To test this research question, the finger print biometric is analysed and the correlation with a three personal(ity) traits are shown. The main research question will be answered with the use of two subquestions.

*Q1:* Is the PII present in quantitative features? Which biometric and which specific feature contain PII?.

*Q2:* What is the probability density function (retrievability) on a specific biometric feature and PII?

An overview of possible privacy threats by biometric features will be provided, which should make people reconsider privacy and inspire further research. The current research is far from complete and could be extended in many ways. Thus, this paper is not about the latest high-tech discoveries or identification methods. This initial research shows some more relations between biometrics and personality.

# 2 Materials and Methods

## 2.1 Literature

This analysis is based on papers from different research fields. All these fields touch on biometrics and personal(ity) traits, and all the papers are written in different styles specific for that field or journal. The fields of Biometrics, Anthropology, and Forensic are involved in this analysis. To be included in the analysis, a paper had to be written in the English or Dutch language, and a full-text be available to us.

The general method for finding studies was a search in the The University Library, Google Scholar, ScienceDirect, and JSTOR for articles. The aim was a search for information about the relation between biometrics and personal(ity) traits.

The information needed about the biometric traits is preferably in the quantified form. In this way information statistics can be used to describe the biometric and determine the prevalence of the personal(ity) trait. Studies that are vague about the test sample and size, are excluded from the analysis.

## 2.2 Data Analysis

### 2.2.1 Biometric Data Description

The physiological biometric feature chosen in this research is the fingerprint. Biometrics in general have multiple visual properties that can be measured. One property of measurement is chosen for comparison. Note that if a biometric appears not to correlate to a personal(ity) trait, this only applies for this specific method of measurement. The property used for fingerprints used is the Total Ridge Count (TRC). This is the count of ridges on a square of $25mm^2$ on the tip of the finger. It is not overwhelmingly clear if this square is rotated for higher ridge count. The ridges are counted on a line from one of the corners to one of the diagonally opposite corners. In general, the average ridge count of one hand is used.

An alternative method that is used to determine the TRC, uses the *core point* and the *triradial point*, which is also known as the *delta point*. Here the ridges are counted between these two points. Unfortunately, this method fails in cases where no or multiple *triradial points* are found. In the case of an *arch* pattern as a fingerprint this method will also fail, because there is no *triradial point* [10, 16]. Due to this, the related papers are only limitedly used in the analysis.

### 2.2.2 Description of used Personally Identifiable Information

The Personally Identifiable Information(PII), investigated in combination with the fingerprint data is:
**Gender**: What is someone's sex, or better specified as the biological gender of a person divided into male and female.
**Ethnicity**: What ethnicity someone has. This is roughly split in African Americans, Caucasians, Orientals, and Latinos. In some cases 'sub'-ethnicities are used, for example Pakistani and Malaysian.
**Sexual Orientation**: The sexual preference of a person. Here, only the distinction is made between heterosexual and homosexual males and females.

### 2.2.3 Data Presentation

All articles were analyzed for biometric, personal(ity) trait, method of gathering and ethnicity. Also important is the sample distribution information ($\mu$ and $\sigma$). The difference between two samples is calculated and illustrated by the $d'$-value. The $d'$ can be calculated by Equation 1, in case of two equal standard deviations ($\sigma_1 = \sigma_2$) replace the denominator with $\sigma$ [18, 19, 20].

$$d' = \frac{(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{2}}} \tag{1}$$

The $d'$ measures the sensitivity between two conditions/groups [20]. The $p$-value shows if there is a difference between the average of two groups. From the distribution information, the $d'$ and Receiver Operating Characteristic (ROC) curve, can be constructed. The retrieval rate can be found by calculating the area under the curve (AUC) of the ROC and is a value between the 0.5 and 1. The bigger the number, the better the result. The ROC's can also be expressed as the true positive rate versus the false positive rate. In this case, when a certain level of exception is wanted, what is the probability of a wrong gender conclusion?

For a better understanding of the topic, an example is described more in-depth: The relationship between the TRC of a fingerprint and the gender of a person. The (sub)hypothesis is: "*females have a higher ridge count than men*". Females have more fine grained bodies than men, thus smaller finger width and smaller ridges, which gives a higher ridge count on a fixed $25\text{mm}^2$ area.

# 3 Results

## 3.1 General results

The outcome of the data-analysis, is shown in Table 1. The table contains the statistical conclusions, the predictability expressed in $d'$ and the AUC for all studies found.

The PII Gender shows a relation with fingerprints. In this specific relation, all results show significant correlations. The $d'$ varies in each specific method and therefor so does the AUC. Ethnicity and Orientation are less strongly related to biometric features. Only one significant result was found and both the $d'$ and AUC are low. This might be due to do the size of the research field or the ethical problems encountered.

## 3.2 Gender recognition based on the Finger Print Ridge Count

In the literature, there are four methods described for collecting the total ridge count of a finger. These are; the *25mm²-method*, the difference between the *core point* and the *triradial-* or *delta point*, the *stretched single centimeter* and the method described by Penrose [9]. There are also two publications where no method is described [15, 17]. An overview of the results can be found in Table 1.

All the found $p$-values are significant and a weighted average $d'$ is 1.68. This indicates that at a hit rate of .8, there is a false hit rate around .2. The results for the *25mm²-method* are really good. The $p$-values are all significant, all except one are even lower than 0.001. The weighted $d'$ is bigger than 2. All, except one, show a high effect size $r$ ($> .8$). A possible explanation for the single research with low results might be its limited sample size.

The *delta-method* has significant $p$-values, but the $d'$ values are low. This is also applicable for the *Penrose-method*. This means that there is a difference between the average of the two groups, but the overlap between the two distributions is too big to successfully identify the gender of a random observation from the sample. The *stretched centimeter method* has a large $d'$, but no $p$-value was available.

The total sample contains participants from several ethnic samples, with an age ranging from 5 till 67 years old. The results show that, as was expected, the ridge count in all populations is higher for the females than for males. However, there are a few exceptions where it is the other way around. A suitable explanation for these occurrences is not found. On the other hand, in some of the research it is not described how the researchers acquired the finger prints or obtained other information about the sample [5, 9, 10, 15, 17]. It is notable that two of the "exceptions" both used the Triradial-method [5, 10]. Another conspicuous observation can be distinguished in the Spanish sample. They have the most fine grained fingertips by far, both on male and female.

Thus, the eight studies found which used the *25mm²-method*, all have high significant results, but can these results be based on coincidence? Meta-analysis can be biased by the file-drawer effect. This is caused by research papers with insignificant results which don't get published and end up in file drawers or archives instead of in journals. Rosenthal's "fail-safe" $N$ is a measure that indicates this number of studies. Formula 2 calculates this number of articles that is needed to bring the overall result to a critical level of significance. Here $N$ is the number of studies in the analysis, $Z_c$ is the critical value of $Z$ and $\bar{Z}_0$ is the mean $Z$ obtained for the $N$ studies [14].

$$N_{fs} = (N/Z_c^2)(N\bar{Z}_0^2 - Z_c^2) \tag{2}$$

For the *25mm$^2$-method*, this value is calculated at 137. So there must be 137 studies with no results to make all the already found results ineffectual. By the rule of thumb the minimal fail-safe $N_{fs}$ has to be $5 \times N + 10$. In our case this would be a minimal number of 50 studies. This shows that our analysis is in the safe zone.

## 3.3 Privacy threat populations and biometrics

After an in depth example where the precision of PII retrieval from biometric features is described, an overview of the best scoring biometric/personal(ity) combinations follows. The best scoring features are in fact the biggest threats to privacy and are marked in yellow, see Table 1.

Table 1: Fingerprint - Statistical Infromation

| Pop | N(♂/♀) | $\mu$ | $\sigma$ | $d'$ | $P$ | AUC | Ref |
|---|---|---|---|---|---|---|---|
| **Gender** | | | | | | | |
| *25mm$^2$-method* | | | | | | | |
| AA | 100/100 | 10.90/12.61 | 1.15/1.43 | 1.32 | Sig*** | 0.824 | [1] |
| Cau | 100/100 | 11.14/13.32 | 1.31/1.24 | 1.71 | Sig*** | 0.866 | [1] |
| IN | 40/40 | 13.18/13.53 | 2.74/2.90 | 0.12 | Sig* | 0.535 | [2] |
| ES | 100/100 | 16.23/17.91 | 1.39/1.47 | 1.17 | Sig*** | 0.797 | [3] |
| IN | 250/250 | 12.80/14.60 | 0.90/0.09 | 2.82 | Sig*** | 0.927 | [8] |
| IN | 100/100 | 11.05/14.20 | 1.11/0.63 | 3.48 | Sig*** | 0.993 | [11] |
| Mix | 150/150 | 11.63/13.98 | - | - | Sig*** | | [12] |
| CN | 100/100 | 11.73/14.15 | 1.07/1.04 | 2.30 | Sig*** | 0.948 | [12] |
| MY | 50/50 | 11.44/13.63 | 0.99/0.91 | 2.31 | Sig*** | 0.949 | [12] |
| *Delta-method* | | | | | | | |
| UK | 825/825 | 14.50/12.72 | 5.11/5.25 | 0.34 | Sig*** | 0.596 | [5] |
| US | 429/457 | 12.48/11.34 | 3.16/3.64 | 0.33 | Sig*** | 0.593 | [10] |
| *Penrose-method* | | | | | | | |
| IR | 100/100 | 16.65/16.09 | 1.98/1.92 | 0.29 | Sig* | 0.580 | [9] |
| *Stretched Single Centimeter* | | | | | | | |
| US | 100/100 | 10.35/12.7 | 0.63/0.83 | 3.20 | - | 0.988 | [13] |
| *No Method Described* | | | | | | | |
| GR | L(43/52) | 13.34/12.34 | 7.4 /5.9 | 0.15 | - | 0.542 | [15] |
| GR | R(43/52) | 14.60/12.99 | 7.0/5.8 | 0.25 | - | 0.570 | [15] |
| IN | 250 | 14.36/10.65 | - | - | - | - | [17] |
| IN | 75 | 14.26/12.50 | 2.22/2.87 | 0.68 | Sig*** | 0.686 | [17] |
| **Ethnicity** | | | | | | | |
| Cau/AA | ♂(100/100) | 11.14/10.90 | 1.31/1.15 | 0.19 | NS | 0.555 | [1] |
| Cau/AA | ♀(100/100) | 13.32/12.61 | 1.24/1.43 | 0.53 | Sig* | 0.646 | [1] |
| CN/MY | ♂(100/50) | 11.73/11.44 | 1.07/0.99 | 0.28 | - | 0.579 | [12] |
| CN/MY | ♀(100/50) | 14.15/13.63 | 1.04/0.91 | 0.53 | - | 0.647 | [12] |
| **Sexual orientation** | | | | | | | |
| *Difference Triradial & Core Point (Hetrosexuals / Homosexuals)* | | | | | | | |
| Mix | ♂(186/66) | 33.0/32.6 | 8.6/7.7 | 0.05 | NS | 0.514 | [4] |
| US | ♂(169/164) | 63.0/61.3 | 15.6/16.2 | 0.11 | NS | 0.530 | [10] |
| US | ♀(117/164) | 57.7/55.2 | 18.7/18.7 | 0.13 | NS | 0.538 | [10] |

$^*p < .05.$ $^{**}p < .01.$ $^{***}p < .001.$

# 4 Discussion

This literature review aimed to explore the privacy threats within biometric features. Each result shows a better retrieval rate than guessing. The finger-gender combination shows the best result, so there is evidence that confirms the hypothesis. Not all results show a significant relation. However, this is not a problem, since if even one study exists which shows a significant result, a threat occurs. This means there is a threat for that particular group, with the use of a certain method for that biometric to that specific personal(ity) trait. Although it is not generalizable, it applies for a smaller group.

Some results are more convincing than others, but there are only a few results where almost all studies found no correlation between the biometric feature and the personal(ity) trait. This can be the result of the single way of measurement for a biometric. Perhaps, there is a correlation available for a biometric, but only when there is a different method of measurement. For the fingerprint, we only used the TRC as method, but the fingerprint has more characteristics such as the patterns of the print (Henry Classification), the Minutiae patterns and perhaps an another (not yet found) method.

## 4.1 Finger/Gender

The (sub)research question "*have females a higher ridge count than men?*" can, based on Table 1, be validated. This applies for the *25mm*$^2$*-method*. All of the studies showed a significant $p$-value, which means that the there is a difference between the average of the two groups. An average $d'$-value of more than 2 and thus a high AUC indicates that the distributions do not overlap much. This means that at a hit rate of .8, there is a miss rate of .09. Besides, the fail-over $N$ shows that this is a stable measurement.

Overall, the ridges are thinner in detail for females, and thus have higher ridge density compared to males. We have seen that female fingerprints tended to have thinner epidermal ridges. The average fingerprint has, depending on the sample, 13 ridges / 25mm$^2$. A lower ridge count is more likely to be masculine, while a higher ridge count has a higher probability to be feminine.

No suitable explanation was found why females in five samples had lower ridge counts than males. As mentioned, in the Results section this can be method related. Another possibility, in the case of the study of Mustanski, is the usage of a mixed sexual orientation sample [10]. In other studies this is not explicitly mentioned. So this can partly imply that, under the assumption that sexual orientation is hormonally based, the construction of a fingerprint is not fully based on gender, but also based on hormones.

Based on the results of the different studies, it can be concluded that it is possible to determine if a fingerprint is male or female. This is based on quantification of the average total ridge count, compared to an overall average. It must be noted that this only works well if you know from which population the print originates. The population, or the continent, should be a parameter in the test on detectability, to determine gender from a fingerprint.

# 5 Future work

The current research can be extended by adding categories in biometrics or in the measurement of the different biometric characteristics. An extension in personal(ity)'s is also possible, think of diseases or habits such as smoking, drinking, and drug use. There is a reasonable chance a correlation exists between smoking, drinking, or drug use and vein thickness, or the retina vein diameter. Another example is the relationship between the palm print and Down Syndrome.

## 5.1 Other use

Apart from the privacy threat, it can be a useful tool as well. For instance to make people aware of the possibility they have a disease, such as Diabetes. In the Netherlands, there are 740.000 patients who have Diabetes, and probably 250.000 people who have it, but are not aware of this[a].

# References


[1] Mark A. Acree. Is there a gender difference in fingerprint ridge density? *Forensic Science International*, 102(1):35 – 44, 1999.

[2] Angela Bell. Loop ridge count differences between genders. LOOP RIDGE COUNT DIFFERENCES, 2008.

[3] Esperanza Gutiérrez-Redomero, Concepción Alonso, Esther Romero, and Virginia Galera. Variability of fingerprint ridge density in a sample of spanish caucasians and its application to sex determination. *Forensic Science International*, 180(1):17 – 22, 2008.

[4] J. A. Y. Hal and D. Kimura. Dermatoglyphic asymmetry and sexual orientation in men. *Behavioral Neuroscience*, 108(6):1203–1206, 1994.

[5] Sarah B. Holt. Genetics of dermal ridges: Frequency distributions of total finger ridge-count. *Annals of Human Genetics*, 20(2):159–170, 1955.

[6] Anil K. Jain, Karthik Nandakumar, and Abhishek Nagar. Biometric template security. *EURASIP J. Adv. Signal Process*, 2008:1–17, 2008.

[7] Erika. McCallister, Tim. Grance, Karen Scarfone, National Institute of Standards, and Technology (U.S.). *Guide to protecting the confidentiality of Personally Identifiable Information (PII): recommendations of the National Institute of Standards and Technology. Special Publication 800-122.* U.S. Dept. of Commerce, National Institute of Standards and Technology, Gaithersburg, MD, 2009.

[8] S. Gungadin MD. Sex determination from fingerprint ridge density. *Internet Journal of Medical Update*, 2(2), 2007.


[a] `http://www.diabetesfonds.nl/artikel/diabetes-cijfers`

[9] M. Mehdipour and D.D. Farhud. A dermatoglyphic study of iranian muslims part i: Finger patterns and ridge-counts. *Iranian Journal of Public Health*, 7(4):196 – 206, 1978.

[10] Brian S. Mustanski, J. Michael Bailey, and Sarah Kaspar. Dermatoglyphics, handedness, sex, and sexual orientation. *Archives of Sexual Behavior*, 31(1):113–122, 02 2002.

[11] Vinod C. Nayak, Prateek Rastogi, Tanuj Kanchan, Stany W. Lobo, K. Yoganarasimha, Smitha Nayak, Nageshkumar G. Rao, G. Pradeep Kumar, B. Suresh Kumar Shetty, and Ritesh G. Menezes. Sex differences from fingerprint ridge density in the indian population. *Journal of Forensic and Legal Medicine*, 17(2):84 – 86, 2010.

[12] Vinod C. Nayak, Prateek Rastogi, Tanuj Kanchan, K. Yoganarasimha, G. Pradeep Kumar, and Ritesh G. Menezes. Sex differences from fingerprint ridge density in chinese and malaysian population. *Forensic science international*, 197(1):67–69, 04 2010.

[13] Edwin A. Ohler and Harold Cummins. Sexual differences in breadths of epidermal ridges on finger tips and palms. *American Journal of Physical Anthropology*, 29(3):341–362, 1942.

[14] Robert G. Orwin. A fail-safe $n$ for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics*, 8(2):157–159, 1983.

[15] D. F. Roberts, V. Luttrell, and C. Pasternak Slater. 6. finger prints in a greek sample. *Man*, 65:21–22, 1965.

[16] Geoff Sanders and Flavie Waters. Fingerprint asymmetry predicts within sex differences in the performance of sexually dimorphic tasks. *Personality and Individual Differences*, 31(7):1181 – 1191, 2001.

[17] P.R. Sharma, A. K. Gautam, and P. K. Tiwari. Dermatoglyphic variations in five ethno-geographical cohorts of indian populations: A pilot study. *The Internet Journal of Biological Anthropolog*, 2(1), 2008.

[18] A. J. Simpson and M. J. Fitter. What is the best index of detectability?. *Psychological Bulletin*, 80(6):481 – 488, 1973.

[19] John A. Swets. Form of empirical rocs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99(2):181 – 198, 1986.

[20] John A. Swets. Indices of discrimination or diagnostic accuracy: Their rocs and implied models. *Psychological Bulletin*, 99(1):100 – 117, 1986.