TECHNICAL REPORT

ISO/IEC TR 22116

First edition

Information technology — A study of the differential impact of demographic factors in biometric recognition system performance

PROOF/ÉPREUVE

Reference number
ISO/IEC TR 22116:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

Automated systems (including biometrics) are increasingly used in decision-making processes. In recent years, systemic performance differentials reflected in several automated decision systems have been reported and hotly debated. In the context of this report, an algorithm exhibiting performance differentials produces statistically different outcomes or decisions for different groups of individuals, for example, based on gender, age, and race/ethnicity. In the context of biometric recognition, this means that probabilities of false positives and/or false negatives can differ among the demographic groups. The impacts of such performance differentials on the affected individuals can range from mere inconvenience in cooperative access control systems, to consequential harms such as varying arrest rates for certain demographic groups based on decisions produced by facial recognition systems.

Although such systems are almost certainly not designed to be explicitly differential against any group, implicit differences can occur independently of the intentions of the system designers. They can be exhibited and propagated at many stages of the decision-making pipeline, including but not limited to training data itself as well as the data processing. Due to the scalability of such systems, a higher quantity of erroneous or inaccurate decisions can be generated than in the typical, human-based processes. Consequently, in recent years, measuring and ensuring the fairness (i.e. lack of differential performance) of such systems has often been discussed in the media and political circles, with research and commercial interest increasing accordingly. With increasing deployments of the technology, it is important to consider whether it performs similarly for all users. This document helps to identify where recognition performance differences related to demographic factors can exist in biometric systems.

Information technology — A study of the differential impact of demographic factors in biometric recognition system performance

1 Scope

This document introduces the effects of population demographics on biometric functions. It:

— establishes terms and definitions relevant to the study of demographic factors in biometric recognition system performance;

— identifies areas where biometric systems can exhibit different performance based on different demographic factors of the individuals submitting the biometric samples;

— explains how different demographic factors can influence the biometric characteristics captured by different biometric modalities and how these influences can affect biometric performance measures;

— presents a case study on existing scientific material that explores the impact of demographic factors on biometric system performance. Only biometric modalities where quantitative information is available on the impact of demographic factors are considered.

Outside of the scope of this document are:

— effects of disease and injury on biometric performance; and

— how religious and cultural norms can affect biometric operations.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 2382-37, *Information technology — Vocabulary — Part 37: Biometrics*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-3 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
age
length of time an individual has lived

3.2
ageing
natural progression of an individual's characteristics over time

Note 1 to entry: The impact of ageing will vary at different ages.

3.3
categorical demographic variable
demographic characteristic that is nominally or ordinally described

EXAMPLE Gender categories consist of "Male", "Female", or "Other".

3.4
detection error trade-off
DET
relationship between false positive and false negative errors of a binary classification system as the discrimination threshold varies

3.5
differential performance
differences in system variables or system processing between different demographic groups

EXAMPLE Differences in comparison scores, feature-level fusion, and/or image-level fusion.

3.6
differential outcomes
difference in system results between different demographic groups

EXAMPLE Differences in match rate.

3.7
ethnicity
state of belonging to a group with a common origin, set of customs or traditions

Note 1 to entry: Scientifically, race can be defined as a group of humans that share biological features related to genetic ancestry. Practically, race is primarily a social construct, i.e. not related to biology but instead related to self-identity. Ethnicity is frequently used as a proxy for self-reported "race". In the context of this report, "ethnicity" and "race" are considered interchangeable and can both be taken to mean a social identity that was reported or assigned to a particular subject. These reported values can, but do not always, correlate with underlying genetic features.

3.8
failure-to-acquire
FTA
failure-to-accept for subsequent comparison the output of a biometric capture process, a biometric sample of the biometric characteristic of interest

3.9
failure-to-enrol
FTE
failure-to-create and store a biometric enrolment data record for an eligible biometric capture subject in accordance with a biometric enrolment policy

3.10
false accept rate
FAR
proportion of transactions with false biometric claims erroneously accepted

3.11
false match rate
FMR
proportion of the completed biometric non-mated comparison trials that result in a false match

3.12
false negative differential
tendency for mated biometric samples from subjects in one demographic group not to match relative to another demographic group

3.13
false non-match rate
FNMR
proportion of the completed biometric mated comparison trials that result in a false non-match

3.14
false positive differential
tendency for non-mated biometric samples from one demographic group to falsely match relative to another demographic group or a tendency for this effect to occur across demographic groups

3.15
false reject rate
FRR
proportion of biometric transactions with true biometric claims erroneously rejected

3.16
gender
classification as male, female or another category based on social, cultural, or behavioural factors

Note 1 to entry: Gender is generally determined through self-declaration or self-presentation and can change over time.

Note 2 to entry: Depending on jurisdiction recognition, it is possible that gender will or will not require assessment by a third party.

3.17
phenotypic demographic variable
observable demographic characteristic of an individual resulting from the interaction of that individual's genotype and the environment

EXAMPLE An individual's skin reflectance.

3.18
sex
state of being male or female as it relates to biological factors such as DNA, anatomy and physiology

4 Symbols and abbreviated terms

For the purposes of this document, the symbols and abbreviated terms given in ISO/IEC 2382-37 apply.

5 Understanding demographic factors in biometric systems

5.1 Introduction

Demographic factors include any characteristics or attributes that apply to a specific group within a population[1]. There are potentially an infinite number of different demographic factors that can be considered in terms of how they impact biometric systems. In order to maintain a manageable scope for this document, the demographic factors considered are limited to those where at least some research is available to evaluate the impact on biometric comparison performance, quality score, failure-to-enrol rate or other performance metrics[1]. The state-of-the-art of biometric system performance can change rapidly, and performance can improve by an order of magnitude over the course of just a few years. Therefore, performance results observed in this report can be overcome or obviated by more recent studies. It is valuable to continuously monitor for publications that provide new insights into aspects of

differential performance. Specifically, this document considers the following demographic factors that have been shown to impact the performance of biometric systems (see Clause 3 for definitions):

— Age and ageing

— Ethnicity

— Gender and sex

The concept of differential performance is closely related with the long-studied biometric concept of the biometric menagerie[2] which divides individuals into individual "animal" categories based on the profile of their mated and non-mated score distributions.

5.2 Biometric system components

The functions of the generalized components of a biometric system, applied in ISO/IEC 19795-1 and ISO/IEC 2382-37, can be affected by the demographics of the target population. Those functions are:

— **Acquisition:** Some applications detect, localize, and acquire biometric characteristics, for example counting faces in a crowded area. Acquisition accuracy can depend on demographics. Some systems can include biometric sample quality estimation in the acquisition process, which can also be affected by the demographics of the target population. A failure-to-acquire (FTA) can also impede downstream biometric processes.

— **Enrolment:** For certain individuals it can be difficult enrolling in a biometric system. This can be caused by sensor or system properties, or by innate characteristics of these individuals, including characteristics associated with demographics. In such cases, the failure-to-enrol (FTE) rate can vary by demographic group.

— **Verification:** In verification processes such as access control, accuracy (false rejection rates or false acceptance rates) can vary. One potential reason for such a variation are demographic effects. In addition, biometric performance can be affected by failure-to-acquire rates which can deviate for different demographic groups.

— **Identification:** In identification systems demographics can affect both false negative and false positive identification rates. Unlike verification applications, demographic effects in identification can consider the demographics of the probe and the gallery biometric samples. Again, biometric performance can be affected by failure-to-acquire rates influenced by demographic properties.

5.3 The influence of demographics on biometric recognition

5.3.1 The influence of sex and gender

Biometric comparison performance, enrolment success and other aspects of performance can be affected by sex and by gender. As explained in Clause 3, these two terms are distinct. For the purposes of statistical analysis when correlating demographic factors with biometric performance, gender is usually recorded based on the gender (recorded as "sex" on most government issued documents; male or female or undefined) listed on the identity documents belonging to an individual.

The broadly generalizable influence of gender on biometric performance can be difficult to characterize as practices associated with gender vary. For example, manual labour can result in friction ridge degradation over time that can impact all phases of biometric operations. However, these activities are associated with different gender identities in different geographic locations. Since manual labour can result in degradation of the quality of fingerprints and lower fingerprint comparison performance, this is an example of a demographic factor that is related to gender but can be more significant for males in some cultures and for females in others.

5.3.2 The influence of age and ageing

Biometric performance can vary substantially with a person's age. Physiological properties such as skin elasticity and bone structure change with age, meaning the process of ageing can change a subject's biometric characteristics, causing a general decline in comparison performance[4][5][6]. Behavioural aspects of very young or very old subjects can impact the comprehension of instructions, thereby affecting the usability of a biometric system[7][8].

Comparison performance depends on both age and ageing. For example, enrolment and subsequent verification of fingerprints can be much less accurate for children than adults[4][5][6].

5.3.3 The influence of race and ethnicity

The performance of some biometric modalities has been shown to vary with race/ethnicity. This is particularly true for modalities and algorithms where biometric features depend on anatomical traits formed under genetic expression. Most studies seem to use broad categories of ethnicity based on the country of origin or nationality of an individual, or sometimes based on their personal declaration of ethnic identity. In a few studies, images of the individuals have been examined and classified by a human into specific ethnic groups. Both self and group classifications are likely sub-optimal in terms of consistency and description of the underlying physiological or behavioural effect.

5.4 Measurement and analysis

Given sufficient amounts of subject-specific demographic data and associated biometric processing results (e.g. comparison scores, transaction times, or recognition decisions), an analysis can be performed to expose the effects of demographics on biometric processing outcomes. At a high level, in an investigation of demographic effects in biometric systems, it is important to consider answering the following questions, prior to execution:

1) Is this study investigating effects in the mated distribution (false negative differentials) or non-mated distribution (false positive differential)?

2) Is this study investigating differential performance (i.e. score level, feature level fusion, low or image level fusion, etc.) effects or differential outcome (i.e. FNMR, TIR, etc.) effects?

3) Is this study using self-reported, categorical demographic variables or more descriptive phenotypes?

4) Are there any uncontrolled demographic variables that are confounded with the demographic variable under study? For example, on average women are shorter than men. In a study comparing the biometric performance of men and women, could any observed effects in the study be due to the confounding variable (height) and not the studied variable (gender or sex)?

Once these questions have been answered, the appropriate study description and statistical techniques for determining the presence and significance of an effect can be selected.

6 Impact of demographic factors on facial recognition systems

6.1 Existing literature on demographic factors impacting facial recognition systems

6.1.1 General notes

There are publications available on the demographic factors impacting the performance of facial recognition systems, based on some databases containing facial images from passports, visa mugshots, and driver's licences. Some factors, such as plastic surgery, the wearing of glasses, use of face covering aids, and significant changes to the face due to makeup or hairstyle are known to affect the comparison performance of facial recognition systems. Such factors are not precisely demographic factors but are influenced by them.