Biometrics: Trust, but Verify

Anil K. Jain, Life Fellow, IEEE, Debayan Deb, Student Member, IEEE, and Joshua J. Engelsma, Student Member, IEEE

Abstract

Over the past two decades, biometric recognition has exploded into a plethora of different applications around the globe. This proliferation can be attributed to the high levels of authentication accuracy and user convenience that biometric recognition systems afford end-users. However, in spite of the success of biometric recognition systems, there are a number of outstanding problems and concerns pertaining to the various sub-modules of biometric recognition systems that create an element of mistrust in their use - both by the scientific community and also the public at large. Some of these problems include: i) questions related to system recognition performance, ii) security (spoof attacks, adversarial attacks, template reconstruction attacks and demographic information leakage), iii) uncertainty over the bias and fairness of the systems to all users, iv) explainability of the seemingly black-box decisions made by most recognition systems, and v) concerns over data centralization and user privacy. In this paper, we provide an overview of each of the aforementioned open-ended challenges. We survey work that has been conducted to address each of these concerns and highlight the issues requiring further attention. Finally, we provide insights into how the biometric community can address core biometric recognition systems design issues to better instill trust, fairness, and security for all.

Index Terms

Trustworthy Biometrics, Recognition Performance, Scalability, Bias and Fairness, Security, Interpretability, Privacy

Fig. 1: Examples where biometrics are introduced for trust. For instance, Amazon employs Amazon One, a biometric recognition system for e-commerce, that lets shoppers pay for their groceries by authenticating them via their palmprints [11]. US-VISIT authenticates international travelers to the United States via their fingerprints [12]. “Touchless” authentication via face recognition is being increasingly employed for entry, exit, and flight boarding [3] for airport security.

1 INTRODUCTION

The Digital Age we live in has accelerated a proliferation of sensitive and personal data needing absolute protection. For instance, most of us now carry access to our bank account, email, business dealings, private message history, personal videos and photos, and much more all within a few taps on the smartphones in our pockets. It goes without saying that such data needs to be secured at all times. At the same time, users want the convenience of being able to access such data in a seamless and safe manner. It is therefore not surprising that virtually all smartphones now come equipped with a biometric authentication system (either face or fingerprint) for highly accurate and convenient unlocking of our phones. In addition, every day, a variety of organizations pose identity-related questions such as, Should John be granted a visa?, Does Alice already have a driver’s license?, and Is Cathy the owner of the bank account?
Consequently, the use of biometric recognition systems has now pervaded into the lives of billions of human-beings all around the globe through a variety of applications (Figure 1).

Biometric recognition, or simply biometrics, refers to automatic person recognition based on an individual’s physical or behavioral traits [13]. The term, Biometrics, is derived from the Greek words bios (life) and metron (measure). Hence, biometrics in the context of person recognition refers to recognition based on measurements of the body (e.g., face, fingerprint and iris). The origin of modern day biometric recognition has its roots in the “Habitual Criminals Act” passed by the British Parliament in 1869 [14]. In particular, the Home Office Committee expressed the need for a reliable person recognition scheme for tracing repeat offenders [15].

“What is wanted is a means of classifying the records of habitual criminal, such as to have as the particulars of the person of any prisoner (whether description, measurements, marks, or photographs) are received, it may be possible to ascertain readily, and with certainty, whether his case is in the register, and if so, who he is.” [15]

In essence, biometrics relies on who you are or how you act as opposed to what you know (such as a password) or what you have (such as an ID card).

Prior to automated biometric recognition systems, reliable identification of fellow beings had been a long-standing problem in human society. In early civilizations, people lived in small, connected communities. However, as humanity became more mobile and populations increased, we needed to start relying on credentials for person recognition. Dating back all the way to ancient Rome, passwords had long been viewed as the ideal method of securing information and gaining access to exclusivity [16]. While passwords may have served their purpose in ancient Rome, in this day and age, passwords, while still in common use, are rife with problems. For example, passwords are prone to social engineering hacks, where someone can access a user’s password by gaining their trust [17]. Alternatively, a malicious individual can observe and log a victim’s typed password characters on a keyboard [18]. Finally, plain-text passwords may be hacked or leaked from an insecure database [19]. Other knowledge-based authentication schemes such as PINs are also prone to such attacks [20]. To combat the limitations imposed by passwords, an alternative authentication scheme involves physical tokens, such as certificates, ID cards, passports and driver’s licenses. Unfortunately, these tokens are also vulnerable to social engineering attacks and theft. Furthermore, in developing countries around the world, many economically disadvantaged individuals lack any type of identification documentation making it difficult for them to access government benefits, healthcare, and financial services. If an individual does possess an official ID document, it may be fraudulent or shared with others [21-23]. Finally, even if identification documentation can be adequately distributed to everyone in a society, it cannot be trusted. For example, Dhiren Barot, an Al-Qaeda fanatic, was issued with nine fake British passports [24].

Not surprisingly, the problems associated with password or token based authentication and identification has led to society exploring a more accurate and reliable method of user authentication and identification management systems which society as a whole can trust. The word “trust” is defined in the Oxford dictionary as 25.

TRUST: “Firm belief in the reliability, truth, ability, or strength of someone or something.”

Thus for biometric recognition to be used in lieu of conventional passwords or as an identity management system, they must be shown to be highly accurate (establishing the reliability and truth portion of the definition) and also robust, or reliable. In other words, biometric recognition systems must be demonstrated to be trustworthy. Subsequently and finally, a firm belief in this trustworthiness must be established with system users to gain their trust.

To date, significant progress has been made in solidifying the accuracy component of a trustworthy biometric recognition system. In particular, while automated biometric recognition systems have now been around for quite some time, recent advances in hardware (e.g., an NVIDIA 3090 GeForce RTX performs at 35.58 TFLOPS) and computer vision algorithms (specifically deep learning [28-30]) have led to biometric recognition systems which now surpass human recognition performance [31]. More specifically, NIST evaluations for fingerprint [32], face [33], and iris [34] search algorithms boast accuracies of FNIR = 0.001, 0.058, and 0.0059 @ FPIR = 0.001, respectively (Table 1).

Although the accuracy and convenience of biometric recognition systems has fueled their replacement of traditional password or token based methods (and more importantly, their widespread use in identity management systems), scientists must begin shifting their attention away from a purely recognition accuracy and convenience driven mindset to concerns voiced by policy makers and the general public about the reliability of biometric recognition systems (first component of the definition of trust). Biometric systems are here to stay and their proliferation in our society will continue to grow. It is also given that biometric systems will make incorrect decisions, albeit small, and, like any security system, will be subjected to attacks by hackers. Therefore, the following concerns must be adequately addressed:

1) Performance: Although biometric recognition system accuracy has matured, are there inputs and ambient noise that will still break the system? How

![Image link](https://www.nvidia.com/en-ph/geforce/graphics-cards/30-series/rtx-3090/)

| Trait     | Evaluation | Gallery Size | Iden. Error | Fingerprint |
|-----------|------------|--------------|-------------|-------------|
| Fingerprint | NIST FpVTE 2012 | 5M²         | 0.001       |
| Face      | Ongoing NIST FRVT | 12M         | 0.058       |
| Iris      | NIST IREX 10 | 500K         | 0.006       |

1 FNIR @ FPIR = 0.001.
2 10-print fusion performance.
will the recognition system perform over time? How will the system scale to millions or even billions of users?

2) **Bias and Fairness:** Does the biometric recognition system work well across all demographic groups? Does the system mis-classify members of one demographic group more than another (e.g., age, gender, race, ethnicity and country of origin)? Why? What are the sources of bias in a biometric recognition system?

3) **Security:** Have biometric recognition systems solved the spoofing (presentation attack) vulnerability? Are biometric recognition systems robust to adversarial perturbations? Can users’ templates stored in the system database be stolen or altered and used to reconstruct a biometric image or glean demographic information? How can we thwart these attack vectors?

4) **Explainability and Interpretability:** Why is the biometric recognition system making the decision it is making? What parts of the input image are being used to make a final decision? What features of the input image are most important in the decision? Will these features enable the model to operate accurately and consistently over time and in different operating conditions?

5) **Privacy:** Even if we have a highly accurate and secure biometric system, how can we protect privacy of end users (and those who are in the training database)? Can we train on decentralized data, e.g., federated learning? Can we perform training or make inference directly on encrypted data? Can the model parameters also be encrypted?

In other words, the trustworthiness of biometric recognition systems must be verified [25].

**VERIFY:** “The process of establishing the truth, accuracy, or validity of something.”

While some work has begun to verify behaviors of biometric recognition systems via studying the aforementioned questions with scientific rigour, we argue that more work remains to be done. To that end, in this paper, we point out each of the major points of attack, question, or concern (Figure 2) on the biometric recognition pipeline. Next, we systematically survey the literature to locate pertinent research aimed at addressing the aforementioned questions. We discuss remaining limitations left by the existing literature. Finally, we summarize recommended steps that can be taken and research that can be pursued (and also how it can be conducted rigorously, fairly, etc.) to build biometric recognition systems which are more trustworthy.

We note that this paper is unique in that it aggregates and examines the main components of a trustworthy biometric recognition system into one manuscript. Indeed many surveys [13, 35–43] have been written in great detail on each one of these topics individually, however we posit that there is benefit in extracting the key points from each of these areas and summarizing them in one place such that researchers can very quickly and easily assess the current state of trustworthy biometric recognition systems. Furthermore, many of the existing survey papers on these individual topics have become outdated. In short, this paper provides the latest and most comprehensive overview of the state of trustworthy biometric recognition systems.

### 2 Recognition Performance Robustness and Scalability

An initial prerequisite to placing trust in any recognition system is that the system is accurate. In biometric recognition systems, we expect that accuracy to be robust to various intrinsic and extrinsic noise in the input biometric signal (Figure 2), and we also expect (in some cases) the system to be scalable to millions or even billions of users. In terms of accuracy, much research has been conducted since Mitchell Trauring’s first paper on automated fingerprint recognition in the journal *Nature* in 1963 [25]. Indeed, modern day biometric recognition systems now boast accuracies in excess of human level performance (Figure 2). However, in spite of this tremendous progress, there are still a number of situations where the biometric recognition system is not yet robust. To examine what these problems are, we first briefly...
A typical biometric recognition system has two stages of operation, namely, the enrollment stage (instance of the trait is captured and linked to user’s credentials) and the recognition stage (a probe or query trait is compared with the enrolled trait(s)). In addition, biometric recognition systems are typically operated under one of two modes: (i) authentication (1:1 verification) and (ii) search (1:N identification). In both stages of enrollment and search, the biometric recognition system utilizes a series of sub-modules in a systematic pipeline (Figure 2). First, a biometric sensor (e.g., fingerprint reader, RGB camera, or IR sensor) acquires the biometric trait (e.g., fingerprint, face, or irises) of a user in digital form. Next, the digitized trait is passed to a feature extractor to generate a compact and salient representation (or feature set) differentiating one user from another. This representation should have high inter-class separability, i.e., different users should have very different representations. In addition, the representation should have very low intra-class variability, i.e., two representations from the same user should be very similar. The representation could be based on hand-crafted features (e.g., fingerprint minutiae or iris hamming codes), learned features (e.g., deep face representations), or a combination of handcrafted features with learned features (e.g., through feature fusion or by guiding deep learning methods via domain knowledge). Finally, when a user needs to be authenticated or identified, a representation extracted from the query sample can be compared to enrolled representation(s) with a matching algorithm. Breakdowns in the biometric recognition system can occur at any one of the aforementioned modules and as such, robustness and scalability must be imparted to each of them.

There are many different biometric traits, that can be utilized in conjunction with the aforementioned pipeline, however, in this paper we focus our attention on the three most popular and widely accepted traits, namely face, fingerprint and iris (Figure 5).

2.1 Noisy Inputs

Despite impressive recognition performance, accuracies of prevailing biometric systems are sensitive to the image acquisition conditions. For example, in unconstrained scenarios, biometric image acquisition may not be well-controlled and subjects may be non-cooperative (or even unaware). **Image Quality:** The quality of a biometric image severely affects biometric recognition performance. For example, Figure 5 shows the increase in error rates when lower quality webcam and profile face photos are matched to the mugshot gallery. In practice, unconstrained face images are of poor image quality (such as those captured from surveillance cameras). In the case of fingerprints, images fed to fingerprint comparison algorithms may contain distortion and motion blur due to variations in pressure applied on the sensor platen, and may have poor contrast due to dry/wet fingers. Studies show that such degraded fingerprint images hamper recognition performance. Finally, iris images which are occluded by eyelashes and eyelids can cause failures in the iris recognition system. Automated person recognition performance on poor quality images is far from desirable and remains an ongoing challenge for the biometric community.

### PIE Variations:

It is now well established that accuracies of face recognition systems are adversely affected by factors including pose, illumination, expression, collectively known as PIE. Fingerprints also suffer from such adverse inputs including non-linear distortion due to finger pressure and orientation and noisy backgrounds or debris. COTS latent fingerprint rank-1 search accuracy against a 100K gallery from an operational database is ≈ 70%, while rolled fingerprint rank-1 search accuracy against a gallery of 1 million fingerprints from the same database is ≈ 99% (54 55). Likewise iris recognition can be influenced by heavy specular reflections on the eyes. While ongoing efforts in mitigating such adverse noise in biometric systems is commendable, further research needs to be conducted for trustworthy and robust biometric systems.

### Aging Effects:

A considerable amount of research has been conducted to study the permanence of various biometric traits, i.e., the trend in recognition rates as a person ages. Longitudinal studies have shown that the time gap between enrollment and gallery images have no significant impact on recognition accuracies of iris and fingerprint matchers. However, a human face undergoes various temporal changes, including skin texture, weight, facial hair,
it is becoming exceedingly difficult to acquire large-scale face datasets with identity labels due to privacy concerns. Furthermore, large-scale datasets can introduce other challenges such as underrepresented subjects (many subjects have few images per subject) [70].

Instead, an alternative approach is to collect a large set of unlabeled images to enhance the traditional supervised training setting. This can be achieved in a semi-supervised learning approach via label propagation [71]. A different line of work explores utilizing a Graph Convolutional Network to cluster unlabeled biometric images; pseudo-labels can then be used for semi-supervised learning [72-75]. Besides increasing the quantity of training data, a heterogeneous unlabeled dataset can also be introduced to augment the diversity of the prevailing labeled dataset, which has been shown to improve model generalizability to challenging and unconstrained images [76].

2.3 Scalability

Given the success of India’s Aadhaar national ID system, it would seem that biometric recognition systems have achieved a remarkable level of scalability [77]. The Aadhaar system boasts over 1.3 billion enrollees based upon de-duplication utilizing all ten fingerprints, face, and both iris images [77]. However, although Aadhaar has been extremely successful in its mission to provide unique and verifiable digital identity to all, open ended questions remain in the scientific literature on the scalability of biometric recognition systems. In particular, very few evaluations exist in the literature to show how biometric recognition systems operate at a scale the size of Aadhaar (an average of 35M biometric authentications per day\(^1\)), the FBI’s NGI program [28] (an average of 860K monthly searches\(^2\)), and DHS surveillance system [29] (more than 350K biometric transactions per day\(^3\)). Disney Parks also employ fingerprint authentication at their entrances which encounters an average of 427K visitors per day\(^4\). If the system is not scalable and false rejects and false matches are introduced, it will cause chaos and ill-will.

---

\(^1\)https://uidai.gov.in/aadhaar_dashboard/auth_trend.php
\(^2\)https://www.fbi.gov/file-repository/ngi-monthly-fact-sheet/view
\(^3\)https://www.dhs.gov/biometrics
\(^4\)https://disneynews.us/disney-parks-attendance

---

2.2 Training Data

Large-scale datasets have massively contributed to the improved robustness and accuracy of biometric recognition systems over the years. With the advent of deep neural networks for person recognition [55, 67, 68], availability of large-scale labeled dataset is paramount. For example, face recognition systems are primarily trained on 8M face images [67, 68] from MS-Celeb-1M dataset, while a deep-learning-based fingerprint matcher is trained on 445K rolled fingerprints from 38,291 unique fingers [55]. Although increasing the number of training images further could potentially improve the overall recognition performance,

etc. [61, 62]. Several studies have analyzed the extent to which facial aging affects the performance of face matchers and two major conclusions can be drawn: (i) Performance decreases with an increase in time lapse between enrollment and query image acquisitions [33, 57, 63, 64], and (ii) performance degrades more rapidly in the case of younger individuals than older individuals [51, 65]. Figure [5] illustrates that state-of-the-art face matchers fail considerably when it comes to matching an enrolled child in the gallery with the corresponding probe over large time lapses (even the best face matchers begin to deteriorate after a time lapse of 10-12 years between the enrollment and probe image (Figures [4] and [5]). Unlike other factors, face aging is intrinsic and cannot be controlled by the subject or the acquisition environment. Therefore, it is essential to enhance the longitudinal performance of biometric systems (specifically, face matchers) in order to instill trust when deployed in real-world applications such as tracing missing children [66].

Fig. 4: (Top row) Examples of low-scoring genuine face image pairs of two subjects from the PCSO longitudinal mugshot dataset [51]. Ages at image acquisitions are given along with similarity scores from COTS for each pair. COTS is a top-performing AFR vendor in the Ongoing NIST FRVT [33]. (Middle row) Fingerprint impressions from one subject in a longitudinal fingerprint dataset [55]. (Bottom row) A subject’s left iris images collected approximately six months apart [58]. False rejects increase as a person’s face ages, whereas, recognition performance of fingerprints and iris has been shown to be stable across large time lapses [58-60].

Fig. 5: (a) Identification error rates of five SOTA AFR vendors when mugshots (high-quality), webcam (medium-quality), and profile (low-quality) faces are compared against a 1.6M mugshot dataset [51]. (b) Identification error rates of six SOTA AFR systems on a 3M mugshot dataset under aging [51].
Theoretically, iris recognition should be incredibly scalable [80]. A few studies have evaluated the search/clustering performance of face recognition against a gallery of 80 million and 123 million, respectively [81, 82]. The large scale galleries were obtained by scraping photos from the web. In a similar fashion, a study was conducted in [83] to ascertain the performance of fingerprint search algorithms against a gallery of 100 million prints. Since there is no publicly available large-scale database for evaluating fingerprint search, the authors in [83] first synthesize a database of 100 million fingerprints which are then used in the search evaluation. A limitation of the approach in [83] is that a domain gap exists between synthetic fingerprints and real fingerprints such that synthetic distractors could artificially inflate the true search performance at scale. This limitation could also exist in the large scale face search studies [81, 82] where even galleries of web scraped real data could have a domain gap with the probes from surveillance video frames. Given these challenges, and the additional increasing privacy concerns over biometric data, a very important ongoing area of research in biometrics is that of large scale synthesis. In particular, if methods can be developed to synthesize biometric images which bridge the domain gap between real and synthetic samples, better estimates on the scalability (both accuracy and speed) of biometric recognition systems can be established and consequently, biometric recognition systems can be made more trustworthy.

3 Security

Aside from the performance robustness and scalability of state-of-the-art (SOTA) biometric recognition systems discussed in the previous section, perhaps the next most important aspect of biometric recognition systems needed to solidify trustworthiness is that of their security or their often perceived lack thereof. When talking about biometric system security, we are specifically referring to those areas of the biometric recognition system which are vulnerable to manipulation and exploitation by various malicious hackers. These “hacks” can be carried out at each of the individual stages of the biometric recognition system as shown in Figure 2. To focus our attention on the most serious threats, we dive down into a few of the major points of security concern within SOTA biometric recognition systems. In particular, security threats exist at (i) the sensor level in the form of presentation attacks, (ii) the feature extraction module via adversarial attacks, and (iii) the database and matching modules with template theft and subsequent template reconstruction attacks. Each of these areas of security concern have been investigated by the biometrics research community. However, points of concern remain unaddressed, particularly with respect to their generalizability to detect new attack types and new sensors not known during their training. In this section, we define each of these attacks, discuss the state-of-the-art in mitigating against these attacks, highlight what remains unsolved, and conclude with what can be done to further enhance the security of biometric recognition systems to instill trust in their continued widespread use.

3.1 Presentation Attacks

In IEC 30107-1:2016(E), presentation attacks (PAs) are formally defined as:

“Presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system.”

PAs can be deployed either as an obfuscation attack (an attempt to hide one’s own identity) or as an impersonation attack (an attempt to mimic someone else). For example, fingerprints could be cut or burned in an attempt to obfuscate one’s identity and thus evade identification [83]. Alternatively, spoofs comprised of common household materials such as playdoh, wood glue, or gelatin can be used by a hacker to create an impersonation of a victim’s fingerprint. More sophisticated attacks include use of high-resolution 3D [84-87] or 2D printing [88], or cadaver fingers [89]. In the domain of face, glasses or a mask could be used for obfuscation, while a replay on a mobile phone could be used for impersonation. Some of the well-known spoof attacks for face, fingerprint, and iris are shown in Figure 6. In all of these examples, the attack against the biometric recognition system is carried out at the sensor level (Figure 2).

PAs have gained notoriety due to several real world examples where they have been shown to fool biometric recognition systems. For example, the German Chaos Computer Club demonstrated with ease the breaking of Apple’s TouchID already in 2013 [90]. Fast forward to today, fingerprint recognition systems are still being thwarted by spoof attacks with some success [91]. Across biometric traits, Apple’s highly touted FaceID was compromised by a 3D mask shortly

---

Footnote:
7 Of course using mega-scale galleries of real data would be best for building trustworthiness, however, in practice, obtaining such datasets from legacy sources is becoming extremely difficult due to privacy concerns and/or the time and cost of collecting such an evaluation dataset.

---

References:
[80] https://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid
[81] https://blog.talosintelligence.com/2020/04/fingerprint-research.html
Table 2: SOTA PAD Performance for Face, Fingerprint, and Iris on known (seen during training) PA types

| Trait     | Competition       | Accuracy        |
|-----------|-------------------|-----------------|
| Fingerprint| LiveDet 2019      | 96.17%¹         |
| Face      | 2020 Celeb-A Spoof Challenge | 100%²         |
| Iris      | 2020 LivDet-Iris Challenge | 97.82%¹       |

¹ Average accuracy reported in [95] and [96].
² TDR @ FDR = 10^-6 reported in [97].

After its deployment by the Vietnamese cybersecurity firm Bkav², all of these successful attacks come twenty years after early successful spoof attacks were shown in [93, 94].

The continued success in spoofing modern day biometric recognition systems is not a consequence of a lack of research into developing presentation attack detection (PAD) systems. Indeed the past couple of decades have seen a plethora of research into developing PAD systems which can automatically detect and flag a spoof attack prior to performing authentication or identification [35, 36, 98, 100]. Typically these approaches are divided into hardware or software based approaches detecting face, fingerprint, and iris spoofs. Hardware approaches deploy additional sensors (e.g. depth, IR cameras, multispectral illumination, etc.) to capture features which differentiate bonafide acquisitions from PAs [101–112]. In contrast, software based solutions extract anatomical, physiological, textural, challenge response, or deep network based features to classify an input sample as live (bonafide) or presentation attack (spoof) [35, 36, 96, 98, 100, 113, 122]. The culmination of these approaches can be seen in the high performances of the various algorithms submitted as part of the IARPA ODIN program⁰ and also the public fingerprint and face liveness competitions (Table 2). However, after years of rigorous research into various PAD approaches, the continued success of spoof attacks against deployed biometric recognition systems leads to the inevitable question, “What can be done to more reliably secure the biometric sensing module from spoof attacks?”.

From our review of the literature, we posit that there are a few different sub-problems of biometric PAD that remain unaddressed. Solving these problems will close the spoofing loopholes remaining and will go a long way towards building trust in biometric recognition systems.

Perhaps the most significant outstanding problem with deployed PAD systems is their lack of generalization to spoofs fabricated from materials different than the spoofs that were used to train the PAD system. This problem is typically referred to as “unseen” or “cross” material generalization. In the domain of fingerprint recognition, multiple studies specifically showed that when a material is left out of training a state-of-the-art spoof detector and then subsequently used for evaluation, the detection accuracy drops below 10% [91, 125]. Similar deterioration of unseen material detection accuracy have been observed in the face domain [124]. In operational settings, the likelihood of a hacker using a spoof made from a novel material can be high and thus, without addressing this problem, spoof detectors remain limited in their applicability. Unfortunately, many papers continue to work on addressing “known-material” spoof detection which already obtains nearly perfect accuracy (Table 2), while ignoring this more challenging problem. There are a number of more recent and promising works that focus specifically on addressing the “unseen material” and “unseen sensor” challenge, however, the accuracy remains insufficient for field deployment [90, 91, 115, 123, 125–137]. Thus, we urge a stronger research push in this direction in an effort to build trustworthy biometric recognition systems.

In addition to the major vulnerability of “unseen materials”, other practical limitations of PAD systems must also be addressed. For example, many PAD systems evaluated in the literature train on one partition of a dataset captured by a particular sensor or camera, and then test on a separate partition of the same dataset (again captured by the same sensor model or camera under the same capture conditions). However, there can be a number of differences in the data distribution observed in the actual deployment scenario such as: sensor model, illumination, subject demographics, and environmental conditions. As such, models reporting near perfect accuracy on intra-dataset; intra-sensor perform quite poorly when deployed into a inter-dataset and inter-sensor scenario. We encourage PAD researchers to examine more difficult evaluation scenarios (cross dataset, cross sensor [132, 138, 145]) which may be more indicative of how the PAD system will perform in the wild.

Finally, from a practical perspective, many of the PAD solutions place little emphasis on the efficiency of the PAD solution. However, many of the biometric recognition systems we use today are deployed on resource constrained devices (such as our smartphones) and as such, many of the deep learning PAD systems are impractical for real world applications. Research needs to be done to prune the parameters of the deep learning based approaches and perhaps combine deep learning approaches with simpler, faster, and lighter weight handcrafted approaches [91, 146].

3.2 Adversarial Attacks

With unrestricted access to the rapid proliferation of face images on social media platforms, such as FaceBook,
SnapChat, Instagram, etc., a community of attackers dedicate their time and efforts to digitally manipulate face images in order to evade automated face recognition (AFR) systems [147]. AFR systems have been shown to be vulnerable to “adversarial faces” resulting from perturbing an input probe [149–152]. Even when the perturbations are imperceptible to the naked eye, adversarial faces can degrade the performance of numerous state-of-the-art (SOTA) AFR systems [149–150] (see Figure 7). For example, face recognition performance of SOTA AFR system, ArcFace [68], drops from a TAR of 99.82% to 0.17% at 0.1% FAR on LFW dataset [45] when the adversarial face generator, AdvFaces [149], is encountered. Note that adversarial images are an attack on the feature extraction module of biometric recognition system (Figure 2).

In contrast to face presentation attacks where the attacker needs to actively participate by wearing a mask or replaying a face photograph/video of the victim, adversarial faces do not require active participation during verification. Given the unattended nature and “touchless” acquisition of AFR systems, an individual may maliciously enroll an adversarial image in the gallery such that at border crossing, his legitimate face image will be matched to a known and benign individual (known as an impersonation attack). An individual may also synthesize adversarial faces in order to safeguard personal privacy (e.g., obfuscate automated face recognition in video conference calls [153]). Also different from face presentation attacks, the adversarial perturbations are extremely subtle and directly inhibit face representations thereby making detection an extremely challenging task.

Given the growing dissemination of “fake news” and “deep fakes”, the research community and social media platforms alike are pushing towards defenses against digital perturbation attacks. In order to safeguard AFR systems against these attacks, numerous defense strategies have been proposed in literature. A common defense strategy, namely adversarial training, is to re-train the classifier we wish to defend with perturbation attacks [148–154, 157]. However, adversarial training has been shown to degrade classification accuracy on real (non-adversarial) images [158–159]. In the case of face recognition, adversarial training drops the accuracy on real images in the LFW dataset [45] from 99.13% to 98.27% [147]. Therefore, a large number of defense mechanisms have been deployed as a pre-processing step where a binary classifier is trained to distinguish between real and perturbed faces [147, 160–172]. Another pre-processing strategy, namely purification, involves automatically removing perturbations in the input image prior to passing it to an AFR system [142–154].

Similar to PAD mechanisms, an adversarial defense system also suffers from poor generalizability to perturbation types that are not encountered during its training (“unseen perturbation types”) [147]. In addition, employing separate pre-processing steps to detect perturbation attacks that inhibit the face feature extraction module is cumbersome and adds computational burden. Further research needs to be conducted to improve the intrinsic robustness of AFR systems to such adversarial perturbations which eliminates the need for separate detectors or purifiers.

Finally, we note that to date, adversarial attacks on the feature extraction module (Figure 2) have been mostly associated with face recognition systems, since most AFRs utilize deep networks for feature extraction. In contrast, most fingerprint and iris recognition systems rely primarily on handcrafted minutiae points or iris hamming codes and are thus assumed to be safe from adversarial attacks. However, this assumption should be treated with caution as deep networks are now being explored for fingerprint and iris recognition systems as well for a number of tasks including: fixed-length representation extraction [55, 188, 189], minutiae extraction [190], minutiae descriptor extraction [54], spoof detection [114], etc. Presumably, any one of these deep network based fingerprint or iris algorithms could also be vulnerable to adversarial attacks. In fact, some work has been done to show that fingerprint PAD systems can be evaded by adversarial attacks [191]. This is concerning since another study showed that these adversarial attacks could be converted back into a physical attack and then deployed as a successful attack on the PAD system [192]. In addition, several successful adversarial attacks have been crafted to evade iris matchers as well [193, 196].

3.3 Template Attacks

Finally, in addition to the security threats that exist at the sensor level in the form of spoof attacks and at the feature extraction module in the form of adversarial attacks, a very serious vulnerability of biometric recognition systems is that of limited template security. In particular, numerous studies have shown that templates extracted by biometric recognition systems (deep face representations [183], minutiae-based fingerprint representations [185, 197], and iriscode features [187, 198, 199]) can be inverted back into the image space with high fidelity (Figure 5). Other studies have shown that “soft” demographic attributes (such

---

12 Adversarial perturbations refer to altering an input image instance with small, human imperceptible changes in a manner that can evade CNN models [148].
as age and gender) are encoded into the biometric templates [200, 201]. This is of serious concern given a number of reported breaches of databases containing biometric templates [37]. Note that a template can be stolen immediately following feature extraction, as it resides in the enrollment database, or even during the matching routine if the template needs to be decrypted to perform the matching. Thus, the biometric recognition system needs to ensure that the templates remain encrypted and secured from hackers at all times.

A plethora of research has been conducted to secure biometric templates [37]. Some of these approaches are based upon cryptography [202] and others are pattern recognition based. For example fuzzy vault cryptosystems have been proposed for fingerprint [203] and iris [204] recognition. Common pattern recognition based approaches include non-invertible transformation functions [205] and cancelable biometrics [206]. Another approach that has been tried is to bind a secret key with a biometric template [207, 208]. Finally, techniques based on deep networks [209] and representation geometry [210] have been proposed. All of these approaches are limited in that they trade off the recognition accuracy of a biometric system for the enhanced security.

A more recent development in biometric template protection is that of homomorphic encryption (HE) systems [211–215]. Homomorphic encryption enables doing basic arithmetic operations directly in the encrypted domain. Because of this, the primary benefit of using HE is that it can protect the template as it resides in the database, and also while it is being compared (assuming the matching function can be reduced to arithmetic operations of addition and multiplication, e.g., the cosine similarity between two face representations). The limitation of HE systems is that it is computationally expensive, especially Fully HE systems which allow for both addition and multiplication operations directly in the encrypted domain. Work has been done to alleviate the computational burden of FHE for biometric matching [216, 218], however, research remains to further speed up this encrypted matching process (e.g., the work in [218] showed encrypted fingerprint search against 100 million gallery in 500 seconds, a 275× speedup over SOTA; the same search in the unencrypted domain would take 10 seconds [55]).

Generally speaking, all of the methods that attempt to better protect the biometric template, seek a compromise along multiple axes of speed, memory, accuracy, and security. Research must continue to minimize the trade-offs and sacrifices that occur in any one of these dimensions. Ideally, a trustworthy biometrics recognition system would secure the template, at all times, while sacrificing very little along any of these axes.

### 3.4 Unifying Security Efforts

As an addendum on the security efforts across sensing, feature extraction, and matching modules, we note that prevailing research efforts focus on mitigating one of the three attack categories at a time: (i) presentation attacks, (ii) adversarial attacks, and (iii) template attacks. Since the exact type of biometric attack may not be known a priori, researchers are encouraged to design generalizable detectors that can defend biometric systems against any of the three attack categories [219] (e.g., in an enrollment scenario, a single detector could quickly check for live vs. spoof, adversarial perturbations, and reconstruction attacks). Such systems will alleviate the computational burden of securing the entire biometric recognition pipeline.

### 4 Explainability and Interpretability

In addition to being accurate and secure, a trustworthy biometric recognition system should also have a certain degree of interpretability such that system designers and agency deploying the system can understand why a decision is made and adjust the system’s decision if needed (i.e., by inserting a human in the loop). Interpretability is also important in courts of law, where fingerprint and face evidence could be used to convict a person [221, 222]. For example, if we are using a face recognition system’s prediction to identify someone as a criminal, we would like to understand why the system thinks the probe and gallery faces appear similar to prevent potential false convictions or false acquittals [223]. However, most deep neural network based models, utilized for face recognition, serve as black boxes that give final decisions on probe samples directly via millions of learned parameters.

To better impart credibility and interpretability to these black box systems, many methods have been proposed in the broader computer vision and machine learning community. One popular direction of research is visualizing the features that are learned in the model [223, 227]. Others focus on the attribution of the decision, either by finding the features [228, 230] or the local regions in images [231, 233] that lead to the final decision. Although the feature

---

[1] https://bit.ly/2HD83Pq
[2] https://wapo.st/39PQuaT
[3] https://wapo.st/2V3kHP5
[4] https://bit.ly/2QOhLM3

---

Interpretability can also greatly aid the judge and the jury in cases where both the prosecution and the defense present conflicting recognition results based on their own proprietary black-boxes.
visualizing methods could be directly applied to the feature extraction module of biometric systems, the attribution methods may be better geared towards classification models used in biometrics (i.e., PAD algorithms) since the goal is to interpret a final classification decision.

Within the biometric modality of face specifically, a few studies have attempted to understand how features are learned and used to compare faces. For example, Yin et al. [200] propose to constrain the learning stage such that features are directly related to different areas of the face. Once the models are trained, saliency maps can be used to visualize which part of the face a filter is looking at (see Figure 9). Experimental results also show that in addition to imparting spatial interpretability, regularizing the spatial diversity of the features enables the model to become more robust to occlusion. A drawback of Yin’s method is that a model needs to be re-trained to obtain such interpretability (i.e., interpretability cannot be extracted from prevailing commodity AFR systems). In response, Stylianou et al. [236] propose a model-agnostic method that visualizes the salient areas that contribute to the similarity between a pair of faces. It is observed that models are indeed focusing on the entire face, but they were not able to provide more fine-grained details, such as which part of the face is contributing to the similarity/dissimilarity between a pair of faces. Therefore, we believe this problem of similarity attribution remains a meaningful yet unsolved problem for future research.

In another line of research, the studies in [200] and [201] provide interpretability to AFRs by studying how facial attributes are encoded in the deep neural network. In particular, the authors in [200] chose four common attributes, namely identity, age, gender and face angle (yaw), and estimated their correlation with face representations. They found that compared to low-level features, high-level deep face representations tend to be more correlated with identity and age while less correlated with gender and face angle. In a similar line of research, the authors in [201] examined the effect of 47 high-level attributes on face recognition performance. They observed that many nuisance factors such as accessories, hair-styles and colors, face shapes, or facial anomalies influenced the face recognition performance.

A more recent direction in interpreting and improving AFRs is through uncertainty estimation. For example, Shi et al. proposed in [238] to represent each input face as a distributional representation in the feature space (rather than a single point or feature vector), where the variance of the distributions represent the uncertainty of the corresponding features. Besides improving the face recognition performance, they showed that the feature uncertainty could also be used to visualize the perception of the model about the input.

While all of the aforementioned methods have certainly helped impart more interpretability to AFRs, there is still much we do not yet know and understand about what information about the input image is being encoded into deep face representations. Having a better understanding of what these encodings are comprised of could help address biasness and other failures in the AFR system. Interpretability also needs to be extended to other modules of the face recognition pipeline (such as spoof detection) where nuisance factors could potentially cause a spoof face to be misclassified as a live face. Therefore, we posit that more work in this area remains to be done in an effort to build trustworthy biometric recognition systems.

We note that much of the interpretability concerns mentioned thus far have been centered around face recognition systems. This is because nearly all face recognition systems employ the use of “black-box” deep networks for encoding and matching. However, as per our earlier discussion on adversarial attacks, deep networks are now being increasingly used for fingerprint and iris recognition systems as well. Thus several studies have begun to more carefully discuss interpretability of deep networks deployed for various tasks within the fingerprint and iris recognition pipeline. For example, the authors in [55] utilize the feature attribution method from [223] to visualize the features being learned by a deep network for fixed-length fingerprint representation.
of a mugshot dataset. Similar bias issues in AFRs were biased performances based on gender, race, and age groups submitted to the NIST FRVT [51] exhibit different levels of algorithms (from academia and industry alike) that were groups (see Figure 14). In fact, all 106 face recognition mographics and lower performance in other demographic recognition performance for users within a subset of demographically biased, it algorithmically provides higher bias in biometrics. When a biometric system is defined to be insignificant, it is problematic that biometric sensors can play in biasness. For instance, a person's age or aging in other biometric modalities (fingerprint [52, 59, 248, 249], or iris [250]). A consistent finding of bias in face recognition across studies in [57, 63, 64] is that the recognition performance is worse for female cohorts (possibly due to the use of cosmetics). The studies of [251, 252] showed a significant attribute estimation accuracy impact based on age, gender and race. In the domain of fingerprint, [59] indicated a non-trivial impact of age on genuine match scores.

**5 DEMOGRAPHIC BIAS AND FAIRNESS**

Another issue of trust with biometric recognition systems that has more recently been brought to light in mainstream media is that of biased performance against certain demographic groups [51, 63, 244, 245], referred to as demographic bias in biometrics. When a biometric system is defined to be demographically biased, it algorithmically provides higher recognition performance for users within a subset of demographics and lower performance in other demographic groups (see Figure 14). In fact, all 106 face recognition algorithms (from academia and industry alike) that were submitted to the NIST FRVT [51] exhibit different levels of biased performances based on gender, race, and age groups of a mugshot dataset. Similar bias issues in AFRs were reported by earlier studies on demographic attribute estimation [246]. It should be acknowledged that the demographic bias shown by the best performing commodity AFR systems in the NIST FRVT on mugshot faces is less than 1.0% across the four groups: Black Male, Black Female, White Male and While Female [51]. Furthermore, every top-tier AFR system studied in NIST FRVT Ongoing is most accurate on Black Males [53]. It should also be noted that the extent of bias across different demographic cohorts cannot be precisely known until proper ground truth adjudication can be done on large scale datasets such as that used in the NIST FRVT (where the level of performance on different demographic groups flipped before and after manual ground truth adjudication) [51].

Since facial regions contain rich information of demographic attributes, most studies on bias are focused on face-based biometrics [51, 63, 64, 244, 247]. However, several studies have also investigated the bias factor of age or aging in other biometric modalities (fingerprint [52, 59, 248, 249], or iris [250]). A consistent finding of bias in face recognition across studies in [57, 63, 64] is that the recognition performance is worse for female cohorts (possibly due to the use of cosmetics). The studies of [251, 252] showed a significant attribute estimation accuracy impact based on age, gender and race. In the domain of fingerprint, [59] indicated a non-trivial impact of age on genuine match scores.

**Biometrics for Lifetime**: Multiple studies have shown the extreme difficulty in performing biometric recognition on the most vulnerable amongst us, namely infants and young children (Figure 12) [52, 248, 249].

Most of the aforementioned studies address algorithmic demographic bias, however, we also highlight the role that biometric sensors can play in biasness. For instance, matching fingerprints from different sensors is a challenging problem [253]. In the case of iris recognition, brown-eyed individuals are more susceptible to sensor issues and therefore, near infra-red sensors are adopted instead of RGB cameras. Finally, a study on AFRs showed that “the magnitude of measured demographic effects depends on

---

**Fig. 12**: Face image and corresponding thumb-print of a 1-week old infant [52]. The thumb-print was captured and matched using a custom 1,900 ppi reader and accompanying high-resolution matcher, since the standard 500 ppi COTS readers and matchers do not have sufficient resolution to capture and match an infant’s fingerprints.

**Fig. 13**: Face recognition system wrongfully identified (a) Robert Williams when the CCTV frame in (b) is searched against a 49M gallery. On arrest, Williams responds, “This is not me. You think all Black men look alike?” [243]. Relying on automated person recognition alone may lead to a lack of trust in the eyes of policymakers and citizens alike. Therefore, it is imperative to have “humans-in-the-loop” where human examiners can verify the decisions made by biometric systems. In addition, biometric systems should also have a “reject” option instead of match/non-match binary decisions.
Deploying biased systems could come with significant consequences, especially against those whom the system does not perform as well on, e.g., being unjustly incarcerated or denial of bail or parole [255–258] (see Figure 13). Therefore, it is crucial to estimate and mitigate demographic bias in biometric recognition systems. Such systems should show no statistically significant difference on the performance amongst different demographic groups of individuals. At the same time, the overall accuracy of the system should not be compromised, ideally.

To mitigate bias in biometric recognition systems, a simple question must first be answered: What factors lead to bias in biometric recognition systems? The answer to that question is multi-faceted. First of all, many state-of-the-art biometric recognition systems are based on deep networks, which rely on large training datasets. These training datasets of human subjects are often biased towards certain demographic cohorts. Secondly, the implementation of biometric recognition systems can be statistically biased during the learning process, for example, by parameter optimization and regularization. For example, a representation extraction network undergoing training is typically trying to satisfy training samples on the average case while potentially placing less weight on under-represented samples leading to biasness. Finally, the fourth factor is what is referred to as **intrinsic bias**, a notion first introduced by [63], stating that subjects in certain demographic groups are inherently more difficult to be recognized.

Given the various sources of bias mentioned above, bias mitigation requires special attention on both data sampling and algorithm design. Early studies on dataset-induced bias include data re-sampling methods (oversampling or undersampling images of certain demographics) [267–269]. Data re-sampling is limited in that useful, diverse information is discarded. Therefore, rather than re-weight the sample distribution in the training set, later studies tackle bias by re-weighting the loss values in objective functions [270, 271], also called **cost-sensitive learning**, based on a sample’s demographic cohort.

The aforementioned works do not take into account the correlation between demographics and identity. As such, [247] proposes a framework to jointly learn unbiased representations for both the identity and demographic attributes by disentangling them. The impact of bias is mitigated by removing sensitive information (demographics or identity) from each component of the disentangled representation. A limitation of [247], is that the overall recognition performance declines. To be practical, algorithms mitigating bias in face recognition should also maintain the overall recognition accuracy. To address this challenge, Wang et al. [272] propose an adaptive margin for faces in each demographic group. Another approach proposed by [273] adapts the network operations by employing dynamic convolutional kernels and attention maps based on the demographic group. Both [272] and [273] manage to improve the performance on under-represented groups while better maintaining the overall accuracy.

Despite recent progress in mitigating demographic bias, this issue has not been completely rectified and still demands further research, especially given the fact that a variety of factors could lead to bias other than the predefined demographic groups that most studies assume. Existing studies need to make sure that overall system accuracy is not compromised via bias reduction. Furthermore, since the majority of the existing studies are concentrated on bias mitigation for face-based biometrics, there is an urgent need for research on other biometric modalities (e.g., fingerprint [274]). Finally, biasness research should also be conducted on algorithms other than the recognition system (i.e., the PAD modules, where biasness could inconvenience users of certain demographics unfairly). Biased biometric recognition systems create an element of mistrust in the general public and as such, removing this bias is a critical step on the path towards trustworthy biometric recognition systems.

### 6 Privacy

A final key area of biometric recognition systems that we posit is necessary in order to build trust is that of user privacy. Note, we explicitly differentiate between **security** (such as the template security previously discussed) and **privacy**. While security is aimed at addressing attacks on the biometric recognition system with the goal of interference, privacy does not necessarily entail an attack. Rather it entails the respect and confidentiality of an individual’s personal identifying information (PII) or data as well as transparency surrounding its use and storage.

A number of high profile laws have been enacted to better ensure privacy. In 2008, the Illinois legislature unanimously passed the Biometric Information Privacy Act (“BIPA”), based on efforts by the ACLU [18]. The Illinois law enables individuals a better control of their own biometric data and prohibits private companies from collecting it unless they:

- Inform the person in writing of what data is being collected or stored.
- Inform the person in writing of the specific purpose and length of time for which the data will be collected, stored and used.
- Obtain the person’s written consent.

[18] [https://www.aclu-il.org/en/campaigns/biometric-information-privacy-act-bipa](https://www.aclu-il.org/en/campaigns/biometric-information-privacy-act-bipa)
Likewise, in 2016 the GDPR \cite{GDPR} (General Data Protection Regulation) passed the European Parliament. The GDPR defined personal data (to include biometric data) as: “... any information that relates to an individual who can be directly or indirectly identified ... including biometric data, ...”. The GDPR further laid out strict guidelines for processing data:

- **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.
- **Purpose limitation** — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
- **Data minimization** — You should collect and process only as much data as absolutely necessary for the purposes specified.
- **Accuracy** — You must keep personal data accurate and up to date.
- **Storage limitation** — You may only store personally identifying data for as long as necessary for the specified purpose.
- **Integrity and confidentiality** — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
- **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

The legal response of BIPA and GDPR can be in part traced to the rapid proliferation of biometric images (especially face) on social media websites such as Facebook, Twitter and Instagram, and their use in training biometric recognition systems without the informed consent of subjects. For example, a face recognition startup, Clearview AI, is currently facing litigation for allegedly amassing a dataset of about 3 billion face images \cite{Clearview} from various social media sites without subjects’ permission. This lack of consent and transparency has led to some cities wanting to curb facial recognition technology\footnote{https://www.wcjb.com/2021/05/05/states-push-back-against-use-of-facial-recognition-by-police/}. In addition, publicly available biometric datasets that were collected without consent are now being retracted \cite{BiometricDatasetRetraction}. To make matters worse, there has been work to show that even generative adversarial networks can leak private information about the dataset on which they were trained \cite{GANPrivacy}.

In the computer vision community, research has been conducted to alleviate these concerns. In particular, a number of studies have explored using homomorphic encryption to perform inference or classification on encrypted data with encrypted model parameters \cite{HomomorphicEncryption}. While these approaches are quite promising as they offer the data/model parameters a high level of security and consequently privacy, they require significant computational burden which limits the size of models in practice. Basically, as with our previous discussion on trade offs of speed, memory, accuracy and security when using fully homomorphic encryption for protecting biometric templates, the same issue applies towards its use in protecting data and model parameters.

An alternative method that has been explored to impart privacy is to train biometric systems in a decentralized manner (i.e., federated learning). In particular, multiple participating clients jointly learn a biometric recognition system without ever sharing their training data with each other. For example, the study in \cite{FederatedLearning} used federated learning for training a face PAD algorithm. Likewise, Aggarwal et al. \cite{AggarwalFederatedLearning} propose to use federated learning to collaboratively learn a global face recognition system, training from face images on multiple clients (mobile devices) in a privacy preserving manner. Only local updates from each mobile device are shared to the server where they are aggregated and used to optimize the global objective function, while the training face images on each mobile device are kept private. Their proposed framework is able to enhance the performance of a pretrained face recognition system namely, CosFace \cite{CosFace}, from a TAR of 81.43% \rightarrow 83.79% on IJB-A dataset \cite{IJB-A} at 0.1% FAR and accuracy from 99.15% \rightarrow 99.28% on...
The Guardian, “Commuters walk by surveillance cameras at a walkway between two subway stations in Beijing.” https://www.theguardian.com/cities/2019/dec/02/big-brother-is-watching-chinese-city-with-2bm-cameras-is-worlds-most-heavily-surveilled. 2019. [Online; accessed 11-April-2021].

PC guide, “Amazon Announces New Contactless Payment System Using Just Your Hand.” https://www.pcguide.com/news/amazon-announces-new-contactless-payment-system-using-just-your-hand/. 2020. [Online; accessed 11-April-2021].

CNN, “How facial recognition is taking over airports.” https://www.cnn.com/travel/article/airports-facial-recognition/index.html. 2019. [Online; accessed 17-April-2021].

Shutterstock, “Credit card with a fingerprint sensor.” https://www.shutterstock.com/image-photo/credit-card-fingerprint-sensor-purchase-biometric-102094742. 2020. [Online; accessed 11-April-2021].

eI Boletín, “CaixaBank, the first bank in the world to use facial recognition in its ATMs.” https://www.eiboletin.com/caixabank-primer-banco-del-mundo-que-utiliza-el-reconocimiento-facial-en-sus-cajeros. 2019. [Online; accessed 11-April-2021].

Techspot, “iPhone X users are inadvertently training Face ID to recognize siblings.” https://www.techspot.com/news/71741-iphone-x-users-inadvertently-training-face-recognize-siblings.html. 2017. [Online; accessed 11-April-2021].

Gerald Nino, “U.S. Customs and Border Protection.” https://commons.wikimedia.org/wiki/File:US-VISIT_(CBP).jpg. 2007. [Online; accessed 11-April-2021].

SRBPost, “Ration Card Aadhaar Card Link Online.” http://www.srbpost.com/sarkari-yojana/ration-card-aadhaar-card-link-online/. 2020. [Online; accessed 11-April-2021].

China Daily, “China’s vision drives iris-recognition technology.” https://www.chinadaily.com.cn/a/201809/14/WS5b960fbca31033f43eb03e8.html. 2018. [Online; accessed 17-April-2021].

The Verge, “Amazon’s palm reading store now online.” https://www.theverge.com/2020/10/1/21496673/amazon-one-palm-reading-vein-recognition-payments-identity-verification. 2020. [Online; accessed 17-April-2021].

Department of Homeland Security, “US-VISIT Face Sheet.” https://www.dhs.gov/xlibrary/assets/usvisits/usvisit_edu_10-fingerprint_collection_fact_sheet.pdf. 2009. [Online; accessed 17-April-2021].

A. K. Jain, A. A. Ross, and K. Nandakumar, Introduction to biometrics. Springer Science & Business Media, 2011.

Victorian Crime & Punishment, “Habitual Criminals.” http://vcp.e2bn.org/justice/page11571-habitual-criminals.html. 2006. [Online; accessed 29-April-2021].

A. Barnett and C. Harrison, Crime and punishment in England: A sourcebook. Routledge, 2005.

T. R. Martin, Ancient Rome: From Romulus to Justinian. Yale University Press, 2012.

IT Pro, “Fraudsters use AI voice manipulation to steal £200,000.” https://www.itpro.co.uk/social-engineering/34308/fraudsters-use-ai-voice-manipulation-to-steal-200000. 2019. [Online; accessed 21-April-2021].

The Quint, “Hackers Can Detect What You’re Typing By Listening To You Type.” https://www.thequint.com/tech-and-auto/tech-news/hackers-can-know-your-password-just-by-listening-to-your-typing. 2019. [Online; accessed 21-April-2021].

Microsoft News, “WhatsApp is leaking mobile numbers of users in plaintext.” https://www.msn.com/en-in/money/news/whatsapp-is-leaking-mobile-numbers-of-users-in-plaintext-claims-an-independent-cybersecurity-researcher/ar-BB1b8s8A9/ii=AAGgbIxN. 2020. [Online; accessed 21-April-2021].

A. J. Aviv, K. L. Gibson, E. Mossop, M. Blaze, and J. M. Smith, “Smudge attacks on smartphone touch screens.” Woot, vol. 10, pp. 1–7, 2010.

World Health Organization, “Prevention not cure in tackling health-care fraud.” https://www.who.int/bulletin/volumes/89/12/11-021211/en/. 2011. [Online; accessed 11-April-2021].

World Food Programme, “WFP demands action after uncovering misuse of food relief intended for hungry people in Yemen.” https://www1.wfp.org/news/wfp-demands-action-after-uncovering-misuse-food-relief-intended-hungry-people-yemen-2018. 2018. [Online; accessed 11-April-2021].

World Food Programme Insight, “These changes show that WFP loves us.” https://insight.wfp.org/these-changes-show-
Deep learning for biometrics. Springer, 2017.

M. Vatsa, R. Singh, and A. Majumdar, Deep Learning in Biometrics. CRC Press, 2018.

C. Lu and X. Tang, “Surpassing human-level face verification performance on lfw with gaussianface,” in AAAI, vol. 29, 2015.

C. I. Watson, G. P. Fiumara, E. Tabassi, S. L. Cheng, P. A. Flanagan, and W. J. Salamon, “Fingerprint vendor technology evaluation,” NIST, 2015.

P. J. Grother, M. L. Ngan, and K. K. Hanaoka, “Ongoing face recognition vendor test (frvt) part 2: Identification,” NIST, 2018.

G. W. Quinn and J. R. Matey, IREX 10: Ongoing Evaluation of Iris Recognition Concept, Evaluation Plan, and API Overview. NIST, 2019.

E. Marasco and A. Ross, “A survey on antispooﬁng schemes for ﬁngerprint recognition systems,” ACM Computing Surveys (CSUR), vol. 47, no. 2, pp. 1–36, 2014.

J. M. Singh, A. Madhun, G. Li, and K. Ramachandra, “A survey on unknown presentation attack detection for ﬁngerprint,” arXiv preprint arXiv:2005.08337, 2020.

A. K. Jain, K. Nandakumar, and A. Nagar, “Biometric template security, EURASIP Journal on advances in signal processing, vol. 2008, pp. 1–17, 2008.

A. K. Jain, K. Nandakumar, and A. Ross, “50 years of biometric research: Accomplishments, challenges, and opportunities,” Pattern recognition letters, vol. 79, pp. 80–105, 2016.

A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen, and S. Yadav, “Some research problems in biometrics: The future beckons,” in ICB, pp. 1–8, IEEE, 2019.

A. K. Jain and A. Ross, “Bridging the gap: from biometrics to forensics,” Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 360, no. 1674, p. 20140254, 2015.

F. Vakhshiteh, A. Nickabadi, and R. Ramachandra, “Adversarial attacks against face recognition: A comprehensive study,” arXiv preprint arXiv:2007.11709, 2020.

P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch, “Demographic bias in biometrics: A survey on an emerging challenge,” IEEE Transactions on Technology and Society, vol. 1, no. 2, pp. 89–103, 2020.

A. Boyd, Z. Fang, A. Czajka, and K. W. Bowyer, “Iris presentation attack detection: Where are we now?,” Pattern Recognition Letters, vol. 138, pp. 483–489, 2020.

B. Donizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Hounam, S. Garcia-Salicetti, and A. Mayoue, “Fingerprint and on-line signature veriﬁcation competition,” in IEEE ICB, pp. 725–732, 2007.

G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

K. Wang and A. Kumar, “Toward more accurate iris recognition using dilated residual features,” IEEE ICB, vol. 14, no. 12, pp. 3233–3245, 2019.

Mordor Intelligence, “Consumer biometrics market growth,” https://www.mordorintelligence.com/industry-reports/consumer-biometrics-market, 2021. [Online; accessed 13-May-2021].

T. Chugh, S. S. Torra, A. K. Jain, and N. G. Paultier, “Benchmarking fingerprint minutiae extractors,” in IEEE BIOSIG, pp. 1–8, 2017.

S. A. Grosz, J. J. Engelsma, and A. K. Jain, “White-box evaluation of fingerprint recognition systems,” arXiv preprint arXiv:2009.00128, 2020.

K. W. Bowyer and M. J. Burge, Handbook of iris recognition. Springer, 2016.

P. Grother, M. Ngan, and K. Hanaoka, “Face recognition vendor test (FRVT) part 3: Demographic effects,” in NIST, 2019.

J. J. Engelsma, D. Deb, K. Cao, A. Bhatnagar, P. S. Sudhish, and A. K. Jain, “Infant-id: Fingerprinters for global good,” IEEE PAMI, 2021.

X. Si, J. Feng, J. Zhou, and Y. Luo, “Detection and rectiﬁcation of distorted ﬁngerprints,” IEEE PAMI, vol. 37, no. 3, pp. 555–568, 2015.

K. Cao, D.-L. Nguyen, C. Tymoszek, and A. K. Jain, “End-to-end latent ﬁngerprint search,” IEEE TIFS, vol. 15, pp. 880–894, 2019.

J. J. Engelsma, K. Cao, and A. K. Jain, “Learning a ﬁxed-length ﬁngerprint representation,” IEEE PAMI, 2019.

L. Tran, X. Yin, and X. Liu, “Disentangled representation learning gan for pose-invariant face recognition,” in CVPR, pp. 1415–1424, 2017.

D. Deb, Best-Rowden and A. K. Jain, “Longitudinal study of automatic face recognition,” IEEE PAMI, vol. 40, no. 1, pp. 148–162, 2017.

M. Johnson, D. Yambay, D. Rissacher, L. Holsopple, and S. Schuckers, “A longitudinal study of iris recognition in children,” in IEEE ISBA, pp. 1–7, 2018.

S. Yoon and A. K. Jain, “Longitudinal study of ﬁngerprint recognition,” Proceedings of the National Academy of Sciences, vol. 112, no. 28, pp. 8535–8560, 2015.

P. J. Grother, J. R. Matey, E. Tabassi, G. W. Quinn, and M. Chumakov, “Irex vi-temporal stability of iris recognition accuracy,” NIST, 2013.

S. R. Coleman and R. Grover, “The Anatomy of the Aging Face: Volume Loss and Changes in 3-Dimensional Topography,” Aesthetic Surgery Journal, vol. 26, pp. S4–S9, 2006.

N. Ramanathan and R. Chellappa, “Modeling age progression in young faces,” in CVPR, 2006.

B. F. Klare, M. J. Burge, J. C. Klontz, W. V. Bruegge, Richard, and A. K. Jain, “Face recognition performance: Role of demographic information,” IEEE TIFS, vol. 7, no. 6, pp. 1789–1801, 2012.

D. Deb, L. Best-Rowden, and A. K. Jain, “Face recognition performance under aging,” in CVPRW, 2017.

D. Deb, N. Nain, and A. K. Jain, “Longitudinal study of child face recognition,” in IEEE ICCP, pp. 225–232, 2018.

D. Deb, D. Aggarwal, and A. K. Jain, “Identifying missing children: Face age progression via deep feature aging,” IEEE ICPR, 2020.

H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “Cosface: Large margin cosine loss for deep face recognition,” in CVPR, 2018.

J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in CVPR, pp. 4690–4699, 2019.

Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, “Ms-celeb-1m: A dataset and benchmark for large-scale face recognition,” in ECCV, pp. 87–102, 2016.

X. Yin, X. Yu, K. Sohn, L. Xu, and M. Chandraker, “Feature transfer learning for face recognition with under-represented data,” in CVPR, pp. 5704–5713, 2019.

J. Wan and Y. Wang, “Cost-sensitive label propagation for semi-supervised face recognition,” IEEE TIFS, vol. 14, no. 7, pp. 1729–1743, 2018.

L. Yang, D. Chen, X. Zhan, R. Zhao, C. C. Loy, and D. Lin, “Learning to cluster faces via confidence and connectivity estimation,” in CVPR, pp. 13369–13378, 2020.

S. Guo, J. Xu, D. Chen, C. Zhang, X. Wang, and R. Zhao, “Density-aware feature embedding for face clustering,” in CVPR, pp. 6698–6706, 2020.

A. RoyChowdhury, X. Yu, K. Sohn, E. Learned-Miller, and M. Chandraker, “Improving face recognition by clustering unconstrained labeled faces in the wild,” ECCV, pp. 119–136, 2020.

Q. Zhang, Z. Lei, and S. Z. Li, “Neighborhood-aware attention network for semi-supervised face recognition,” in IEEE IJCNN,
Anil K. Jain is a University distinguished professor in the Department of Computer Science and Engineering at Michigan State University. His research interests include pattern recognition and biometric authentication. He served as the editor-in-chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence and was a member of the United States Defense Science Board. He has received Fulbright, Guggenheim, Alexander von Humboldt, and IAPR King Sun Fu awards. He is a member of the National Academy of Engineering, and The World Academy of Sciences, and foreign fellow of the Indian National Academy of Engineering and Chinese Academy of Sciences.

Debayan Deb received his B.S. degree in computer science from Michigan State University, East Lansing, Michigan, in 2016. He is currently working towards a PhD degree in the Department of Computer Science and Engineering at Michigan State University, East Lansing, Michigan. His research interests include pattern recognition, computer vision, and machine learning with applications in biometrics.

Joshua J. Engelsma graduated magna cum laude with a B.S. degree in computer science from Grand Valley State University, Allendale, Michigan, in 2016. He is currently working towards a PhD degree in the Department of Computer Science and Engineering at Michigan State University. His research interests include pattern recognition, computer vision, and image processing with applications in biometrics. He won the best paper award at the 2019 IEEE International Conference on Biometrics (ICB), and the 2020 Michigan State University College of Engineering Fitch Beach Award.