Journal of *Cybersecurity and Privacy*

MDPI

Review

Biometric Systems De-Identification: Current Advancements and Future Directions

Md Shopon *, Sanjida Nasreen Tumpa, Yajurv Bhatia, K. N. Pavan Kumar and Marina L. Gavrilova *

Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada, sanjidanasreen.tumpa@ucalgary.ca (S.N.T.); yajurv.bhatia@ucalgary.ca (Y.B.); pavankumar.karkekopp@ucalgary.ca (K.N.P.K.)

* Correspondence: md.shopon@ucalgary.ca (M.S.); mgavrilo@ucalgary.ca (M.L.G.)

Abstract: Biometric de-identification is an emerging topic of research within the information security domain that integrates privacy considerations with biometric system development. A comprehensive overview of research in the context of authentication applications spanning physiological, behavioral, and social-behavioral biometric systems and their privacy considerations is discussed. Three categories of biometric de-identification are introduced, namely complete de-identification, auxiliary biometric preserving de-identification, and traditional biometric preserving de-identification. An overview of biometric de-identification in emerging domains such as sensor-based biometrics, social behavioral biometrics, psychological user profile identification, and aesthetic-based biometrics is presented. The article concludes with open questions and provides a rich avenue for subsequent explorations of biometric de-identification in the context of information privacy.

Keywords: human identity; privacy preservation; biometric security; de-identification; gait recognition; emotion recognition; social behavioral biometrics; personality traits estimation; risk analysis; threat assessment; secure surveillance

Citation: Shopon, M.; Tumpa, S.N.; Bhatia, Y.; Pavan Kumar, K.N.; Gavrilova, M.L. Biometric Systems De-Identification: Current Advancements and Future Directions. *J. Cybersecur. Priv.* **2021**, *1*, 470–496. https://doi.org/10.3390/jcp1030024

Academic Editor: Peter Corcoran

Received: 27 July 2021
Accepted: 26 August 2021
Published: 31 August 2021

1. Introduction

We live in a deeply interconnected society where aspects of someone's personal and social life, professional affiliations, hobbies, and interests become part of a public profile. A notable example where different facets of a person's life become publicized is their social network profiles or digital identities. The intricate relationships between online personalities and our physical world have useful applications in the areas of decision making, information fusion, artificial intelligence, pattern recognition, and biometrics. Extensive studies have evaluated intelligent methods and information fusion techniques in the information security domain [1,2]. Recent advancements in machine learning and deep learning present new opportunities to extract new knowledge from the publicly available data [3] and, thus, pose new threats to user privacy. This review article examines how integrating de-identification with other types of auxiliary information, which may be available directly or indirectly, can impact the performance of existing biometric identification systems. Analytical discussions on the de-identification of biometric data to protect user privacy are presented. This article also provides insights into the current and emerging research in the biometric domain and poses some open questions that are of prime import to information privacy and security researchers. The answers to these questions can assist the development of new methods for biometric security and privacy preservation in an increasingly connected society.

Privacy is an essential social and political issue, characterized by a wide range of enabling and supporting technologies and systems [4]. Amongst these are multimedia, big data, communications, data mining, social networks, and audio-video surveillance [5,6]. Along with classical methods of encryption and discretionary access controls, de-identification became one of the primary methods for protecting the privacy of multimedia content [7].

De-identification is defined as a process of removing personal identifiers by modifying or replacing them to conceal some information from public view [8]. However, de-identification has not been a primary focus of biometric research despite the pressing need for methodologies to protect personal privacy while ensuring adequate biometric trait recognition.

There is no agreement on a single definition for what de-identification truly is in the literature on the subject. For instance, Meden et al. [9] defined de-identification as follows: "The process of concealing personal identifiers or replacing them with suitable surrogates in personal information to prevent the disclosure and use of data for purposes unrelated to the purpose for which the data were originally collected". However, Nelson et al. [10] proposed the following definition: "De-identification refers to the reversible process of removing or obscuring any personally identifiable information from individual records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them. It involves the provision of additional information to enable the extraction of the original identifiers by, for instance, an authorized body". While the primary goal of de-identification is to protect user data privacy, its implementation is strikingly different depending on the application domain or the commercial value of the designed system. In the subsequent sections, we explore the differences among de-identification methodologies in depth, create a taxonomy of de-identification methods, and introduce new types of de-identification based on auxiliary biometric features.

This article summarizes fragmented research on biometric de-identification and provide a unique classification based on the mechanisms that achieve de-identification. Thus, it makes the following contributions:

1. For the first time, a systematic review is presented with a circumspect categorization of all de-identification methodologies based on the modalities employed and the types of biometric traits preserved after de-identification.
2. Four new types of emerging modalities are presented where de-identification is desirable and beneficial, namely sensor-based, emotion-based, social behavioral biometrics-based, and psychological traits-based de-identification.
3. A new paradigm for the design and implementation of multi-modal de-identification is proposed by considering the categories of traditional, soft, and auxiliary biometric traits and their de-identification.
4. A list of applications in the domains of cybersecurity, surveillance, risk analysis, mental health, and consumer applications is presented, where de-identification can be of critical importance in securing the privacy of biometric data.

2. Taxonomy of Biometric De-Identification

We start by introducing the definitions of *biometrics*, *traditional biometrics*, *soft biometrics*, *social behavioral biometrics*, and *emerging biometrics*.

Definition 1. *Biometrics: Biometrics are human characteristics that can be used to digitally identify a person to grant access to systems, devices, or data [11].*

Definition 2. *Traditional Biometrics: Traditional biometrics are defined as well-established biometrics that are categorized as either physiological (face, fingerprint, and iris) or behavioral (gait, voice, and signature) [11].*

Definition 3. *Soft Biometrics: The term soft biometrics is used to describe traits such as age, gender, ethnicity, height, weight, emotion, body shape, hair color, facial expression, linguistic and paralinguistic features, and tattoos, etc., that possess significantly lower uniqueness than traditional biometric traits [12].*

Definition 4. *Social Behavioral Biometrics: Social Behavioral Biometrics (SBB) is an identification of an actor (person or avatar) based on their social interactions and communication in different social settings [13].*

Definition 5. *Emerging Biometrics: Emerging biometrics are new biometric measures that have shown the prospect of enhancing the performance of the traditional biometrics by fusing these new biometric modalities with established ones [14].*

Thus, social behavioral biometrics can be considered as one example of emerging biometrics. Sensor-based, emotion-based, or psychological traits-based user identification are others examples of new identification types.

Current research into de-identification is highly dispersed. Hence, there has been no consistent method of classifying different approaches and reconciling various definitions of de-identification. In this review article, we categorize biometric de-identification into three classes based on the biometric type and the ability of a biometric system to identify a subject. The categories are as follows:

1. Complete de-identification;
2. Soft biometric preserving de-identification;
   (a) Utility lost de-identification;
   (b) Utility retained de-identification;
3. Traditional biometric preserving de-identification.

The proposed classifications of de-identification are discussed below.

1. **Complete De-identification:**

   Complete de-identification is the first category of de-identification research. A known problem of pair-wise constraint identification refers to a situation where a system can determine that two de-identified faces in a video belonging to the same individual by using hairstyle, clothing, dressing style, or other soft biometric features [5]. Thus, in addition to traditional biometric de-identification, soft biometric de-identification is also necessary. We define complete de-identification as a process where the biometric modality of a person is entirely de-identified, for instance, by being fully masked or obscured. Neither the identity of a person based on this biometric modality nor soft biometrics of the de-identified person can be recognized. This is true for human identification through visual inspection, as well as for a more common computer-based biometric system. Complete de-identification is used in mass media or police video footage, where sensitive information needs to be hidden [15].

2. **Soft Biometric Preserving De-identification:**

   Soft biometric preserving de-identification is the second proposed category. It is a process of de-identifying a particular traditional biometric trait, while the soft biometric traits remain distinguishable. The purpose of such de-identification methods is to remove the ability to identify a person using the given biometric, while still retaining soft biometric traits. For example, this type of de-identification would prevent face recognition technologies from identifying an individual, while still retaining their gender or age information [16], making it possible for a user to post a natural-looking video message on a public forum anonymously.
   We further subdivide this category into ***utility lost*** and ***utility retained de-identification***. As established above, in this group of methods, soft biometric traits are preserved, while the key traditional biometric trait/traits is/are obscured. The main difference is that in ***utility retained de-identification***, the biometric system is able to establish the identity of a person using the obscured key traditional biometric. In ***utility lost de-identification***, this is no longer possible for a human observer or for a computer [17].

3. **Traditional Biometric Preserving De-Identification:**

   Traditional biometric preserving de-identification is the third proposed category. It encompasses methods where only the soft biometric traits are obscured, while the key traditional biometric traits are preserved. Both human inspectors and biometric recognition systems are able to identify an individual based on their key biometric trait, whereas the soft biometric traits, such as height or a hair color, are rendered

non-identifiable [18]. For example, the face of an individual remains the same, while the height or a hair color is changed.
Figure 1 depicts the proposed classification of biometric de-identification.

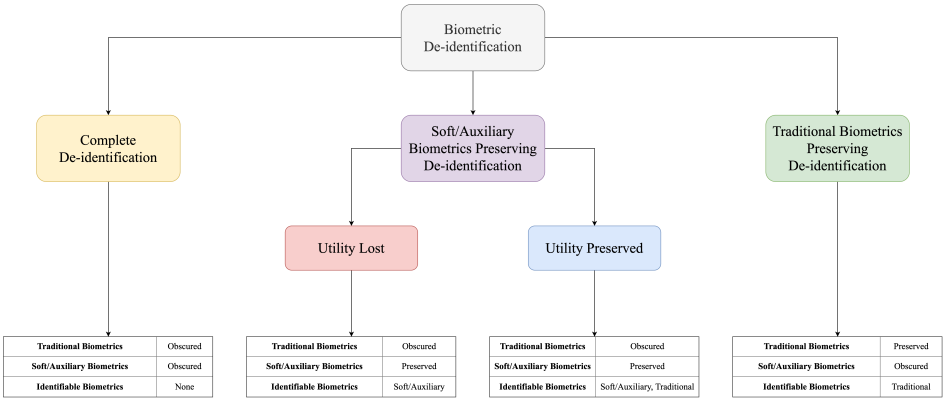

Figure 1. Taxonomy of biometric de-identification.

Finally, it is worth noticing that the traditional multi-modal biometric identification system can be classified according to the above taxonomy as having the following characteristics. Key traditional biometrics are preserved, soft biometrics are preserved, and identification can be performed from both traditional and soft biometrics.

In addition to the above-mentioned categories, the de-identification methods can be either ***reversible*** or ***non-reversible*** [6].

Definition 6. *Reversible De-identification: In reversible de-identification, the system is developed such that the modified biometric traits can be reversed back to their original form [6].*

Definition 7. *Irreversible De-identification: In irreversible de-identification, the transformation is intentionally developed not to be reversible [6].*

Recent developments have expanded our traditional understanding of biometric traits from physiological and behavioral to social, temporal, emotional, sensor-based, and other auxiliary traits [19].

Definition 8. *Auxiliary Biometric Trait: All biometric traits that are not unique enough on their own for person identification can be considered as auxiliary biometric traits. Thus, spatio-temporal patterns, idiosyncratic communication styles, personality types, emotions, age, gender, clothing, and social network connectivity are all examples of auxiliary biometric traits [19].*

For example, a person's emotion can be considered as an auxiliary trait, while their face or gait are treated as a traditional biometric trait. Either emotion or a key trait (or both) can be obscured in order to retain user privacy, based on the application domain and the main purpose of the biometric system.

Based on the above discussion, we propose an expansion of the taxonomy of biometric de-identification to include the aforementioned auxiliary categories of emerging biometric traits. The category definitions include the previously introduced taxonomy, where the notion of soft biometrics is expanded to include emerging auxiliary traits:

1. Complete de-identification;
2. Auxiliary biometric preserving de-identification;
   (a) Utility lost de-identification;
   (b) Utility retained de-identification;
3. Traditional biometric preserving biometric de-identification.

This classification is reflected in Figure 1.

3. Comprehensive Classification of Existing De-identification Methods

Privacy of biometric data is of paramount importance [20]. The research on biometric de-identification originated about a decade ago, with earlier works considering complete de-identification as the primary method of ensuring user privacy. This section summarizes key findings in the domains of traditional and soft biometric de-identification and classifies existing research studies into the proposed de-identification categories.

1. ***Complete de-identification*** refers to the modification of the original biometric trait such that the identifying information is lost. A comprehensive review of methods for visual data concealment is found in the work by Padilla-Lopez et al. [21]. They include filtering, encryption, reduction through k-same methods or object removal, visual abstraction, and data hiding. Korshunov et al. [22] used blurring techniques for complete face de-identification. Their process applied a blurring filter on the localized portion of a face. They also applied the pixelization and masking method to measure its impact on face de-identification. Subsequently, they performed a masking operation on a face image to fully prevent a biometric recognition system from identifying a face. Another work by Cichowski et al. [23] proposed a reversible complete de-identification method based on reallocating pixels in an original biometric. Recently, a complete generative de-identification system for full body and face was developed, utilizing an adaptive approach to de-identification [24]. In 2020, an interesting study [25] considered an effect of video data reduction on user awareness of privacy. Chriskos et al. [26] used hypersphere projection and singular value decomposition (SVD) to perform de-identification.

Behavioral biometric de-identification is also a popular topic. In one of the earlier works on speaker anonymization, Jin et al. [27] proposed a speaker de-identification system to prevent revealing the identity of the speaker to unauthorized listeners. The authors used a Gaussian Mixture Model (GMM) and a Phonetic approach for voice transformation and compared the performance. Magarinos et al. [28] suggested a speaker de-identification and re-identification model to secure the identity of the speaker from unauthorized listeners. In [29], the authors developed a speaker anonymization system by synthesizing the linguistic and speaker identity features from speech using neural acoustic and waveform models. Patino et al. [30] designed an irreversible speaker anonymization system using the McAdams coefficient to convert the spectral envelope of voice signals. Most recently, Turner et al. [31] presented a voice anonymization system that improved the anonymity of the existing x-vector by learning the distributional properties of the vector space. The generated anonymous voices were highly dissimilar and diverse from the original speakers while preserving the intra-similarity distribution. One of the most recent works on gait de-identification was carried out in 2019 by Tieu et al. [32]. They developed a gait de-identification system based on Spatio-Temporal Generative Adversarial Network (ST-GAN). The network incorporated noise in the gait distribution to synthesize the gait sequences for anonymization. In [33], the authors proposed a method that produced fully anonymized speech by adopting many-to-many voice transformation techniques based on variational autoencoders (VAEs). The method changed speaker's identity vectors of the VAE input in order to anonymize the speech data. The summary of complete de-identification research is presented in Table 1.

Table 1. Summary of complete de-dentification methods.

Authors	Year	De-Identified Biometrics	Unchanged Biometrics	Purpose of De-Identification	Dataset	Accuracy of Identification
Jin et al. [27]	2009	Voice	None	Completely unidentifiable speaker	WSJ0 corpus	Fully unrecognizable
Cichowski and Czyzewski [23]	2011	Face	None	Completely unidentifiable face	Custom dataset	Fully unrecognizable
Korshunov et al. [22]	2012	Full body	None	Completely unidentifiable full body	Custom dataset	Fully unrecognizable
Brkic et al. [24]	2017	Full body and face	None	Completely unidentifiable full body and face	Clothing Co-Parsing, Human3.6M	Fully unrecognizable
Magarinos et al. [28]	2017	Voice	None	Completely unidentifiable speaker	Spanish Albayzin database	Fully unrecognizable
Chriskos et al. [26]	2018	Face	None	Completely unidentifiable face	Custom dataset	Fully unrecognizable
Tieu et al. [32]	2019	Gait	None	Completely unidentifiable gait	CASIA-B	Fully unrecognizable
Fang et al. [29]	2019	Voice	None	Completely unidentifiable speaker	VoxCeleb, VCTK corpus	Fully unrecognizable
Patino et al. [30]	2020	Voice	None	Completely unidentifiable speaker	LibriSpeech, VCTK	Fully unrecognizable
Turner et al. [31]	2020	Voice	None	Completely unidentifiable speaker	VoicePrivacy challenge 2020	Fully unrecognizable
Yoo et al. [33]	2020	Voice	None	Completely unidentifiable speaker	Voice conversion challenge (VCC) 2016 corpus	Fully unrecognizable

2 ***Soft biometrics preserving de-identification*** aims to remove traditional biometric traits from the data while retaining the soft or auxiliary traits. For example, in gait recognition, clothing information can be retained, while gait patterns are rendered unrecognizable. The majority of research in this category has been performed on face biometric. Table 2 summarizes research focusing on *soft biometric preserving utility lost de-identification* and Table 3 summarizes research focusing on *soft biometric preserving utility retained de-identification*.

The k-Same technique is the commonly used method for soft biometrics preserving utility lost de-identification [16]. This method determines the similarity between faces based on a distance metric and creates new faces by averaging image components, which may be the original image pixels or eigenvectors, and is shown to be more effective than pixelation or blurring. Gross et al. [34] proposed the k-Same-M approach for face de-identification while preserving facial expression as soft biometrics by incorporating the Active Appearance Model (AAM). An active appearance model is a computer vision algorithm for matching a statistical model of object shape and appearance to a new image. Meden et al. [35] advanced this research further by proposing the k-Same-Net, which combined the k-anonymity algorithm with generative neural network architecture. Their method could de-identify the face while preserving its utility, natural appearance, and emotions. Du et al. [36] explicitly preserved race, gender, and age attributes in face de-identification. Given a test image, the authors computed the attributes and selected the corresponding attribute-specific AAMs. Meng et al. [37] adopted a model-based approach, representing faces as AAM features to avoid ghosting artifacts. Their approach identified k faces that were furthest away from a given probe face image. The algorithm calculated the average of the k furthest faces and returned it as a de-identified face, keeping the facial expression unchanged. Wang et al. [38] proposed a face de-identifying method using multi-mode discriminant analysis with AAM. By using orthogonal decomposition of the multi-attribute face image, they established the independent subspace of each attribute,

obtained the corresponding parameter, and selectively changed the parameters of other attributes in addition to the expression. Their system only preserves the facial expression and obscures all the other attributes. The k-Same furthest algorithm guarantees that the face de-identified by it can never be recognized as the original face, as long as the identity distance measure used to recognize the de-identified faces is the same as that used by the k-Same furthest method.

Apart from the k-same based method, several other approaches were proposed for soft biometrics preserving de-identification. Bitouk et al. [39] introduced an interesting idea of face-swapping, where a new face was blended with an original face and then lighting and contrast were adjusted to create a naturally looking de-identified face. The system preserved the body shape and pose of the person. Li and Lyu [40] used a neural style transfer method for face-to-face attribute transfer. The target was to preserve the consistency of non-identity attributes between the input and anonymized data while keeping the soft biometrics unaffected by this transfer. Brkic et al. [41] adopted a neural style transfer technique for face, hairstyle, and clothing de-identification, keeping the body shape preserved. Another work by the same authors focused on transferring a style content of an input image to the target image and performed a full-body and face obfuscation, while the shape of the subject remains identifiable [42]. Yang et al. [43] proposed an identity concealing method that preserved the pose, hair color, and facial expression. Their method added an adversarial identity mask to the original face image to remove the identity. Chi and Hu [44] used Facial Identity Preserving (FIP) features to preserve the aesthesis of the original images, while still achieving k-anonymity-based facial image de-identification. The main characteristic of the FIP features was that the conventional face descriptors significantly reduced intra-identity variances while maintaining inter-identity distinctions. In [45], the authors designed a gait de-identification system by using 2D Gaussian filtering to blur the human body silhouettes in order to conceal human body posture information while preserving the activity information. Not all work on de-identification involved videos and facial images. Malhotra et al. [46] proposed an adversarial learning-based perturbation algorithm, which modified the fingerprint of the finger-selfie. The algorithm prevented the finger-selfie to identify the person without authorization, entirely retaining the visual quality. Zhang et al. [47] proposed an iris de-identification system that prevented iris-based identification; however, it preserved the iris' biological features of an eye image fully. The iris area was detected based on the Hough transform and the transformation of iris information was performed by using the adopted polar coordinate transform. In [48], the authors designed a de-identification system using a face-swapping technique, Deepfake. The system retained the body and the face key points were almost unchanged, which were useful for medical purposes. Aggarwal et al. [49] proposed an architecture for face de-identification using conditional generative adversarial networks. This proposed method successfully preserved emotion while obscuring the identifiable characteristics in a given face image.

The focus of *soft biometric preserving utility retained de-identification* is to retain the ability of a system to preserve a person's soft biometric traits while concealing the primary biometric traits and retaining some face identification abilities. While soft biometrics and other auxiliary biometrics are preserved, key biometric recognition is still possible based on the modified primary traits. Yu et al. [15] utilized several abstract operations to de-identify a person's body shape. Recent methods demonstrated that new approaches comprising advanced image processing techniques can be superior to pixelation and blurring. The authors proposed another method to hide the subject's identity, while preserving his/her facial expression. Jourabloo et al. [50] proposed a joint model for face de-identification and attribute preservation of facial images by using the active appearance model (AAM) and k-same algorithm. The authors estimated the optimal weights for k-images instead of taking the average Hao et al. [51] proposed Utility-Preserving Generative Adversarial Network (UP-GAN), which aimed to provide a significant obscuration by generating faces that only depended on the non-identifiable facial features. Nousi et al. [17] proposed

an autoencoder-based method for de-identifying face attributes while keeping the soft biometrics (age and gender) unchanged. This method obscures the face attributes; however, they can still be identified by a face recognition system.

Table 2. Summary of soft biometrics preserving utility lost de-identification methods.

Authors	Year	De-Identified Traditional Biometrics	Unchanged Soft Biometrics	Purpose of De-Identification	Dataset	Accuracy of Identification
Newton et al. [16]	2005	Face	Facial attributes	Prevent face recognition but preserve facial attributes	FERET	Fully unrecognizable
Gross et al. [34]	2006	Face	Facial expression	Prevent face recognition but preserve expression	CMU Multi-PIE, FERET	Fully unrecognizable
Bitouk et al. [39]	2008	Face	Pose, body shape	Prevent face recognition but preserve body pose and shape	Proprietary dataset	Fully unrecognizable
Ivasic-Kos et al. [45]	2014	Gait	Activity	Prevent identification by gait but preserve activity	i3DPost database	Fully unrecognizable
Meng et al. [37]	2014	Face	Facial expression	Prevent face recognition but preserve expression	IMM	Fully unrecognizable
Du et al. [36]	2014	Face	Gender, age, race	Prevent face recognition but preserve gender, age, and race	MORPH	Fully unrecognizable
Chi and Hu [44]	2015	Face	Face geometry	Prevent face recognition but preserve geometry	MultiPIE	Fully unrecognizable
Brkic et al. [41]	2017	Face	Shape of the body	Altering face, hairstyle, and clothing but preserve body shape	ChokePoint, LFW	Fully unrecognizable
Brkic et al. [42]	2017	Face	Shape of the body	Altering face, hairstyle, and clothing but preserve body shape	Human3.6m, CDNet2014	Fully unrecognizable
Meden et al. [35]	2018	Face	Emotions	Prevent face recognition but preserve emotions	RaFD, CK+,XM2VTS	Fully unrecognizable
Zhang et al. [47]	2018	Iris	Biological features	Prevent iris recognition but biological features preserved	Custom dataset	Fully unrecognizable
Wang et al. [38]	2018	Face	Emotions	Prevent face recognition but preserve emotions	CK+	Fully unrecognizable
Li and Lyu [40]	2019	Face	Emotions	Prevent face recognition but preserve emotions	LFW, PIPA	Fully unrecognizable
Malhotra et al. [46]	2020	Fingerprint	Shape	Prevent fingerprint recognition but shape preserved	ISPFDv1, SMPF	Fully unrecognizable
Yang et al. [43]	2020	Face	Hair color, expression	Prevent face recognition but preserve expression	LFW, Megface	Fully unrecognizable
Zhu et al. [48]	2020	Face	Face shape and body shape	Swapping faces to protect privacy	Parkinson patients' dataset, Deep Fake Detection Dataset	Fully unrecognizable
Aggarwal et al. [49]	2020	Face	Emotion	Prevent face recognition but preserve expression	Radboud Faces Dataset	Fully unrecognizable

With the widespread use of video surveillance systems, the need for better privacy protection of individuals whose images were recorded increased. Thus, in 2011, Agrawal et al. [52] developed full-body obscuring video de-identification system. In their work, they preserved the activity, gender, and race of an individual. Meng et al. [53] used the concept of k-anonymity to de-identify faces while preserving the facial attributes for the face recognition system. In addition to that, emotion is also preserved in the de-identified

image. Their method was also applicable for video de-identification. Bahmaninezhad et al. [54] proposed a voice de-identification system to preserve the speaker's identity by modifying speech signals. The authors used a convolutional neural network model to map the voice signal, keeping the linguistic and paralinguistic features unchanged. Gafni et al. [55] proposed a live face de-identification method that automated video modification at high frame rates. Chuanlu et al. [56] proposed a utility preserving facial image de-identification using appearance subspace decomposition method. They showed that the de-identified faces preserved expressions of the original images while preventing face identity recognition. The system kept the perception, such as pose, expression, lip articulation, illumination, and skin tone identical. The summary of the above discussion is presented in Table 3.

Table 3. Summary of soft biometric preserving utility retained de-identification methods.

Authors	Year	De-Identified Traditional Biometrics	Unchanged Soft Biometrics	Purpose of De-Identification	Dataset	Accuracy of Identification
Yu et al. [15]	2008	Full body, face	Facial expression	Preserving expressions while gradually hiding visual information	Custom dataset	Fully unrecognizable
Agrawal et al. [52]	2011	Full body	Activity, gender, race	Preserving activity, gender, and race that do not reveal identity	CAVIAR, BEHAVE	Fully unrecognizable
Jourabloo et al. [50]	2015	Face	Facial attributes	Preserving facial features that do not reveal identity	FaceTracer, FaceScrub	Fully unrecognizable
Meng et al. [53]	2017	Face	Dynamic facial expression, behavior, gender, age, clothing	Preserving facial features that do not reveal identity	UNBC-McMaster Shoulder Pain Expression Database	Fully unrecognizable
Bahmanine-zhad et al. [54]	2018	Voice	Linguistic and paralinguistic features	Preserving speaker features that do not reveal identity	Voice conversion challenge 2016	Fully unrecognizable
Hao et al. [51]	2019	Face	Facial features, shape	Preserving facial features that do not reveal identity	UTKFace	Fully Fully unrecognizable
Gafni et al. [55]	2019	Face	Pose, expression, lip articulation, illumination, skin tone	Preserving facial features that do not reveal identity	LFW, CelebA, PubFig	Fully unrecognizable
Nousi et al. [17]	2020	Face	Age, gender, emotion	Preserving facial features that do not reveal identity	CelebA	Fully unrecognizable
Chuanlu et al. [56]	2020	Face	Expression	Preserving facial features and expression that do not reveal identity	Cohn-Kanade Dataset	Fully unrecognizable

Traditional biometrics preserving de-identification focuses on de-identifying soft biometrics such as gender, age, race, and ethnicity while preserving as much utility as possible from a traditional biometric trait (such as face, iris, and voice, etc.). Othman and Ross [57] introduced the method where a face image was modified but it remained recognizable, while the gender information was concealed. Lugini et al. [58] devised an ad-hoc image filtering-based method for eliminating gender information from the fingerprint images of the subjects, retaining the matching performance as it is. In [59], the authors proposed an automatic hair color de-identification system preserving face biometrics. The system segmented the image hair area and altered basic hair color for natural-looking de-identified images.

Mirjalili and Ross [60] proposed a technique that perturbed a face image so that the gender was changed by using a gender classifier while the face recognition capacity was

preserved. They used a warping technique to simultaneously modify a group of pixels by using the Delaunay Triangulation application on facial landmark points. The authors have experimented the system using two gender classifiers namely, IntraFace [61] and Commercial-off-The-Shelf (G-COTS) software. Mirjalili et al. [62] extended this idea by putting forward a convolutional autoencoder, which could modify an input face image to protect the privacy of a subject. They suggested an adversarial training scheme that was expedited by connecting a semi-adversarial module of a supplementary gender classifier and a face matcher to an autoencoder. The authors further tackled the generalizability of the proposed Semi Adversarial Networks (SANs) through arbitrary gender classifiers via the establishment of an ensemble SAN model, which generates a different set of modified outputs for an input face image. Later, in 2020, Mirjalili et al. [18] proposed a GAN-based SAN model, called PrivacyNet, which is further advanced to impart selective soft biometric privacy to several soft biometric attributes such as gender, age, and race. They showed that PrivacyNet provides a condition for users to decide which attributes should be obfuscated and which ones should remain unchanged. Chhabra et al. [63] proposed an adversarial perturbation-based de-identification algorithm, which anonymized k-facial attributes to remove gender, race, sexual orientation, and age information, preserving the identification ability of the face biometrics. Terhörst et al. [64] proposed a soft biometrics privacy-preserving system to hide binary, categorical, and continuous attributes from face biometric templates using an incremental variable elimination algorithm. Wang [65] applied face morphing to remove the gender identification attributes from face images. The identification ability of the face images as face biometrics was preserved.

An interesting direction of research is focused on tattoo de-identification. A unique tattoo in this case can be considered as a form of a soft biometric. In [66], the authors created a system to detect tattoos and de-identified it for privacy protection in still images. Hrkać et al. [67] designed a system to distinguish between tattooed and non-tattooed areas using a deep convolutional neural network. The neural network grouped the patches into blobs and replaced the pixel color inside the tattoo blob with the surrounding skin color to de-identify it. Another interesting type of soft biometric is clothing color. In 2018, Prinosil [68] proposed a method to de-identify clothing color as soft biometrics, keeping the traditional biometrics preserved. The system used silhouette splitting and clothing color segmentation algorithms. The components of the HSV color space of the segmented clothing were modified for de-identification. Pena et al. [69] proposed a method for two face representations that obscured facial expressions associated with emotional responses while preserving face identification. The summary of those methods is presented in Table 4.

For all discussed categories, the evaluation protocols for de-identification systems consist of human evaluation, re-identification, and diversity methods.

Definition 9. *Human Evaluation: Typically, in this evaluation method, experts are asked to recognize the de-identified person by performing a visual inspection [15].*

Definition 10. *Re-identification: Re-identification refers to identifying a particular person by using a classification method. Before performing de-identification, a classification method is used to classify the biometric data (images, videos, and signals) that will be de-identified. After performing the de-identification, the same method is used to check whether it can re-identify the data successfully [9].*

Definition 11. *Diversity: This evaluation protocol is used to show how diverse the de-identified face images are from the enrolled template database. As some of the above-mentioned methods use existing face images to de-identify a sample face image, this evaluation protocol determines how likely it is for a biometric recognition software to falsely match the de-identified version of the image with another identity that exists in the template database [17].*

Table 4. Summary of traditional biometric preserving de-identification methods.

Authors	Year	De-Identified Soft Biometrics	Unchanged Traditional Biometrics	Purpose of De-Identification	Dataset	Accuracy of Identification
Othman and Ross [57]	2014	Gender	Face	Concealing gender information	MUCT	Accuracy 95% using face biometrics, gender unidentifiable
Marčetic et al. [66]	2014	Tattoo	All traditional biometrics	Preventing unauthorized identification by tattoo	Custom dataset	Rate of false positive 4.90%, rate of false negative 29.05% for tattoo localization
Lugini et al. [58]	2014	Gender	Fingerprint	Concealing gender information	Custom dataset	Gender estimation reduced from 88.7% to 50.50%
Prinosil et al. [59]	2015	Hair color	Face	Concealing hair color	Custom dataset	Accuracy of face recognition not reported
Hrkać et al. [67]	2016	Tattoo	All traditional biometrics	Preventing unauthorized identification by tattoo	Custom dataset prepared from ImageNet	Tattoo identification accuracy 80–83%
Mirjalili and Ross [60]	2017	Gender	Facial biometrics except for gender	Preserving facial features and concealing gender	MUCT, LFW	76.6% error in gender prediction in MUCT, and 90.0% error in LFW dataset
Mirjalili et al. [62]	2018	Gender	Facial biometrics except for gender	Preserving facial features and concealing gender	CelebA-test, MUCT, LFW, AR-face	39.3% error in gender prediction in CelebA-test, 39.2% error in MUCT, 72.5% error in LFW, and 53.8% in AR-face dataset
Prinosil [68]	2018	Clothing color	All traditional biometrics	Concealing clothing color	Custom dataset	Average precision and recall for the upper body clothing was 95%, 90% and 83%, 86% for lower body clothing
Chhabra et al. [63]	2018	Gender, race, sexual orientation, age	Face	Preventing subjects from being profiled	MUCT, LFWcrop, CelebA	0.2% accuracy in gender, 0.1% accuracy in emotion recognition after de-identification
Terhörst et al. [64]	2019	Gender and age	Face	Preventing subjects from being profiled	Color FERET	Gender recognition accuracy 64.7%, age recognition accuracy 47.5%
Mirjalili et al. [18]	2020	Race, age, gender	Facial biometrics except for gender, race, and age	Controllable soft biometrics concealing	CelebA, MORPH, MUCT, RaFD, UTK-face	Gender recognition error increased to 60%, race recognition error increased to 20%
Wang [65]	2020	Gender	Face	Preserving facial features and concealing gender	CMU Multi-PIE database	Degree of gender obfuscation ranges from 0.077 to 0.298 (FaceVacs) and from 0.009 to 0.170 (ShuffleNet)
Pena et al. [69]	2020	Expression	Face	Preserving facial features and concealing expression	CelebA	Emotion recognition accuracy dropped 40% while identity recognition accuracy dropped 2%

One should keep in mind that attacks against de-identified systems are still possible. Thus, biometric systems must be tested against a possible attacker that can attempt to match de-identified images to originals. According to Newton et al. [16], there are three types of attacks and corresponding protocols to test whether de-identification of biometric information is effective in retaining data privacy. The attacks are as follows:

1. Matching original images to de-identified images;
2. Matching de-identified images to original images ;
3. Matching de-identified images to de-identified images.

In the first protocol, the attacker tries to match the original images with de-identified images, and this protocol is named naive recognition [16]. The gallery set in naive recognition protocol only includes the original images, and the probe set includes the de-identified images. The de-identified images are compared with the original images using standard face recognition software. No significant modification is performed on the original face set by the attacker in naive recognition.

In the second protocol, de-identified images are matched with the original images by the attacker, and this protocol is named reverse recognition [16]. In this protocol, it is assumed that the attacker has access to the original face images that were de-identified. The purpose for the attacker is to match one-to-one de-identified image with the original image set. Using principal component analysis, it is possible for an attacker to determine a one-to-one similarity between the de-identified image and the original image.

In the third protocol, the attacker tries to match the de-identified images to de-identified images, and this method is called parrot recognition [16]. Consider a scenario where the attacker already has the original face set of the de-identified images. In parrot recognition, the same distortion or modification is made to the original images as to the de-identified images. For instance, if blurring or pixelation is used for de-identifying an image set, the attacker can perform the same blurring or pixelation technique on the gallery set.

Thus, in order to fully validate the efficiency of de-identification, the above three types of attacks should be investigated, and the de-identified system performance should be tested against them.

4. Emerging Types of Biometric De-Identification

We now introduce additional types of de-identification related to emerging biometric research domains. These include sensor-based de-identification, social behavioral biometrics, emotion-based biometrics, and personality traits de-identification.

4.1. Sensor-Based Biometric De-Identification

Definition 12. *Sensor-based Biometric De-identification: Sensor-based biometric de-identification can be defined as the introduction of perturbation in sensor-based biometric data to obfuscate either traditional or auxiliary biometric traits or both of them.*

Some of the common sensor-based biometrics involves gait sequences and brain signals. Motions of a subject's body joints, while they are walking, represent their gait sequence, and they can be captured using RGB cameras or wearable sensors such as an accelerometer and a gyroscope or a marker-based sensor such as Vicon or a marker-less sensor such as Kinect or a combination thereof [70,71]. Brain signals are captured using an Electroencephalogram (EEG). EEG measures electrical impulses from several electrodes that are attached to the subject's scalp. The device can directly measure neuronal activity and is the most effective method for measuring neurons [72]. In the gait recognition domain, a biometric de-identification system can be designed by considering the gait as a primary behavioral biometric and the estimations of age, gender, emotion, or activity as auxiliary biometrics [71,73,74]. Furthermore, spatial and temporal features extracted over the gait sequence can act as the distinguishing characteristics for the identification of primary and auxiliary biometrics. For brain signal de-identification, a person's identity can remain recognizable while the information about their underlying emotions can be obfuscated.

Widespread deployment of sensors in both indoor and outdoor settings resulted in the application development based on biometric characteristics in domains such as kinesiology, physical rehabilitation, smart-home design, and search-and-rescue operations [19,75,76]. The appropriate architectural design of the biometric system can enable primary biometric identification and auxiliary biometric estimation. Therefore, perturbations need to be introduced in the data in order to obfuscate either the primary biometric trait or

auxiliary biometric traits or both to ensure biometric de-identification. Prior research conceals auxiliary biometric traits while preserving primary biometric traits within the data by introducing a deep learning-based neural style transfer [77]. Obscuring auxiliary biometric traits such as age, gender, activity, and emotion, while retaining the ability to identify a person using their gait can be a topic of future work in sensor-based biometric de-identification. Additionally, perturbing gait sequences to prevent gait-based identification while preserving the auxiliary biometric traits can be another future direction of research. The performance of the de-identification methods of each of the future works can be evaluated by using the established primary and auxiliary biometric identification and estimation methodologies.

The methods for identifying the primary biometric or for estimating the auxiliary biometric traits are available in the literature [6]. A deep learning-based approach, such as Generative Adversarial Network (GAN) [78], can be utilized to obtain the optimal perturbation scheme for sensor-based biometric data. In this method, the generator architecture of the network would be responsible for the perturbation and the discriminator architecture would handle the estimation of the primary and auxiliary biometric traits. The architecture of a gait-based behavioral biometric de-identification system is shown in Figure 2. The GAN is trained for the person identification task using either the primary biometric traits or auxiliary biometric traits depending on the desired de-identification mode. The random gait sequences which are perturbed using the generator network are passed into two discriminators, which are distinctly responsible for primary biometric de-identification and auxiliary biometric de-identification. The two discriminators are responsible for different tasks: one is to determine the person's Identity based on gait sequence and another is to estimate age, gender, or emotion from the gait. In the current system, both discriminators are executed. However, there could be a different system envisioned where only one of the discriminators is invoked.

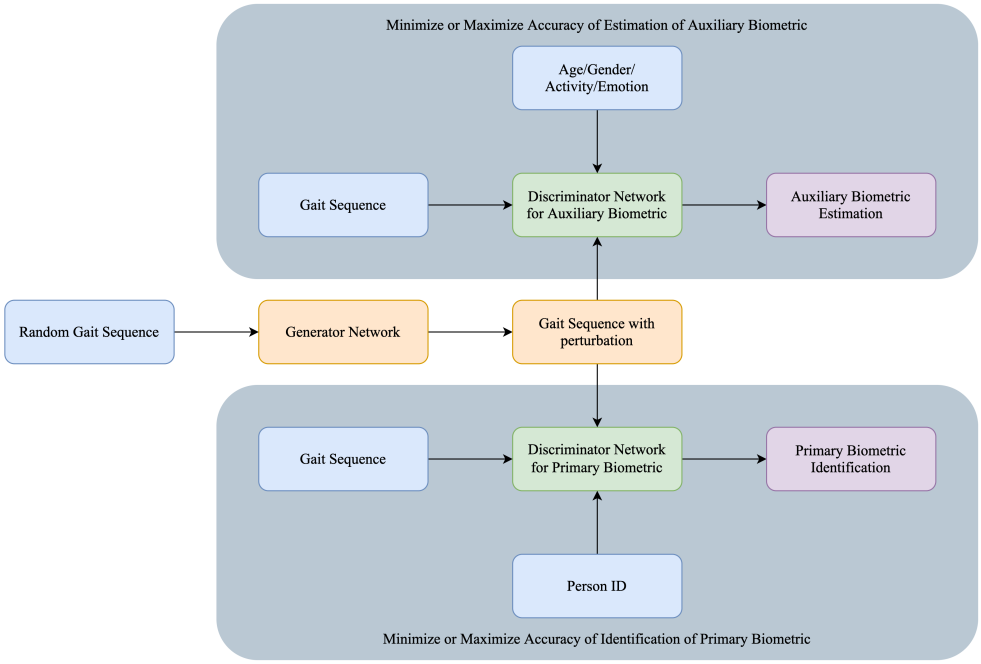

Figure 2. Architecture of gait-based biometric system that can identify both primary gait and auxiliary biometric traits.

In [79], the researchers proposed a method for person identification through gait videos. They found that wearing accessories introduce variations in an individual's gait patterns. Hence, they designed the identification system to handle gait sequences of a

person wearing a jacket, holding a bag, or having a specific type of footwear. Hence, another approach to de-identify gait sequences can be used to alter the appearance of the subject by adding artificial accessories using GNNs. This might preserve the original gait information for emotion recognition while perturbing the soft biometric traits. Table 5 summarizes the above mentioned sensor-based identification and de-identification research studies.

Table 5. Summary of sensor-based identification and de-dentification methods.

Authors	Year	Biometrics	De-Identified Biometrics	Unchanged Biometrics	Accuracy of Recognition
Ahad et al. [71]	2012	Gait	None	All biometrics	24.23% prediction error for gender estimation and 5.39 mean absolute error for age estimation
Iwashita et al. [79]	2013	Gait	None	All biometrics	94.0% on gait recognition
Brkić et al. [77]	2016	Full body	All traditional and soft biometrics	None	Qualitative evaluation
Xu et al. [80]	2017	Gait	None	All biometrics	8.92% mean absolute error on age estimation
Bari et al. [73]	2019	Gait	None	All biometrics	98.08% on person identification
Ahmed et al. [74]	2019	Gait	None	All biometrics	86.67% on emotion recognition

4.2. Emotion-Based De-Identification

Definition 13. *Emotion-based de-idetification: Emotion-based biometric de-identification can be defined as the introduction of perturbation in emotion to obfuscate either traditional or auxiliary biometric traits or both of them.*

Emotions are one of the most common auxiliary data that are frequently extracted from a human face; however, they can also be deduced from gait and speech [81]. For instance, the authors of [82] proposed a novel method to de-identify faces and the soft biometrics while retaining emotions. They highlighted the difference between their proposed method and naive approaches, such as blurring, pixelization, blindfolding, and inversion of the face images. Their adaptive filtering algorithm smoothed the facial details until the software-based authentication rate fell to approximately half of the original and the human recognition rate.

Thus, the authors of [40] masked original faces with donor-faces to de-identify an image of the original subject. The results show that emotions such as disgust, surprise, and neutrality are preserved 100% of the time, while anger and sadness are preserved more than 98% of the time. Lastly, fear and happiness are preserved only 79% of the time. Similarly, other works used Generative Neural Networks (GNNs) to mask original faces by using donor faces while preserving emotion [35].

The above research studies aimed to preserve emotion while concealing identities. A dual problem of concealing emotion while preserving identity is also possible for consideration. The authors of [83] used Cycle Generative Adversarial Networks (Cycle GANs) to transform a person's voice to hide emotions while retaining the ability for personal identification and speech recognition. Another less common parameter that can be estimated from a face is the body mass of a person [84].

Biometrics such as gait, Electroencephalogram (EEG), and Electrocardiography (ECG) are also gaining popularity for the emotion recognition problem and being researched for personal identification [72,79,85]. Since recognition methods involving these biometric traits are not studied as extensively as facial biometrics, experiments aimed at de-identification of these biometric traits have rarely been conducted. The particular biometric features that play a vital role in person identification are still uncertain; hence, not many have attempted to leverage those features. In [86], features responsible for human activity recognition were compared by using different machine learning methods. In [74], novel techniques for identifying the most significant gait features for emotion recognition were proposed. Such works can be extended to learn important features required for

gait-based person identification. Therefore, the features exclusively important for identification can be suppressed to achieve de-identification. Recently, many works attempt to identify person age from their biometrics. Notably, a recent attempt based on gait is presented in [80]. De-identifying age while preserving gait can be a new direction of research. Table 6 demonstrates the works that were performed on emotion-based identification and de-identification.

Future work in the domain of emotion-based de-identification can include investigations of other biometrics such as voice, signature, or a communication style in the presence of emotion-revealing traits.

Table 6. Summary of emotion-based identification and de-identification methods.

Authors	Year	Biometrics	De-Identified Biometrics	Unchanged Biometrics	Accuracy of Recognition
Letournel et al. [82]	2015	Face	Face	Expression	56.4% on re-identification
Jyotishi et al. [85]	2016	ECG Signal	None	All biometrics	97.3% on person identification
Meden et al. [35]	2018	Face	Face	Emotion	0.016% on re-identification
Li et al. [40]	2019	Face	Face	Emotion	16.5% on re-identification
Aloufi et al. [83]	2019	Voice	Emotion	All remaining biometrics	4.00% on re-identification
Ahmed et al. [74]	2019	Gait	None	All biometrics	86.67% on emotion recognition

4.3. Social Behavioral Biometrics-Based De-Identification

Definition 14. *Social Behavioral Biometrics-based De-identification: Social behavioral biometrics-based de-identification can be defined as obscuring either traditional or auxiliary social behavioral biometric traits or both of them to hide the identity of the users.*

As social beings, people communicate and interact with each other. Online social networking (OSN) platforms have evolved to become important extensions of the social fabric. Platforms such as Facebook, Instagram, Snapchat, LinkedIn, and Twitter, etc., emulate various facets of everyday social interactions within the personal, professional, and public realms of our society. According to the definition of Social Behavioral Biometrics (SBB), these social interactions possess many unique features that can be used as the person's biometric signature [13]. Social behavioral patterns provide important biometric cues and hold discriminating capabilities with regards to an individual's identity [13]. The area of social behavioral biometrics aims to model distinguishing characteristics that manifest within a subject's soft-biometric traits such as the patterns in their behaviors, social interactions, and communications. Over recent years, increased adoption and usage of online social platforms has meant that its users leave an ever-increasing trail of digital footprints in the form of the content they share or the patterns in their interactions with other users and the platform. Therefore, privacy preservation of these identifiable digital footprints is required in order to protect users' privacy. SBB-based de-identification refer to the original SBB traits and prevent person-identification.

The concept of Social Behavioral Biometrics (SBB) was introduced by Sultana et al. in 2015 [13]. The weighted networks are generated from the shared URLs, hashtags, retweeted, replied acquaintances, and the tweeting pattern of the users. Li et al. proposed a user identification method across social networks based on the k-hop ($k > 1$) friendship networks by considering the uniqueness of friendship networks [87]. Brocardo et al. proposed a method using the Gaussian–Bernoulli deep belief network to capture the writing style of the users obtained from the lexical, syntactic, and application-specific features for continuous user authentication of Twitter [88]. More recently, Tumpa et al. proposed an SBB system for user identification using users' linguistic profiles by applying score and rank level fusion [89].

Social Behavioral Biometrics de-identification is a new research avenue. For complete de-identification, all traditional and auxiliary SBB features must be obscured or masked. For example, one of the traditional SBB features is linguistic profiles. The linguistic profile of a

user can be masked by hiding the writing style of a user, which also changes the sentiment and emotion of the written contents [90]. Thus, both traditional and auxiliary features are obscured. In the case of auxiliary biometrics preserving de-identification, the sentiments of a user's tweets can be preserved while changing the vocabularies of the tweets. The identity of the user cannot be identified by using the traditional biometric, namely linguistic profile as this profile depends on the user's vocabulary for identification. However, the tweets deliver the same messages with the exact sentiments as the auxiliary biometrics are preserved. If the tweets of a user can be changed in such a way that a machine is able to retrieve the original tweets but a human cannot, then this de-identification is considered to be an auxiliary biometrics preserving utility retained de-identification. For the traditional biometric preserving de-identification, the sentiment from a tweet can be removed so that others will obtain the information expressed in the tweet but will not understand the sentiment of the user from that tweet. The examples are discussed considering linguistic profile as traditional biometric and sentiment as auxiliary biometric. A similar idea can be applied by considering the reply, retweet, URL, or hashtag network as traditional and tweeting behavior or emotion as auxiliary biometrics. The de-identification of SBB systems will help to preserve the privacy of the users without interrupting the legal use of information. Table 7 summarizes the works that were performed on social behavioral biometrics identification.

Table 7. Summary of social behavioral biometrics-based identification methods.

Authors	Year	Biometrics	De-Identified Biometrics	Unchanged Biometrics	Accuracy of Recognition
Wu et al. [90]	2006	Writing style	None	All	72.43% on person identification
Brocardo et al. [88]	2019	Writing style	None	All	23.49% on user verification
Tumpa et al. [89]	2020	Linguistic profile	None	All	99.45% on person identification
Li et al. [87]	2020	Twitter friendship networks	None	All	94.83% on user identification

4.4. Psychological Traits-Based De-Identification

Definition 15. *Psychological Traits-based De-identification: Psychological traits-based biometric de-identification can be defined as the introduction of perturbation in psychological traits to obfuscate either traditional or auxiliary biometric traits or both of them.*

Personality models have been used extensively by clinical psychologists to study the underlying factors influencing an individual's behavioral patterns [91]. While the users' personality traits have been shown to influence the language used to express themselves and the structure of their social network [92], this concept can be applied to the domain of social behavioral biometric recognition. In the interest of protecting user's privacy on OSN platforms, it is important to study the de-identification of personality traits from social network data. Social network data collected for user identification may also contain information regarding the users' psychological traits. Moreover, the psychological traits information may also be essential for user identification system. In such a scenario, psychological traits-based de-identification refers to the manipulation and storage of social network data in such a way that the personality traits information of users is obfuscated from the stakeholders in a user identification system development process and third parties while preserving the social behavioral user recognition capability from the data.

Patterns in social media activity and the contents of social media posts can be analyzed to predict the user's psychological traits. Research has also suggested that personality expressed through OSN platforms can represent unique, permanent, and predictive models of human behavior which can further be used for soft-biometric recognition [93]. Automated systems that classify the psychological traits of individuals via social network data use two prevalent personality scales: the big-five model and the Myers–Briggs Type Indicator (MBTI). The groundwork for applying personality traits-aware social computing systems within the domain of social intelligence and cybersecurity was first established

by Wang et al. in 2007 [94]. They demonstrated that the semantic characteristics of an individual's online language can reflect their underlying psychological state. Moreover, recent advances in the field of natural language processing (NLP) have produced powerful language models that can rapidly extract rich feature sets from textual data to be used for further classification tasks [95]. Using such distributed representations of text has shown to be instrumental in deciphering authors psychological traits from a relatively short corpus of their posts on OSN platforms [96]. A classifier trained to predict users' psychological traits based on the discussed models can embed information about the user's personality traits in the low-dimensional representation of data [97].

The count-based metric, such as Term Frequency-Inverse Document Frequency (TF-IDF), was used to extract characteristic features from tweets and was first used in [13]. This work demonstrated that the TF-IDF measure can be applied to the number of occurrences of replies and retweets in order to denote it as a friendship network. The frequency of overlap in the URLs and hashtags shared by users can be considered as their contextual profile. Additionally, temporal patterns can be extracted from a user's posting behavior to build a real-valued representation of a user profile. Follow-up studies aimed at closed-set user recognition on OSNs, with focus on user tweets and linguistic and stylistic signals [98]. Recently, many neural networks have been trained to efficiently learn representations of graph data to be further used for tasks such as node classification, link prediction, and user identity linkage [99]. These ideas allow generic and reusable input representation methodologies to be formalized for analyzing social network data. Representations of user-profiles can further be utilized to predict their personality traits and discern their identity.

For psychological traits de-identification, the first step is to convert textual, image, and/or graph data into real-valued vector representations to be processed by the subsequent individual component such as psychological traits classification and content-based feature extraction. After obtaining the intermediate representation for each user during enrollment, one can choose to preserve only the low-dimensional representations of the OSN users and discard the original content from their posts. Thereby, any psychological traits information and social behavioral biometric traits embedded in its content are obfuscated. During verification, the trained individual components are used to extract the representation of a test example, and a similarity-based decision-making component is employed to provide the user identification functionality. Table 8 depicts a summary of the above-mentioned works on psychological trait-based identification.

Table 8. Summary of psychological trait-based identification methods.

Authors	Year	Biometrics	De-Identified Biometrics	Unchanged Biometrics	Accuracy of Recognition
Arnoux et al. [96]	2017	Twitter tweets	None	All	Average correlation of 33% on 5 different personality traits
Kumar et al. [97]	2019	Twitter tweets	None	All	Average 76.55% accuracy on trait classification
Theóphilo et al. [98]	2019	Twitter tweets	None	All	65% on authorship attribution

4.5. Multi-Modal De-Identification System

Many existing biometric systems deployed in real-world scenarios are primarily unimodal, which means only one biometric trait is relied upon. In unimodal biometric systems, intraclass variation, non-universality, interclass variation, and noisy data are some of the problems that can result in high False Acceptance Rates (FAR). Multi-modal biometrics refers to the utilization of two or more biometric traits in an identification or verification system. Incorporating multiple biometric traits in a biometric identification system increases the accuracy rate of the system, which is technically true. It also uncovers how multi-modal systems can mitigate the effect of weaker modalities (one of their primary uses) [100,101].

Multi-modal biometric systems relying on information fusion techniques have seen widespread commercial use [102]. Hariprasath and Prabakar [103] proposed a multi-

modal identification system with palmprint and iris score level fusion and Wavelet Packet Transform. Murakami and Takahashi [104] utilized face, fingerprint, and iris biometric modalities with Bayes decision rule-based score level fusion techniques to identify these modalities. Ayed et al. [105] developed biometric system using face and fingerprint by using Local Binary Patterns (LBP) and Gabor wavelet. Next, a weighted sum-based match level fusion allowed for an increase in accuracy. Recently, deep learning has gained a great interest in biometric de-identification, partly due to its ability to apply style transfer to obscure visual information [106,107]. Architectures such as autoencoders, neural style transfer, Generative Adversarial Networks (GAN), and Recurrent Neural Networks (RNN) have been useful for enhanced accuracy of identification [99].

In order to perform multi-modal biometric de-identification, the biometric modalities to be de-identified must be chosen. The biometric modalities of interest should be extracted individually from the raw data. After extracting the biometric modalities, the type of de-identification should be selected. Finally, the de-identified biometric modalities need to be combined by using information fusion techniques. Depending on the type of de-identification, the identification or verification will be performed. The general framework for multimodal de-identification is depicted in Figure 3. We are not aware of developed multi-modal de-identified system, while research on cancelable multi-modal systems and privacy on the cloud has been conducted recently [108–110].

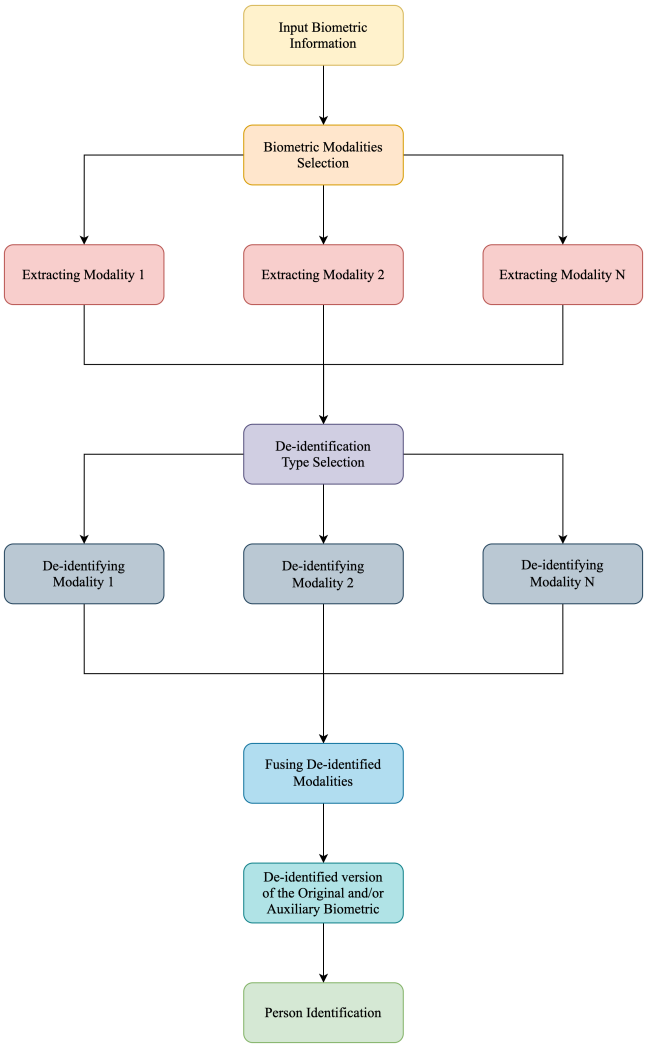

Figure 3. Flowchart of the multi-modal de-identification system.

5. Application Domains

This section summarizes the above discussion by providing a gamut of applications of emerging de-identification research.

Cybersecurity: Gathering intelligence by surveilling suspects in cyberspace is necessary to maintain a secure internet [111]. Government-authorized agents have been known to survey the social networks, disguising themselves among malicious users. Social behavioral biometrics-based de-identification can aid security agents in the covert observation and anonymous moderation of cyberspaces.

Continuous Authentication: Continuous authentication refers to a technology that verifies users on an ongoing basis to provide identity confirmation and cybersecurity protection [112]. Social behavioral biometrics (SBB) authenticates users on social networking sites continuously without any active participation of the users. The templates of users' writing patterns and acquaintance networks information must be stored in the database for SBB authentication. Instead of storing the identifying templates directly, SBB-based de-identification techniques can be applied to the templates to ensure account security and user privacy.

Protecting Anonymity: Authorized officials often publish case studies and written content of cybercrime victims to create public awareness [113]. In such cases, social networking portals and blogs are used as convenient media to disseminate information. Typically, the identities of the victims are kept anonymous. However, the content written by the victim and their social behavioral patterns may still contain identifying information. Therefore, de-identification of these published materials helps protect user anonymity when their identity must be kept confidential.

Multi-Factor Authentication: Leveraging the discriminative ability of an individual's social data and psychological information, a multi-factor authentication system can be implemented [114]. As a remote and accessible biometric, aesthetic identification can also provide additional security if the primary modality is suspected to be compromised. De-identification in this context would preserve the security of the system when storing a user's preference template.

Video Surveillance: Anonymization of primary or auxiliary biometric data protects the privacy of the subjects. If the original biometric is perturbed such that primary biometric identification is successful while the auxiliary biometric traits are not easily recognizable, or vice versa, this solution can be integrated with surveillance methods [115]. In such a situation, the de-identification of primary biometric can ensure the data privacy of individuals who appear in the footage but are not persons-of-interest.

Risk Analysis: The ability to estimate a person's emotional state using the facial biometric or gait analysis finds potential applications in threat-assessment and risk analysis [116]. Analysis of emotional state can be applied in the surveillance of public places in order to estimate the threat posed by an individual based on continuous monitoring of their emotional state. Based on the necessity of data protection, primary biometrics can be obscured while preserving the auxiliary information about emotions.

Health Care: Individuals can exhibit postural problems which could be diagnosed through static posture and gait analysis [117]. In such a case, primary gait biometric can be readily de-identified while preserving auxiliary biometric traits, such as age, gender, activity, and emotion.

Mental illness: Many applications predict and identify mental and/or physical illnesses by monitoring user emotions [74]. De-identifying any sensitive patient biometric data using the methods in the applications discussed above would ensure patient privacy, which could increase their willingness to opt-in for such services.

Adaptive Caregiving: The ability of an intelligent system to analyze user emotion information and exhibit realistic interactions has high potential [74]. De-identification of identity while still recognizing client emotions can preserve client privacy.

Advertisement: One reason why many social media companies mine their users' data is to identify customer interests and gain insights that can drive sales [118]. Naturally,

this raises concerns with regards to user data ownership and privacy. De-identifying the corresponding sensitive data while still understanding user's preferences towards certain products can supplement data mining.

Entertainment: Another possible usage of social behavioral information is adaptive entertainment experiences [119]. For instance, movies and/or video games that change the narrative based on the user's emotional responses can be created. However, such applications require the storage and analysis of user information. Users might be more willing to participate when user data are protected and anonymized.

Psychology: Personality traits can be revealed from the digital footprints of the users [19]. A personality trait de-identification system can be used to protect sensitive user information and implement privacy-preserving user identification systems. Furthermore, this concept can be applied in user behavior modeling problems such as predicting the likeliness to take a particular action, for example, clicking on a particular ad. Moreover, personality traits-based de-identification can be used in conjunction with other privacy-preserving measures such as data anonymization to further ensure user privacy protection within OSNs.

Consumer Services: Replacement of traditional identification cards by biometrics is the future of many establishments, such as driver license offices or financial services. De-identification of some real-time information obtained by security cameras for identity verification would ensure additional protection relative to sensitive user data [120].

6. Open Problems

The domain of biometric de-identification remains largely unexplored and has many promising avenues for further research. The impact of the perturbation in the original data on the identification of primary biometric and the estimation of auxiliary biometric can be further investigated. Moreover, the design of innovative deep learning architectures for sensor-based biometric de-identification can result in the development of a practical solutions for privacy preserving video surveillance systems. The acceptable obscureness of biometric data while preserving other biometric is open to discussion. Since certain behavioral biometrics may change over time, the procedure to adapt with the updated behavioral biometric in biometric de-identification requires further analysis in the future.

De-identification approaches for gait and gesture rely heavily on the blurring technique. In this scenario, retaining the naturalness of the de-identified video after the individual's characteristic walking patterns are obscured is one of the main challenges in gait and gesture de-identification. This represents one of the interesting open problems in the domains of gait and gesture de-identification.

Research in emotion-preserving de-identification has been more prevalent with faces than with any other biometric. For gait, EEG, and ECG, which are the most significant features for person identification, are unknown. Hence, the first step with these biometrics will be to identify the biometric features that are crucial for personal identification. Consequently, methods must be developed to obscure any personally identifiable information while retaining the features that represent the subject's emotion in the data. Additionally, face emotion-based de-identification research has produced some promising results. Hence, increasing person identification error is a likely future research direction for emotion preservation-based facial emotion recognition systems.

In the domain of social behavioral biometrics, de-identifying friendship and acquaintance networks is an open problem. The technique of changing the linguistic patterns of social media tweets while preserving emotions and information, and vice versa, has not been explored previously. The reversibility to the original SBB traits after de-identification and subsequent measures to increase the difficulty of reverse-engineering those traits are other interesting problems to explore.

There are many open problems in applying the concept of psychological trait-based de-identification within the domain of privacy-preserving social behavioral biometrics. While clinical research indicates the permanence of psychological traits among adults,

they change over time due to significant life events and circumstances. Considering time dependencies and their effect on data preservation is another interesting open problem.

Psychological traits factorize a wide range of human behaviors into a fixed number of labels. Therefore, any de-identification of psychological traits may result in the loss of a nuanced representation of user-generated content. This loss of information may reduce the accuracy of the downstream prediction task. Mitigating this unwanted effect is one of the open problems. Secondly, the degree to which a dataset is de-identified may not be directly measurable. As humans may not be capable of inferring psychological traits from user content, it is difficult to ascertain if the information regarding psychological traits is truly obfuscated from automated systems. This is another interesting problem that should be investigated further.

Finally, multi-modal biometric de-identification has not been explored before. Common multi-modal biometric authentication systems involve combining traditional biometric traits with emerging biometric traits using information fusion. One potential open problem is to design a multi-modal de-identification system that conceals soft biometric traits. As there can be several fusion methods for combining biometric modalities, experiments aimed at finding the most suitable architecture in the context of an applied problem are needed. For multi-modal de-identification, some applications may require all the biometric traits to be obscured, while some may need only particular traits to be modified. Formalizing the underlying principles for the optimal design of multi-modal biometric systems offers a rich avenue for future investigations.

7. Conclusions

This article provided a comprehensive overview of the domain of biometric information de-identification for ensuring user data privacy. For the first time, a systematic review of all de-identification methodologies based on the modalities employed and the types of unchanged biometric traits was presented. Analytical discussions on how physiological, behavioral, and social-behavioral biometric data can be protected in various applications were carried out. By drawing on the most recent developments in the biometric security domain, the paper also introduced new de-identification paradigms: social biometric de-identification, sensor-based de-identification, emotion-based de-identification, and psychological traits de-identification. Multitudes of potential applications of the de-identification concept in public health and safety domains were described. Finally, the article formulated a set of open questions concerning the future direction of investigations in this vibrant research field. Answers to those questions will assist not only in the establishment of the new methods in the biometric security and privacy domains but also provide insights into the future emerging topics in big data analytics and social networking research.

Author Contributions: Conceptualization, M.S.; Funding acquisition, M.L.G.; Investigation, M.S. and S.N.T.; Methodology, M.S., S.N.T., Y.B., and K.N.P.K.; Supervision, M.L.G.; Validation, S.N.T. and M.L.G.; Writing—original draft, M.S., S.N.T., Y.B., and K.N.P.K.; Writing—review and editing, M.S., S.N.T., Y.B., K.N.P.K. and M.L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NSERC Discovery Grant #10007544 and NSERC Strategic Planning Grant #10022972.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the National Sciences and Engineering Research Council of Canada for partial support of this research study in the form of the NSERC Discovery Grant, NSERC Strategic Planning Grant, and the Department of National Defense's Innovation for Defense Excellence and Security (IDEaS) program Canada.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AAM	Active Appearance Model;
ECG	Electroencephalogram;
FAR	False Acceptance Rates;
FIP	Facial Identity Preserving;
GAN	Generative Adversarial Network;
GMM	Gaussian Mixture Model;
GNN	Generative Neural Networks;
LBP	Local Binary Patterns;
MBTI	Myers-Briggs Type Indicator;
NLP	Natural Language Processing;
OSN	Online Social Networking;
RNN	Recurrent Neural Networks;
SAN	Semi Adversarial Networks;
SBB	Social Behavioral Biometrics;
ST-GAN	Spatio-Temporal Generative Adversarial Network;
SVD	Singular Value Decomposition;
TF-IDF	Term Frequency-Inverse Document Frequency;
UP-GAN	Utility-Preserving Generative Adversarial Network.

References

1. Jain, L.C.; Halici, U.; Hayashi, I.; Lee, S.; Tsutsui, S. *Intelligent Biometric Techniques in Fingerprint and Face Recognition*; CRC Press: Boca Raton, FL, USA, 1999; Volume 10.
2. Jain, L.C.; Martin, N. *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications*; CRC Press: Boca Raton, FL, USA, 1998; Volume 4.
3. Tsihrintzis, G.A.; Jain, L.C. *Machine Learning Paradigms: Advances in Deep Learning-Based Technological Applications*; Springer Nature: Berlin, Germany, 2020; Volume 18.
4. Baaziz, N.; Lolo, N.; Padilla, O.; Petngang, F. Security and privacy protection for automated video surveillance. In Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 15–18 December 2007; pp. 17–22.
5. Chen, D.; Chang, Y.; Yan, R.; Yang, J. Protecting personal identification in video. In *Protecting Privacy in Video Surveillance*; Springer: Berlin, Germany, 2009; pp. 115–128.
6. Ribaric, S.; Ariyaeeinia, A.; Pavesic, N. De-identification for privacy protection in multimedia content: A survey. *Signal Process. Image Commun.* **2016**, *47*, 131–151. [CrossRef]
7. Garfinkel, S.L. *De-Identification of Personal Information*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2015.
8. Ribaric, S.; Pavesic, N. An overview of face de-identification in still images and videos. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 4, pp. 1–6.
9. Meden, B.; Peer, P.; Struc, V. Selective face deidentification with end-to-end perceptual loss learning. In Proceedings of the 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), San Carlos, Costa Rica, 18–20 July 2018; pp. 1–7.
10. Nelson, G.S. Practical implications of sharing data: A primer on data privacy, anonymization, and de-identification. In Proceedings of the SAS Global Forum Proceedings, Dallas, TX, USA, 26–29 April 2015; pp. 1–23.
11. Jain, A.; Hong, L.; Pankanti, S. Biometric identification. *Commun. ACM* **2000**, *43*, 90–98. [CrossRef]
12. Dantcheva, A.; Elia, P.; Ross, A. What else does your biometric data reveal? A survey on soft biometrics. *IEEE Trans. Inf. Forensics Secur.* **2015**, *11*, 441–467. [CrossRef]
13. Sultana, M.; Paul, P.P.; Gavrilova, M. Social behavioral biometrics: An emerging trend. *Int. J. Pattern Recognit. Artif. Intell.* **2015**, *29*, 1556013. [CrossRef]
14. Chauhan, S.; Arora, A.; Kaul, A. A survey of emerging biometric modalities. *Procedia Comput. Sci.* **2010**, *2*, 213–218. [CrossRef]
15. Yu, X.; Chinomi, K.; Koshimizu, T.; Nitta, N.; Ito, Y.; Babaguchi, N. Privacy protecting visual processing for secure video surveillance. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1672–1675.
16. Newton, E.M.; Sweeney, L.; Malin, B. Preserving privacy by de-identifying face images. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 232–243. [CrossRef]
17. Nousi, P.; Papadopoulos, S.; Tefas, A.; Pitas, I. Deep autoencoders for attribute preserving face de-identification. *Signal Process. Image Commun.* **2020**, *81*, 115699. [CrossRef]

18. Mirjalili, V.; Raschka, S.; Ross, A. PrivacyNet: Semi-adversarial networks for multi-attribute face privacy. *IEEE Trans. Image Process.* **2020**, *29*, 9400–9412. [CrossRef]
19. Tumpa, S.N.; Kumar, K.P.; Sultana, M.; Hsu, G.S.J.; Yadid-Pecht, O.; Yanushkevich, S.; Gavrilova, M.L. Social Behavioral Biometrics in Smart Societies. In *Advancements in Computer Vision Applications in Intelligent Systems and Multimedia Technologies*; IGI Global: Hershey, PA, USA, 2020; pp. 1–24.
20. Natgunanathan, I.; Mehmood, A.; Xiang, Y.; Beliakov, G.; Yearwood, J. Protection of privacy in biometric data. *IEEE Access* **2016**, *4*, 880–892. [CrossRef]
21. Padilla-López, J.R.; Chaaraoui, A.A.; Flórez-Revuelta, F. Visual privacy protection methods: A survey. *Expert Syst. Appl.* **2015**, *42*, 4177–4195. [CrossRef]
22. Korshunov, P.; Cai, S.; Ebrahimi, T. Crowdsourcing approach for evaluation of privacy filters in video surveillance. In Proceedings of the ACM Multimedia 2012 Workshop on Crowdsourcing for Multimedia, Nara, Japan, 29 October 2012; pp. 35–40.
23. Cichowski, J.; Czyzewski, A. Reversible video stream anonymization for video surveillance systems based on pixels relocation and watermarking. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 1971–1977.
24. Brkic, K.; Sikiric, I.; Hrkac, T.; Kalafatic, Z. I know that person: Generative full body and face de-identification of people in images. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1319–1328.
25. Boyle, M.; Edwards, C.; Greenberg, S. The effects of filtered video on awareness and privacy. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, New York, NY, USA, 23–27 October 2021; pp. 1–10.
26. Chriskos, P.; Zhelev, R.; Mygdalis, V.; Pitas, I. Quality Preserving Face De-identification Against Deep CNNs. In Proceedings of the 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP), Aalborg, Denmark, 17–20 September 2018; pp. 1–6.
27. Jin, Q.; Toth, A.R.; Schultz, T.; Black, A.W. Voice convergin: Speaker de-identification by voice transformation. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 3909–3912.
28. Magarinos, C.; Lopez-Otero, P.; Docio-Fernandez, L.; Rodriguez-Banga, E.; Erro, D.; Garcia-Mateo, C. Reversible speaker de-identification using pre-trained transformation functions. *Comput. Speech Lang.* **2017**, *46*, 36–52. [CrossRef]
29. Fang, F.; Wang, X.; Yamagishi, J.; Echizen, I.; Todisco, M.; Evans, N.; Bonastre, J.F. Speaker anonymization using x-vector and neural waveform models. *arXiv* **2019**, arXiv:1905.13561.
30. Patino, J.; Tomashenko, N.; Todisco, M.; Nautsch, A.; Evans, N. Speaker anonymisation using the McAdams coefficient. *arXiv* **2020**, arXiv:2011.01130.
31. Turner, H.; Lovisotto, G.; Martinovic, I. Speaker Anonymization with Distribution-Preserving X-Vector Generation for the VoicePrivacy Challenge 2020. *arXiv* **2020**, arXiv:2010.13457.
32. Tieu, N.D.T.; Nguyen, H.H.; Nguyen-Son, H.Q.; Yamagishi, J.; Echizen, I. Spatio-temporal generative adversarial network for gait anonymization. *J. Inf. Secur. Appl.* **2019**, *46*, 307–319. [CrossRef]
33. Yoo, I.C.; Lee, K.; Leem, S.; Oh, H.; Ko, B.; Yook, D. Speaker Anonymization for Personal Information Protection Using Voice Conversion Techniques. *IEEE Access* **2020**, *8*, 198637–198645. [CrossRef]
34. Gross, R.; Sweeney, L.; De la Torre, F.; Baker, S. Model-based face de-identification. In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA, 17–22 June 2006; pp. 161–161.
35. Meden, B.; Emeršič, Ž.; Štruc, V.; Peer, P. k-Same-Net: k-Anonymity with generative deep neural networks for face deidentification. *Entropy* **2018**, *20*, 60. [CrossRef]
36. Du, L.; Yi, M.; Blasch, E.; Ling, H. GARP-face: Balancing privacy protection and utility preservation in face de-identification. In Proceedings of the IEEE International Joint Conference on Biometrics, Clearwater, FL, USA, 29 September–2 October 2014; pp. 1–8.
37. Meng, L.; Sun, Z. Face de-identification with perfect privacy protection. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; pp. 1234–1239.
38. Wang, X.; Xiong, C.; Pei, Q.; Qu, Y. Expression preserved face privacy protection based on multi-mode discriminant analysis. *CMC Comput. Mater. Contin.* **2018**, *57*, 107–121. [CrossRef]
39. Bitouk, D.; Kumar, N.; Dhillon, S.; Belhumeur, P.; Nayar, S.K. Face swapping: automatically replacing faces in photographs. In *ACM SIGGRAPH 2008 Papers*; Association for Computing Machinery: New York, NY, USA, 2008; pp. 1–8.
40. Li, Y.; Lyu, S. De-identification without losing faces. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Paris, France, 3–5 July 2019; pp. 83–88.
41. Brkić, K.; Hrkać, T.; Kalafatić, Z.; Sikirić, I. Face, hairstyle and clothing colour de-identification in video sequences. *IET Signal Process.* **2017**, *11*, 1062–1068. [CrossRef]
42. Brkić, K.; Hrkać, T.; Kalafatić, Z. Protecting the privacy of humans in video sequences using a computer vision-based de-identification pipeline. *Expert Syst. Appl.* **2017**, *87*, 41–55. [CrossRef]
43. Yang, X.; Dong, Y.; Pang, T.; Zhu, J.; Su, H. Towards privacy protection by generating adversarial identity masks. *arXiv* **2020**, arXiv:2003.06814.

44. Chi, H.; Hu, Y.H. Face de-identification using facial identity preserving features. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 586–590.
45. Ivasic-Kos, M.; Iosifidis, A.; Tefas, A.; Pitas, I. Person de-identification in activity videos. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; pp. 1294–1299.
46. Malhotra, A.; Chhabra, S.; Vatsa, M.; Singh, R. On privacy preserving anonymization of finger-selfies. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 26–27.
47. Zhang, H.; Zhou, H.; Jiao, W.; Shi, J.; Zang, Q.; Sun, J.; Zhang, J. Biological features de-identification in iris images. In Proceedings of the 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), Yichang, China, 16–18 October 2018; pp. 67–71.
48. Zhu, B.; Fang, H.; Sui, Y.; Li, L. Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New Orleans, LA, USA, 13 February 2018 2020; pp. 414–420.
49. Aggarwal, A.; Rathore, R.; Chattopadhyay, P.; Wang, L. EPD-Net: A GAN-based Architecture for Face De-identification from Images. In Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 21–24 April 2020; pp. 1–7.
50. Jourabloo, A.; Yin, X.; Liu, X. Attribute preserved face de-identification. In Proceedings of the 2015 International Conference on Biometrics (ICB), Phuket, Thailand, 19–22 May 2015 ; pp. 278–285.
51. Hao, H.; Güera, D.; Reibman, A.R.; Delp, E.J. A utility-preserving gan for face obscuration. *arXiv* **2019**, arXiv:1906.11979.
52. Agrawal, P.; Narayanan, P. Person de-identification in videos. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 299–310. [CrossRef]
53. Meng, L.; Sun, Z.; Collado, O.T. Efficient approach to de-identifying faces in videos. *IET Signal Process.* **2017**, *11*, 1039–1045. [CrossRef]
54. Bahmaninezhad, F.; Zhang, C.; Hansen, J.H. Convolutional Neural Network Based Speaker De-Identification. Odyssey, 2018; pp. 255–260. Available online: https://www.semanticscholar.org/paper/Convolutional-Neural-Network-Based-Speaker-Bahmaninezhad-Zhang/f2cd2f81b188166058ea04b454a4c59135d744a5 (accessed on 25 August 2021)
55. Gafni, O.; Wolf, L.; Taigman, Y. Live face de-identification in video. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 9378–9387.
56. Chuanlu, L.; Yicheng, W.; Hehua, C.; Shuliang, W. Utility Preserved Facial Image De-identification Using Appearance Subspace Decomposition. *Chin. J. Electron.* **2021**, *30*, 413–418. [CrossRef]
57. Othman, A.; Ross, A. Privacy of facial soft biometrics: Suppressing gender but retaining identity. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2014; pp. 682–696.
58. Lugini, L.; Marasco, E.; Cukic, B.; Dawson, J. Removing gender signature from fingerprints. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; pp. 1283–1287.
59. Prinosil, J.; Krupka, A.; Riha, K.; Dutta, M.K.; Singh, A. Automatic hair color de-identification. In Proceedings of the 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Delhi, India, 8–10 October 2015; pp. 732–736.
60. Mirjalili, V.; Ross, A. Soft biometric privacy: Retaining biometric utility of face images while perturbing gender. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 564–573.
61. De la Torre, F.; Chu, W.S.; Xiong, X.; Vicente, F.; Ding, X.; Cohn, J. Intraface. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1, pp. 1–8.
62. Mirjalili, V.; Raschka, S.; Namboodiri, A.; Ross, A. Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images. In Proceedings of the 2018 International Conference on Biometrics (ICB), Gold Coast, Australia, 20–23 February 2018; pp. 82–89.
63. Chhabra, S.; Singh, R.; Vatsa, M.; Gupta, G. Anonymizing k-facial attributes via adversarial perturbations. *arXiv* **2018**, arXiv:1805.09380 .
64. Terhörst, P.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Suppressing gender and age in face templates using incremental variable elimination. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.
65. Wang, S.; Kelly, U.M.; Veldhuis, R.N. Gender obfuscation through face morphing. In Proceedings of the 2021 IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, 6–7 May 2021; pp. 1–6.
66. Marcetic, D.; Ribaric, S.; Struc, V.; Pavesic, N. An experimental tattoo de-identification system for privacy protection in still images. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014.
67. Hrkać, T.; Brkić, K.; Ribarić, S.; Marčetić, D. Deep learning architectures for tattoo detection and de-identification. In Proceedings of the 2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE), Aalborg, Denmark, 6–8 July 2016; pp. 1–5.
68. Prinosil, J. Clothing Color Based De-Identification. In Proceedings of the 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 4–6 July 2018; pp. 1–5.

69. Peña, A.; Fierrez, J.; Morales, A.; Lapedriza, A. Learning emotional-blinded face representations. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3566–3573.
70. Tao, W.; Liu, T.; Zheng, R.; Feng, H. Gait analysis using wearable sensors. *Sensors* **2012**, *12*, 2255–2283. [CrossRef] [PubMed]
71. Ahad, M.A.R.; Ngo, T.T.; Antar, A.D.; Ahmed, M.; Hossain, T.; Muramatsu, D.; Makihara, Y.; Inoue, S.; Yagi, Y. Wearable sensor-based gait analysis for age and gender estimation. *Sensors* **2020**, *20*, 2424. [CrossRef]
72. Ismail, W.W.; Hanif, M.; Mohamed, S.; Hamzah, N.; Rizman, Z.I. Human emotion detection via brain waves study by using electroencephalogram (EEG). *Int. J. Adv. Sci. Eng. Inf. Technol.* **2016**, *6*, 1005–1011. [CrossRef]
73. Bari, A.H.; Gavrilova, M.L. Artificial neural network based gait recognition using kinect sensor. *IEEE Access* **2019**, *7*, 162708–162722. [CrossRef]
74. Ahmed, F.; Bari, A.H.; Gavrilova, M.L. Emotion recognition from body movement. *IEEE Access* **2019**, *8*, 11761–11781. [CrossRef]
75. Tang, Y.; Teng, Q.; Zhang, L.; Min, F.; He, J. Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. *IEEE Sens. J.* **2020**, *21*, 581–592. [CrossRef]
76. Ahmed, F.; Polash Paul, P.; Gavrilova, M.L. Kinect-based gait recognition using sequences of the most relevant joint relative angles. *J. WSCG* **2015**, *23*, 147–156.
77. Brkić, K.; Sikirić, I.; Hrkać, T.; Kalafatić, Z. De-identifying people in videos using neural art. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016; pp. 1–6.
78. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*. Available online: https://papers.nips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf (accessed on 15 August 2021).
79. Iwashita, Y.; Uchino, K.; Kurazume, R. Gait-based person identification robust to changes in appearance. *Sensors* **2013**, *13*, 7884–7901. [CrossRef] [PubMed]
80. Xu, C.; Makihara, Y.; Ogi, G.; Li, X.; Yagi, Y.; Lu, J. The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 1–14. [CrossRef]
81. El Ayadi, M.; Kamel, M.S.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit.* **2011**, *44*, 572–587. [CrossRef]
82. Letournel, G.; Bugeau, A.; Ta, V.T.; Domenger, J.P. Face de-identification with expressions preservation. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27 September–1 October 2015; pp. 4366–4370.
83. Aloufi, R.; Haddadi, H.; Boyle, D. Emotionless: Privacy-preserving speech analysis for voice assistants. *arXiv* **2019**, arXiv:1908.03632.
84. Wen, L.; Guo, G. A computational approach to body mass index prediction from face images. *Image Vis. Comput.* **2013**, *31*, 392–400. [CrossRef]
85. Jyotishi, D.; Dandapat, S. An LSTM-Based Model for Person Identification Using ECG Signal. *IEEE Sens. Lett.* **2020**, *4*, 1–4. [CrossRef]
86. Li, F.; Shirahama, K.; Nisar, M.A.; Köping, L.; Grzegorzek, M. Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors. *Sensors* **2018**, *18*, 679. [CrossRef] [PubMed]
87. Li, Y.; Su, Z.; Yang, J.; Gao, C. Exploiting similarities of user friendship networks across social networks for user identification. *Inf. Sci.* **2020**, *506*, 78–98. [CrossRef]
88. Brocardo, M.L.; Traore, I.; Woungang, I. Continuous authentication using writing style. In *Biometric-Based Physical and Cybersecurity Systems*; Springer: Berlin, Germany, 2019; pp. 211–232.
89. Tumpa, S.N.; Gavrilova, M.L. Score and Rank Level Fusion Algorithms for Social Behavioral Biometrics. *IEEE Access* **2020**, *8*, 157663–157675. [CrossRef]
90. Wu, C.H.; Chuang, Z.J.; Lin, Y.C. Emotion recognition from text using semantic labels and separable mixture models. *ACM Trans. Asian Lang. Inf. Process. (Talip)* **2006**, *5*, 165–183. [CrossRef]
91. Goldberg, L.R. The structure of phenotypic personality traits. *Am. Psychol.* **1993**, *48*, 26. [CrossRef] [PubMed]
92. Ning, H.; Dhelim, S.; Aung, N. PersoNet: Friend recommendation system based on big-five personality traits and hybrid filtering. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 394–402. [CrossRef]
93. Saleema, A.; Thampi, S.M. User Recognition Using Cognitive Psychology Based Behavior Modeling in Online Social Networks. In *International Symposium on Signal Processing and Intelligent Recognition Systems*; Springer: Berlin, Germany, 2019; pp. 130–149.
94. Wang, F.Y.; Carley, K.M.; Zeng, D.; Mao, W. Social computing: From social informatics to social intelligence. *IEEE Intell. Syst.* **2007**, *22*, 79–83. [CrossRef]
95. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
96. Arnoux, P.H.; Xu, A.; Boyette, N.; Mahmud, J.; Akkiraju, R.; Sinha, V. 25 tweets to know you: A new model to predict personality with social media. In Proceedings of the International AAAI Conference on Web and Social Media, Montréal, Québec, Canada, 11–15 April 2016; Volume 11.
97. Kumar, K.P.; Gavrilova, M.L. Personality traits classification on twitter. In Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 18–21 September 2019; pp. 1–8.

98. Theóphilo, A.; Pereira, L.A.; Rocha, A. A needle in a haystack? harnessing onomatopoeia and user-specific stylometrics for authorship attribution of micro-messages. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2692–2696.
99. Khamparia, A.; Singh, K.M. A systematic review on deep learning architectures and applications. *Expert Syst.* **2019**, *36*, e12400. [CrossRef]
100. Sanjekar, P.; Patil, J. An overview of multimodal biometrics. *Signal Image Process.* **2013**, *4*, 57.
101. Monwar, M.M.; Gavrilova, M.; Wang, Y. A novel fuzzy multimodal information fusion technology for human biometric traits identification. In Proceedings of the IEEE 10th International Conference on Cognitive Informatics and Cognitive Computing (ICCI-CC'11), Banff, AB, Canada, 18–20 August 2011; pp. 112–119.
102. Yang, F.; Ma, B. A New Mixed-Mode Biometrics Information Fusion Based-on Fingerprint, Hand-geometry and Palm-print. In Proceedings of the Fourth International Conference on Image and Graphics (ICIG 2007), Sichuan, China, 22–24 August 2007; pp. 689–693.
103. Hariprasath, S.; Prabakar, T. Multimodal biometric recognition using iris feature extraction and palmprint features. In Proceedings of the IEEE-International conference on Advances in Engineering, Science And Management (ICAESM-2012), Tamil Nadu, India, 30–31 March 2012; pp. 174–179.
104. Murakami, T.; Takahashi, K. Fast and accurate biometric identification using score level indexing and fusion. In Proceedings of the 2011 International Joint Conference on Biometrics (IJCB), Washington, DC, USA, 11–13 October 2011; pp. 1–8.
105. Ayed, N.G.B.; Masmoudi, A.D.; Masmoudi, D.S. A new human identification based on fusion fingerprints and faces biometrics using LBP and GWN descriptors. In Proceedings of the Eighth International Multi-Conference on Systems, Signals & Devices, Sousse, Tunisia, 22–25 March 2011; pp. 1–7.
106. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
107. Sundararajan, K.; Woodard, D.L. Deep learning for biometrics: A survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–34. [CrossRef]
108. A. K. Jain, A. A. Ross, K.N. *Introduction to Biometrics*; Springer Science & Business Media: Berlin, Germany, 2011.
109. Sudhakar, T.; Gavrilova, M. Cancelable biometrics using deep learning as a cloud service. *IEEE Access* **2020**, *8*, 112932–112943. [CrossRef]
110. Paul, P.P.; Gavrilova, M.; Klimenko, S. Situation awareness of cancelable biometric system. *Vis. Comput.* **2014**, *30*, 1059–1067. [CrossRef]
111. Sarker, I.H.; Kayes, A.; Badsha, S.; Alqahtani, H.; Watters, P.; Ng, A. Cybersecurity data science: an overview from machine learning perspective. *J. Big Data* **2020**, *7*, 1–29. [CrossRef]
112. Deutschmann, I.; Nordström, P.; Nilsson, L. Continuous authentication using behavioral biometrics. *IT Prof.* **2013**, *15*, 12–15. [CrossRef]
113. Jones, L.M.; Finkelhor, D.; Beckwith, J. Protecting victims' identities in press coverage of child victimization. *Journalism* **2010**, *11*, 347–367. [CrossRef]
114. Dasgupta, D.; Roy, A.; Nag, A. Multi-factor authentication. In *Advances in User Authentication*; Springer: Berlin, Germany, 2017; pp. 185–233.
115. Sreenu, G.; Durai, M.S. Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *J. Big Data* **2019**, *6*, 1–27. [CrossRef]
116. Mason, J.E.; Traoré, I.; Woungang, I. Applications of gait biometrics. In *Machine Learning Techniques for Gait Biometric Recognition*; Springer: Berlin, Germany, 2016; pp. 203–208.
117. Ahmed, F.; Bari, A.H.; Sieu, B.; Sadeghi, J.; Scholten, J.; Gavrilova, M.L. Kalman filter-based noise reduction framework for posture estimation using depth sensor. In Proceedings of the 2019 IEEE 18th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), Milan, Italy, 23–25 July 2019; pp. 150–158.
118. Bhowmik, A.; Gafur, S.R.; Rafid, A.; Azad, S.; Mahmud, M.; Kaiser, M.S. User Awareness for Securing Social Networks. In *Securing Social Networks in Cyberspace*; CRC Press: Boca Raton, FL, USA, 2021; pp. 3–15.
119. Wong, K.K.W. Player adaptive entertainment computing. In Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts, Perth, Australia, 19–21 September 2007; pp. 13–13.
120. Mrityunjay, M.; Narayanan, P. The de-identification camera. In Proceedings of the 2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Hubli, India, 15–17 December 2011; pp. 192–195.