Review on Age Identification Mechanism Based on Biometrical Data

Author: Reza Ghasemi

February 20, 2023

1 Abstract

The gradual increase in computing power has made biometric systems widely popular. Most smartphones and devices offer biometrics for authentication purposes. In addition to authentication, biometrics could also be used to estimate additional information. This is known as "soft biometrics" and features could indicate age, gender, ethnicity, or other information about subjects. The goal is to review the literature for each subsequent feature, discuss the details of the approaches used, and discern the validity of each feature for age identification. The gradual increase in computing power has made biometric systems widely popular. Most smartphones and devices offer biometrics for authentication purposes. In addition to authentication, biometrics could also be used to estimate additional information. This is known as soft biometrics and features could indicate age, gender, ethnicity or other information about subjects. The goal is to review the literature for each subsequent feature and to discuss details of approaches used and to discern validity of each feature for age identification.

2 Introduction

The term "biometrics" is derived from the Greek words "bio" meaning life, and "metric" meaning to measure. The gradual increase in computation power has made biometrics a popular choice, especially in recent years.

The ideas behind current automated systems are not entirely novel but are based on pre-existing notions known to humankind. Before the modern era, our ancestors used fingerprints and handprints as forms of authentication. It has been stated that, in ancient Babylon, fingerprints were used as a means of authentication on contracts. Similarly, in ancient China, government documents included fingerprints [1]. Furthermore, the Persian historian Rashid al-Din Hamadani reported the use of fingerprints as signatures in 1303, stating, "Experience shows that no two individuals have fingers exactly alike" [2].

On a daily basis, users rely on biometric systems to authenticate themselves. Currently, most smartphones offer biometric solutions as a substitute for the traditional PIN/password approach [3]. Users are no longer required to remember long sequences of characters or experience the loss of a token. Some argue that biometrics are generally better for user security [4].

However, similar to any other technology, there is a trade-off. Although they might offer a higher degree of security and convenience, at the same time, the privacy of the individual may be entirely eroded if data has not been stored properly. When using passwords, after a data breach, you simply have to change your password. Recovery from an attack on biometric data, on the other hand, will be more difficult [5].

Recently, a military device containing the fingerprints and iris scans of 2,632 individuals (US military members and residents of Iraq and Afghanistan) was purchased in an auction. The device was not properly sanitized, and the buyer had access to all the information stored on it [6].

Typically, biometric systems are divided into three categories: biological, morphological, and behavioral [7]. Biological biometrics refer to traits at a molecular level, like the DNA. Morphological (*e.g.* fingerprint) refers to the structure of the human body, and more physical traits and behavioral traits are based on patterns of behavior like gait, speech, or handwriting.

By definition, biometrics refer to any metric used to measure the physiological or behavioral characteristics of an individual. The most well-known traits are fingerprints and facial features. In fingerprint scanners, patterns of ridges and valleys are captured and compared with the template stored within the system. In face recognition systems, facial contours are analyzed. Numerous features could be analyzed to identify or verify the identity of an individual, each feature having its own set of advantages and disadvantages.

In recent years, soft biometrics have been employed for other applications besides authentication. They are not so distinctive to be used for distinguishing individuals from each other, but they could be used for estimation of information such as age, gender, ethnicity, etc. The purpose of this paper is to review the literature to understand whether they have been studied for estimating age and whether they have been effective or not.

A biometric system is composed of multiple components: Initially, a sensor is used to capture the characteristics, then features are extracted, and depending on whether the aim is identification or verification, features will be matched to make the final decision.

3 Objective

The main objective of this work is to review the literature for different biometric features and analyze their effectiveness for the task of age estimation.

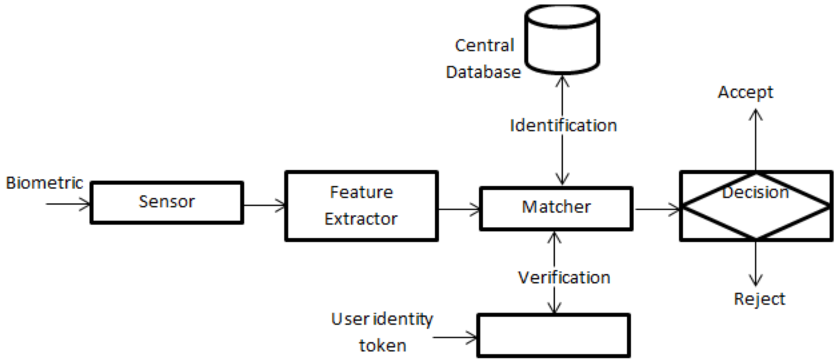

Figure 1: A biometric system is composed of various components.

4 Facial Features

Biometric systems based on facial features have become widespread and offer numerous advantages compared to other means of recognition like fingerprint scanners. Facial features could be captured from a distance; hence, cooperation is not needed. Moreover, specialized hardware is not needed, and a normal camera would be sufficient to acquire the image. It should be mentioned that the US government is actively trying to improve image acquisition from long-range distances in challenging environments for advanced security and military purposes [8, 9]. The ability to capture images from a distance without the subject's knowledge raises ethical concerns that, if misused, could have a negative impact on social liberties and privacy [10, 11, 12, 13].

Facial recognition systems are usually composed of multiple stages. In the first stage, given an image, face is detected and cropped. Then, face alignment by facial landmarks and normalization occur to improve pose, size, etc. In the third phase, features from the face are extracted, which will represent the face. The features will be stored within the system and later, in the last stage, be used for matching and making a decision. The steps can be seen in Figure 2.

Figure 2: If the score computed based on facial features is higher than a specific threshold, the subject will be accepted.

4.1 Age estimation using facial features

The features of the face are not only limited to verification and identification purposes. There has been extensive research for the purpose of soft biometrics, age estimation being one of them. This could prove useful if we want to limit access to a specific product for a certain age range or improve the customer experience. Face contains numerous demographic identities [14]. The features of the face may depend on various factors, including gender, ethnicity, etc. The features of the face are normally divided into two categories: global (the whole face) and local (regions of the face). Local features used could be wrinkles, aging spots, etc. Aznar-Casanova et al. showed that the number of wrinkles could estimate the age group of an individual [15]. For instance, a greater number of wrinkles could indicate an older individual. Local features are useful to estimate age groups, whereas global features are better for estimating exact ages. Of course, age estimation is not extremely simple, as various factors could introduce challenges. A skin disease could cause spots on the surface of the face, or living in a polluted area may cause skin aging [16]. The features of the face may also differ depending on race or gender [17].

Most of the studies done so far indicate that deep learning produces better results for age estimation [18]. Needless to say, if the network has more layers and parameters, the final result will be more accurate. Unconstrained datasets used will be more prone to errors as they contain images from different poses, different degrees of illumination, etc. So it is more likely to have better results if constrained datasets are used. But of course, we should try to find a solution that achieves good accuracy with unconstrained datasets since, in real-life settings, we would be faced with similar scenarios as if we were using an unconstrained dataset.

Given an input image or video, the face must be detected and cropped. Then, pre-processing occurs (to improve alignments, illumination, pose, etc.) before feature extraction. Feature extraction could be done either automatically by using deep learning or manually by an algorithm. Some studies used the manual approach [19], while others used deep learning [20, 21]. The manual approach has the advantage of not requiring a large number of samples compared to deep learning, where it is considered to be data-hungry. One drawback of manual algorithms is the need to define features beforehand. Unlike the manual approach, deep learning could learn them automatically.

Methods used for age estimation could be classified into multiple categories, which have been employed in different studies so far. The techniques include classification, regression, ranking, and hybrid. In classification, age is considered a class. For instance, we could have three different classes. The classes must be defined beforehand; for instance, we could have the following classes: children, adults, and the elderly. Some researchers have treated age estimation as a classification problem [22, 23]. Classification is useful where we have a small number of classes. To estimate the exact age, other methods like regression should be preferred. Regression considers age to be a numerical value, and multiple studies have used regression [20, 25]. Ranking was introduced to overcome the disadvantages of regression and classification and should be used for comparison. For instance, if we want to find out which subject is younger between two subjects, we could use the ranking technique [26]. This method, compared to other methods, is considered to be less accurate. Hybrid, as the name suggests, combines classification and regression. It is more complicated to implement, but it also achieves better results, and it has been employed by researchers for age estimation [27].

One essential ingredient for training a model is using an appropriate dataset.

The correct choice would have a crucial impact on the outcome. As it was mentioned before, datasets could be divided into two categories, namely unconstrained and constrained datasets [18]. Unconstrained datasets contain facial images from different poses and under different conditions, but constrained datasets consist of frontal and upright facial images. It should be noted that the samples in the datasets do not necessarily have equal distributions. Nonetheless, many datasets exist that are available to researchers; some examples include: AgeDB, UTKFace, IMDB-wiki, etc.

Once the estimation is done, we must evaluate its performance. Two common parameters used for this purpose are MAE (mean absolute error) and CS (cumulative score) [18]. An additional parameter that could be used is the Gaussian error, which can be computed in the following manner:

$$GaussianError = 1 - \Sigma_{n=1}^{N} exp(-\frac{(\hat{y_n} - y_n)^2}{2\sigma_n^2}) \tag{1}$$

In this equation, N represents the number of samples in the test set, $\hat{y_n}$ indicates the estimated age value of the n-th sample, σ_n is the standard deviation, and y_n is the ground truth.

4.2 Methodology summary

Overall, facial features do not necessitate complex systems, and simple cameras embedded in phones and laptops are sufficient to capture the image from which facial features are extracted. Nonetheless, age estimation is faced with numerous challenges as many variables have a potential influence on the outcome. Pollution, gender, race, and health status are just a few of the variables that must be taken into account in the study. Despite whether we treat age estimation as a regression, classification, ranking, or hybrid problem, the number of parameters and the dataset used are crucial.

Researchers should aim to find a solution that performs well using unconstrained datasets as they simulate real-life scenarios. Even though in the studies done so far, deep learning has achieved better performance compared to manual algorithms, traditional approaches should not be completely abandoned, and more study is required in this field.

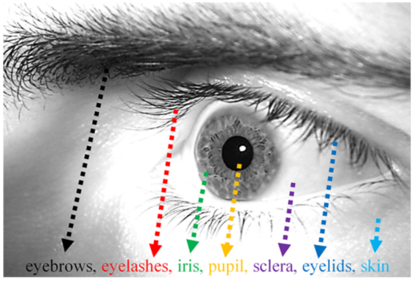

Figure 3: Pupil refers to the circular orifice that is at the center of the iris. Image credit: Karakaya (2021)

5 Pupil

Features extracted from the periocular area (iris, retina, pupil, skin texture, etc.) are rich in features that can be extracted and used for various purposes, [28]. One of the many advantages of periocular features is that they are considered to be non-intrusive. Many biometric features, while being extremely accurate (*e.g.* DNA), are not easy to collect, which makes them unsuitable for applications where the subject is not accommodating.

In certain scenarios, we may even wish to classify a subject without notifying him. Take, for instance, a terrorist who is under suspicion. We plan to estimate his age to ensure it is him before taking any action against him. Furthermore, close contact with dangerous prisoners in a prison may pose a risk to individuals, and we may want to limit physical contact as much as possible.

Besides, periocular features like the pupil do not require physical contact. Lack of physical contact makes them a more hygienic choice compared to fingerprints, hence more suitable to be used during pandemics. Moreover, the pupil reveals additional information, like the emotional state of subjects [29].

Furthermore, it has been observed that the dilation of pupils among male and female subjects differs. An experiment showed that the pupils of male heterosexual individuals started to dilate after an image of a female individual was shown to them. The same phenomenon occurred when heterosexual females were exposed to images of male individuals. What made women different from men was that their pupils dilated even when they were shown images of female subjects [30]. Hence, the pupil could be used as an indication of gender. Moreover, pupils gradually change over time. It has been observed that younger subjects have larger pupil sizes compared to children and old subjects, which makes it a useful trait for age estimation [31].

5.1 Age estimation using pupil

A recent study examined the pupil as a soft biometric to determine its effectiveness [32]. The subjects in this study range in age from 17 to 80. The total

Table 1
Age range of participants in GANT dataset.

Age range	# participants
a (17–18)	11
b (21–30)	57
c (31–40)	9
d (41–50)	16
e (51–60)	8
f (61–70)	9
g (71–80)	1

Table 1: Image credit: Cascone et al. (2020b)

number of subjects was 112, with 72 men and 39 women. As it can be seen from Table 1, the age range was divided into seven groups. Each subject was shown images (landscapes, human faces, and blank screens) in random order. For the learning approach, supervised learning was used, and different classifiers were employed to observe which had the best performance.

In this scheme, two elements play a major role in determining the output: the extracted features and the selected parameters. Also, to avoid overfitting, the k-fold method was adopted, and eighty percent of the dataset was used as the training set and the rest as the test set. Preprocessing removed eye images with low confidence. If one of the eyes (*e.g.* the left eye) had high confidence but the other had low confidence, both were filtered out. For the task of age estimation, the MLP (multilayer perceptron) had the best performance out of all classifiers on both pupils and right pupil. The Gaussian Process and Linear Support Vector performed best for the left pupil.

The pupils in this study have also been studied for gender classification due to the different reactions of males and females to stimuli. However, the result for age classification proved to be superior to using this feature for gender classification, and the classifiers used achieved comparable and satisfactory results, as shown in the Table 2.

	PUPIL - LEFT (PL)				PUPIL - RIGHT (PR)				PUPIL - BOTH (PLR)			
CLASSIFIER	ACCUR.	PRECIS.	RECALL	F1	ACCUR.	PRECIS.	RECALL	F1	ACCUR.	PRECIS.	RECALL	F1
MLP	0.8018	0.7277	0.9145	0.8105	**0.8369**	**0.8362**	*0.8059*	0.8208	**0.8277**	**0.7616**	*0.9145*	**0.8311**
DT	0.7988	0.756	*0.8355*	*0.7938*	0.8079	0.7214	0.9539	0.8215	0.8079	0.7214	0.9539	0.8215
GNB	*0.7957*	0.7361	0.8717	0.7982	0.7851	0.69	0.9737	0.8076	0.8049	0.728	0.9243	0.8145
ADA	0.8018	0.7364	0.8914	0.8065	0.7851	0.69	0.9737	0.8076	0.8034	0.7238	0.9309	0.8144
BC	0.8018	**0.7605**	*0.8355*	0.7962	0.7851	0.69	0.9737	0.8076	0.8003	0.7125	0.9539	0.8158
QDA	0.8018	0.7266	0.9178	0.811	0.7851	0.69	0.9737	0.8076	0.7927	0.7069	0.9441	0.8085
LDA	0.7988	0.7287	0.9013	0.8059	0.7759	0.6796	0.977	0.8017	0.7912	0.7072	0.9375	0.8062
LinSVC	**0.8034**	0.7285	0.9178	**0.8122**	0.779	0.6828	0.977	0.8038	0.7881	0.696	0.9638	0.8083
KNN	0.7988	0.7228	0.9178	0.8087	*0.7652*	*0.6667*	**0.9868**	*0.7958*	0.7851	0.6945	0.9572	0.805
GaussPro	**0.8034**	0.7285	0.9178	**0.8122**	0.7698	0.6734	0.977	0.7973	0.7851	0.6909	0.9704	0.8071
SGD	0.8003	0.7294	0.9046	0.8076	0.8201	0.7397	0.9441	**0.8295**	0.7835	0.6893	0.9704	0.806
SVM	0.8018	0.7266	0.9178	0.811	0.7805	0.6852	0.9737	0.8043	0.7713	0.6742	**0.9803**	0.7989
GB	0.8003	0.7224	0.9243	0.811	*0.7683*	0.6743	0.9671	*0.7946*	0.7652	*0.6682*	**0.9803**	0.7947
RF	*0.7957*	*0.7146*	**0.9309**	0.8086	0.8064	0.7185	0.9572	0.8209	*0.7591*	*0.6629*	0.977	*0.7899*

Table 2: In this table, the best performance is highlighted using bold, whereas the worst performance is shown using italic. Image credit: Cascone et al. (2020)

5.2 Methodology summary

Pupil is a feature that does not require contact and is an appealing option to use, especially during pandemics. It could be employed not simply as a means of identification and verification but also for gender and age classification since we know that pupils differ among females and males and at different ages. The results indicate that the pupil is a promising feature to be used for age estimation.

6 Ear

The ear as a biometric feature has gained more attention from researchers in recent years. It was first introduced by Bertillon in 1896 as a possible feature to be used for identification [33]. It has the capability of being used in conjunction with other biometric features (*e.g.* facial features) to improve accuracy. Especially if we have a limited view of facial features due to issues with illumination, etc. The facial features may also be affected by facial hair and make-up, whereas the ears remain mostly unaffected.

It should be mentioned that that ear has its shortcomings. For instance, it is not a reliable feature to be used in outdoor environments, and it has been mostly studied in closed and controlled settings. Moreover, ear recognition has proven to be a challenging problem [34]. Furthermore, it may have problems with ear individuality, which, while highly variable between individuals, has not been scientifically proven to be unique for each individual. Also, it could be partially or fully occluded by elements like strands of hair, hats, etc.

A biometric system based on the ear can capture images, whether in 3D or 2D form. 3D captures the three-dimensional features of the eye, while 2D offers less information. The ear recognition system is composed of multiple phases: segmentation, pre-processing, feature extraction, matching, and decision-making.

The goal of segmentation is to locate the ear given an input. This input could be an image or a video. Detection of ear reduces the error rate as we are not bothered by futile features. Then, normalization must occur, where we perform enhancements to improve the fidelity (*e.g.* enhancing illumination). After normalization, features must be extracted, and a mathematical model is generated that represents salient features of the ear. The template will be stored within the system, and each time we have to perform matching, we use it for comparison and make a decision based on this template. Biometric systems operate in two modes: verification and identification. For verification, there is an identity (a claim by the subject) that must be verified. The claim can be verified by capturing the features and comparing them with the template already stored in the system. In identification, there is no claim to be validated; instead, the identity must be discovered.

The studies have utilized various databases that are available to researchers, to name a few: WVU (West Virginia University), USTB (University of Science and Technology Beijing), UND (University of Notre Dame), and FERET. For segmentation, numerous techniques have been used, including: Haar-based (Viola Jones) [35], hybrid [36], shape-based [37] and other approaches [38]. Viola-Jones uses the well-known classifier called AdaBoost to detect regions of interest (ear), and it does not suffer from low-quality input. But unfortunately, the training time could take several weeks.

Various algorithms for recognizing ears have been proposed over the years [38]. To provide two examples of the algorithms that were used, we can first refer to Gabor filters. This algorithm was used in a study by Yaqubi et al. (2008). Another algorithm is Force Field (FF), used by Hurley et al. (2005), which can be seen in Figure4.

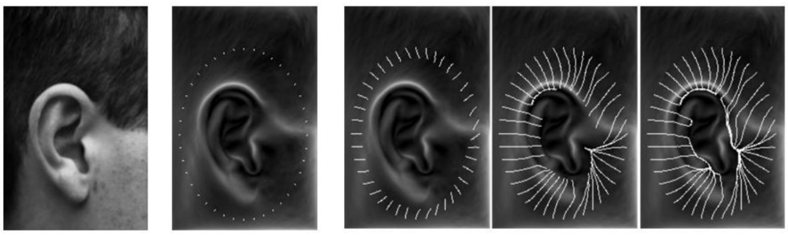

Figure 4: Image is treated as an array of Gaussian attractors. Image credit: Li Yuan et al. (2006)

As we mentioned earlier, several factors could affect the precision rate. To resolve this, we have various methods and techniques available. For instance, to remove strands of hair partially or fully covering ears, we may generate a thermogram image (see Figure 5).

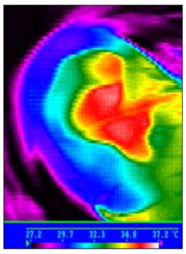

Figure 5: A thermogram of the image generated by Burge and Burger (2000).

In a thermogram image, heat surfaces indicate which regions are lower in temperature. Lower temperature regions represent the subject's hair, while higher temperature regions represent the subject's ear. To remove occlusions, we simply remove low-temperature regions.

6.1 Age estimation using ear

In the context of age estimation, ears have not been studied in depth. Most studies have focused on gender classification [39, 40, 41]. The first study used geometric and appearance-based features of the ears to estimate age [42].

The appearance-based method uses a deep convolutional neural network for representation and classification. The models used in this study are AlexNet, VGG-16, GoogLeNet, and SqueezeNet. The final result for age estimation showed 52% accuracy, but the result for gender classification was much higher. Geometric features instead are based on eight landmarks of the ear, from which 16 measurements were calculated to generate the feature vector (see Figure 6). The classifiers used include: logistic regression (different classes are linearly discriminated), random forests (ensemble machine learning algorithm based on sub-trees), and support vector machines (which aim to find the decision boundary between different classes and neural networks). The neural network used had three hidden layers and the number was chosen empirically.

From the total number of features (sixteen), six were considered to be more important. This was decided by the random forest model. Also, to ensure zero mean and unit variance, different geometric features were normalized.

As it was mentioned previously, multiple architectures were utilized for appearance-based approaches. AlexNet, for example, is not as deep as GoogLeNet.But this architecture has fewer parameters compared to AlexNet, whereas more layers exist (22 layers). Furthermore, VGG-16 was used instead of VGG-19 in VGG architectures.

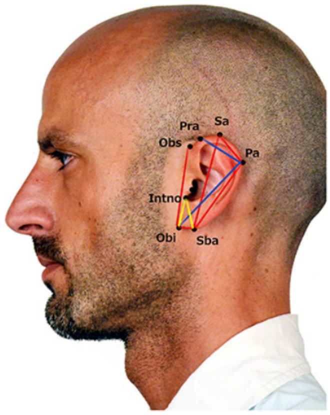

Figure 6: Eight landmark features have been identified, and the distance between them has been calculated. . Image credit: Yaman et al. (2018b)

The 80:20 rule was used to train the dataset for training and test data. The total number of profile images (338) was processed by OpenCV. Due to the small number of subjects, data augmentation was used. More specifically, techniques like flipping were applied to increase the number of samples. For the purpose of generalization, Gaussian blurring and dropout were applied.

It is clear from Figure 7 that appearance-based learning performs better compared to geometric approaches. GoogLeNet had the best performance among the appearance-based methods. The winner of the geometric-based approaches was three hidden layers of NN. A simple solution to further improve the result would be to use a combination of both. In the study, researchers noted that a possible motive for having less accuracy for age estimation could be due to the small number of samples used.

Approach	Test Result
Logistic Regression	43% (Geometric)
Random Forest	34% (Geometric)
SVM	39% (Geometric)
3 hidden layers NN	43% (Geometric)
AlexNet	45% (Appearance)
VGG-16	39% (Appearance)
GoogLeNet	52% (Appearance)
SqueezeNet	39% (Appearance)

Figure 7: The results for different classifiers and the corresponding accuracy for each approach. Image credit: Yaman et al. (2018)

Subsequently, another study was published by them to discover whether increasing the number of samples had any significant effect on the accuracy or not [43]. In this work, a multimodal system was implemented where features with no relevance (*e.g.* hair, background area) were removed to improve accuracy.

FERET, a public dataset, and profile and ear images were used as input for the dataset. By combining them, accuracy was improved by 9%. Still, compared to gender classification, age estimation had lower accuracy.

6.2 Methodology summary

The ear has the potential to be used in combination with the face, especially when the face could be occluded by various factors. Unfortunately, the number of studies related to ear is quite limited, so it is difficult to determine the effectiveness of ear as a soft biometric at the moment. Furthermore, it should be mentioned that, until now, most studies have been done in controlled and closed environments. Factors like hair or hijab could also occlude the ear, whether partially or completely, resulting in less accurate results.

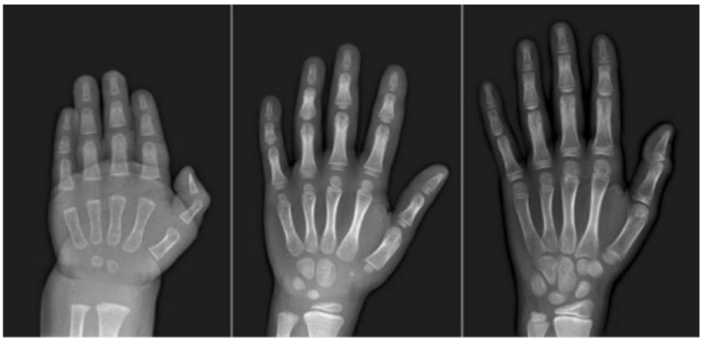

Figure 8: Hand bones undergo changes over time. Images are captured using X-ray scanners and compared to estimate the age. Image credit: Lee et al. (2020)

7 Hand Bones

Age estimation is not done entirely for the purpose of authentication. According to UNICEF, a quarter of children under the age of 5 are not officially registered [44]. Due to an increased rate of migratory activity in the 21st century, individuals not possessing an official birth registration date or a valid identity document may be faced with potential difficulties when entering a new country for the first time. Moreover, a refugee child may lack certain privileges due to an incorrect age estimation. To provide an example, in some judicial systems (*e.g.* France), a child under the age of 10 will not be incarcerated for any reason [45].

Simultaneously, children without birth registration may not be able to take advantage of numerous benefits (*e.g.* education, health services). Also, in some countries, a criminals punishment for an offense against a younger child compared to an older child would be more severe. Moreover, jihadist and militia groups have used children as soldiers and continue to do so [46]. As a result, there are valid reasons to estimate a child's age in order to reduce crime and ensure a safer life not only for the child but also for society.

One way to estimate age is to analyze x-ray images of bones. For this task, hands are preferred over other body parts due to their high bone count and low level of radiation. Traditionally, two atlases (GP and TW) have been used, but both techniques are manual and costly. They also suffer from the fact that different specialists could have different estimations. Bones also vary between genders and races, and many additional factors like nutrition could have an impact on the density of the bone.

The idea is to observe morphological changes in hand bones at different ages and compare them with the atlases. One possible solution to reduce human labor and cost would be to use machine learning models to estimate the age. In the subsequent section, age estimation from hand bones will be covered in more detail.

7.1 Age estimation using hand bones

As mentioned previously, the age of an individual could be estimated by assessing the bones. Using this method, there is always a margin for error, especially for older children. Nonetheless, for a better age estimation, it may be necessary to perform dental and physical examination to ensure estimation was more accurate [45].

In traditional bone assessment methods, x-rays of the hands and wrists are captured, and two well-known techniques are used to estimate the age. Greulich and Pyle (GP) atlas, published in 1959, is one of them. It was based on a study done in 1939, where the goal was to assess skeletal maturity [47]. Due to its speed and simplicity, it has become one of the most widely used references, but it suffers from the assumption that skeletal maturity is uniform in healthy children. This approach has significant intra- and inter-observer variability, indicating its subjectivity [48].

Crawley (2007) states that it does not take into consideration various factors that could affect skeletal maturity, including socio-economic discrepancies, ethnicities, and nutrition. Many researchers believe this technique should be used with caution, especially for individuals with different nationalities [49].

An alternative technique used to assess bone age is the Tanner-Whitehouse (TW) method. There are different variations (TW2, TW3), but all are based on radiographs taken from the left hand and wrist. Chronologically, TW3 is more recent, while TW2 was developed using radiographs collected from average-class children in the UK between 1950 and 1960. It cannot be applied to subjects from different racial and ethnic backgrounds. Moreover, it is believed to not provide useful information for any subject older than 16 [45]. Pinchi et al. (2014) explored different bone assessment methods to discover their strengths and weaknesses and recommended TW3 be used instead of GP for females.

Other than the aforementioned atlases, alternative techniques for age estimation have been proposed over the years. X-ray of the shoulder area is proposed to be the best indicator for chronological age. Furthermore, Risser's test is considered a reliable method for subjects between 12 and 16. Although the level of radiation emanated is not considered to be harmful to children, some believe that exposing them to radiation solely for the estimation of their age is unethical [45].

A more recent approach has tried to propose a quantitative approach instead of relying on the subjective opinion of the specialist [50]. Remy et al. (2021) used 1003 hand radiographs of subjects younger than 21. To consider variability in this study, subjects from the south and north of France were included. The training set and test set consisted of 768 samples and 235 samples, respectively.

This method follows a two-step procedure. In the initial stage, the most likely age must be determined. The subject falls into one of the following categories: [1-12], [13-21]. Then, age must be estimated using one of the formulas written below:

	MAD (female)	CCC
Caffenet	12.272	0.904
GoogleNet	8.890	0.941
Resnet	15.366	0.855

Table 5: This table indicates results for different architectures used in the study by Lee et al. (2020).

$$Age_{1-12} = 0.353 * LG \cdot MC1 - 0.337 * DW \cdot PH5 + 0.392 * PW \cdot MC4 - 5.491 \quad (2)$$

$$Age_{13-21} = 0.555 * DW \cdot MC2 + 0.685 * DW \cdot PH5 - 0.284 * PW \cdot MC2 + 6.979 \quad (3)$$

Note that if the subject does not belong to one of the groups above, the general formula is used:

$$Age_{1-21} = 0.853 * DW.MC4 + 1.231 * PW.PH5 - 12.063 \quad (4)$$

Gradually, the precision will degrade as the children continue to grow older. This becomes evident around age 13-15. Nonetheless, the overall rate of correct classification was shown to be 89.8%. Researchers suggested using gender-dependent estimation formulas to further improve the results.

A further study by Lee et al. (2020) used different deep learning architectures to estimate the age. Using deep learning strategies would reduce labor costs and increase the speed of estimation. This study also included subjects in their infancy and very early childhood. Instead of incorporating the entire image of the hand, the palm regions proved to be sufficient for age estimation. To train the model, for each gender, 1,000 images were used, and the rest was used as training data. The dataset used in this study came from the RSNA challenge.

Age estimation is considered by the authors to be a regression problem since it is a continuous variable. The following formula was used to estimate the age:

$$\hat{a} = M(I) \quad (5)$$

I represents the input image in this equation, M is a regression model, and $\hat{a}$ represents the estimated age. To evaluate the performance, mean absolute difference and concordance correlation coefficient measures were used.

As it can be seen from Table 5, GoogLeNet had the lowest error rate. Further research revealed that gender had a significant impact on performance (Table 6), with gender-aware models outperforming others.

	MAD (male, female)
Caffenet, gender-agnostic	16.933
Caffenet, gender-aware	12.649 (12.272, 13.026)
GoogleNet, gender-agnostic	12.642
GoogleNet, gender-aware	10.380 (10.186, 10.573)

Table 6: Gender-aware models have better performance compared to gender-agnostic models. Image credit: Lee et al. (2020)

In this study, the error level was 9.35 months. It should be mentioned that researchers assumed that chronological age would be available, making this work useful for cases of growth problem diagnosis.

7.2 Methodology summary

Overall, estimating age based on bones has proven to be a challenging task since it is dependent on many factors like nutrition, gender, race, etc. Bone does not seem to be a reliable approach for age estimation by itself. Using it in combination with different features may be more acceptable. It should be noted that age assessment using bones is best done to classify a subject rather than provide an exact age estimation. Future studies should try to consider variability (*e.g.* different races) in their work to have a more realistic estimation.

8 Gait

Gait simply refers to the patterns of movement in an individual. It fits into a category of biometrics known as behavioral biometrics. This category is concerned with analyzing behaviors and physical patterns. Gait has numerous positive traits. One such positive trait is that it could be used from a distance, and no cooperation from the subject is needed. For the purpose of surveillance, where we do not wish to notify the subject, gait is a suitable choice to be used. It could also be used from different angles, unlike facial features, where the individual must be facing the camera.

Gait recognition systems are typically composed of multiple phases. In the first phase, gait data is acquired. This can be done with video cameras. To ease the processing, a silhouette image is created. In the third phase, feature extraction occurs. A more detailed view can be seen in Figure 9.

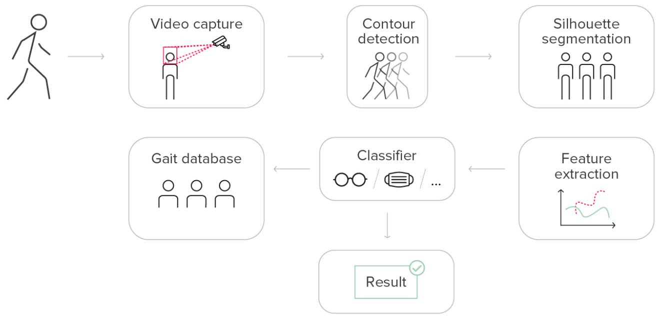

Figure 9: Image credit: RecFaces. (2022, October 13). Gait Recognition System: Deep Dive into This Future Tech. https://recfaces.com/articles/what-is-gait-recognition

8.1 Age estimation using gait

So far, most of the studies using gait have focused on hard biometrics rather than soft biometrics [51, 52]. Gait offers great possibilities in terms of being used as a soft biometric [53, 54, 55, 56]. For instance, a robot assistant could recommend personalized services which are based on age or gender. Moreover, since no direct contact is required, it is a suitable choice for dangerous environments (*e.g.* prison).

The gait cycle has been used in studies to estimate age [57, 58, 59]. Unfortunately, most of them rely on a full gait cycle. For real-time systems, this raises issues as we do not have the full cycle and must wait until the whole cycle is captured. This unnecessary latency will have a negative impact on system performance.

A real-time system may be useful when we have a robot that is supposed to offer personalized suggestions to subjects based on their age group.

To make gait features suitable for real-time online applications, latency must be reduced, and this can be done using a single image. Although this is not a trivial task, since body shape and temporal pose alterations could indicate gender or age [60]. For instance, females have a smaller degree of arm swings compared to male individuals, and children have a larger head-to-body ratio. Without the full cycle, learning whether an individual is male or female becomes a challenging task. Here, instead of having access to the complete cycle, we have a single image, and we must determine age or gender based on that.

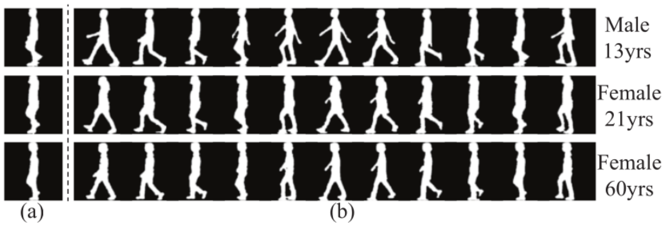

Figure 10: This figure demonstrates the difference in amount of information which is available in a single image (a) in comparison to a full cycle (b). Image credit: Xu et al. (2021)

As it can be seen in Figure 10, motion information (*e.g.* arm swing, stride) cannot be observed from a single image. One possible solution would be to build the full gait cycle from a single shot retrieved from the subject, as proposed by Xu et al. (2020). This snapshot contains pre-and post-motion information. To reconstruct the cycle, a silhouette sequence is generated. To learn the distinctive features, a sequence is given as the input to the network. Then, distinctive features will be used to determine the age of the subject.

The model by Xu et al. (2020) can handle different angles. The OU-MVLP dataset, which contains 10,307 samples ranging in age from 2 to 87, was used for training and testing the model. 5,153 subjects from this dataset were used for training purposes, and 5,154 subjects were used for testing purposes. Additionally, two systems (a stand-alone system and a client-server system) were implemented. This was done to understand whether this model would work for a real-time system or not. The stand-alone system requires a GPU (Graphics Processing Unit), but the client-server system could be used even via tablets. For a client-server system, the client would send the captured image to the server via the web API. The server, on the other end, will process the image and return estimations to the client. One factor that could have a negative effect would be communication speed.

It was observed that, compared to regression and classical approaches, deep learning approaches like CNN performed better [61]. Moreover, to reduce errors, a label distribution was used instead of relying on single age values. Xu et al. (2020) extracted the silhouette image from an image that was captured by a normal camera. The input will undergo feature extraction and be normalized.

In the training phase, a pair of silhouette images from different phases of the same subject are used. PA-GCR was used to reconstruct the complete cycle. PA-GCR consists of four components: an encoder, a phase estimator, a feature transformer, and a decoder.

A full gait cycle will be generated in the output where different phases have been specified. The full cycle will be given to GaitSet, which is a sequence-based gait recognition network. As it can be seen from Figure 11, discriminative features are learned. This occurs by using a fully connected layer to estimate the age and gender.

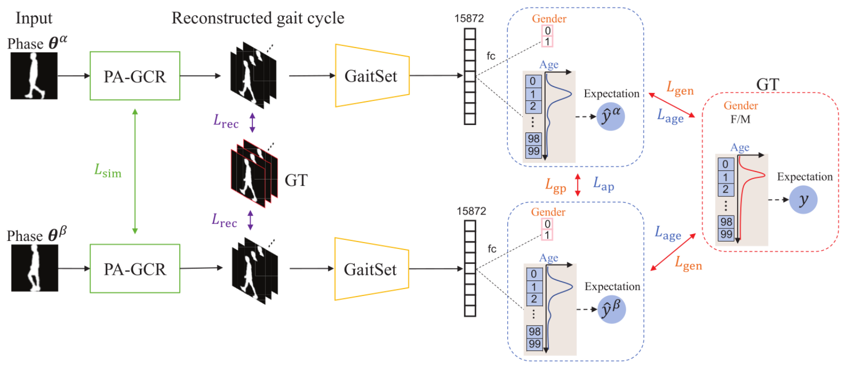

Figure 11: Different stages of the model proposed by Xu et al. (2021)

In their work, Jensen-Shannon (JS) divergence was adopted due to its symmetry, making it a suitable choice for handling pairs of images. Assuming P and Q to be two probability distributions, and KL to be Kullback Leibler divergence among the two, JS divergence is defined in the following manner:

$$JS(P||Q) = \frac{1}{2}KL(P|||P + \frac{Q}{2}) + \frac{1}{2}KL(Q|||P + \frac{Q}{2}) \tag{6}$$

For the purpose of evaluation, mean absolute error (MAE) and cumulative score (CS) are used. MAE is defined in the following manner where N_s indicates the number of subjects in the test sample, and $\hat{y}_i$ and y_i show the estimated age and ground truth age of the i-th test sample, respectively.

$$MAE = \frac{1}{N_s}\Sigma_{i=1}^{N_s}\|\hat{y}_i - y_i\| \tag{7}$$

The CS score for y-year is defined below, where N_s represents the number of samples where the absolute error estimation is within y years.

$$CS(y) = \frac{N_s(y)}{N_s} \tag{8}$$

An additional evaluation metric used is mean cross-entropy (MCE):

$$MCE = \frac{1}{N_s} \Sigma_{i=1}^{N_s} \Sigma_{y=1}^{100} a_i(y) \log(\hat{a}_i(y)) \tag{9}$$

$a_i(y)$ and $\hat{a}_i(y)$ represent the i-th test sample's ground truth and estimated probability of age label y, respectively.

Age estimation is a challenging task, especially if we only have a single image. This explains why we need to reconstruct the full cycle to have additional information. Otherwise, we would be relying on information from the shape. Age estimation for children and teenagers is less challenging due to their body shape. For instance, children have a larger head-to-body ratio, which could be observed even from a single frame.

Method	MAE	CS(1)	CS(5)	CS(10)	MCE	CCR
GEINet	9.11	15.70	46.73	65.82	-13.60	N/A
GaitSet	9.02	**16.39**	47.59	66.01	-5.04	92.72
Ours	**8.39**	15.84	**48.00**	**68.40**	**-4.27**	**94.27**

Table 3: Evaluation metrics for different methods The best results have been shown in bold. Image credit: Xu et al. (2021)

Needless to say, performance from side angles reveals more information compared to front and back angles. The reason is that both body shape and motion patterns are more visible from side angles (see Table 4).

Metrics	0°	15°	30°	45°	60°	75°	90°	180°	195°	210°	225°	240°	255°	270°	Mean
MAE	8.91	8.83	8.27	8.48	8.35	8.08	7.83	9.24	8.86	8.61	8.48	8.11	7.88	7.74	8.39
MCE	-4.28	-4.36	-4.29	-4.34	-4.27	-4.21	-4.19	-4.42	-4.34	-4.30	-4.34	-4.19	-4.16	-4.16	-4.27
CCR	91.5	93.0	94.5	94.6	95.0	94.9	95.7	93.0	93.6	94.9	94.8	94.4	94.6	94.9	94.3

Table 4: Side angles reveal more information regarding body shape and motion patterns; therefore, performance is higher from side angles. Image credit: Xu et al. (2021)

8.2 Methodology summary

Gait is a behavioral feature that could be used to estimate valuable information such as gender. Gait has been studied extensively over the years, and it has also been studied for both gender and age classification. Many studies have been done using both traditional techniques like regression and deep learning techniques like CNN.

Studies indicate great potential for gait to be used as a soft biometric. One of the works mentioned before indicated that gait analysis could even be implemented as a real-time system. Although gait gives us the capability to learn additional information about subjects from a distance, it should be used with caution as inappropriate use could threaten the privacy of the individuals. So far, studies demonstrate its effectiveness to be used as a soft biometric.

9 Electrocardiogram (ECG)

Electrocardiogram or more commonly known as ECG, was first invented in 1901. It is mainly used to measure heart activity and detect any abnormalities. Since numerous wearable devices now support ECG capability, collecting ECG has become easy and cost-effective. It is already known and accepted that during an individuals lifespan, the 12-lead ECG may undergo changes. Therefore, males and females may demonstrate different outcomes [62]. Moreover, ECG could also be used as a biomarker, revealing useful information in identifying potential health risks and diseases. If an individual is chronologically 30 years old, but marked as an older person, valuable insight could be inferred.

It should be noted that males and females age differently. Compared to their male counterparts, female hearts age more slowly. This remains true until menopause, after which aging increases, reaching the same rate as male individuals. A large discrepancy between biological age and estimated age could be used as a measurement of health [63].

9.1 Age estimation using ECG

ECG has previously been studied for soft biometrics, and various approaches have been explored by researchers. Ball et al. (2014) used the Bayesian model; Starc et al. (2012) used multilinear regression to estimate cardiac ages. Attia et al. (2019) used convolutional neural networks (CNN) to predict age and gender using 12-lead ECG signals. Selecting neural networks instead of high-level statistical techniques has both benefits and impediments. For instance, neural networks fail in explainability.

Lack of explainability remains a big challenge. A patient may wish to know why a certain operation is being performed, and if we do not understand how the model works, we cannot provide such information to the patient. At the same time, neural networks require a considerable amount of data, introducing further difficulties like data verification.

Despite all the shortcomings, we should not discard neural networks entirely. Neural networks have positive traits, which makes them a suitable option in many scenarios. To provide an example, neural networks are not hindered by the biases since they learn associations through repeated exposure; therefore, results could be more accurate.

The age predicted by Attia et al. (2019) was within seven years of the reported biological age. This model was built using the Keras framework. Different networks for gender and age were used in this study, as the output for sex is binary (male or female) and for age is a single output (continuous number). 12-lead ECG signals were used as inputs. To optimize the weights, Adam was used. The architecture for this network can be seen in Figure 12

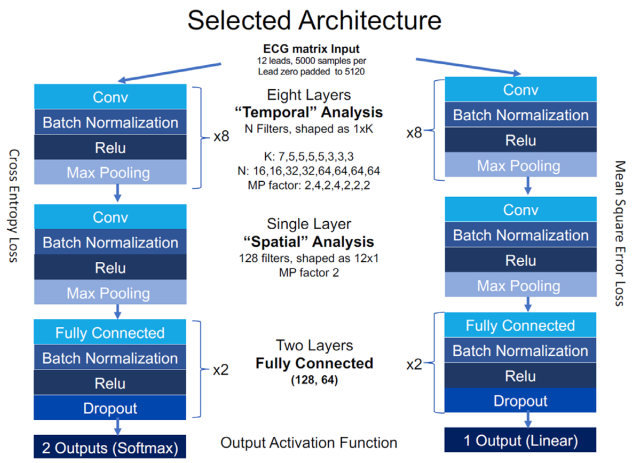

Figure 12: The architecture is made up of convolutional, batch normalization, and max pooling stacks. Image credit: Attia et al. (2019)

As already mentioned, the output for age is considered to be a continuous value. For evaluation, mean absolute error (MAE) and R squared were used. In total, 774,783 patients were used, and the final result indicated that the mean absolute error was equal to 6.9 ± 5.6 years, and R squared was equal to 0.7.

The accuracy for younger subjects compared to older subjects (> 55 years old) was higher. Lower accuracy for older subjects may be explained by variations in hormonal levels. The variations could create difficulties for the network and affect its performance.

It should be mentioned that a number of medical episodes (*e.g.* cardiac transplantation) may cause major deviations from the reported chronological age. However, after stabilization, CNN can approximate predict reported chronological age (see Figure 13).

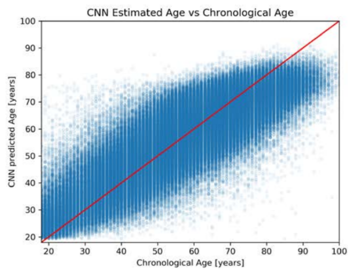

Figure 13: Blue represents estimated age, and red indicates chronological age. Image credit: Attia et al. (2019)

The study by Attia et al. (2019) suffers from a lack of explainability, and the authors cannot explain which factors determined the outcome. Different solutions have been proposed to improve explainability [64]. With possible advancements in upcoming years, using CNN may become a more sensible choice.

9.2 Methodology summary

At the moment, there is no superior solution based on ECG. The models proposed are primarily used in medical settings where the patients chronological age is reported beforehand and the estimated age is used as a biomarker. For commercial applications, ECG does not seem promising at the moment, especially compared to other features that provide better results. But with possible advancements in AI approaches, ECG could possibly gain more attention.

10 Iris

Iris refers to the area surrounding the pupil that is colored. This area is protected from external factors that could cause damage to it, by the cornea. It is considered to be highly accurate, hence the Department of Defense (DoD) of the USA has adopted it for both soldiers and detainees in Afghanistan and Iraq.

It is believed by some to remain unchanged throughout an individuals life. This claim was initially made by Daughman, who stated the iris remains immutable throughout life, although its immutability has not been scientifically proven [65]. Iris is also more hygienic compared to fingerprints and handprints, where physical contact is required. Privacy-conscious individuals consider iris scanners to be a threat to their privacy since samples could be collected from a distance without the subjects knowledge [66]. Furthermore, it has been demonstrated that recreation of images from digital codes of the iris is possible [67].

Iris recognition systems are composed of multiple stages. In the first stage, an image is acquired. Following image acquisition, segmentation occurs where the iris is detected. Parts like eyelashes or eyelids that could affect accuracy must be eliminated. The image will be divided into blocks, with each block becoming a feature value. Feature values represent the image and are used to make decisions.

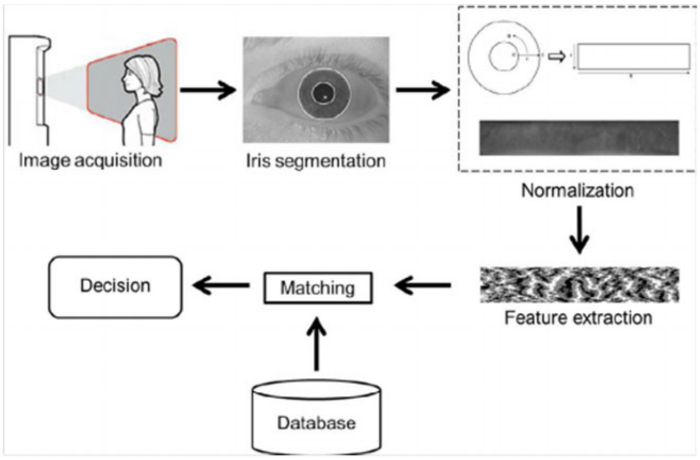

Figure 14: Iris recognition systems are composed of multiple stages. The acquired image will be processed and used for comparison. Image credit: Thakkar, D. (2018, August 15). An Overview of Biometric Iris Recognition Technology and Its Application Areas. Bayometric. https://www.bayometric.com/biometric-iris-recognition-application/

10.1 Age estimation using iris

Unlike facial features, studies on the iris for age estimation are quite limited. Using the iris for age estimation would be useful, since it can be collected even with normal cameras. For instance, a website could try to estimate users age to decide whether or not an age-restricted item should be shown. Gradually, pupil dilation decreases with age; hence, the performance of iris recognition declines as well [68].

Sgroi et al. (2013) divided the subjects into two groups: young and elderly. The total number of subjects consisted of 98 individuals, and the number of images used was 600, which were acquired using LG IrisAccess 4000. To consider different races and eye colors, five races and six eye colors were included in the study. For the model, Random Forest with 300 trees was used. To determine the best classification rate, a parameter sweep was employed. The result for correct classification was 64.68%. The distribution for match and non-match scores can be seen in Figure 15.

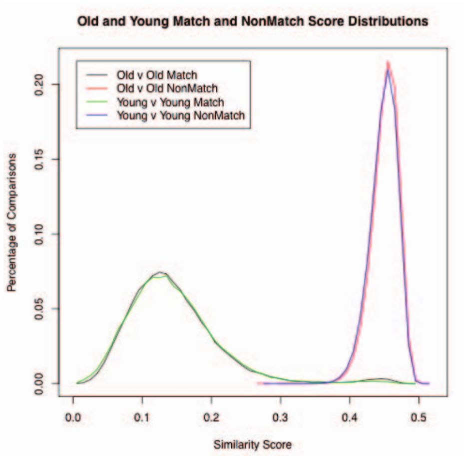

Figure 15: The score distribution for match and non-match for young and old subjects. Image credit: Sgroi et al. (2013)

To further test the results, two validation sets were used. The first validation set consisted of five new images from young subjects and five new images from older subjects (10 in total). The classification rate was equal to 71.1%. The second validation set used 10 new images of only young subjects. The classification rate was 64.5%. The authors reached an accuracy of 64% for age estimation.

Later work by Erbilek et al. divided subjects into three groups consisting of "young", "middle" and "old". The dataset used in this study had more subjects, that is 210 and their age ranged between 18-73.

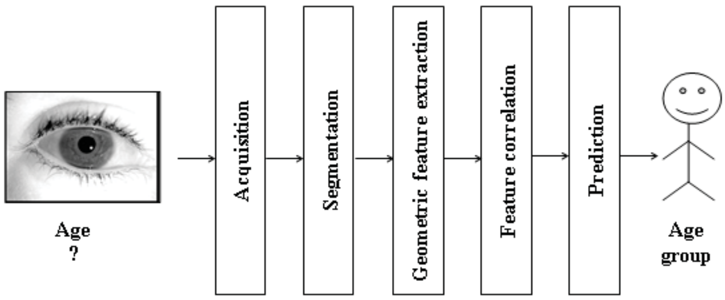

Figure 16: The proposed model consists of multiple stages where the acquired image is processed to extract distinguishing features. Image credit: Erbilek et al. (2013)

As it can be seen in Figure 16, geometric features must be extracted from the iris. In total, 12 features are extracted, and later a correlation test is applied where relevant features are selected. For correlation, Spearman's rank correlation is used.

For prediction, different classifiers were employed to assess their performance. Classifiers used included SVM, MLP, KNN, Jrip, Decision Tree, Fusion, and negotiation.

SVM aims to find the surface having more samples. To do this, the problem space is divided using hyperplanes. MLP has multiple layers, where intermediate layers are responsible for extracting features and passing them to the output layer, where classification output is generated. For KNN, a modified version of hamming distance was used to consider possible rotational inconsistencies, and the value for k is equal to 1. In the decision tree, the problem is simply divided into sub-problems.

To train the classifier, 72% of the subjects were used for training, and the rest were used for testing. The following table indicates the results obtained from different classifiers:

Classifier		Accuracy (%)
SVM		62.06
MLP		61.80
Jrip		62.50
KNN		52.41
Decision Tree (J48)		51.09
Fusion	Sum	64.11
	Vote	62.94
Negotiation	Game theory	72.65
	Sensitivity	75.09

Table 5: Results for different classifiers used. Image credit: Erbilek et al. (2013)

As it can be seen from Table 5, a multi-classifier approach had better results. The aforementioned studies have several flaws. For instance, both have neglected minor subjects (under 18).

In a more recent study by Rajput and Sable (2020), minor subjects were included, and only three statistical features from iris images were used. The age range was between 2.5 and 75. After image acquisition, the Hough transform was applied for localization, and for normalization, a rubber sheet model was used to convert the iris into a rectangular structure.

The features are extracted in three steps. In the first step, 2D multi-wavelet transform is applied. This transformation supports properties such as symmetry, orthogonality, and compact feature support. In the second stage, radon transform is applied to the features from the first stage. Equation (10) indicates this transformation.

$$R(r, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot \delta(r - x\cos(\theta) - y\sin(\theta)) dx \cdot dy \tag{10}$$

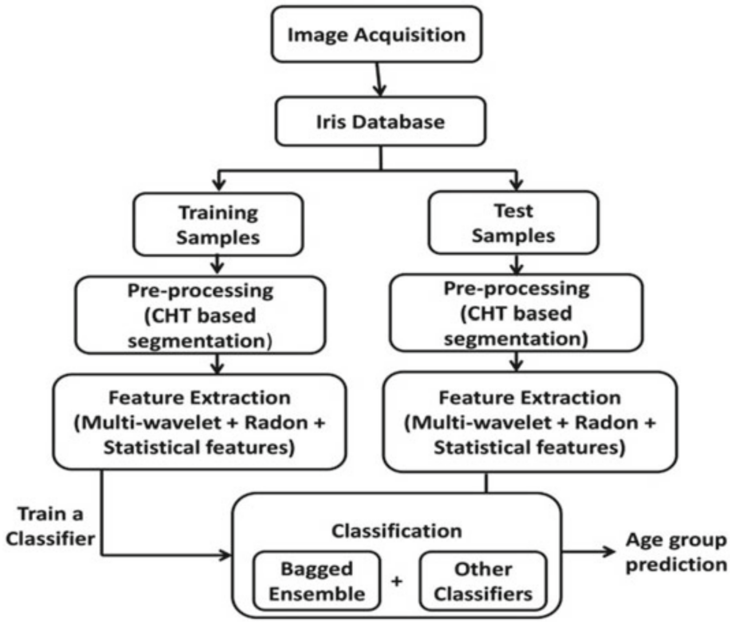

Figure 17: Rajput and Sable have proposed a new model where only three statistical features are extracted. Image credit: Rajput and Sable (2020b)

In this transformation, $R(r, \theta)$ represents the radon transform, with input image $f(x, y)$ and Dirac delta function $\delta(r - x\cos(\theta) - y\sin(\theta)$. The large feature matrix computed by the equation must be reduced. To do so, iris images are transformed and represented by three features. Equation (11) indicates the transformation used.

$$F = (M_d, S, S^2) \tag{11}$$

Rajput and Sable employed different classifiers and observed their performance. For evaluation, various metrics were used. Accuracy states the total number of correct predictions compared to the total number of samples. Further metrics used are F1, recall, and precision, which you can see below.

$$F1 = 2 * (Precision * Recall)/(Precision + Recall) \tag{12}$$

$$Precision = (TruePositives)/((TruePositives) + (FalsePositives)) \tag{13}$$

$$Recall = (TruePositives)/((TruePositives) + (FalseNegatives)) \tag{14}$$

It is evident from Figure 18 that the bagged ensemble classifier achieved the best performance out of all classifiers. The overall accuracy was reported to be 83.7%.

Classifier	Training accuracy (%)	Testing accuracy (%)	Precision	Recall	F1 score
Cosine KNN	75.4	65.16	0.6633	0.4587	0.5418
Fine Gaussian SVM	74.9	59.55	0.59	0.436	0.5
Decision tree (medium)	74.9	69.66	0.7953	0.537	0.641
Ensemble (bagged)	**74.6**	**83.7**	**0.875**	**0.7863**	**0.8284**
Linear discriminant	73.4	55.6	0.3726	0.487	0.422

Figure 18: Different classifiers and the corresponding results. Image credit: Rajput and Sable (2020b)

10.2 Methodology summary

Iris is a periocular feature that has become quite popular. Compared to other features like hand bones and ECG, there is more literature covering the iris. The studies indicate the iris is an effective feature for age estimation. Evidently, researchers should aim to consider variability by including different genders and eye colors. The number of subjects used in the study is deemed important as one of the studies used a small number of subjects.

11 Other biometrics

In addition to the aforementioned papers, other biometrics have been employed by the researchers to predict age. The fingerprint is one of the most well-known and traditional features used for identification. A fingerprint is simply a representation of the epidermis consisting of ridges and valleys, and it is assumed that it remains unchanged during the lifetime.

To identify an individual, systems refer to existing patterns in fingerprints. Whorl, arch, and loop are examples of patterns. Each pattern could have various sub-types; for instance, arches could be plain or tented. Moreover, biometric systems could be based on various technologies (*e.g.* optical, capacitance). A wide array of elements such as skin condition, pressure or sensor noise could affect performance of a fingerprint scanner.

To analyze the arrangement of patterns known as minutiae, algorithms are used. Hence, when a subject has to be verified, after acquiring the image from his fingerprint, minutiae must be compared with the template stored within the system. Further studies have indicated the possibility of fingerprints being used for soft-biometric applications. Falohun et al. (2016) developed a system to estimate age and gender, adopting DWT+PCA for age estimation instead of Back Propagation neural networks. The authors believe that previous research demonstrated that Back Propagation neural network and ANN are not effective.

GENDER ACCURACY	AGE ACCURACY
FEMALES : 80.00% MALES : 72.86%	82.14%

Figure 19: The result of the system developed by Falohun et al. (2016)

	Age Classification			
Sl. No	***Age Groups***	***Total fingerprints-60***	***Accuracy***	***Over all Accuracy***
1	6,7	34	56%	60%
2	8-12	32	53%	
3	13-15	32	53%	
4	16-19	31	52%	
5	20-30	34	56%	
6	30-50	33	55%	
7	Above 50	36	60%	

Figure 20: T Abraham and M (2016) reported age classification results.

A total of 140 subjects were used in this study, and the final result can be seen in Figure 19. The fingerprint of subjects with a larger body structure compared to other individuals in the same age group may be different, thereby affecting the precision.

In another study by T Abraham and M (2016), a model based on wavelet transform and SVD was proposed. SVD is an information-oriented method based on principal component analysis (PCA). The overall accuracy reported by the authors indicated 60% success rate, where individuals above the age of 50 had the maximum success rate while the worst performance was marked for subjects in the range of 16–19, as shown in Figure 20. To further improve the results, the authors mentioned that increasing the number of subjects would bring better results. Additionally, using digital images instead of the optical images originally used in the study may have produce better results.

In a more recent work, Jayakala and Sudha (2022) employed ResNet50 to estimate the age. The authors used the house database, where the subjects ranged from 1 to 60. To train and evaluate the classifier, 1000 images were used for training and 200 images were used for testing. The researchers achieved 93% accuracy.

An alternative approach that has gained more attention in recent years is DNA methylation (DNAm), which consists of adding a methyl group ($-CH_3$) to the $5'$ carbon of cytosines positioned next to guanines (CpG nucleotides). It is a tissue-specific technique, and analyzing the literature shows a wide variety of tissues have been employed by researchers, but most studies have focused on blood. Additional sources that could be used include buccal swabs, saliva, semen, skeleton, and others [69]. One possible application would be to estimate the age of a criminal based on evidence left at the crime scene in order to narrow the list of suspects.

DNA has the ability to be used in both living and dead people. It remains effective even when it is applied to different races and populations [70]. It should be mentioned that DNA methylation could still be affected by multiple factors, including: lifestyle, environment, ancestry, and diseases. In the studies, various DNA age-related markers have been employed, such as ELOVL2, TRIM59, EDARADD, PDEC4, NPTX2, ITGA2B, and KCNAB3 [71]. Moreover, to consider the influence of races, we could retrain the model or try to find markers that are consistent among different origins [72].

Numerous studies have been done to estimate age based on DNA markers. To provide an example, Bocklandt et al. (2011) used a regression model with an average accuracy of 5.2 years. In this study, researchers relied on DNA methylation on two sites. Ambroa-Conde et al. (2022) used logistic regression and had 88.59% correct classification rate for the training set and 83.69% correct classification rate for the test set. Individuals ranged from 21 to 86 years old, and a total of 184 samples were used.

In spite of the fact that DNAm offers numerous advantages, it is still difficult to collect when compared to other features. Additionally, collecting tissues from a distance would not be feasible. To further improve future studies, different factors that could affect DNA methylation must be addressed to have a better estimation.

12 Discussion

In this paper, various features were selected and discussed. Initially, a brief introduction of how the underlying system for verification and identification operates was given, and then papers related to age estimation were discussed. Some features had inherent qualities that allowed them to be used even from a distance (*e.g.* iris), while others required close contact. Hence, some features are considered to be impractical for challenging environments. Fingerprints, while widely used and common are susceptible to pressure, skin quality, etc.

Researchers have employed traditional machine learning algorithms and considerably good results were achieved, while more recent work has used deep learning algorithms. Unfortunately, deep learning algorithms inherently lack explainability, which poses a problem, particularly in medical settings where patients may be curious about why a decision was made. But, it is worth mentioning that most medical papers already assume that chronological age is known and estimated age is used simply to evaluate health status of an individual.

ECG, was one such feature that could reveal valuable information about the patients health. While deep learning has produced positive results in some medical settings, it may be replaced by traditional approaches due to its lack of explainability. Also, with possible advancements in the upcoming years, existing issues may be solved, making deep learning a more practical choice.

Furthermore, DNA methylation was mentioned, a powerful feature that could be used in both living and dead subjects. Among its numerous positive qualities, studies indicate its effectiveness between different populations around the globe. A less common feature, especially for commercial purposes, was hand bones. While two atlases already exist, both suffer from pitfalls. Recent papers demonstrate that age estimation from hand bones is extremely challenging due to many variables (*e.g.* diet).

Currently, no superior classifier for age estimation exists, and researchers have employed various models and evaluated them. One solution commonly used is to fuse multiple classifiers or combine features to get more accurate results. To conclude this section, a table containing information related to the studies mentioned in this paper can be found on the next page.

Age estimation			
Author(s)	Feature	Method	Evaluation
Duan et al. (2018)	Facial	CNN2ELM	66.49% (Accuracy)
Hasan and Mahdi (2021)	Facial	BPNN	93.22% (Accuracy)
Pei et al. (2020)	Facial	SIAM	4.74 (MAE)
Xie and Pun (2020)	Facial	DOEL3groups	5.69204 (MAE)
Ali et al. (2015)	Facial	Naïve Bayesian	72.48% (Precision)
Wan et al. (2018)	Facial	GPR	2.93 (MAE)
Liu et al. (2020)	Facial	CR-MT	6.04 (MAE)
Cascone et al. (2020)	Pupil	MLP	0.8277 (Accuracy)
Yaman et al. (2018)	Ear	GoogLeNet	52% (Accuracy)
Lee et al. (2020)	Hand Bones	GoogLeNet	0.941 (CCC)
Sakata et al. (2019)	Gait	CNNs	5.60 (MAE)
Sakata et al. (2020)	Gait	GEINet (modified)	5.43 (MAE)
Lu and Tan (2010)	Gait	ML-KNN	3.86 (MAE)
Makihara et al. (2011)	Gait	GPR	8.2 (MAE)
Xu et al. (2021)	Gait	CNN	8.39 (MAE)
Attia et al. (2019)	ECG	CNN	6.9 ± 5.6 (MAE)
Sgroi et al. (2013)	Iris	Random Forest	64.68 % (Accuracy)
Erbilek etal. (2013)	Iris	SVM	62.06% (Accuracy)
Rajput and Sable (2020)	Iris	Ensemble (bagged)	0.875 (Precision)
Falohun et al. (2016)	Fingerprint	DWT+PCA	82.14% (Accuracy)
Abraham and M (2016)	Fingerprint	SVD	60% (Accuracy)
Jayakala and Sudha (2022)	Fingerprint	ResNet50	93% (Accuracy)
Bocklandt et al. (2011)	DNAm	Regression	5.2 years (Accuracy)
Ambroa-Conde et al. (2022)	DNAm	Logistic Regression	3.55 (MAE)

13 Conclusion

The aim of this work was to introduce soft biometrics and review the literature to observe the effectiveness of selected features for age estimation. Age estimation remains a challenging task as it depends on many factors. Some studies neglected certain age groups, open environments, and variability. Nonetheless, few works seemed promising. Gender has been studied more extensively compared to age, but the author believes that in the upcoming years, the number of studies on age estimation will likely increase.

To have a better estimation of age, subsequent features should try to simulate the real world as much as possible. This can be done by considering different genders, races, nationalities, and other factors that play a crucial role. It should be noted that soft biometrics should be used with caution, as many features do not require direct physical contact, which could threaten privacy. More studies are required to discover privacy-friendly solutions.

14 References

References

[1] Song, Houbing; Fink, Glenn A.; Jeschke, Sabina (2017). Security and Privacy in Cyber-Physical Systems: Foundations, Principles, and Applications. John Wiley & Sons. p. 189. ISBN 978-1119226048.

[2] Cole, Simon (2001). Suspect Identities: A history of fingerprinting and criminal identification. Cambridge, Massachusetts: Harvard University Press. pp. 60–61. ISBN 978-0674004559.

[3] Password-free smartphones are no longer the stuff of science fiction — they're everywhere. (2017, December 27). Business Insider Nederland. https://www.businessinsider.nl/smartphone-biometrics-are-no-longer-the-stuff-of-science-fiction-2017-12

[4] Columbus, L. (2020, March 8). Why Your Biometrics Are Your Best Password. Forbes. https://www.forbes.com/sites/louiscolumbus/2020/03/08/why-your-biometrics-are-your-best-password/

[5] Biometric data breach: Database exposes fingerprints and facial.. (n.d.). https://us.norton.com/blog/emerging-threats/biometric-data-breach-database-exposes-fingerprints-and-facial-recognition-data

[6] Belanger, A. (2022, December 27). Military device with biometric database of 2K people sold on eBay for $68. Ars Technica. https://arstechnica.com/tech-policy/2022/12/military-device-with-biometric-database-of-2k-people-sold-on-ebay-for-68/

[7] Agnihotri, N. (2022, July 23). What are different types of biometric sensors? Engineers Garage. https://www.engineersgarage.com/different-types-biometric-sensors-morphological-behavioral/

[8] Rashid, K. (n.d.). National Lab Develops Face and Body Recognition Technology at COD. The Courier. https://codcourier.org/national-lab-develops-face-and-body-recognition-technology-at-cod/

[9] David, C. I., Brogan, J., Barber, N., Aykac, D., Baird, S., Burchfield, N., Dukes, C., Duncan, A., Ferrell, R., Goddard, J., Jager, G., Larson, M., Murphy, B., Johnson, C., Shelley, I., Srinivas, N., Stockwell, B., Thompson, L., Yohe, M., . . . Bolme, D. S. (2022). Expanding Accurate Person Recognition to New Altitudes and Ranges: The BRIAR Dataset. Cornell University - ArXiv. https://doi.org/10.48550/arxiv.2211.01917

[10] Johnson, K. (2023, January 10). Iran Says Face Recognition Will ID Women Breaking Hijab Laws. WIRED. https://www.wired.com/story/iran-says-face-recognition-will-id-women-breaking-hijab-laws/

[11] Burgess, M. (2022, April 6). Europe Is Building a Huge International Facial Recognition System. WIRED. https://www.wired.com/story/europe-police-facial-recognition-prum/

[12] Johnson, K. (2023b, February 2). Get Used to Face Recognition in Stadiums. WIRED. https://www.wired.com/story/get-used-to-face-recognition-in-stadiums/

[13] Racial Discrimination in Face Recognition Technology. (2020, October 26). Science in the News. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

[14] Bunke, H., & Kandel, A. (2008). Applied Pattern Recognition. Studies in Computational Intelligence.

[15] J. Aznar-Casanova, N. Torro-Alves, and S. Fukusima, "How much older do you get when a wrinkle appears on your face? Modifying age estimates by number of wrinkles," Aging, Neuropsychol. Cogn., vol.17, no. 4, pp. 406–421, 2010.

[16] A. M. Albert, K. Ricanek, and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," Forensic Sci. Int., vol. 172, no. 1, pp. 1–9, 2007.

[17] Duan, M., Li, K., & Li, K. (2018). An Ensemble CNN2ELM for Age Estimation. IEEE Transactions on Information Forensics and Security, 13(3), 758–772. https://doi.org/10.1109/tifs.2017.2766583

[18] A. S. Al-Shannaq and L. A. Elrefaei, "Comprehensive Analysis of the Literature for Age Estimation from Facial Images," IEEE Access, vol.7, pp. 93229–93249, 2019.

[19] N. F. Hasan and S. Q. Mahdi, "Facial Features Extraction Using LBP for Human Age Estimation Based on SVM Classifier," 2020 Int. Conf. Comput. Sci. Softw. Eng., pp. 50–55, 2020.

[20] W. Pei and H. Dibeklio, "Attended End-to-End Architecture for Age Estimation From Facial Expression Videos," IEEE Trans. IMAGE Process., vol. 29, pp. 1972–1984, 2020.

[21] J. C. Xie and C. M. Pun, "Deep and Ordinal Ensemble Learning for Human Age Estimation from Facial Images," IEEE Trans. Inf. Forensics Secur., vol. 15, no. 8, pp. 2361–2374, 2020.

[22] N. da V. L. Young Ho Kwon, "Classification from Facial Images," Comput. Vis. Image Understan, 1994.

[23] S. M. Ali, Z. A. Darbar, and K. N. Junejo, "Age estimation from facial images using biometric ratios and wrinkle analysis," 2015 5th Natl. Symp. Inf. Technol. Towar. New Smart World, NSITNSW 2015, 2015.

[24] W. Pei and H. Dibeklio, "Attended End-to-End Architecture for Age Estimation From Facial Expression Videos," IEEE Trans. IMAGE Process., vol. 29, pp. 1972–1984, 2020.

[25] J. Wan, Z. Tan, Z. Lei, G. Guo, and S. Z. Li, "Auxiliary Demographic Information Assisted Age Estimation with Cascaded Structure," IEEE Trans. Cybern., vol. 48, no. 9, pp. 2531–2541, 2018.

[26] O. F. E. Osman and M. H. Yap, "Computational Intelligence in Automatic Face Age Estimation: A Survey," IEEE Trans. Emerg.Top. Comput. Intell., vol. 3, no. 3, pp. 271–285, 2019.

[27] N. Liu, F. Zhang, and F. Duan, "Facial Age Estimation Using a Multi-Task Network Combining Classification and Regression," IEEE Access, vol. 8, pp. 92441–92451, 2020.

[28] A. Casanova, L. Cascone, A. Castiglione, M. Nappi, C. Pero, Eye-movement and touch dynamics: a proposed approach for activity recognition of a web user, in: 2019 15th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), 2019, pp. 719–724.

[29] M. Zhao, H. Gao, W. Wang, J. Qu, L. Chen, Study on the identification of irritability emotion based on the percentage change in pupil size, 2020, pp. 20–24.

[30] K.D. Suschinsky, M.L. Lalumiere, M.L. Chivers, Sex differences in patterns of genital sexual arousal: measurement artifacts or true phenomena? Arch. Sex. Behav. 38 (4) (2009) 559–573.

[31] M. Guillon, K. Dumbleton, P. Theodoratos, M. Gobbe, C.B. Wooley, K. Moody, The effects of age, refractive status, and luminance on pupil size, Optom. Vis. Sci. 93 (9) (2016) 1093.

[32] Cascone, L., Medaglia, C., Nappi, M., & Narducci, F. (2020). Pupil size as a soft biometrics for age and gender classification. Pattern Recognition Letters, 140, 238–244. https://doi.org/10.1016/j.patrec.2020.10.009

[33] Hurley, D., ArbabZavar , B., and Nixon , M. 2007. Handbook of Biometrics. Springer, Chapter 7 The Ear as a Biometric, 131–150.

[34] Ž. Emeršič et al., "The unconstrained ear recognition challenge," 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 2017, pp. 715-724, doi: 10.1109/BTAS.2017.8272761.

[35] Islam, Syed & Bennamoun, Mohammed & Davies, Rowan. (2008). Fast and Fully Automatic Ear Detection Using Cascaded AdaBoost. 1-6. 10.1109/WACV.2008.4544023.

[36] Watabe, D., Sai, H., Sakai, K., & Nakamura, O. (2008). Ear biometrics using jet space similarity. 2008 Canadian Conference on Electrical and Computer Engineering, 001259-001264.

[37] Zhou, J., Cadavid, S., & Abdel-Mottaleb, M. (2010). Histograms of Categorized Shapes for 3D ear detection. 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 1-6.

[38] Abaza, A., Ross, A., Herbert, C., Harrison, M.F., and Nixon, M.S. 2011. A Survey on Ear Biometrics ACM Trans. Embedd. Comput. Syst. 39, 3, Article xx (September 2013), 33 pages.

[39] Khorsandi, Rahman and Mohamed Abdel-Mottaleb. "Gender classification using 2-D ear images and sparse representation." 2013 IEEE Workshop on Applications of Computer Vision (WACV) (2013): 461-466.

[40] Gnanasivam, P., Muttan, S. (2013). Gender Classification Using Ear Biometrics. In: S, M., Kumar, S. (eds) Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012). Lecture Notes in Electrical Engineering, vol 222. Springer, India. https://doi.org/10.1007/978-81-322-1000-9_13

[41] Lei, Jiajia et al. "Gender classification using automatically detected and aligned 3D ear range data." 2013 International Conference on Biometrics (ICB) (2013): 1-7.

[42] Yaman, D., Eyiokur, F. I., Sezgin, N., & Ekenel, H. K. (2018). Age and gender classification from ear images. 2018 International Workshop on Biometrics and Forensics (IWBF). https://doi.org/10.1109/iwbf.2018.8401568

[43] Yaman, D., Eyiokur, F. I., & Ekenel, H. K. (2019). Multimodal Age and Gender Classification Using Ear and Profile Face Images. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). https://doi.org/10.1109/cvprw.2019.00296

[44] Despite significant increase in birth registration, a quarter of the world's children remain 'invisible.' (n.d.). https://www.unicef.org/press-releases/despite-significant-increase-birth-registration-quarter-worlds-children-remain

[45] Smith, Terry & Brownlees, Laura. (2011). Age Assessment Practices: A Literature Review and Annotated Bibliography.

[46] Staff, R. (2019, May 10). Nigerian militia frees nearly 900 children used in war on Boko Haram: UNICEF. U.S. https://www.reuters.com/article/us-nigeria-security-children-idUSKCN1SG12P

[47] W.W. Greulich, S.I. Pyle, Radiographic Atlas of Skeletal Development of the Hand and Wrist, Stanford University Press, 1959.

[48] Lynnerup, N., Belard, E., Buch-Olsen, K., Sejrsen, B., & Damgaard-Pedersen, K. (2008). Intra- and interobserver error of the Greulich-Pyle method as used on a Danish forensic sample. Forensic science international, 179(2-3), 242.e1–242.e2426. https://doi.org/10.1016/j.forsciint.2008.05.005

[49] D. Zabet, C. Rérolle, J. Pucheux, N. Telmon, P. Saint-Martin, Can the Greulich and Pyle method be used on French contemporary individuals? Int. J. Leg. Med. 129 (2015) 171–177, https://doi.org/10.1007/s00414-014-1028-7

[50] Remy, F., Saliba-Serre, B., Chaumoitre, K., Martrille, L., & Lalys, L. (2021). Age estimation from the biometric information of hand bones: Development of new formulas. Forensic science international, 322, 110777. https://doi.org/10.1016/j.forsciint.2021.110777

[51] Mark S. Nixon, Tieniu N. Tan, and Rama Chellappa. Human Identification Based on Gait. Int. Series on Biometrics. Springer-Verlag, Dec. 2005.

[52] J. Han and B. Bhanu. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2):316– 322, 2006.

[53] X. Li, S.J. Maybank, S. Yan, D. Tao, and D. Xu. Gait components and their application to gender recognition. Trans. on Systems, Man, and Cybernetics, Part C, 38(2):145–155, Mar. 2008

[54] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. IEEE Trans. on Image Processing, 18(8):1905–1910, Aug. 2009.

[55] C. Kalaiselvan and A. Sivanantha Raja. Robust gait-based gender classification for video surveillance applications. Applied Mathematics and Information Sciences, 11:1207–1215, 2017.

[56] Jiwen Lu, Gang Wang, and Thomas S. Huang. Gait-based gender classification in unconstrained environments. In Proc of the 21st International Conference on Pattern Recognition, pages 3284–3287, 2012.

[57] Atsuya Sakata, Noriko Takemura, and Yasushi Yagi. Gait-based age estimation using multi-stage convolutional neural network. IPSJ Transactions on Computer Vision and Applications, 11(4):1–10, 2019.

[58] Atsuya Sakata, Yasushi Makihara, Noriko Takemura, Daigo Muramatsu, and Yasushi Yagi. How confident are you in your estimate of a human age? uncertainty-aware gait-based age estimation by label distribution learning. In IEEE International Joint Conference on Biometrics, Sep. 2020.

[59] Jiwen Lu and Yap-Peng Tan. Gait-based human age estimation. IEEE Trans. on Information Forensics and Security, 5(4):761–770, Dec. 2010.

[60] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi. Gait-based age estimation using a whole-generation gait database. In Proc. of the Int. Joint Conf. on Biometrics (IJCB2011), pages 1–6, Washington D.C., USA, Oct. 2011.

[61] C. Xu et al., "Real-Time Gait-Based Age Estimation and Gender Classification from a Single Image," 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2021, pp. 3459-3469, doi: 10.1109/WACV48630.2021.00350.

[62] Ball, R. L., Feiveson, A. H., Schlegel, T. T., Starc, V., & Dabney, A. R. (2014). Predicting "heart age" using electrocardiography. Journal of personalized medicine, 4(1), 65–78. https://doi.org/10.3390/jpm4010065

[63] Stein, P.; Kleiger, R.; Rottman, J. Differing effects of age on heart rate variability in men and women. Am. J. Cardiol. 1997, 80, 302–305.

[64] Islam, Sheikh Rabiul & Eberle, William & Ghafoor, Sheikh & Ahmed, Mohiuddin. (2021). Explainable Artificial Intelligence Approaches: A Survey.

[65] Browning, K., & Orlans, N.M. (2014). Biometric Aging - Effects of Aging on Iris Recognition.

[66] Iris Recognition. (2019, October 26). Electronic Frontier Foundation. https://www.eff.org/pages/iris-recognition

[67] BBC News. (2012, July 26). Black Hat: Iris scanners "can be tricked" by hackers. https://www.bbc.com/news/technology-18997580

[68] Rajput, M. R., & Sable, G. S. (2020). Age Group Estimation from Human Iris. Advances in Intelligent Systems and Computing, 519–529. https://doi.org/10.1007/978-981-15-2475-2_48

[69] Ambroa-Conde, A., Girón-Santamaría, L., Mosquera-Miguel, A., Phillips, C., Casares De Cal, M., Gómez-Tato, A., Álvarez-Dios, J., De La Puente, M., Ruiz-Ramírez, J., Lareu, M., & Freire-Aradas, A. (2022). Epigenetic age estimation in saliva and in buccal cells. Forensic Science International: Genetics, 61, 102770. https://doi.org/10.1016/j.fsigen.2022.102770

[70] Becker, J., Böhme, P., Reckert, A. et al. Evidence for differences in DNA methylation between Germans and Japanese. Int J Legal Med 136, 405–413 (2022). https://doi.org/10.1007/s00414-021-02736-3

[71] Correia Dias, H., Cunha, E., Corte Real, F., & Manco, L. (2022). Challenges and (Un)Certainties for DNAm Age Estimation in Future. Forensic Sciences, 2(3), 601–614. https://doi.org/10.3390/forensicsci2030044

[72] Becker, J., Böhme, P., Reckert, A., Eickhoff, S. B., Koop, B. E., Blum, J., Gündüz, T., Takayama, M., Wagner, W., & Ritz-Timme, S. (2021). Evidence for differences in DNA methylation between Germans and Japanese. International Journal of Legal Medicine, 136(2), 405–413. https://doi.org/10.1007/s00414-021-02736-3