Age identification using biometrical data

ADVANCED TOPICS IN COMPUTER AND NETWORK SECURITY

Reza Ghasemi

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO MATEMATICA

A.Y. 2022/2023

Biometrics

Università degli Studi di Padova

Definition: measurements whether behavioral or physical to identify individuals.

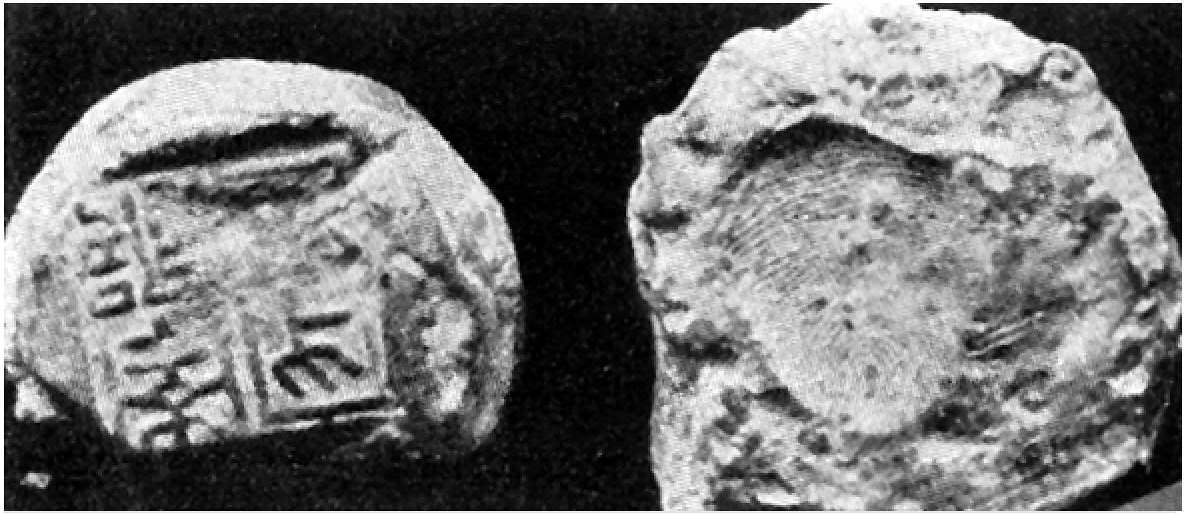
Various features could be employed:

- Gait
- Iris
- Voice
- Odor

DIPARTIMENTO MATEMATICA

Biometrics: History

Different forms used since ancient times..

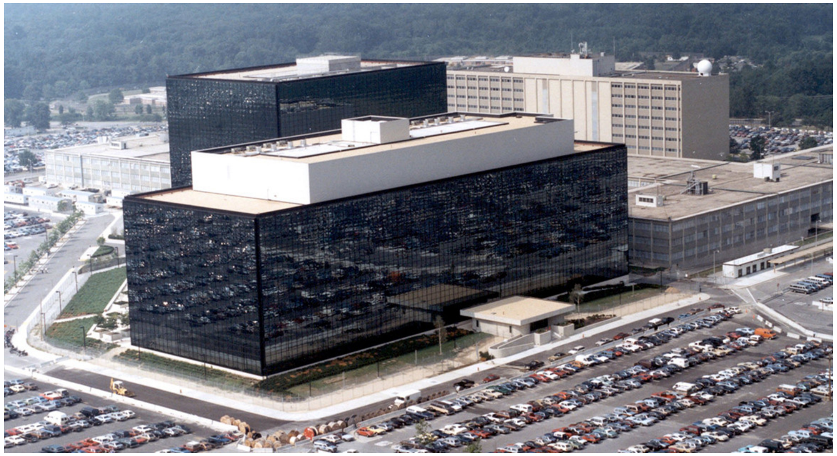


”Experience shows that no two individuals have fingers exactly alike.”

Rashid al-Din Hamadani in 1303

Biometrics: Advantages

- More convenient
- Improved security
- More efficient (e.g. faster login time)



Biometrics: Disadvantages

- It could affect privacy and security negatively
- Biometric data cannot be changed like passwords (data breach)

The Intercept_

THE TALIBAN HAVE SEIZED U.S. MILITARY BIOMETRICS DEVICES

Biometric collection and identification devices were seized last week during the Taliban's offensive.

Ken Klippenstein, Sara Sirota

August 17 2021, 10:11 p.m.

TECHNOLOGY The New York Times

For Sale on eBay: A Military Database of Fingerprints and Iris Scans

German security researchers studying biometric capture devices popular with the U.S. military got more than they expected for $68 on eBay.

Soft Biometrics

Not sufficiently unique to be used for identification

Additional information is provided such as:

- Age
- Gender
- Height

Useful for commercial purposes where **recommendation may differ based on age and gender**

Soft Biometrics (cont.)

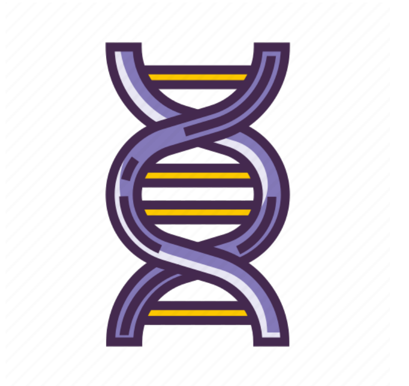
According to UNICEF, a quarter of children under the age of 5 are not registered.

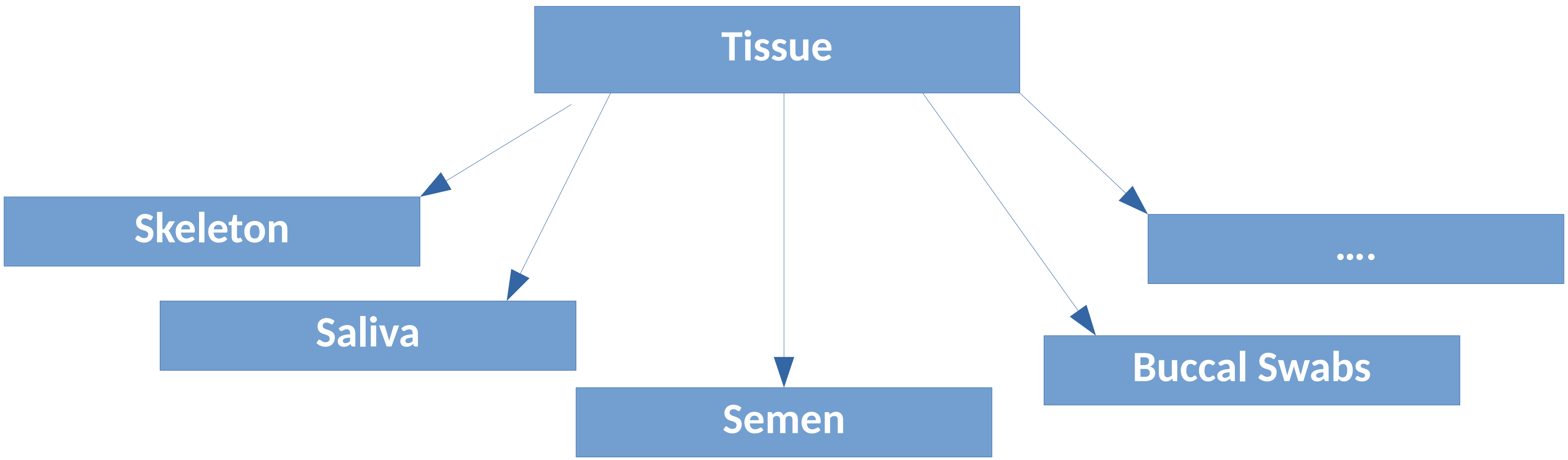
Possible issues:

- Child trafficking
- Children used in wars
- Issues entering countries as refugees
- Lacking certain priviledges due to incorrect estimation (e.g. child may be imprisoned)

Soft Biometrics: DNA methylation

A methyl group ($-CH_3$) is added to 5′ carbon of cytosines positioned next to guanines (CpG)





DNAm: Advantages

Università degli Studi di Padova

Studied mostly for **forensic applications**

It can be used to reduce number of suspects (soft biometrics)

It can be **used with both living and dead subjects**

Could be used as **a biomarker** (e.g. mortality)

Consistent among different populations

DIPARTIMENTO MATEMATICA

Soft Biometrics: DNA methylation

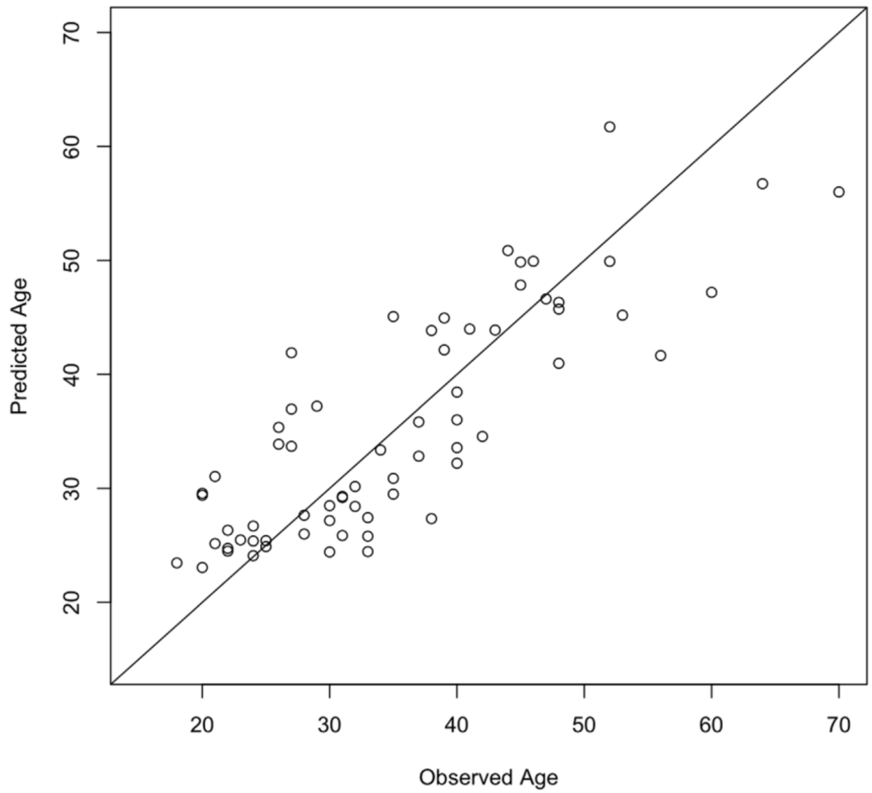
Tissue	Model	CpG number	MAE	MAE_{mean}	RMSE	%CP±PI
Saliva	Model 1	9 CpGs	±3.17	±4.79	6.46	76.55 %
	Model 2	8 CpGs with *ASPA* excluded	±2.98	±4.66	6.4	77.11 %
	Model 3	8 CpGs with *FHL2* excluded	±3.29	±4.76	6.47	75.49 %
	Model 4	8 CpGs with cg10501210 excluded	±3.79	±5.04	6.76	75.59 %
	Model 5	7 CpGs with cg10501210 and *FHL2* excluded	±3.85	±5.17	6.93	76.61 %
	Model 6	7 CpGs with cg10501210 and *ASPA* excluded	±3.96	±4.97	6.69	74.45 %
	Model 7	7 CpGs with *FHL2* and *ASPA* excluded	±3.31	±4.69	6.37	78.74 %
	Model 8	6 CpGs with cg10501210, *FHL2* and *ASPA* excluded	±4.02	±5.10	6.91	78.74 %
Buccal swab	Model 1	9 CpGs	±3.85	±5.01	6.35	75.47 %
	Model 2	8 CpGs with *ASPA* excluded	±4.41	±5.09	6.45	74.91 %
	Model 3	8 CpGs with *FHL2* excluded	±4.13	±4.90	6.24	75.53 %
	Model 4	8 CpGs with cg10501210 excluded	±4.45	±5.27	6.66	76.64 %
	Model 5	7 CpGs with cg10501210 and *FHL2* excluded	±4.89	±5.52	6.94	74.42 %
	Model 6	7 CpGs with cg10501210 and *ASPA* excluded	±4.22	±5.15	6.63	73.86 %
	Model 7	7 CpGs with *FHL2* and *ASPA* excluded	±4.16	±4.99	6.36	75.47 %
	Model 8	6 CpGs with cg10501210, *FHL2* and *ASPA* excluded	±4.72	±5.43	6.86	76.64 %
Combined (saliva and buccal swabs)	Model 1	9 CpGs	±3.31	±4.57	6.06	74.38 %
	Model 2	8 CpGs with *ASPA* excluded	±3.66	±4.75	6.20	74.99 %
	Model 3	8 CpGs with *FHL2* excluded	±3.67	±4.78	6.32	74.36 %
	Model 4	8 CpGs with cg10501210 excluded	±3.78	±5.05	6.52	77.72 %
	Model 5	7 CpGs with cg10501210 and *FHL2* excluded	±4.18	±5.43	6.96	77.28 %
	Model 6	7 CpGs with cg10501210 and *ASPA* excluded	±3.77	±4.93	6.39	80.96 %
	Model 7	**7 CpGs with *FHL2* and *ASPA* excluded**	**±3.54**	**±4.79**	**6.23**	**76.08 %**
	Model 8	6 CpGs with cg10501210, *FHL2* and *ASPA* excluded	±5.23	±5.93	7.54	74.06 %

Credit: Ambroa-Conde et al. (2022)

Soft Biometrics: DNA methylation

Bocklandt et al. (2011) used **regression** to estimate age

Average accuracy reported: 5.2 years



Soft Biometrics: Gait

Università degli Studi di Padova

Gait refers to **patterns of movements** of individuals.

A **behavioral feature** that has many advantages, including:

- Different angles (unlike face)
- No cooperation is needed

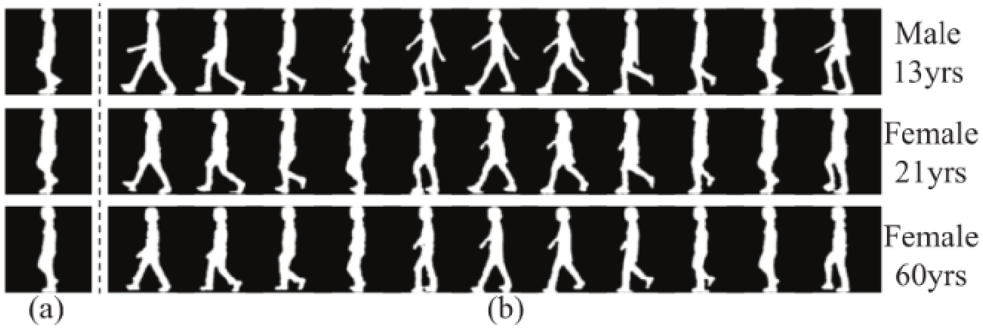
It has been used for age estimation, **most studies have relied on full gait cycle**.

Full gait cycle is not practical for real-time systems

Dipartimento Matematica

Soft Biometrics: Gait

Motion information contained in a full cycle is **extremely valuable**. (e.g. Females have smaller arm swings)



Credit: Xu et al. (2021)

Xu et al. Proposed a solution where **full cycle is created from a single shot**.

It can handle different angles

The model was implemented to demonstrate **it could be used as a real time system**

Soft Biometrics: Gait

Age estimation for children is easier than adults due to **smaller head-body ratio**.
(Single shot is sufficient)

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**

Credit: Xu et al. (2021)

Soft Biometrics: Hand Bones

Università degli Studi di Padova

Analyzing x-ray images of hand bones could be used to estimate ag

Why hands?

- High bone count
- Low level of radiation

Two common approaches

- **Greulich and Pyle (GP)**
- **Tanner-Whitehouse (TW)**

Soft Biometrics: Hand Bones

Università degli Studi di Padova

Issues with GP and TW

- **Manual**
- **Costly**
- **Subjective**

Additional elements could play a role such as:

- Gender
- Race
- Diet
- Socio-economic factors

Dipartimento Matematica

Soft Biometrics: Hand Bones

Lee et al. (2020) used **different deep learning architectures** to estimate the age.

Deep learning could **reduce costs** and **increase speed** of estimation. (compared to GP and TW)

Age estimation treated as regression problem.

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

Credit: Lee et al. (2020)

Soft Biometrics: Hand Bones

Università degli Studi di Padova

Later work indicated that **gender had a substantial effect on age estimation**.

	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)

Credit: Lee et al. (2020)

Gender-aware models have better performance compared to gender-agnostic models

Dipartimento Matematica

Soft Biometrics: Fingerprints

Fingerprint could be used to estimate age of a subject

Falohun et al. (2016) used **DWT+PCA**

The authors believed that **Back Propagation neural network (BPNN)** and **ANN** are **not effective for age estimation**

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

Credit: Falohun et al. (2016)

DIPARTIMENTO MATEMATICA

Soft Biometrics: Fingerprints

T Abraham and M (2016) proposed a model based on **wavelet transform** and **SVD**

	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%	

Credit: T Abraham and M (2016)

Subjects **above the age of 50 had the maximum success rate** while the **worst performance was reserved for subjects in the range of 16–19**

Soft Biometrics: Fingerprints

UNIVERSITÀ DEGLI STUDI DI PADOVA

Jayakala and Sudha (2022) employed **ResNet50** to estimate the age

Subjects used in the study <u>ranged between 1 to 60</u>

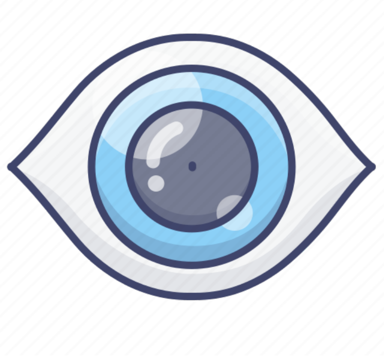
1000 images used for training and 200 images used for testing.

Authors claimed **93% accuracy** in age estimation

DIPARTIMENTO MATEMATICA

Periocular area is rich in useful features:

- Skin texture
- Iris
- Pupil
- Retina



Advantages:

- Non intrusive
- Easy to collect
- Long-distance friendly
- Hygienic

Soft Biometrics: Pupil

Università degli Studi di Padova

Pupils can reveal valuable and interesting information like:

- Emotional state
- Age
- Gender
- Health status

Dilation of pupil **differs** between males and females.

Heterosexual females pupils dilate whether we show them a picture of a male or female subject. (gender estimation)

Moreover, **pupils dilation changes with time** (age estimation)

Soft Biometrics: Pupil

Cascone et al. (2020) achieved **better results for age rather than gender**

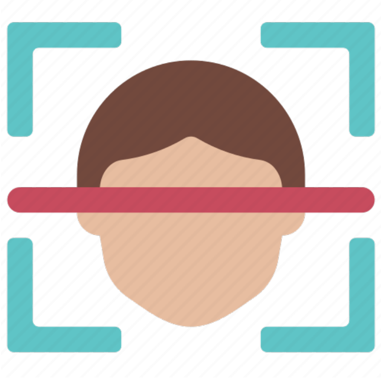
Several classifiers were trained and evaluated:

	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*

Image credit: Cascone et al. (2020)

Biometrics: Facial Features

One of the most common and widespread biometric features



Advantages:

- No cooperation
- No special hardware

IARPA is actively trying to improve biometric systems:

Expanding Accurate Person Recognition to New Altitudes and Ranges: The BRIAR Dataset

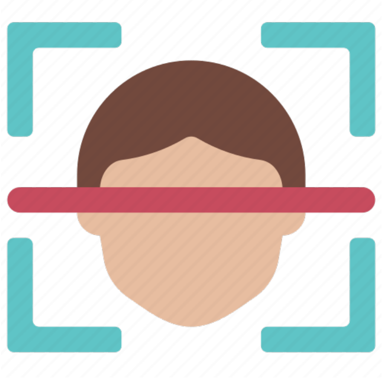
David Cornett III, Joel Brogan, Nell Barber, Deniz Aykac, Seth Baird, Nick Burchfield, Carl Dukes, Andrew Duncan, Regina Ferrell, Jim Goddard, Gavin Jager, Matt Larson, Bart Murphy, Christi Johnson, Ian Shelley, Nisha Srinivas, Brandon Stockwell, Leanne Thompson, Matt Yohe, Robert Zhang, Scott Dolvin, Hector J. Santos-Villalobos, David S. Bolme

Oak Ridge National Laboratory
P.O. Box 2008. Oak Ridge, TN 37831
{cornettdciii,bolmeds}@ornl.gov

Soft Biometrics: Facial Features

Università degli Studi di Padova

Useful when showing an item which is age restricted to users.



Features of the face could indicate gender, age, ethnicity, etc.

Local features of the face (e.g. wrinkles) could be used to estimate age according to Aznar-Casanova et al. (2010)

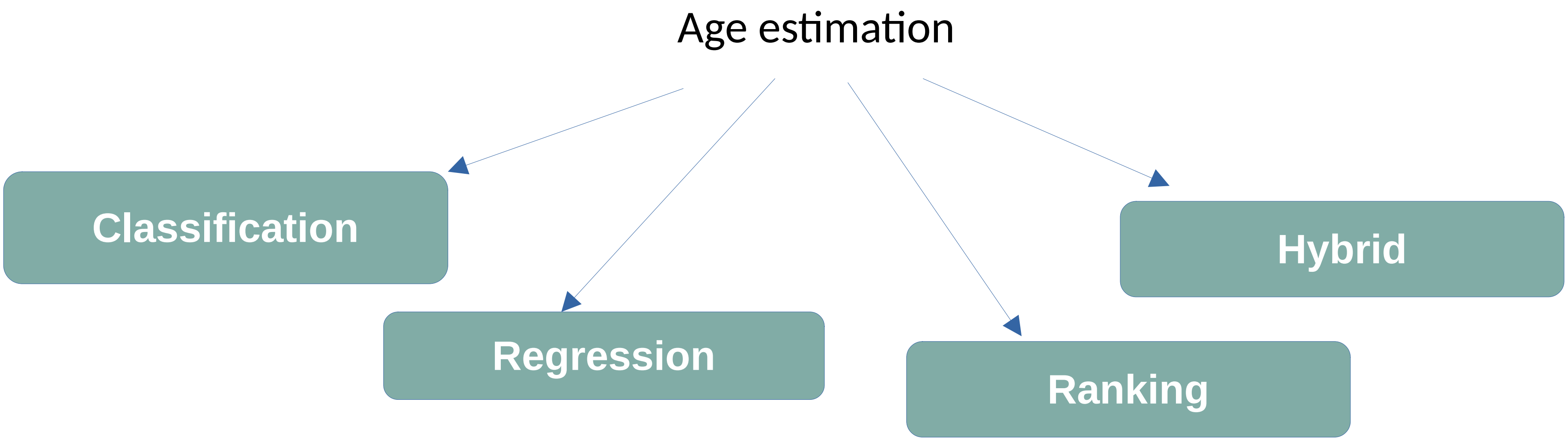
Various factors could affect age estimation:

- Skin diseases
- Living in polluted cities
- Race
- Gender
- ...

Dipartimento Matematica

Soft Biometrics: Facial Features

Most studies indicate deep learning achieves better results.



Datasets that have been used: AgeDB, UTKFace, IMDB-wiki, etc.

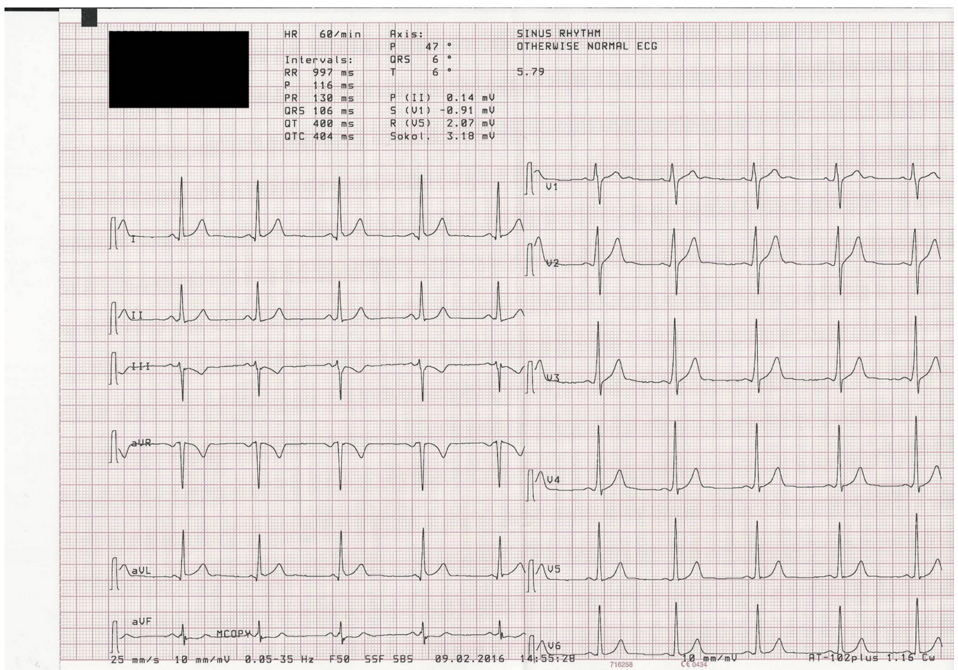
Soft Biometrics: ECG

Università degli Studi di Padova

Mainly used to **measure heart activity** and **detect abnormalities**

Numerous **wearable-devices support ECG-capability**

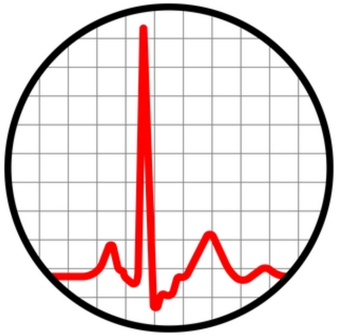
Well known that **12-lead ECG undergo change** during life



DIPARTIMENTO MATEMATICA

Soft Biometrics: ECG (cont.)

Various approaches have been used:



- Starc et al. (2012) → Multi-linear regression
- Ball et al. (2014) → Bayesian
- Attia et al. (2019) → CNN

Attia et al. (2019) treated age as a **continous value**.

Deep learning vs. Traditional algorithms: Lack of explainability remains a challenge

Soft Biometrics: ECG (cont.)

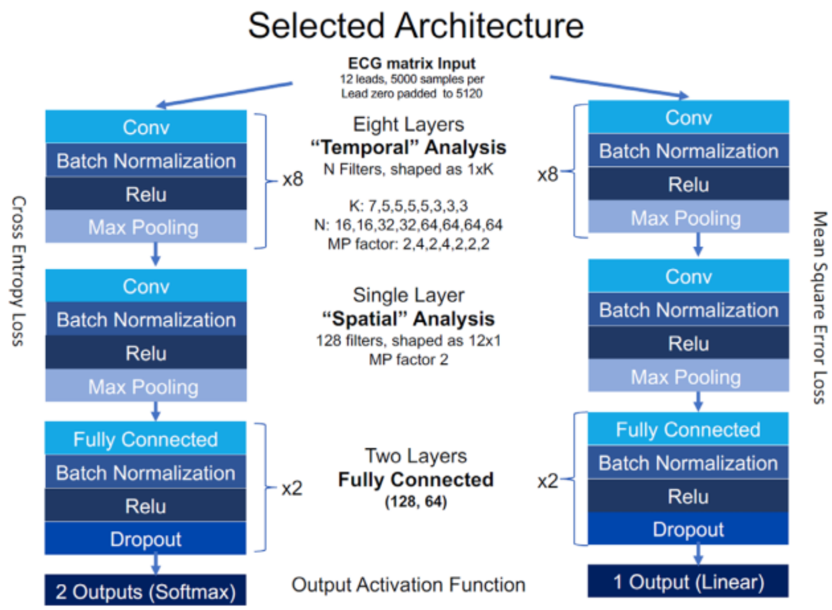


Image credit: Attia et al. (2019)

MAE reported was 6.9 ± 5.6 years

Results more accurate for younger subject compared to elderly (>55 y/o)

Alterations in hormonal level may explain inferior results for elderly

Ear Biometrics

First introduced in 1896 by **Bertillon** for identification purposes.

Ear's individuality has not been scientifically proven.

Could be used with other features (e.g. face) where:

- Issues with illumination
- Occlusions due to hair, limited view, ...

Studied have mostly focused on **closed** and **controlled** environments.

DIPARTIMENTO MATEMATICA

Ear Biometrics (cont.)

Various datasets are available:

- WVU (West Virginia University)
- USTB (University of Science and Technology Beijing)
- UND (University of Notre Dame)
- FERET
- ...

For segmentation, different approaches exist:

- Viola-Jones
- Shape-based
- Hybrid
- ...

VJ does **not suffer from image quality (+)** but it is **time consuming (-)**.

DIPARTIMENTO MATEMATICA

Ear Biometrics (cont.)

To further improve accuracy, futile features (e.g. hair) must be removed.

A solution would be to use a thermogram of ear.

Regions lower in temperature indicate hair where higher temperature regions are related to ear.

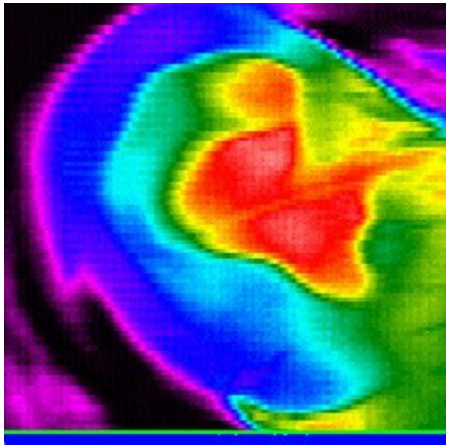


Image credit: Burge and Burger (2000)

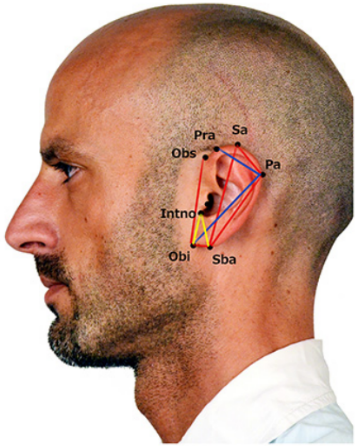
Soft Biometrics: Ear

Mostly studies have focused on gender rather than age.

Yaman et al. (2018) Used two approaches:

- Geometric
- Appearance based

Geometric features: based on **eight landmarks** from which eighteen measurements were derived.



Soft Biometrics: Ear

Apperance based approaches had better results compared to geometric approaches

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)

Image credit: Yaman et al. (2018)

To further improve the results, in another study **number of samples were increased** By the authors which resulted in 9% improvement for accuracy.

Soft Biometrics: Iris

Sgroi et al. (2013) used **Random Forest** with **300 trees**.

The correct classification rate reported **64.68%**

Erbilek et al. (2013) used various classifiers and reported the following results:

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

Soft Biometrics: Iris

Università degli Studi di Padova

Rajput and Sable (2020) proposed a new model to estimate age.

Only three statistical features were selected to be used

Different classifiers were evaluated with the following results:

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

Q & A