



PREDICTING AGE, GENDER & ETHNICITY WITH NEURAL NETWORKS & XGBOOST

BIG DATA ANALYSIS – PROJECT PRESENTATION

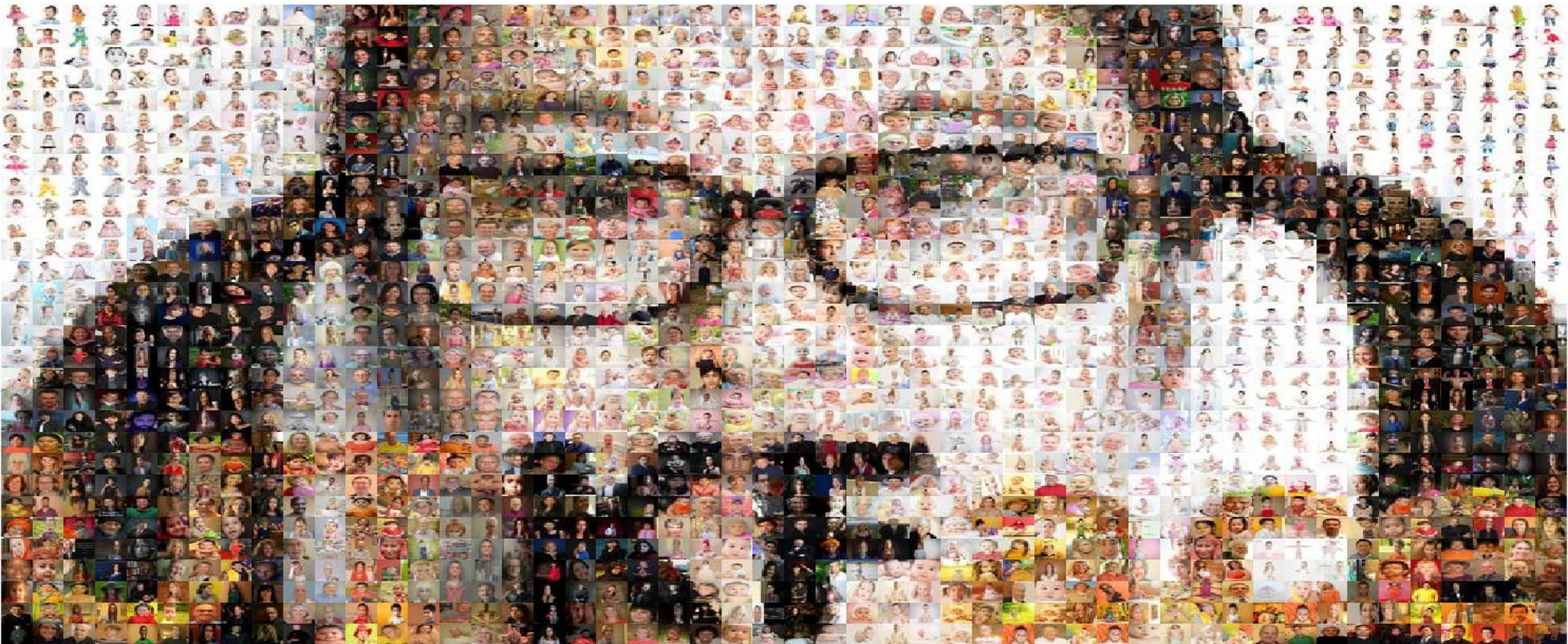
By *Tai Thorsen, Magnus Ibh & Andrew Khoudi*

June 12th 2019

All group members contributed evenly

UNIVERSITY OF COPENHAGEN





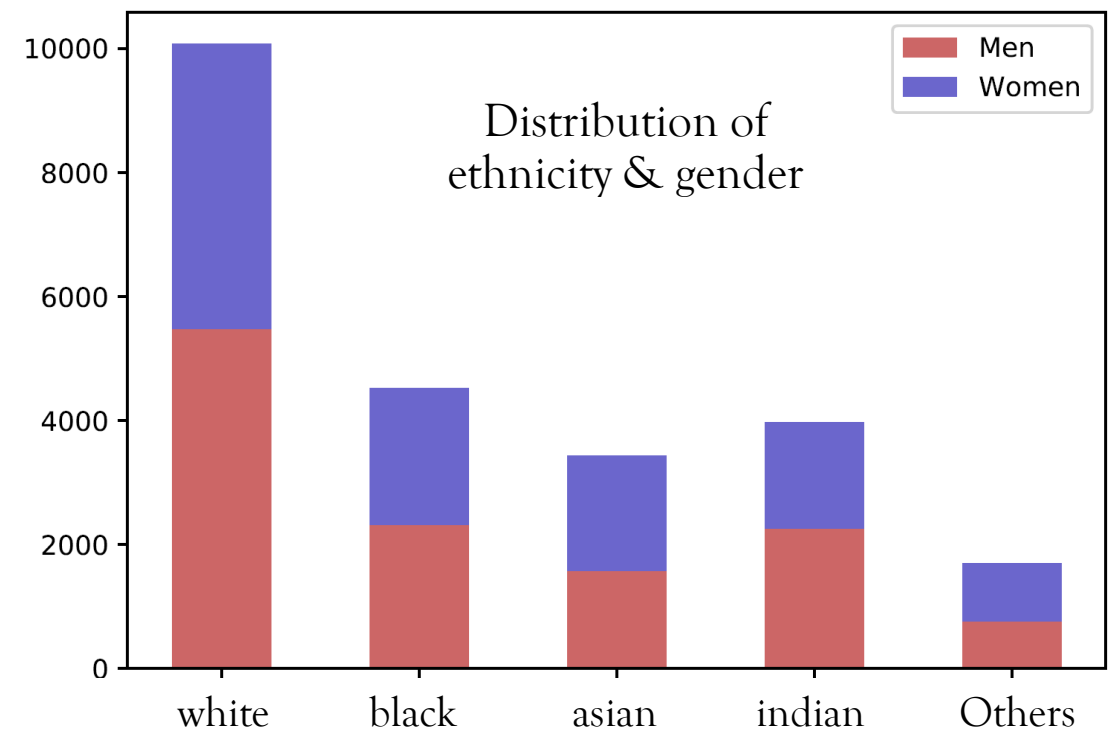
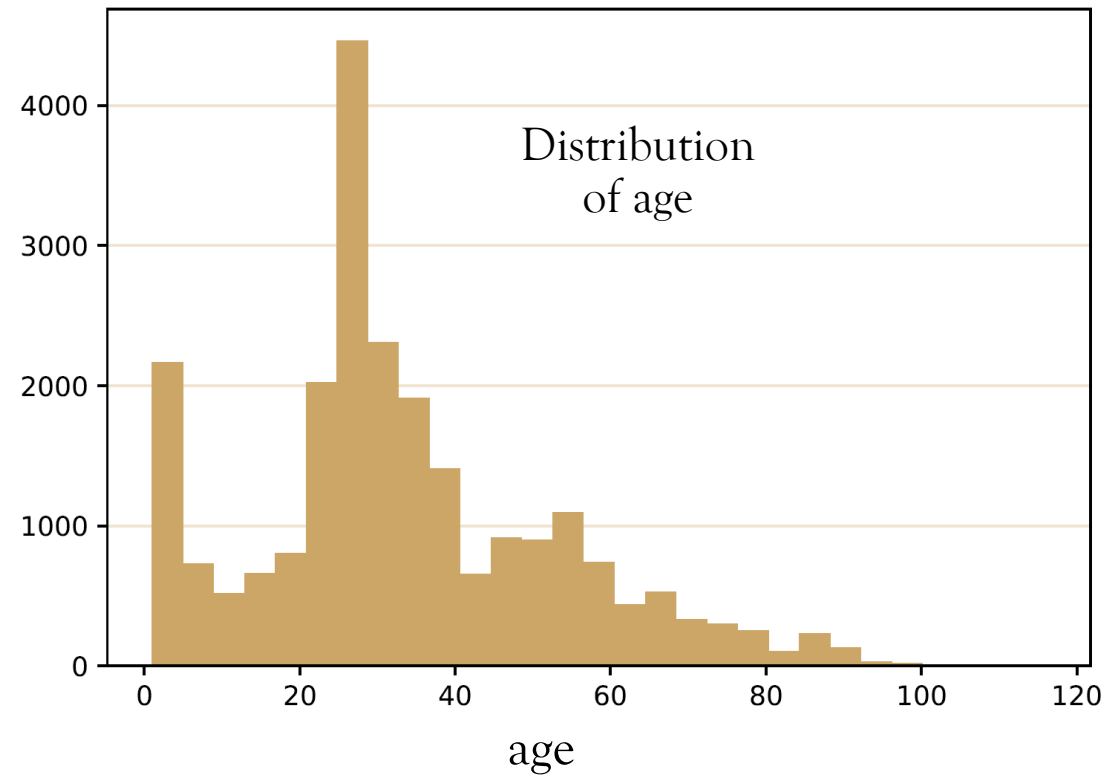
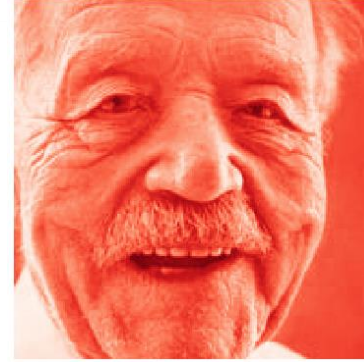
Gender

Ethnicity

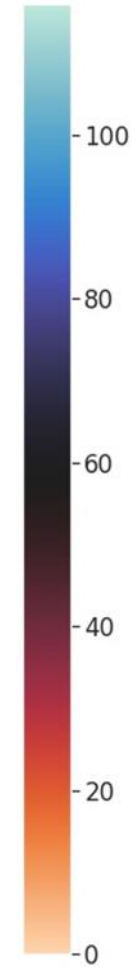
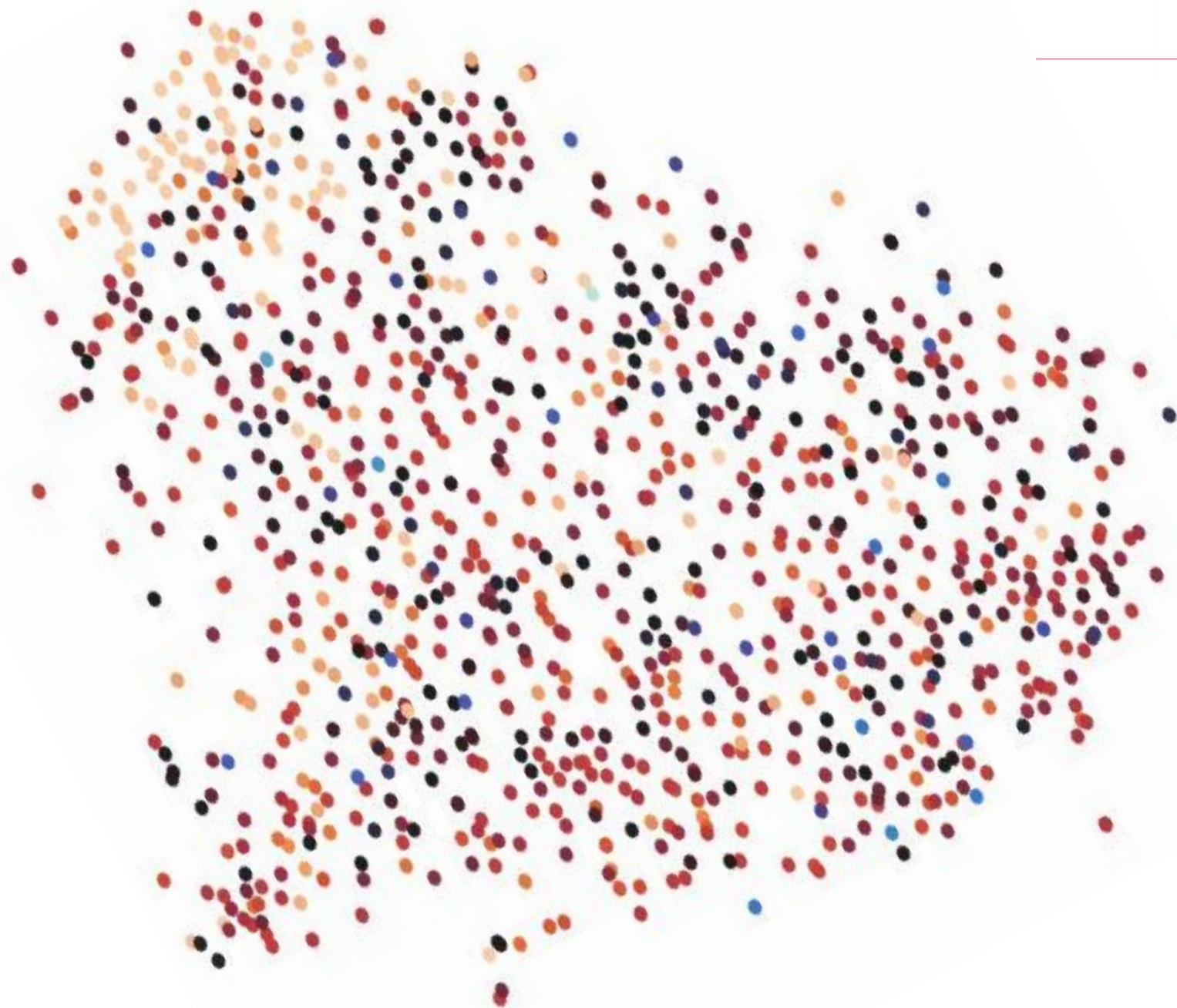
Age



Data

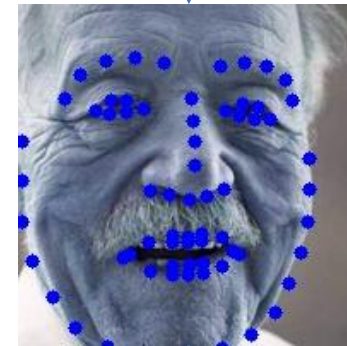


T-SNE



Age

Different approaches

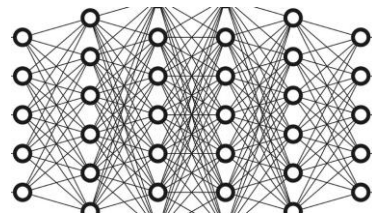
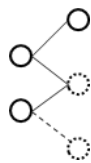


PCA

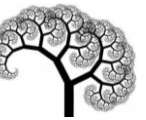
From scratch

"Transfer learning"

Convolutional
Neural
Network

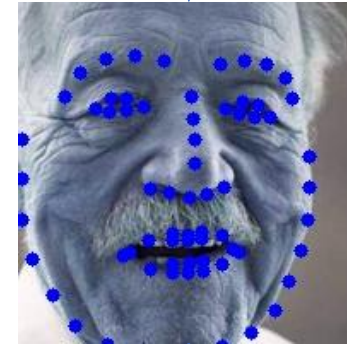


Decision Tree



XGBoost

Different approaches

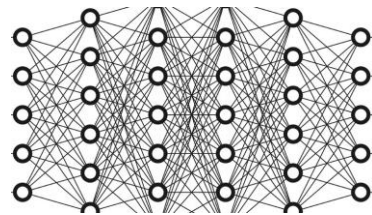
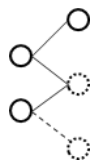


PCA

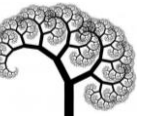
From scratch

"Transfer learning"

Convolutional
Neural
Network

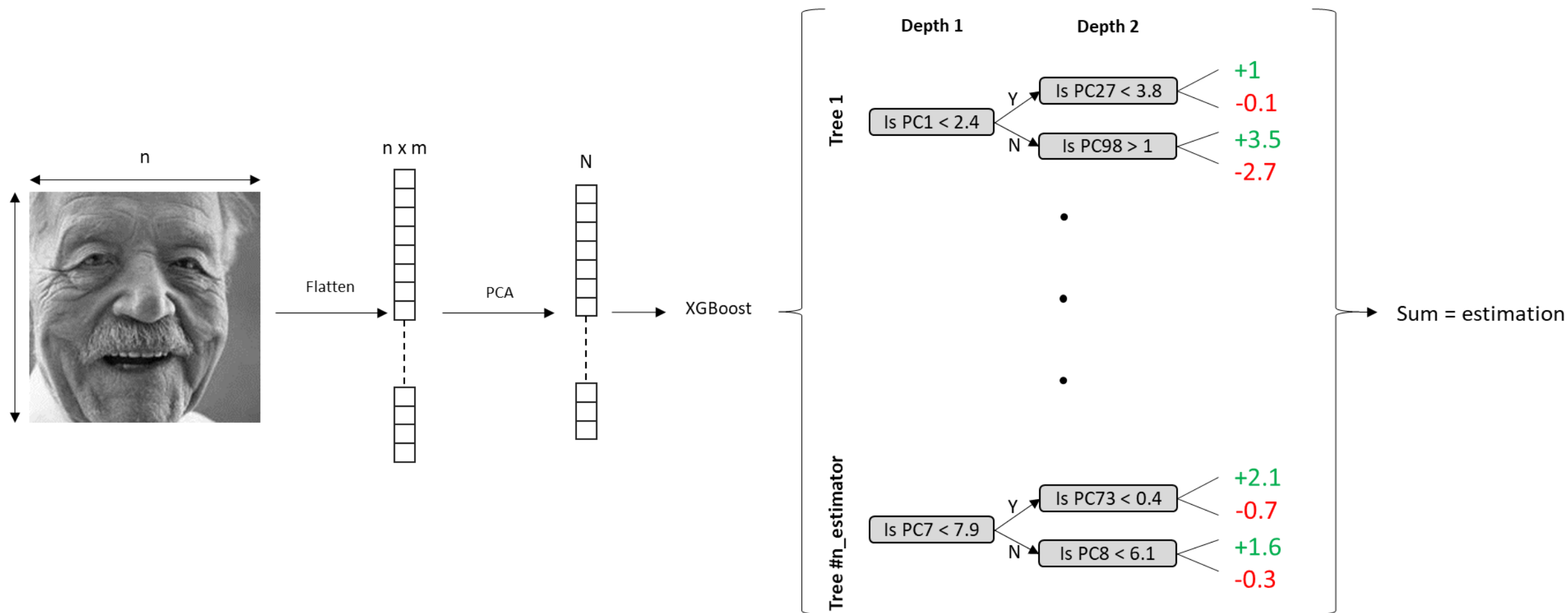


Decision Tree

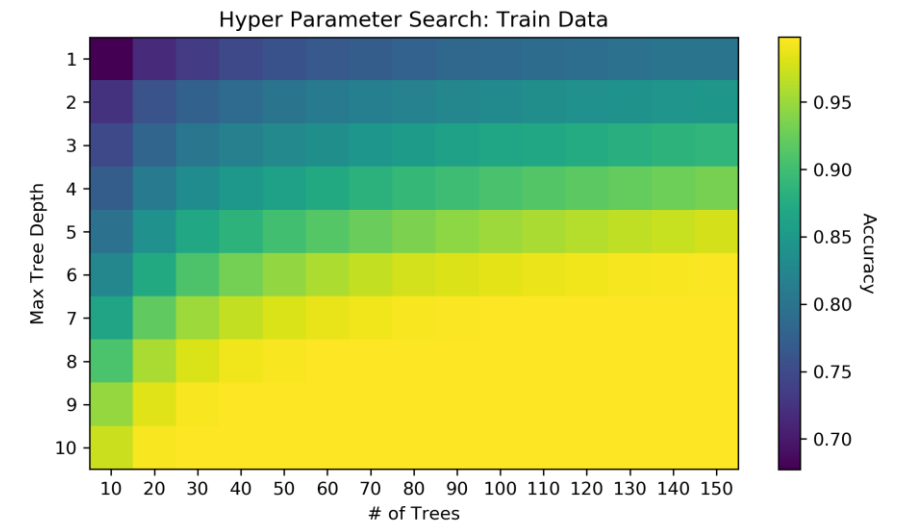
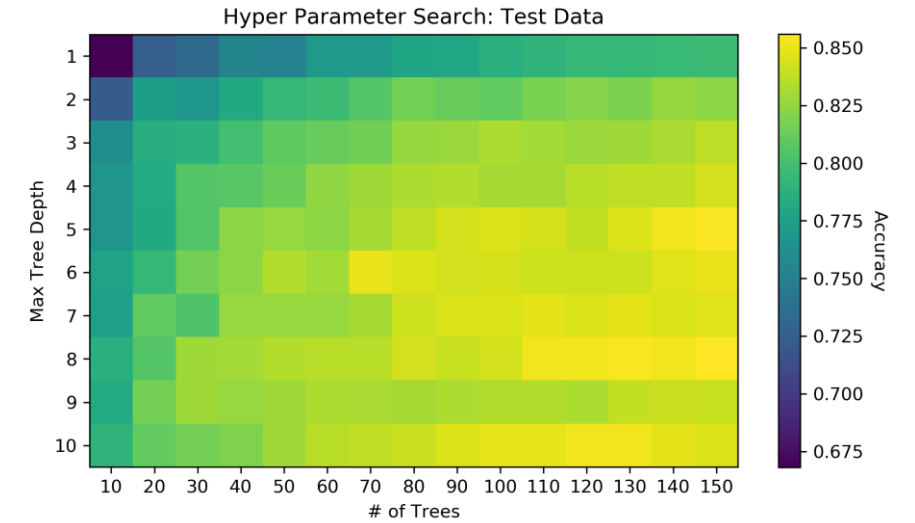
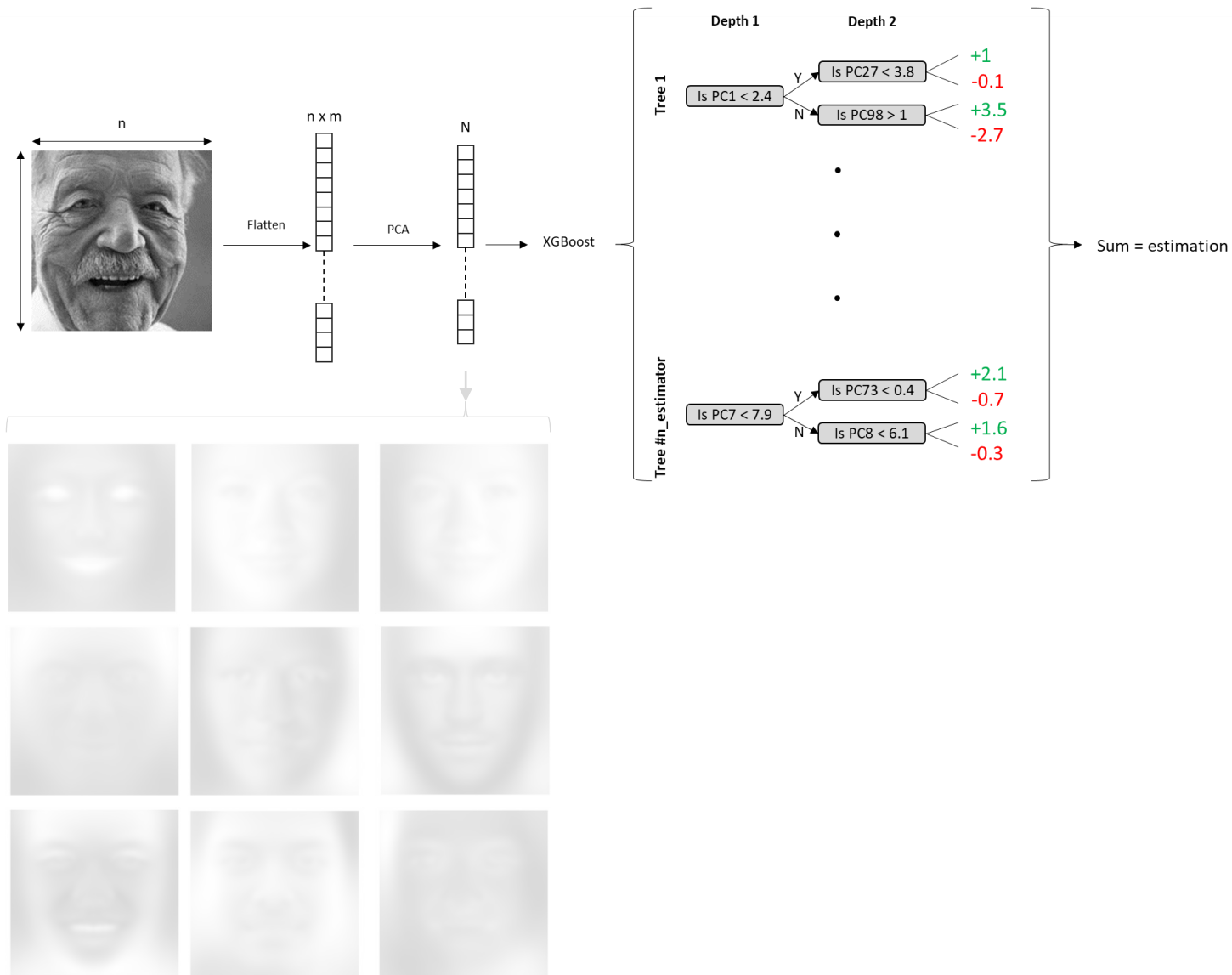


XGBoost

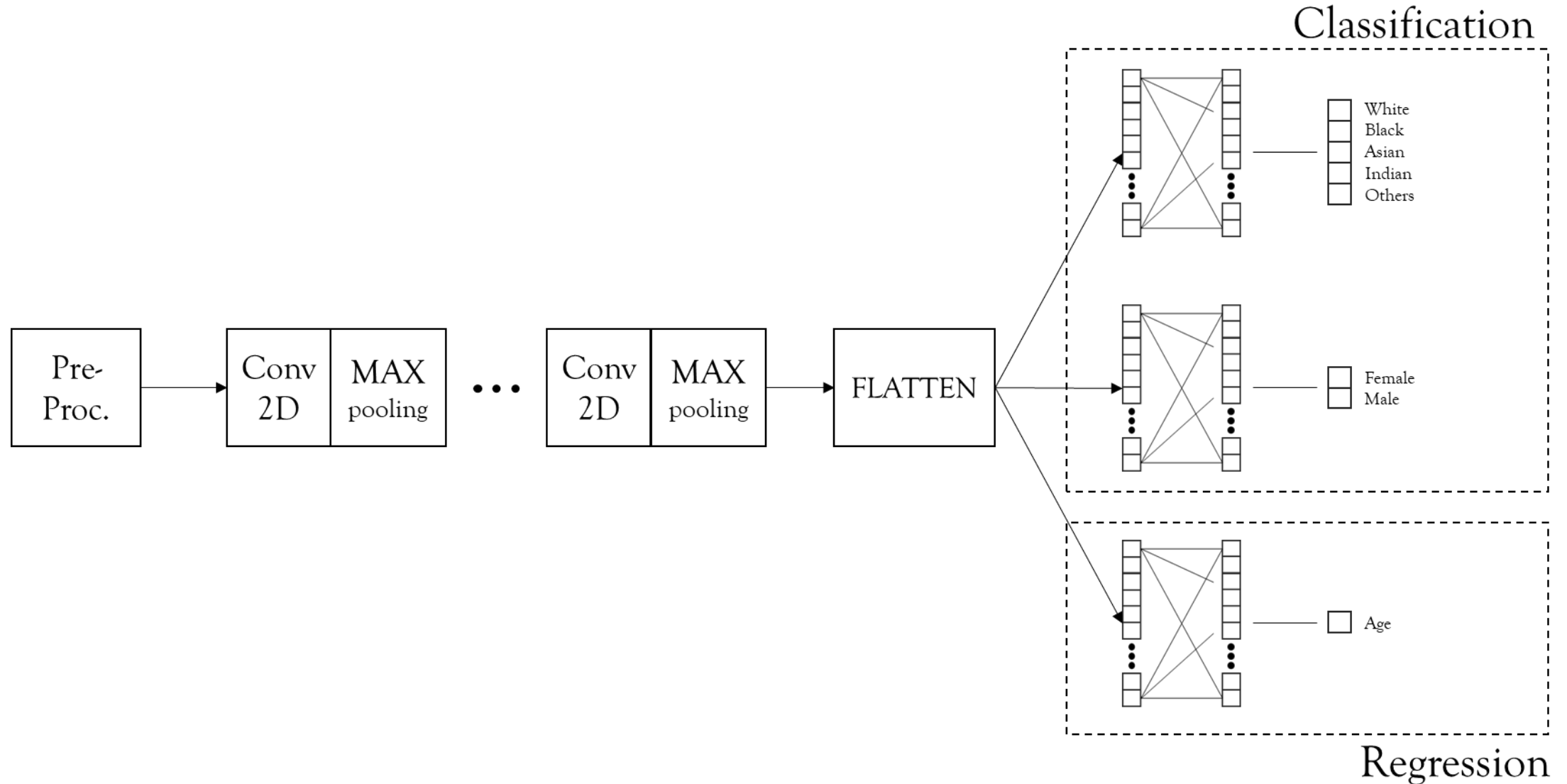
Method XGBoost



Method XGBoost



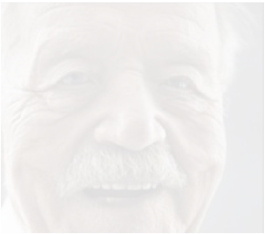
Method Convolutional Neural Network



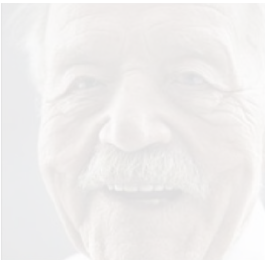
Method Convolutional Neural Network

Image Augmentation

original



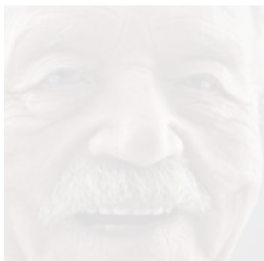
flipped



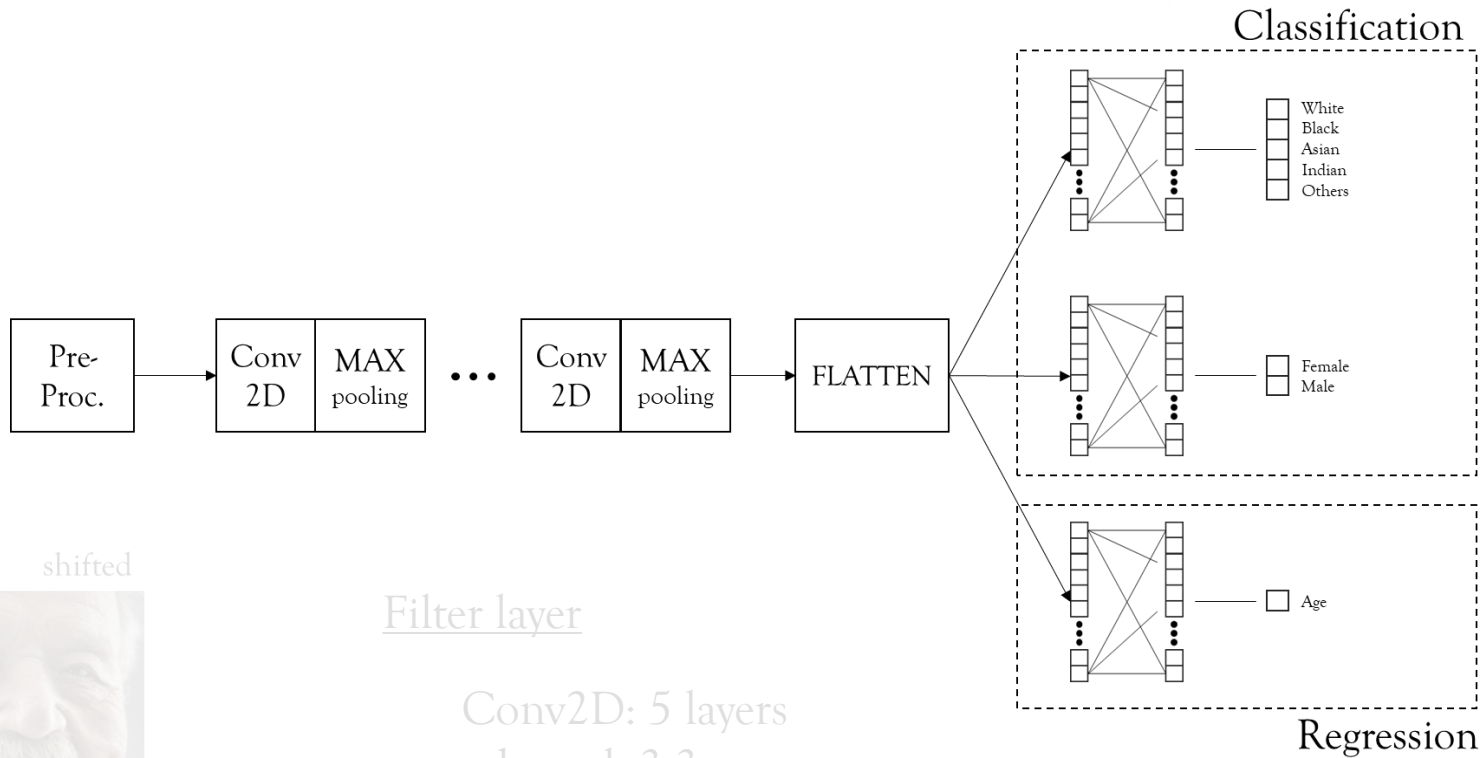
shifted



noise added



zoomed



Filter layer

Conv2D: 5 layers
kernel: 3,3
filters: 32-64

MaxPooling: 3 layers
pool size: 2,2
stride: 2,2

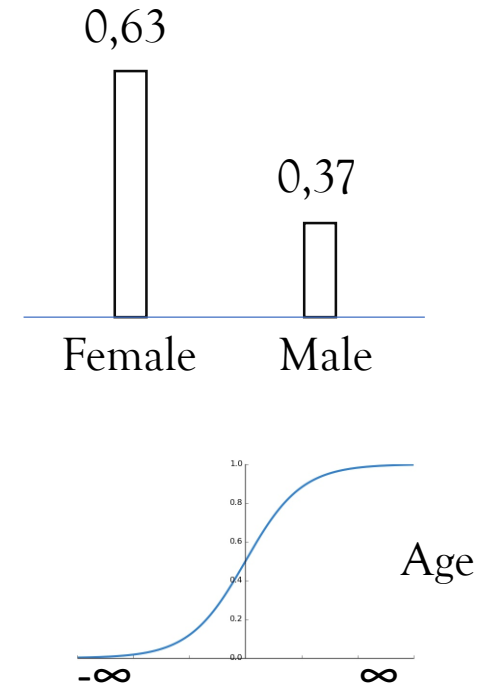
Output layer

3 models: Ethnicity, Gender & Age

Ethnicity & Gender: Categorical crossentropy

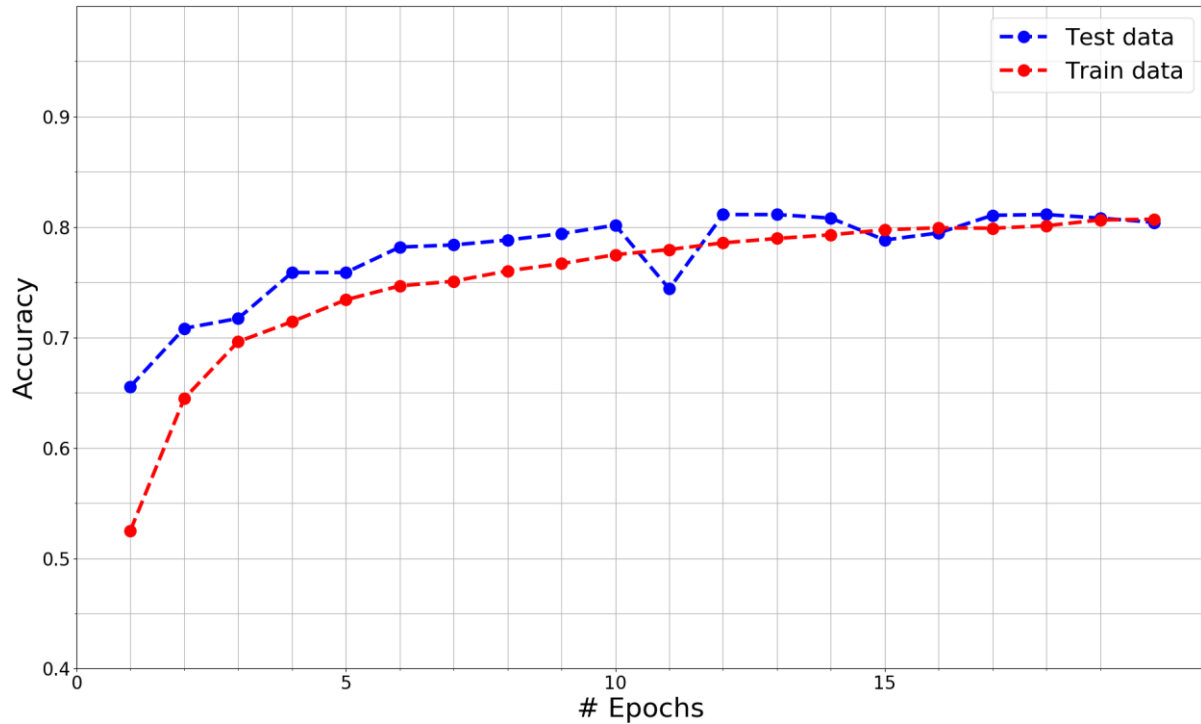
Age: LogCosh Error

Adjustable parameters: app. 93,000 (128 nodes)

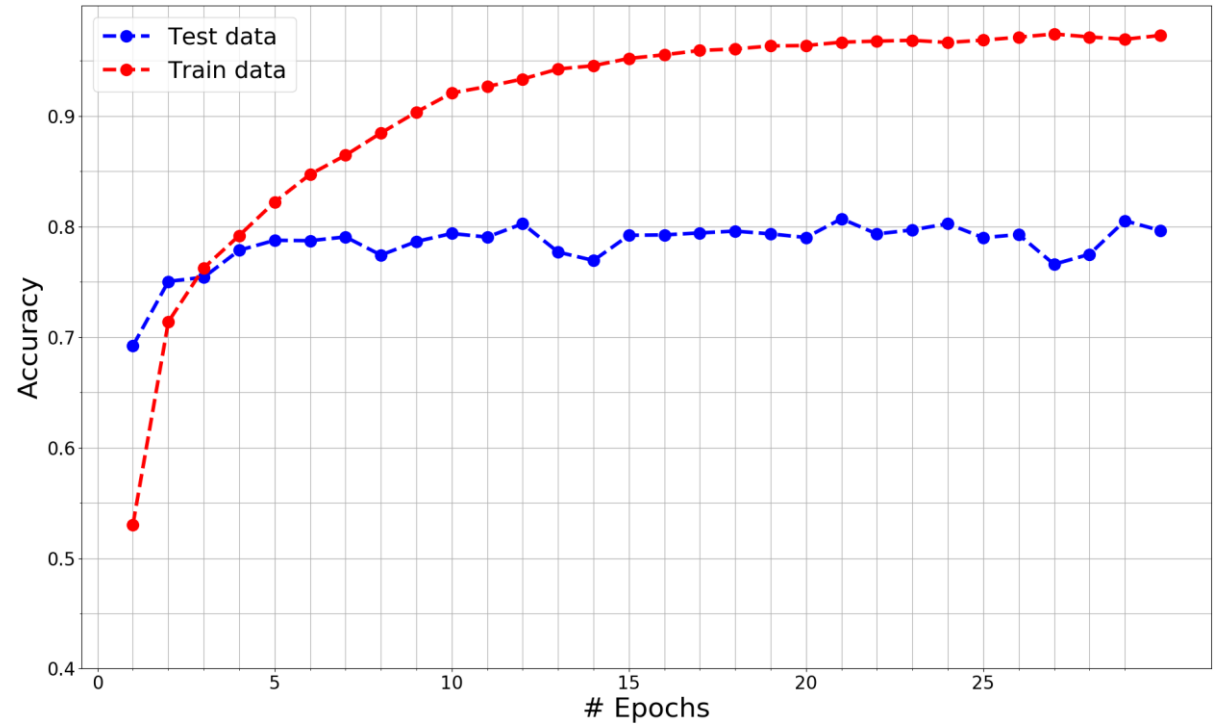


Accuracy of 2 different CNN models: **Ethnicity prediction**

Single-Output CNN-model with Image Augmentation

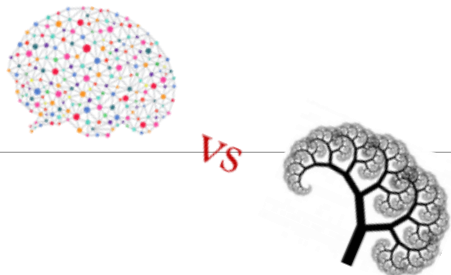


Multiple-Output CNN-model without Image Augmentation



Confusion Matrix

Convolutional
Neural Network

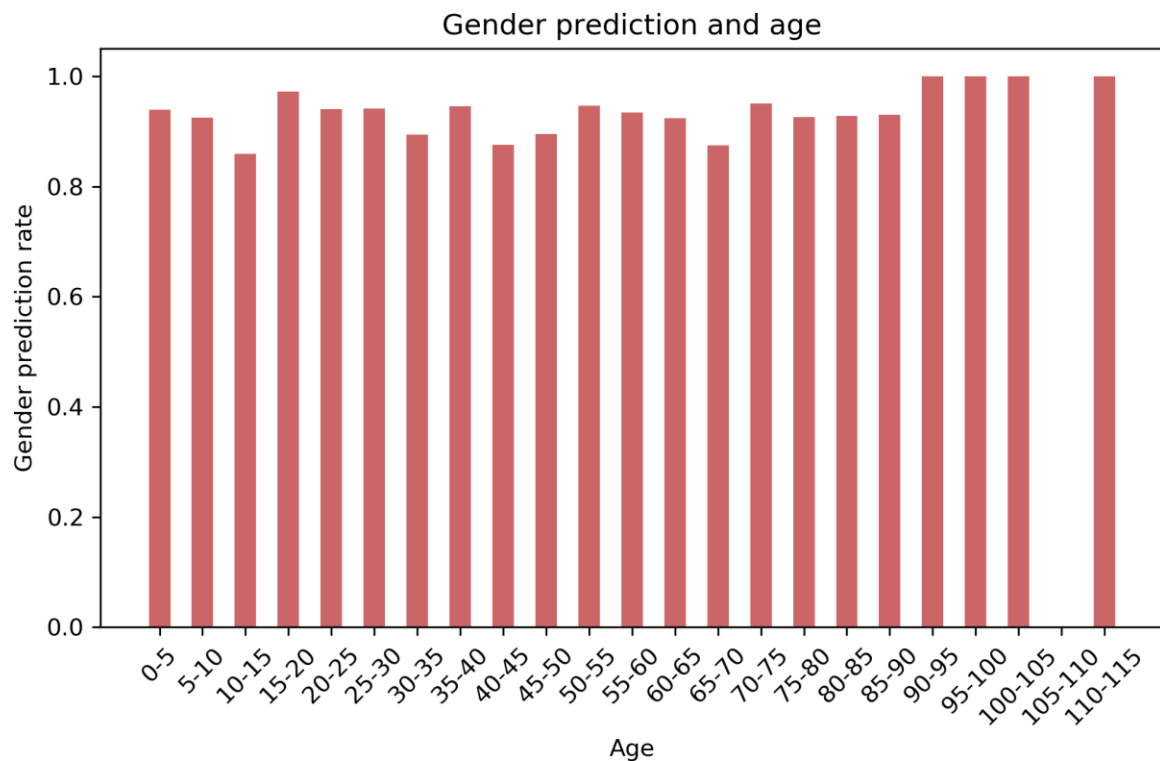


	White	Black	Asian	Indian	Others	Recall
White	1855	12	51	80	21	92%
Black	30	754	18	51	18	87%
Asian	33	95	544	15	3	79%
Indian	92	68	48	611	3	74%
Others	116	56	42	51	74	22%
Precision	87%	77%	77%	76%	62%	81%

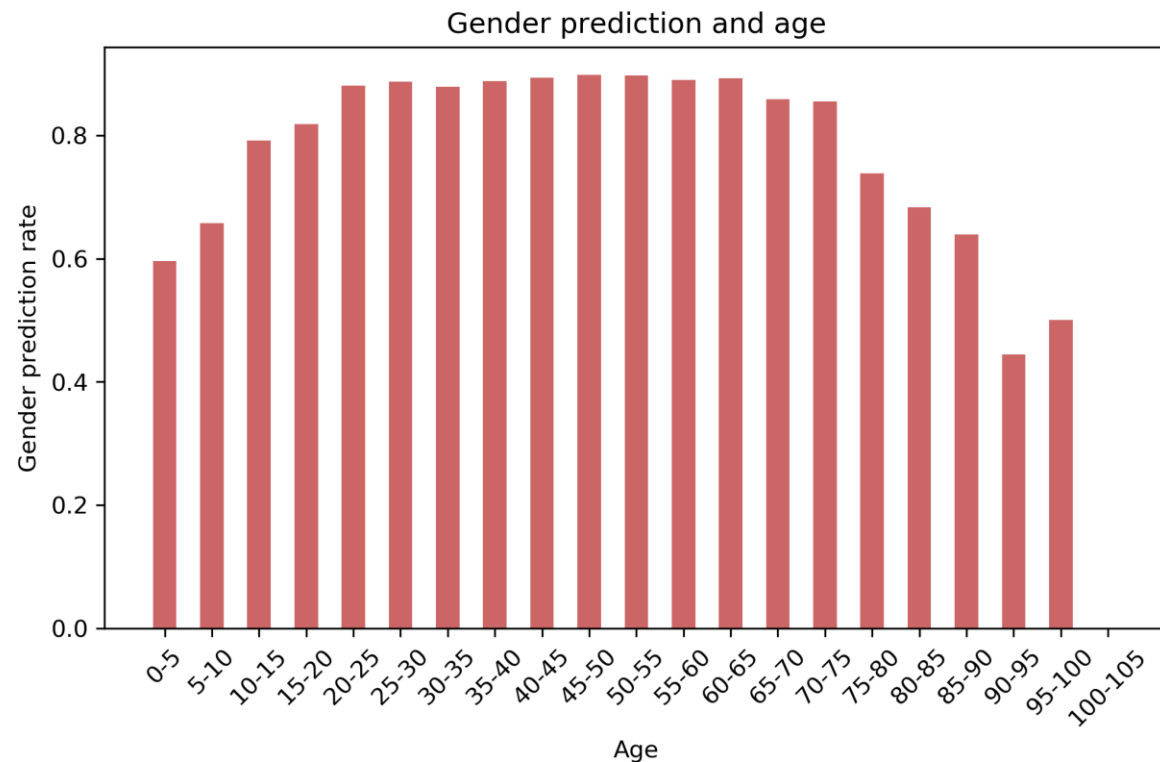
xgBoost

	White	Black	Asian	Indian	Others	Recall
White	1761	94	59	96	12	87%
Black	186	638	25	60	3	70%
Asian	238	31	380	38	5	55%
Indian	257	62	27	424	5	55%
Others	216	30	19	52	23	7%
Precision	66%	75%	75%	63%	48%	68%

Convolutional Neural Network



xgBoost



Summary		CNN	xgBoost
Accuracy	Ethnicity	81%	68%
	Gender	93%	84%
	Age (Mean Abs Error)	8,6	10,6
Model complexity	Size/# adj. Parameters	300MB/93,000	1,5MB/51,000
	Training time (epochs @CPU16GB)	2400 sec.	240 sec.

How to improve...	Challenges	Possible solutions
Unbalanced data	Non-uniform distribution (holes)	More data by import More data by Image Augmentation
Processing Power	CPU: > 40 min. pr Epoch	Access to Cloud GPU (Google Cloud Computing Services)
Model accuracy	Low prediction accuracy (specially Age)	Image Augmentation Transfer Learning

END OF SLIDE SHOW!

Confusion Matrix

Convolutional
Neural Network
grayscale images

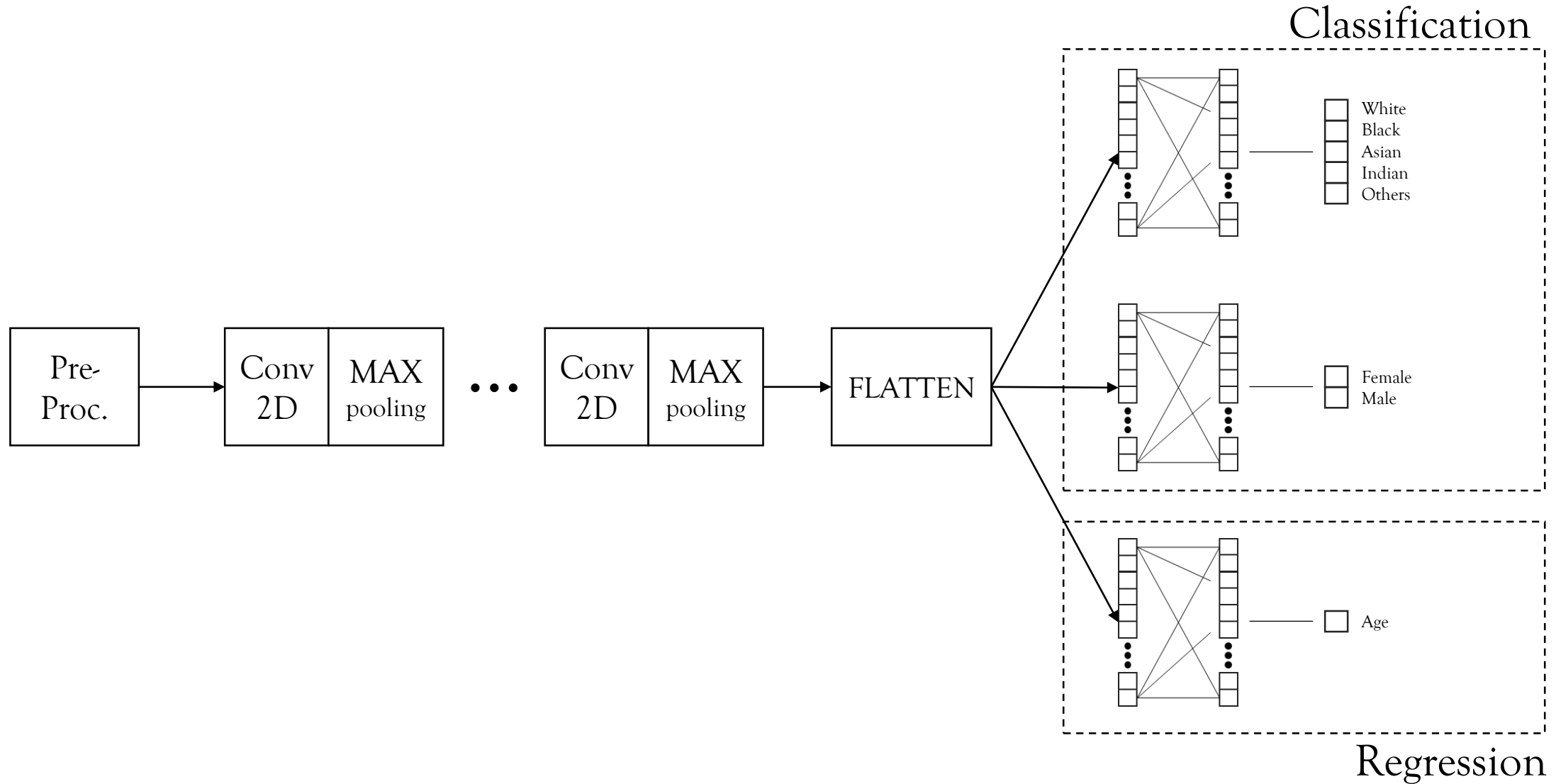


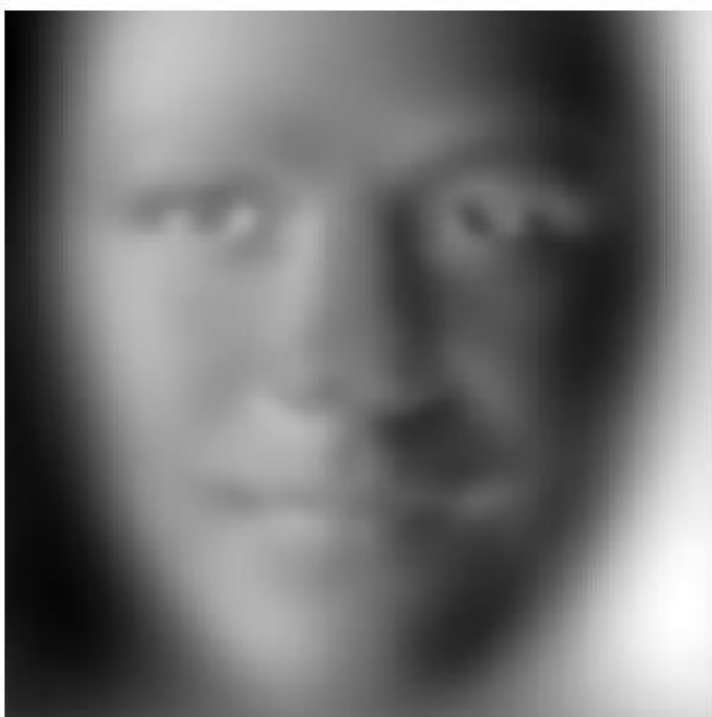
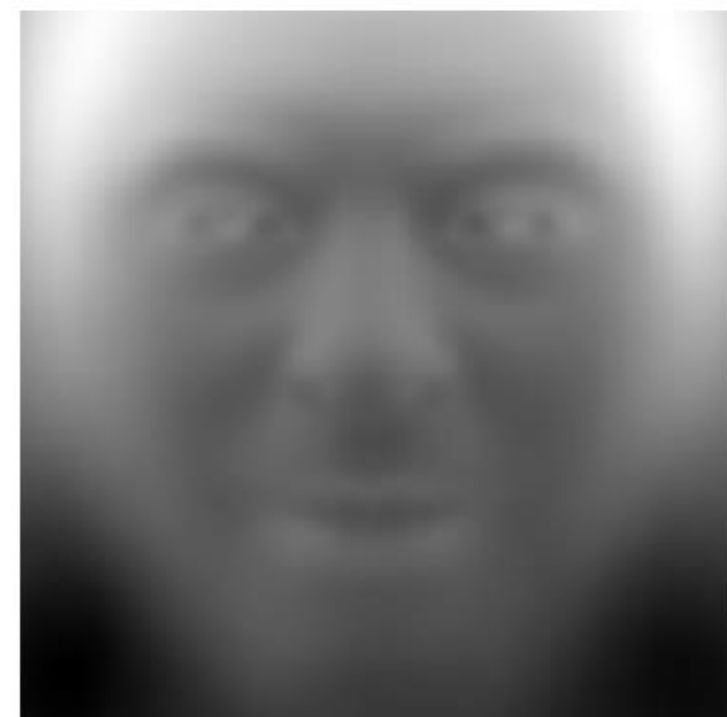
Convolutional
Neural Network
color images

	White	Black	Asian	Indian	Others	Recall
White	1843	20	86	37	0	93%
Black	158	641	64	62	0	69%
Asian	52	2	611	10	0	91%
Indian	241	21	43	500	0	62%
Others	248	10	57	33	2	1%
Precision	73%	92%	71%	78%	100%	76%

	White	Black	Asian	Indian	Others	Recall
White	1855	12	51	80	21	92%
Black	30	754	18	51	18	87%
Asian	33	95	544	15	3	79%
Indian	92	68	48	611	3	74%
Others	116	56	42	51	74	22%
Precision	87%	77%	77%	76%	62%	81%

Method Convolutional Neural Network





Method XGBoost

