Longitudinal Evaluation of Child Face Recognition

Surendra Singh (Clarkson University), Stephanie Schuckers (UNC Charlotte)

Problem

Challenge in child face recognition due to non-linear cranial growth

Deep Neural Network (DNN) models for adults may not always be applicable to children.

Objective

Analyze DNN performance on the YFA (young face aging) dataset (age up to 8 years).

- Study specific changes in face features, e.g., nose, mouth and eyes.
- Identify unique physiological factors contributing to children's facial development.
- Enhance accuracy and effectiveness of face recognition (FR) systems for children.

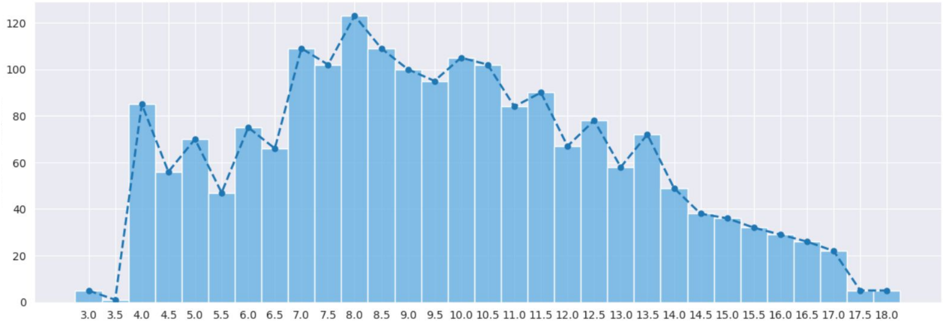
Value

- Understanding **Challenges** in Child Face Recognition.
- **Benchmarking** FR Performance Based on Growth in Children.
- Discovering Specific Changes in Facial Features Impacting Matching Performance.

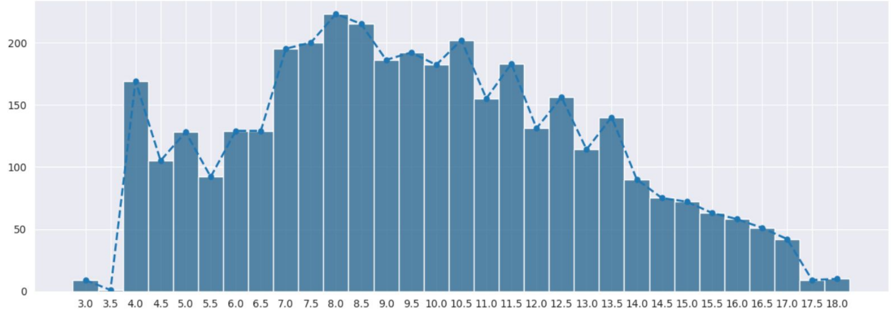
YFA Database (Young Face Aging)

The Young Face Aging (YFA) Database

- Contains face images of children aged 3-18 years.
- 330 subjects, with an average of six collections per subject over eight years.
- Images of the same subject were collected every six months for eight years.
- The first collection image was used for enrollment and verified against each subsequent collection over the eight-year period.
- The database includes 60 subjects with a total of 1,322 samples collected over eight years.
- Collected in a controlled environment with consistent indoor lighting, neutral expressions, and minimized pose variations.
- Manual annotation to exclude extremely blurry images and challenging poses.



Number of subjects at each age in the YFA Database.



Number of images at each age in the YFA Database.

Prior work

Database	Longest time gap	Time interval	Accuracy	Model
ECLF[11]	3 years	6 months	TAR at 0.1% FAR	FaceNet: 84.55 PFE: 98.90 ArcFace: 99.38 COTS: **99.62**
ITWCC-D1[12]			TAR at 0.1% FAR	FR Model: COTS FR-A: 0.676 FR-B: 0.598 FR-C: 0.463 FR-D: 0.434 FR-E: 0.759 FR-F: 0.738 FR-G: 0.718 FR-H: 0.695
NITL[13]	2 years	1 year	TAR at 0.1% FAR	COTS: **60.94**
CLF[14]	3 years	3 year	TAR at 0.1% FAR	COTS: 49.33 FaceNet: **59.80**
CMBD[15]			Rank-1 Accuracy	PCA: 38.8 LBP : 28.8 LDA : 71.3 Fine-tuned VGG-Face: 83.0 Triplet CNN : 72.7 Proposed CNN: **85.1**
YFA	8 years	6 months	TAR at 0.1% FAR	MagFace: **95.48**

Prior work on YFA database

Face Recognition In Children: A Longitudinal Study [1]

TAR @ 0.1% FAR

Model	**ΔT = 6M**	**ΔT = 12M**	**ΔT = 18M**	**ΔT = 24M**	**ΔT = 30M**	**ΔT = 36M**
Facenet-V1	95.8	94.8	92.5	84.3	82.7	76.0
ArcFace	87.6	88.1	85.3	84.8	86.3	81.1
ArcFace-Focal	97.6	98.3	95.4	92.7	93.1	91.6
MagFace	**98.2**	**98.3**	**98.0**	**97.2**	**97.3**	**94.9**

In prior work, we evaluated multiple open-source DNN-based face recognition models and found that MagFace performed the best. Therefore, we selected MagFace for our further analysis.

MagFace[2] : Training database: MS1M-V2 [3] (5.8M images, 85k identities)
Evaluation database: LFW [4], CFPFP [5], AgeDB-30 [6], CALFW [7], CPLFW [8], IJBB [9] and IJB-C [10]

Experimental Setup and Overall Results

- Use of MTCNN for accurate face detection and alignment.
- Feature extraction using MagFace [12]. Input image size 112x112.

Model	TAR @0.1% FAR	Threshold	TAR @0.01% FAR	Threshold
MagFace	95.48	0.45	82.25	0.56

Gender-Based TAR Performance @ 0.1% FAR

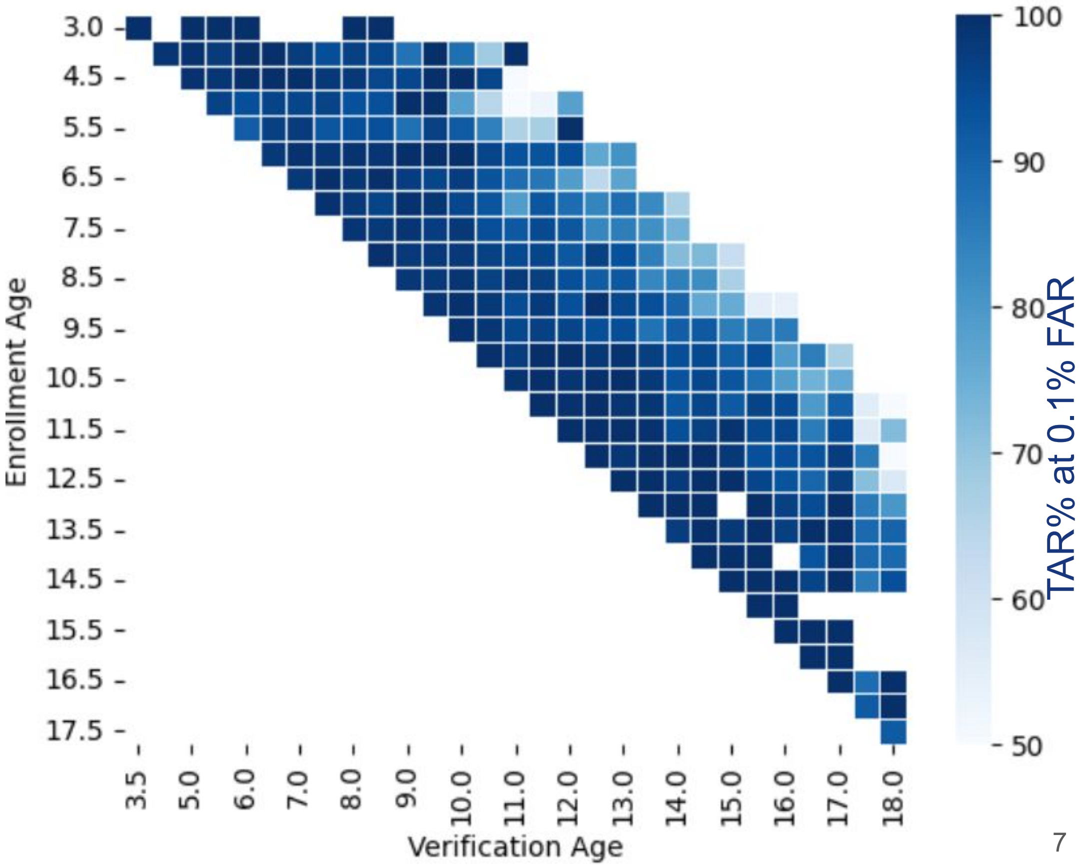
Gender	TAR	Threshold
Male	94.87	0.45
Female	95.80	0.45

Methodology

Categorization of images into enrollment and verification samples based on age increments of 6 months

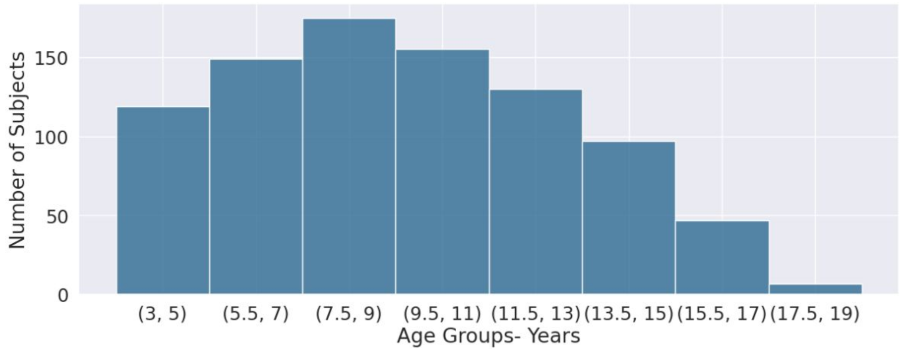
Enrollment Stage: Any image in the dataset can be an enrollment image, categorized by age brackets.

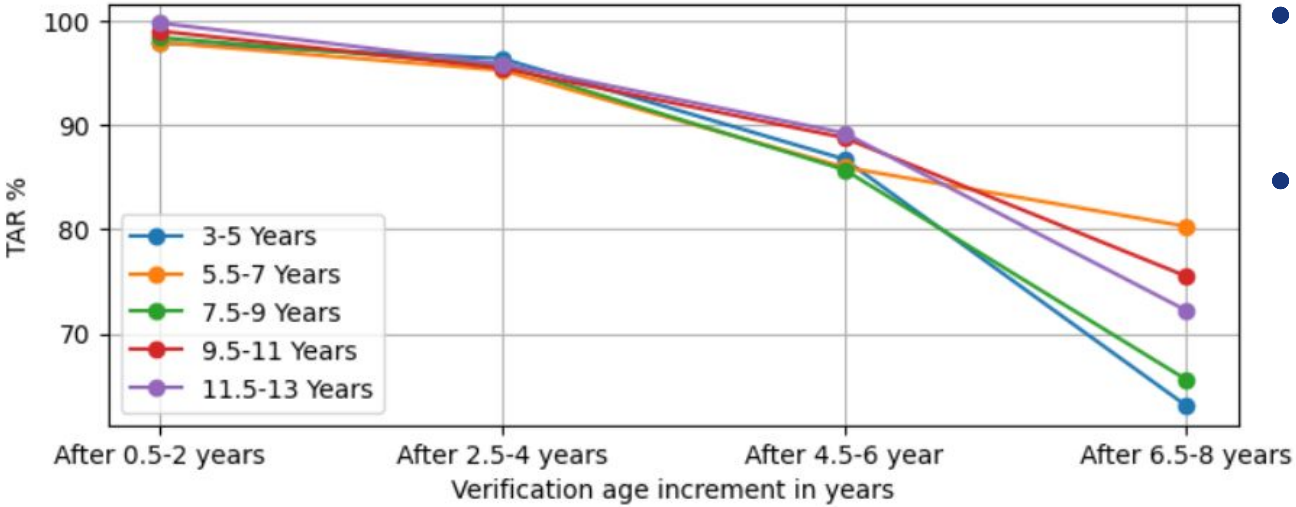
Verification Stage: Subsequent collections used for verification, evaluating performance over time.



Analysis for age groups

- Investigation of age-related trends in face recognition performance.
- Analyzing the impact of enrollment age on True Acceptance Rate (TAR) over time.





- Recognition accuracy declines significantly beyond a 4-year age gap.
- The (3-5) and (7.5-9) age groups experience the largest drop in TAR over time.

Bootstrapping for performance evaluation

Bootstrap

- Statistical technique for estimating sampling distributions.
- Resamples data to approximate uncertainty without assuming distribution.
- Provides robust estimates of parameters (e.g., confidence intervals).

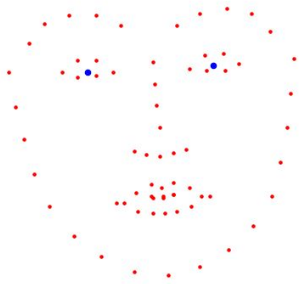
To analyze TAR variability across different age groups and verification periods.

- Utilized bootstrap resampling for robust estimation.
- Significant TAR differences observed across age groups

	TAR% at 0.1% FAR for different time gaps between enrollment and verification							
Enrollment Age	0.5-2 Years	N	2.5-4 Years	N	4.5-6 Years	N	6.5-8 Years	N
3-5 Years	97.8 [96.8, 98.7]	88	96.3 [95.1, 97.5]	65	86.7 [84.4, 88.8]	55	63.1 [60.2, 66.1]	20
5.5-7 Years	97.9 [97.0, 98.7]	133	95.2 [93.8, 96.5]	82	85.9 [83.6, 87.9]	63	80.2 [77.8, 82.5]	43
7.5-9 Years	98.3 [97.6, 99.0]	161	95.7 [94.4, 96.9]	99	85.7 [83.6, 88.1]	88	65.6 [62.5, 68.4]	51
9.5-11 Years	98.9 [98.3, 99.5]	141	95.4 [94.1, 96.6]	112	88.7 [86.4, 90.5]	89	75.5 [72.9, 77.9]	39
11.5-13 Years	99.7 [99.4, 100.0]	114	95.8 [94.6, 97.0]	65	89.2 [87.2, 91.0]	41	72.2 [69.4, 75.1]	4

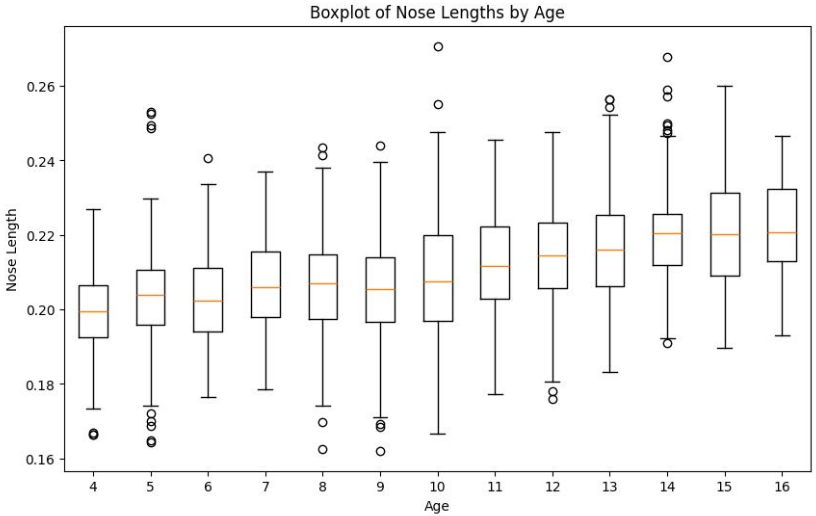
Facial Feature Growth Analysis

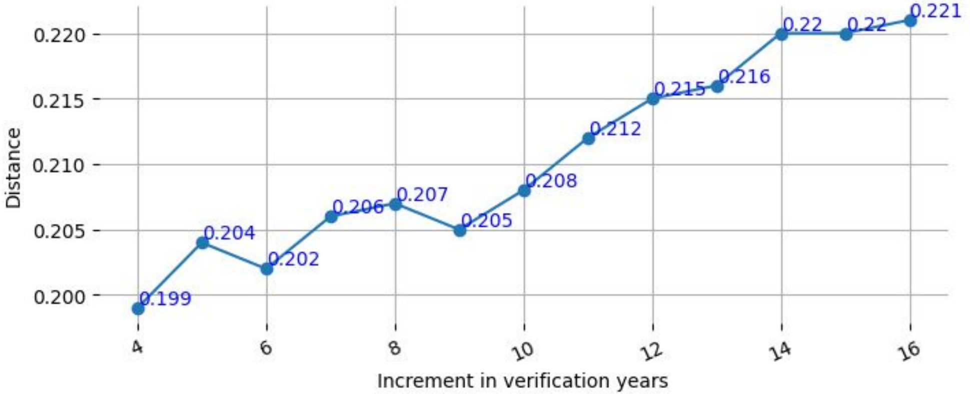
- Measured growth in facial distances (e.g., nose length, chin size, mouth width).
- Normalized distances using inter-eye distance.
- Significant growth observed in features between ages 4-16, impacting recognition accuracy.



Face Features - Relative nose length to distance between the eyes

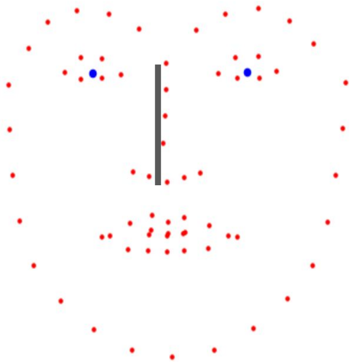
As children age, the relative distance from the nose to the eyes increases. We are continuing to analyze other facial features, such as the mouth and chin.





Distance calculation is by median.

Distance equation: $\dfrac{\text{Distance between feature points (in pixel)}}{\text{Distance between middle point of eyes}}$



Face Cropping & Recognition Accuracy

- Evaluated multiple face cropping algorithms: MTCNN, RetinaFace, OpenCV, DLIB, etc.
- MTCNN and RetinaFace provided the highest recognition accuracy for children.
- OpenCV showed strong performance for adults but lower accuracy for children.

Face cropping with difference face cropping a: Dlib, b: Mediapipe, C: MTCNN, D: OpenCV, e: Retinaface, f: SSD

	Adult		Child	
Face Cropping	TAR @ 0.1 FAR	TAR @ 0.01 FAR	TAR @ 0.1 FAR	TAR @ 0.01 FAR
DLIB	96.9	91.31	25.15	15.31
Mediapipe	56.04	38.76	11.94	7.67
MTCNN	98.58	98.34	**95.48**	82.25
Open CV	99.12	99.07	94.89	77.33
Retinaface	98.38	93.64	95.25	82.65
SSD	**99.26**	89.90	7.28	5.15

Conclusions and Future work

Key Findings:

- Average TAR:
  - (0.5-2) years: 98.52%
  - (2.5-4) years: 95.68%
  - Significant decline post 4-year age difference.
- Notable TAR declines observed in younger age groups (e.g., 3-5 years: 63.1%)

Conclusion:

- Accuracy fluctuations highlight the complexity of age-related biometric performance and the need for adaptive recognition models.
- Study limitations include uneven age distribution, lighting inconsistencies, demographic biases, and reliance on a single face-matching algorithm (MagFace).
- Findings contribute to understanding child facial recognition and its implications for applications like missing child identification.

Future Work

- Expand dataset diversity to improve generalizability across age groups and ethnic backgrounds.
- Investigate deep learning-based models to adaptively account for age progression in child biometrics.
- Assess performance of different face-matching algorithms and explore fine-tuning approaches.
- Conduct real-world evaluations to address operational challenges in uncontrolled environments.

Reference

[1] Bahmani, K., & Schuckers, S. (2022, April). Face recognition in children: a longitudinal study. In *2022 International Workshop on Biometrics and Forensics (IWBF)* (pp. 1-6). IEEE.

[2] Meng, Q., Zhao, S., Huang, Z., & Zhou, F. (2021). Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14225-14234).

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4690–4699, 2019. 1, 2, 3, 4, 5, 6, 11, 12

[4] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007

[5] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1–9. IEEE, 2016

[6] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. AgeDB: the first manually collected, in-the-wild age database. In IEEE Conference on Computer Vision and Pattern Recognition Workshop, pages 51–59, 2017. 5

[7] Tianyue Zheng, Weihong Deng, and Jiani Hu. Crossage LFW: A database for studying cross-age face recognition in unconstrained environments. arXiv preprint arXiv:1708.08197, 2017.

[8] Tianyue Zheng and Weihong Deng. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. Beijing University of Posts and Telecommunications, Tech. Rep, 5, 2018

[9] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K Jain, James A Duncan, Kristen Allen, et al. IARPA Janus benchmark-B face dataset. In IEEE Conference on Computer Vision and Pattern Recognition Workshop, pages 90–98, 2017

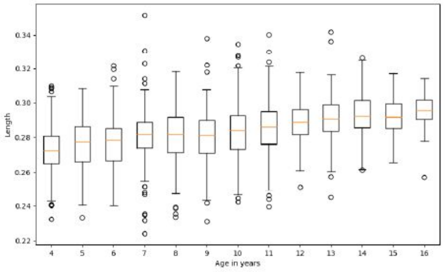
[10 Brianna Maze, Jocelyn Adams, James A Duncan, Nathan Kalka, Tim Miller, Charles Otto, Anil K Jain, W Tyler Niggel, Janet Anderson, Jordan Cheney, et al. IARPA Janus benchmark-C: Face dataset and protocol. In International Conference on Biometrics, pages 158–165. IEEE, 2018.]

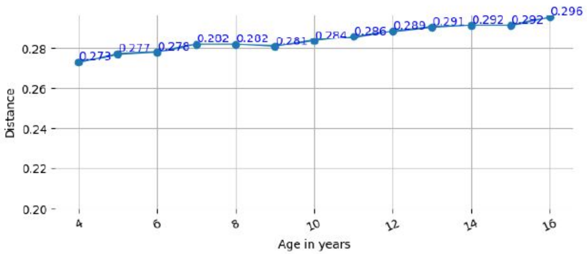
Reference

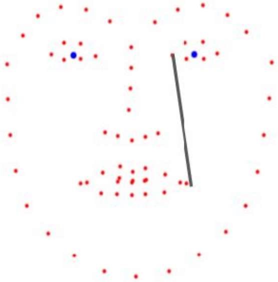
[11] P. K. Chandaliya, Z. Akhtar, and N. Nain, "Longitudinal analysis of mask and no-mask on child face recognition," in Proceedings of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing, 2022, pp. 1–7.
[12] N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme, and M. King, "Face recognition algorithm bias: Performance differences on images of children and adults," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2019, pp. 0–0.
[13] L. Best-Rowden, Y. Hoole, and A. Jain, "Automatic face recognition of newborns, infants, and toddlers: A longitudinal evaluation," in 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE, 2016, pp. 1–8.
[14] D. Deb, N. Nain, and A. K. Jain, "Longitudinal study of child face recognition," in 2018 International Conference on Biometrics (ICB). IEEE, 2018, pp. 225–232.
[15] S. Siddiqui, M. Vatsa, and R. Singh, "Face recognition for newborns, toddlers, and pre-school children: a deep learning approach," in 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 3156–3161.
[16] K. Bahmani and S. Schuckers. Face recognition in children: A longitudinal study. In 2022 International Workshop on Biometrics and Forensics (IWBF), pages 1–6. IEEE, 2022.

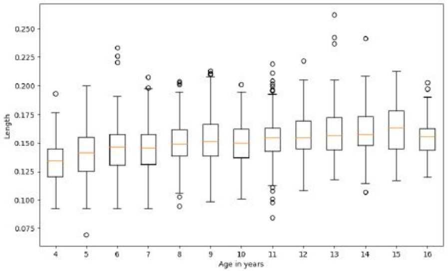
Backup Slide

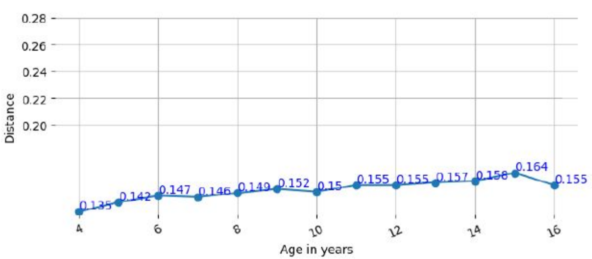
Face features

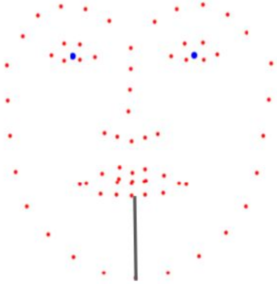


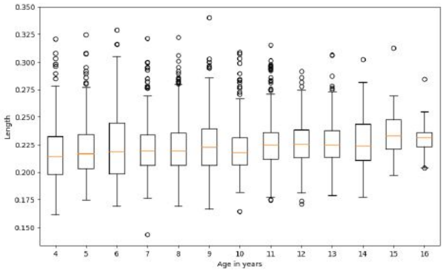


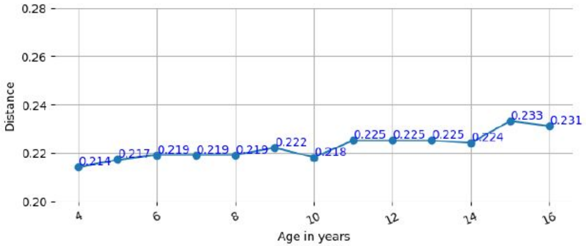


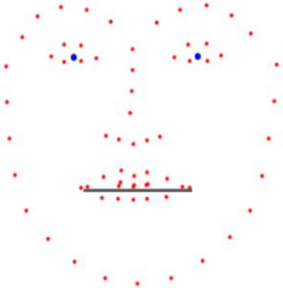












Face features

Longitudinal Growth Measurements of Facial Features Between Ages 4 and 6, Highlighting Changes in Nose Length, Mouth Dimensions, Chin Size, and Horizontal Mouth Length.

Feature	Age 4	Age 16	Distance Difference
Nose length	0.199	0.221	0.022
Distance betwenn left corner of mouth to left corner of mouth	0.272	0.298	0.026
Distance betwenn right corner of mouth to right corner of mouth	0.273	0.296	0.023
Chin size	0.135	0.155	0.020
Horizontal mouth length	0.214	0.231	0.017

Male: Longitudinal Growth Measurements of Facial Features

Feature	Age 4	Age 16	Distance Difference
Nose length	0.199	0.218	0.020
Distance betwenn left corner of mouth to left corner of mouth	0.271	0.302	0.031
Distance betwenn right corner of mouth to right corner of mouth	0.273	0.298	0.025
Chin size	0.131	0.157	0.026
Horizontal mouth length	0.209	0.230	0.021

Female: Longitudinal Growth Measurements of Facial Features

Feature	Age 4	Age 16	Distance Difference
Nose length	0.198	0.227	0.029
Distance betwenn left corner of mouth to left corner of mouth	0.272	0.289	0.017
Distance betwenn right corner of mouth to right corner of mouth	0.271	0.294	0.023
Chin size	0.136	0.147	0.011
Horizontal mouth length	0.214	0.232	0.018

Experimental Setup and Overall Results

- Use of MTCNN for accurate face detection and alignment.
- Cropped faces resized to 224x224 pixels for consistency in analysis.
- Feature extraction using MagFace [12]. Input image size 112x112.

Model	TAR @0.1% FAR	Threshold	TAR @0.01% FAR	Threshold
MagFace	95.48	0.45	82.25	0.56

Age-Based TAR Performance @ 0.1% FAR

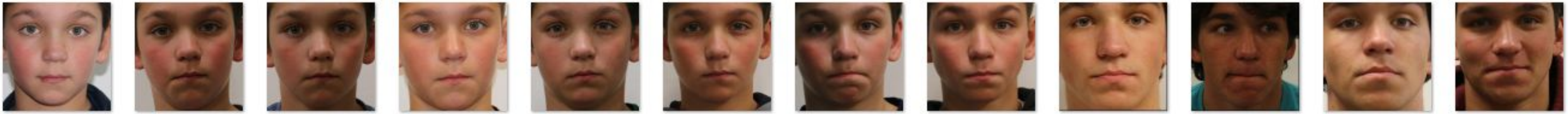
Age Gap	TAR	Subjects
2 years	98.52%	323
4 years	95.68%	199
6 years	87.24%	146
8 years	71.32%	126

Gender-Based TAR Performance @ 0.1% FAR

Gender	TAR	Threshold
Male	94.87	0.45
Female	95.80	0.45

DLIB

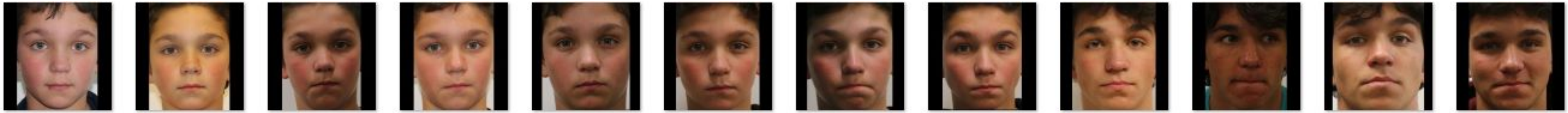
Mediapipe



MTCNN

OpenCV

Retinaface



SSD