Comparative Statistical Analysis of Pediatric Face Recognition Performance

Synthesizing Evidence from Two Longitudinal Studies Using Advanced Statistical Methods for Bounded Data

IA650: Data Mining

Summer 2025

Aaron W. Storey

Study Overview & Context

CLF Study

Children Longitudinal Face

- 919 subjects (age 2-18)
- 3,682 face images
- 6-year study period
- Annual collection intervals
- 3 algorithms tested:
  - COTS-A
  - FaceNet
  - Fusion approach

YFA Study

Young Face Aging

- 330 subjects (age 3-18)
- 3,831 face images
- 8-year study period
- 6-month collection intervals
- MagFace algorithm
- Extended temporal coverage

Research Objective

Synthesize findings through advanced statistical methods to establish robust evidence for system performance

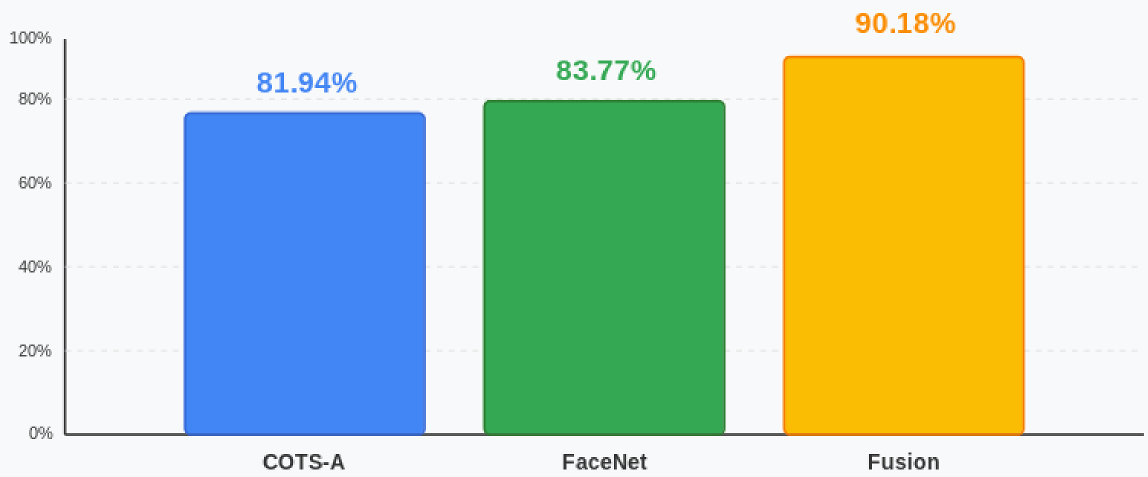
Dataset Characteristics

Characteristic	CLF Dataset	YFA Dataset
Total Subjects	**919**	**330**
Total Images	**3,682**	**3,831**
Age Range	2-18 years	3-18 years
Maximum Time Span	6 years	8 years
Collection Interval	Annual	6-month
Gender Distribution	66% Male, 34% Female	Not specified
Recognition Algorithms	COTS-A, FaceNet, Fusion	MagFace

Complementary Datasets

CLF provides multi-algorithm comparison with longer individual follow-up
YFA offers extended temporal coverage with regular collection intervals

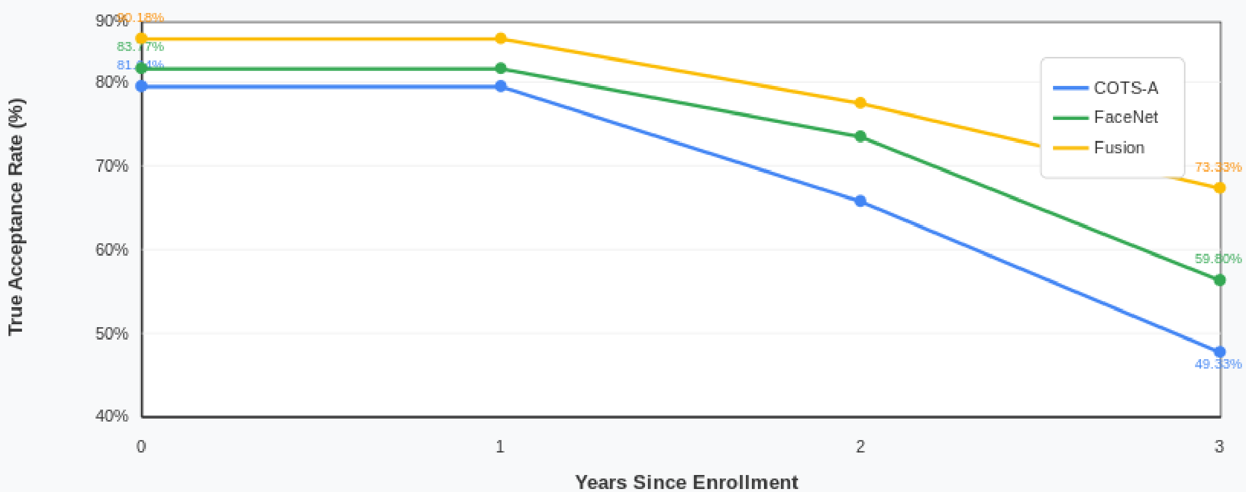
CLF Study – Initial Performance Metrics



Key Finding

Fusion approach achieves highest accuracy at 90.18% TAR
All systems demonstrate operational viability for short-term applications

CLF Study – Temporal Degradation Patterns

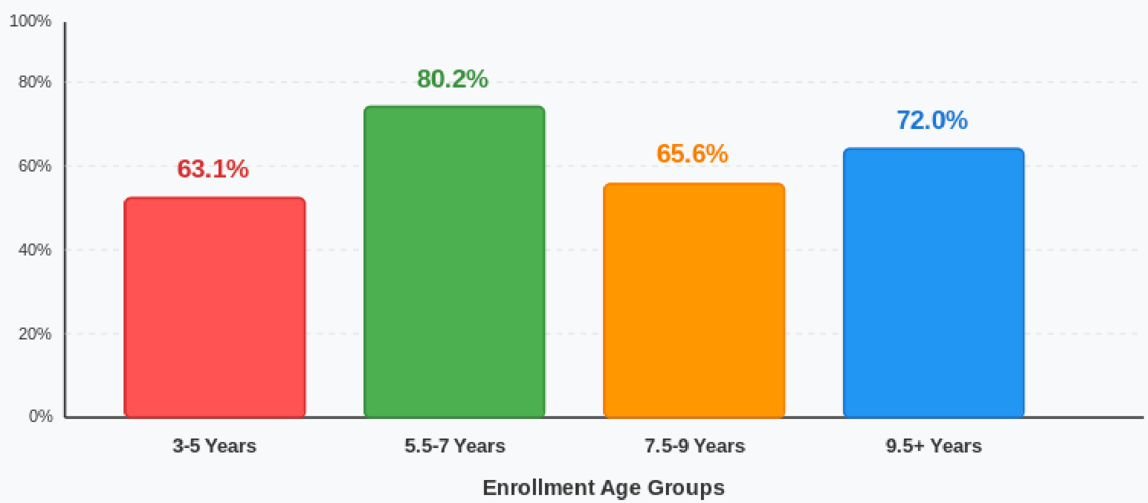


Critical Findings

Substantial degradation across all systems: 17-33 percentage point decline over 3 years

Genuine scores decrease by 0.22-0.24 standard deviations per year

YFA Study – Extended Timeline Results



Critical Age-Related Patterns

Early childhood (3-5 years) shows most severe degradation: 37% accuracy loss
Mid-childhood (5.5-7 years) demonstrates relative stability - optimal enrollment window

YFA Study – Age-Stratified Performance

Current Gaps

- No integrated analysis across datasets
- Limited statistical methods for bounded data
- Inconsistent age stratification approaches
- Lack of operational guidelines
- Algorithm-specific findings not generalized
- Boundary effects underestimated
- No synthesis of temporal patterns

Research Opportunities

- Synthesize 7,500+ images across studies
- Apply beta regression for accurate modeling
- Establish age-specific protocols
- Compare multiple algorithm approaches
- Develop robust confidence intervals
- Create operational decision framework
- Validate patterns across datasets

Value Proposition

Complementary datasets + Advanced statistical methods = Robust operational guidance
First comprehensive framework for bounded longitudinal biometric data

Research Gaps and Opportunities

Primary Hypothesis

Algorithm selection and data collection intervals significantly influence accuracy degradation rates, with younger age groups experiencing disproportionately higher performance decline

Five Key Research Questions

RQ1: Algorithm Comparison

How do accuracy degradation rates differ across COTS-A, FaceNet, Fusion, and MagFace algorithms?

RQ2: Age Vulnerability

Which age groups show highest vulnerability to performance degradation over time?

RQ3: Temporal Modeling

Can beta regression provide more accurate degradation estimates than standard methods?

RQ4: Collection Impact

How do 6-month vs annual collection intervals affect observed degradation patterns?

RQ5: Operational Thresholds

What re-enrollment intervals maintain 85% accuracy for different age groups?

Statistical Approach

- Beta regression for bounded outcomes (TAR $\in$ [0,1])
- Hierarchical modeling for subject and age-group effects
- Bootstrap methods (10,000 replications) for robust inference

Proposed Statistical Methodology

Beta Regression Framework

Why Beta Regression?

- TAR bounded between 0 and 1
- Natural accommodation of proportion data
- No arbitrary transformations needed
- Captures heteroscedastic variance structure
- Accurate inference near boundaries

Hierarchical Structure

Three-Level Model:

Level 1: Observations ~ Beta(μ, φ)
Level 2: Subject-specific effects
Level 3: Age-group effects

Accounts for:

- Within-subject correlation
- Between-age-group heterogeneity

Bootstrap Inference

Parametric Bootstrap (10,000 replications)

- Robust confidence intervals
- No reliance on asymptotic approximations
- Essential for boundary proximity

Cross-Dataset Synthesis

Meta-Analytic Approach

- Harmonization to common metrics
- Inverse-variance weighting
- DerSimonian-Laird heterogeneity estimation

Methodological Advantages

✓ Addresses fundamental limitations of standard approaches for bounded data
✓ Provides reliable uncertainty quantification essential for operational decisions
✓ Framework extends beyond face recognition to any bounded longitudinal data

Expected Outcomes & Applications

Age-Specific Re-enrollment Intervals

(Expected to maintain 85% system accuracy)

- Early Childhood (3-5 yrs) — 2.1 years
- Middle Childhood (6-9 yrs) — 3.8 years
- Adolescent (10-18 yrs) — 4.2 years

Missing Child Identification

Expected Capabilities:

- Recent cases (<2 years): High confidence
- Medium-term (2-4 years): Accuracy/efficiency trade-offs quantified
- Extended gaps (>4 years): Realistic expectations established

Methodological Contributions

- First comprehensive framework for bounded longitudinal biometric data
- Template for future studies involving proportion data near boundaries
- Validation of beta regression advantages over traditional methods

Operational Impact

Evidence-based protocols will directly inform deployment strategies for:

Border security • Healthcare systems • Missing children databases
School security • Child welfare programs • International humanitarian efforts

Expected Contributions & Research Significance

Expected Research Contributions

Statistical Foundation
First rigorous framework for bounded biometric data

Operational Guidance
Evidence-based re-enrollment protocols by age

Comparative Insights
Cross-algorithm performance patterns identified

Methodological Advance
Beta regression validated for biometric applications

Broader Research Significance

This research addresses critical gaps in pediatric biometric system evaluation by:

- Establishing the first comprehensive statistical framework for bounded longitudinal data
- Providing actionable guidance for real-world deployment in child safety applications
- Creating a replicable methodology extending beyond face recognition to other biometric modalities

As pediatric biometric systems expand into critical applications,
rigorous evaluation becomes essential for protecting vulnerable populations
This research provides the foundation for responsible technology deployment

References

Primary Dataset Studies

Deb, D., Nain, N., & Jain, A. K. (2018). Longitudinal study of child face recognition. In *2018 International Conference on Biometrics (ICB)* (pp. 225-232). IEEE.

Singh, S., Bahmani, K., & Schuckers, S. (2024). Longitudinal evaluation of child face recognition and the impact of underlying age. *International Workshop on Biometrics and Forensics (IWBF)*.

Supporting Child Face Recognition Studies

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

Best-Rowden, L., Hoole, Y., & Jain, A. K. (2016). Automatic face recognition of newborns, infants, and toddlers: A longitudinal evaluation. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*(pp. 1-8). IEEE.

Ricanek Jr, K., Bhardwaj, S., & Sodomsky, M. (2015). A review of face recognition against longitudinal child faces. In*BIOSIG*(pp. 15-26).

Srinivas, N., Ricanek, K., Michalski, D., Bolme, D. S., & King, M. (2019). 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*.

Statistical Methodology References

Best-Rowden, L., & Jain, A. K. (2017). Longitudinal study of automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(1), 148-162.

Goldstein, H. (2010). *Multilevel statistical models* (4th ed.). John Wiley & Sons.

Yoon, S., & Jain, A. K. (2015). Longitudinal study of fingerprint recognition. *Proceedings of the National Academy of Sciences*, 112(28), 8555-8560.