# Rigr AI Facial Age Estimation — Model Performance (v3)

> **CONFIDENTIAL** — This document is shared under NDA. Do not distribute.

## Model Overview

The Rigr AI Facial Age Estimation model is a deep learning system that predicts the age of individuals from facial images. It combines a face detector with a vision transformer-based age estimator. The model produces a point estimate (predicted age) along with a calibrated uncertainty value for each detected face.

The current production model (v3) was developed through extensive architectural search, hyperparameter optimization, and dataset curation, achieving significant improvements over prior generations.

---

## Headline Results

Evaluated on our held-out internal benchmark test set:

| Metric | Value |
|---|---|
| **Overall MAE** | **1.64 years** |
| **Weighted MAE** | **0.93 years** |
| **Ages 0–18 MAE** | **0.87 years** |
| **Ages 13–18 MAE** | **1.26 years** |
| **Ages 18+ MAE** | 3.48 years |
| **Predictions within ±0.5 years** | 35.8% |
| **Predictions within ±1.0 years** | 68.9% |
| **Predictions within ±2.0 years** | 82.1% |
| **$R^2$ Score** | 0.961 |

*Weighted MAE applies higher weight to age ranges where accuracy is most critical (younger demographics), tapering for older age groups.*

### Performance by Age Range

| Age Range | MAE | Within ±1 year | Within ±2 years |
|---|---|---|---|
| **0–2** | **0.41** | 95.3% | 97.7% |
| **2–5** | **0.72** | 88.3% | 97.0% |
| **5–8** | **0.77** | 86.1% | 96.4% |
| **8–13** | **1.06** | 73.6% | 93.1% |
| **13–18** | **1.26** | 69.9% | 85.0% |
| **0–18 (all minors)** | **0.87** | 81.9% | 94.1% |
| 18–21 | 1.86 | 60.2% | 76.6% |
| 21–26 | 2.28 | 49.5% | 68.4% |
| 26–31 | 2.41 | 46.9% | 64.2% |
| 31–40 | 3.90 | 30.6% | 46.2% |
| 40–50 | 4.89 | 22.1% | 33.8% |
| 50–60 | 5.16 | 24.6% | 34.9% |
| 60–75 | 5.12 | 20.8% | 34.0% |

---

## Per-Age Accuracy (Ages 0–17)

Independent evaluation against leading commercial competitors, measured as MAE per true age:

| True Age | Rigr.AI MAE | Leading Competitors MAE |
|---|---|---|
| 0 | **0.5** | 4.3 |
| 1 | **0.8** | 1.8 |
| 2 | **0.8** | 2.5 |
| 3 | **1.1** | 2.1 |

| True Age | Rigr.AI MAE | Leading Competitors MAE |
|---|---|---|
| 4 | **1.1** | 2.0 |
| 5 | **1.1** | 3.0 |
| 6 | **1.2** | 1.8 |
| 7 | **1.3** | 3.0 |
| 8 | **1.4** | 4.1 |
| 9 | **1.3** | 2.5 |
| 10 | **1.2** | 3.4 |
| 11 | **1.3** | 2.8 |
| 12 | **1.1** | 2.6 |
| 13 | **1.1** | 3.0 |
| 14 | **1.2** | 2.6 |
| 15 | **1.1** | 3.1 |
| 16 | **1.4** | 3.0 |
| 17 | **1.6** | 2.6 |

Rigr.AI outperforms leading competitors at every age from 0 to 16, with particularly strong advantages for younger children (ages 0–5) and the pre-teen/early-teen range (ages 8–15) where competitor error rates are 2–3x higher.

---

## Evaluation Methodology

### Metrics Suite

The model is evaluated against a comprehensive set of metrics spanning regression accuracy, uncertainty calibration, and age-bucketed performance:

**Regression metrics:** - Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), $R^2$ score - Accuracy-within-threshold at ±0.5, ±1.0, and ±2.0 years - Weighted MAE/MSE with configurable per-bucket importance weights

**Uncertainty calibration metrics:** - Coverage at 1σ, 2σ, and 3σ confidence intervals, measured against theoretical Gaussian expectations (68.3%, 95.5%, 99.7%) - Calibration error (deviation from expected coverage) - Mean absolute z-score (standardized prediction error normalized by predicted uncertainty)

**Age-bucketed metrics:** All regression metrics are computed independently across granular age buckets (e.g., 0–2, 2–5, 5–8, ..., 75+) as well as operationally relevant aggregate ranges (0–13, 13–18, 0–18, 18+). This ensures the model is not merely accurate on average, but performs consistently across the full age spectrum.

### Test Set Composition

The benchmark test set comprises images drawn from multiple independent real-world sources with verified age labels. The test set is strictly disjoint from all training and validation data, with deduplication enforced through both cryptographic (SHA-256), perceptual hashing and manual review to prevent any data leakage across splits.

---

## Uncertainty Estimates

Each prediction includes a calibrated uncertainty value (in years) representing one standard deviation (1σ) of the model's predicted probability distribution over ages. A tightly concentrated distribution yields low uncertainty, while a spread distribution yields high uncertainty.

### Interpreting the ± Value

The `uncertainty` field returned by the API is a **1σ confidence bound**. For a prediction of **age 15.2 ± 1.1**:

| Confidence Level | Range | Meaning |
|---|---|---|
| **±1σ** (±1.1 years) | 14.1 – 16.3 | True age falls in this range ~68% of the time |
| **±2σ** (±2.2 years) | 13.0 – 17.4 | True age falls in this range ~93% of the time |
| **±3σ** (±3.3 years) | 11.9 – 18.5 | True age falls in this range ~98% of the time |

Put another way: **the true age will exceed the ± bound roughly 1 in 3 times**. This is by design — the ± value is a probabilistic confidence interval, not a hard guarantee. For higher confidence, multiply the uncertainty by 2 (for ~93% coverage) or 3 (for ~98% coverage).

### Calibration Validation

We validate that the model's uncertainty estimates are meaningful by checking actual prediction errors against the predicted confidence intervals. A perfectly calibrated model would match the theoretical Gaussian coverage rates:

| Interval | Theoretical (Gaussian) | Actual (v3) | Calibration Error |
|---|---|---|---|
| **±1σ** | 68.27% | **68.48%** | 0.21% |
| **±2σ** | 95.45% | **93.42%** | 2.03% |
| **±3σ** | 99.73% | **98.48%** | 1.25% |

The 1σ calibration is near-perfect — only 0.2 percentage points from the theoretical ideal. At wider intervals (2σ, 3σ), the model is marginally conservative (slightly undercovering by ~2% and ~1.2%), meaning the predicted uncertainty occasionally underestimates the true error at extreme confidence levels.

### What Good vs. Poor Confidence Looks Like

The average predicted uncertainty varies significantly by age range, reflecting the inherent difficulty of estimation at different ages - our faces change as we age, but the rate at which we mature and age becomes more variable as we get older:

| Age Range | Avg. Predicted Uncertainty (σ) | MAE |
|---|---|---|
| 0–2 | 0.5 years | 0.41 |
| 2–5 | 0.9 years | 0.72 |
| 5–8 | 1.1 years | 0.77 |
| 8–13 | 1.4 years | 1.06 |
| 13–18 | 1.5 years | 1.26 |
| 18–21 | 2.2 years | 1.86 |
| 21–26 | 2.9 years | 2.28 |
| 26–31 | 3.4 years | 2.41 |
| 31–40 | 4.3 years | 3.90 |
| 40–50 | 5.5 years | 4.89 |
| 50–60 | 6.1 years | 5.16 |
| 60–75 | 6.5 years | 5.12 |

The model correctly signals its own limitations: where it is less accurate (older ages), it reports higher uncertainty.

---

## Approach Highlights

- **Vision transformer backbone** fine-tuned with parameter-efficient methods, preserving the rich feature representations learned during large-scale pre-training while adapting to the age estimation domain.
- **A composite loss** enabling the model to learn that adjacent ages share visual characteristics, to make accurate predications and to learn well-calibrated uncertainty estimation.
- **Curated training data** a large image set from multiple real-world sources, supplemented with targeted synthetic data to address coverage gaps in underrepresented age ranges. Rigorous deduplication across all sources ensures data quality.
- **Stratified training** with custom batch sampling to ensure balanced age representation during optimization, preventing the model from biasing toward overrepresented age groups.
- **Optimized inference** with models compiled to ONNX format with dynamic batching, supporting both CPU and GPU deployment with sub-second per-image latency.

---

## Deployment

The model is served via a REST API (see [API Reference]) and is available as: - A **cloud-hosted service** at `api.age.rigr.ai` - **Pre-built Docker containers** for on-premise or air-gapped deployment, supporting both CPU and CUDA-enabled GPU inference

---

## Contact

For technical questions, evaluation data, or partnership inquiries:

**Age support**: age@email.rigr.ai