

Effect van beeldresolutie op leeftijdsschatting met DeepFace en InsightFace

Shiyar Jamo

Master Applied Artificial Intelligence
Amsterdam University of Applied Sciences
Amsterdam, Nederland
shiyar.jamo@hva.nl

Abstract—Automatische leeftijdsschatting wordt breed toegepast voor leeftijdsverificatie, waarbij invoerbeelden sterk variëren in resolutie. Dit onderzoek evalueert de invloed van beeldresolutie op de nauwkeurigheid van leeftijdsschatting met DeepFace en InsightFace. Hiervoor zijn 1000 afbeeldingen uit de IMDB-Clean dataset in zeven resoluties (64×64 tot 1080×1080 pixels) verwerkt, in totaal 7000 testbeelden. De prestaties zijn gemeten met de Gemiddelde Absolute Fout (MAE), de Standaarddeviatie (SD) en de Mediaan Absolute Fout (MedAE).

De resultaten laten zien dat resolutie significant invloed heeft. Beide frameworks presteren het best bij 224×224 pixels, met een MAE van 10,83 jaar (DeepFace) en 7,46 jaar (InsightFace). Bij lagere resoluties (64 × 64) stijgt de MAE met 14.1% (DeepFace) en 16.9% (InsightFace) ten opzichte van de optimale resolutie. Bij zeer hoge resoluties (1080 × 1080) neemt de fout opnieuw toe: met 3.6% bij DeepFace en zelfs 30.8% bij InsightFace.

Index Terms—DeepFace, InsightFace, leeftijdsschatting, beeldresolutie, invoerresolutie, CNN, VGG-Face, ArcFace, leeftijdsvoorspelling, IMDB-Clean

I. INLEIDING

Automatische leeftijdsschatting met AI kan in veel situaties nuttig zijn. Voorbeelden zijn zelfscankassa's en verkoopautomaten voor alcohol of tabak, online leeftijdscontroles bij games en sociale media, en evenementen met minimumleeftijd. Voor dit soort toepassingen bestaan er pretrained modellen zoals DeepFace [1] en InsightFace [2], die leeftijden kunnen voorspellen op basis van gezichtsafbeeldingen. Deze afbeeldingen kunnen verschillende resoluties hebben, van lage resolutie (bijv. 64px) tot hoge resolutie (bijv. 1080px).

Wanneer afbeeldingen worden ingevoerd in AI-frameworks voor leeftijdsvoorspelling, voeren veel systemen zoals DeepFace en InsightFace interne preprocessing uit, zoals het resizen of croppen van afbeeldingen naar een vaste inputgrootte. Zo gebruikt ArcFace (InsightFace) standaard 112×112 pixels [4] en VGG-Face (DeepFace) 224×224 pixels als invoerresolutie [3]. (zie Tabel I).

TABLE I
ARCHITECTUREN VAN VEELGEBRUIKTE
GEZICHTSHERKENNINGSMODELLEN [1]

Model	Input Shape	Embedding Dim.	Parameters	Aantal Layers
FaceNet-128d	160 × 160 × 3	128	22M	447
VGG-Face	224 × 224 × 3	4096	134M	36
ArcFace	112 × 112 × 3	512	34M	162
Dlib	150 × 150 × 3	128	63M	34
OpenFace	96 × 96 × 3	128	3M	166

Deze modellen zijn oorspronkelijk ontwikkeld voor gezichtsherkenning [3], [4], maar worden in frameworks zoals DeepFace hergebruikt als feature extractors voor aanvullende taken, waaronder leeftijdsvoorspelling [1]. Daardoor lijkt het in eerste instantie alsof de oorspronkelijke resolutie van een afbeelding geen effect heeft op de prestaties van de modellen. Toch is het onbekend of de invoerresolutie alsnog invloed heeft op de nauwkeurigheid, ondanks dat de frameworks zelf resizen.

In dit onderzoek wordt DeepFace versie 0.0.95 en InsightFace versie 0.7.3 gebruikt.

Eerder onderzoek laat zien dat resolutie een duidelijke invloed kan hebben op de prestaties van neurale netwerken in het algemeen [5], maar er is nog geen specifiek onderzoek uitgevoerd naar de impact van beeldresolutie op de nauwkeurigheid van leeftijdsvoorspelling binnen de frameworks DeepFace en InsightFace.

In dit onderzoek wordt daarom onderzocht of de resolutie van invoerbeelden nog steeds invloed heeft op de nauwkeurigheid van leeftijdsvoorspelling, ondanks dat DeepFace en InsightFace de beelden intern al resizen naar vaste formaten. Indien resolutie géén rol meer speelt, zouden alle invoerresoluties dezelfde nauwkeurigheid moeten opleveren. Als er wel verschillen zichtbaar zijn, wijst dit erop dat de externe resolutie toch nog effect heeft op de uiteindelijke nauwkeurigheid.

Onderzoeksvraag: Heeft de resolutie van invoerafbeeldingen invloed op de nauwkeurigheid van leeftijdsschatting (MAE, SD en MedAE) met DeepFace en InsightFace, ondanks hun interne preprocessing?

II. LITERATURE REVIEW

A. Inleiding tot leeftijdsschatting

Leeftijdsschatting op basis van gezichtsafbeeldingen is een belangrijk probleem binnen computer vision en patroonherkenning. Het doel is doorgaans om de *chronologische leeftijd* van een persoon te voorspellen op basis van een foto van het gezicht. Daarnaast bestaat er een variant: het schatten van de *apparent age*, oftewel hoe oud iemand lijkt op de foto [6].

Onderzoekers hebben dit probleem benaderd door gebruik te maken van diepe neurale netwerken. Vaak wordt eerst een groot aantal gezichtsafbeeldingen zonder apparent-age labels gebruikt om een algemene gezichtsrepresentatie te leren, waarna de netwerken verder worden afgestemd met een dataset

die wél voorzien is van apparent-age labels [6]. Toch blijft de betrouwbaarheid van leeftijdsvoorspelling sterk afhankelijk van de kwaliteit en resolutie van de inputbeelden [11].

Leeftijdsschatting kan op twee manieren worden benaderd [6], [11]:

- 1) Het voorspellen van de *chronologische leeftijd*, oftewel de werkelijke biologische leeftijd;
- 2) Het voorspellen van de *apparent age*, oftewel hoe oud iemand lijkt op de foto.

In de praktijk wordt vaak *apparent age* gebruikt, omdat dit aansluit bij hoe mensen zelf leeftijd inschatten, en omdat deze labels eenvoudiger te verkrijgen zijn via menselijke beoordelaars [19].

B. Regression vs Classification

De state-of-the-art benaderingen voor leeftijdsschatting zijn grofweg in twee categorieën te verdelen:

- 1) Regression-based modellen: Deze modellen behandelen leeftijd als een continu getal en proberen de exacte leeftijd te voorspellen (bijv. 24 jaar). Het verlies wordt vaak gemeten met metrieken zoals Gemiddelde absolute fout (MAE), wat de gemiddelde absolute afwijking van de voorspelde leeftijd tot de werkelijke leeftijd weergeeft [7], [8].
- 2) Classification-based modellen: Deze modellen behandelen leeftijd als een klasse of interval (bijvoorbeeld 20–25 jaar, 26–30 jaar). Het model voorspelt dus in welke leeftijdsgroep iemand valt [8]. Deze aanpak kan robuuster (handiger) zijn dan regressie, omdat het voor mensen van verschillende leeftijden soms lastig te onderscheiden is wie ouder is. Bijvoorbeeld: een persoon van 28 kan qua uiterlijk sterk lijken op iemand van 30. Door beide leeftijden in dezelfde klasse (26–30 jaar) te plaatsen, wordt die verwarring verminderd.

Frameworks zoals DeepFace [1] en InsightFace [2] implementeren voorgetrainde versies van dergelijke state-of-the-art modellen (zoals ArcFace [4]), waardoor onderzoekers en ontwikkelaars ze direct kunnen gebruiken zonder de complexiteit en rekenkosten van het zelf trainen van een model vanaf nul.

C. Verwerkingspipeline voor leeftijdsschatting

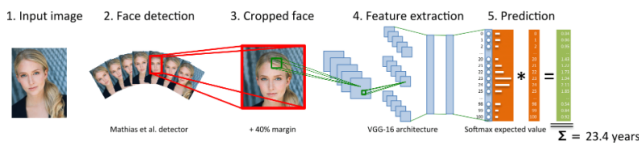


Fig. 1. Verwerkingspipeline voor leeftijdsschatting [11].

De meeste leeftijdsschattingmodellen volgen een vergelijkbare verwerkingspipeline (Fig. 1). Eerst wordt een gezicht gedetecteerd in de invoerafbeelding en vervolgens bijgesneden met een marge rondom het gezicht. Daarna worden de gezichtsbeelden genormaliseerd en door een convolutioneel neurale netwerk (bijvoorbeeld VGG-16 of ResNet) gestuurd om kenmerken (*features*) te extraheren. Op basis van deze

representaties wordt uiteindelijk de leeftijd voorspeld, meestal via een regressie- of classificatielaag.

Een cruciale stap in deze pipeline is het **croppen/resizen** van de invoerbeelden, omdat dit bepaalt welke visuele details behouden blijven voor de feature extractor. Deze stap bepaalt welke visuele details behouden blijven voor de feature extractor en is daarmee sterk afhankelijk van de resolutie van de oorspronkelijke invoerbeelden.

D. Datasets voor leeftijdsschatting

Voor het trainen en evalueren van leeftijdsschattingmodellen zijn grote, gelabelde datasets essentieel. Deze datasets verschillen in omvang, leeftijdsbereik, diversiteit en labelkwaliteit, wat directe invloed heeft op de prestaties van deep learning-modellen.

IMDB-WIKI [9] is met meer dan 500.000 afbeeldingen de grootste publieke dataset voor leeftijds- en geslachtslabels, verzameld van IMDb en Wikipedia. De dataset bevat leeftijden van 0 tot boven de 100 jaar, maar kent een scheve verdeling met een oververtegenwoordiging van jongvolwassenen (15–35 jaar). De labels zijn gebaseerd op de geschatte opnamedatum van de foto, wat ze chronologisch maakt, maar ook ruisgevoelig door foutieve metadata.

Om deze ruis te verminderen is de opgeschoonde variant **IMDB-Clean** ontwikkeld [10]. Deze subset verwijdert foutieve of corrupte afbeeldingen (bijvoorbeeld zwartbeelden, scènes in plaats van gezichten, of verkeerde leeftijdslabels). De dataset bevat daardoor consistentere gelabelde voorbeelden en is betrouwbaarder voor experimenteel onderzoek, zoals dit onderzoek.



Fig. 2. Vergelijking van voorbeelden uit IMDB-WIKI (boven) en de opgeschoonde IMDB-Clean dataset (onder). Foutieve of irrelevante beelden zijn verwijderd, waardoor de kwaliteit verbetert.

UTKFace [21] bevat circa 20.000 afbeeldingen met leeftijden van 0 tot 116 jaar, voorzien van labels voor leeftijd, geslacht en etniciteit. Door de brede spreiding en balans over leeftijdsgroepen is UTKFace veel gebruikt als benchmark in eerder onderzoek. De resolutie van de beelden (200×200 px) is echter relatief laag.

FairFace [12] richt zich specifiek op het verminderen van bias. De dataset omvat meer dan 100.000 afbeeldingen en is gebalanceerd naar etniciteit en geslacht. Leeftijd is hier echter ingedeeld in categorieën (bijvoorbeeld 0–2, 3–9, 10–19, enz.) in plaats van exacte leeftijden, wat de bruikbaarheid voor regressie beperkt.

VGGFace2 [14] bevat 3,31 miljoen hoge-resolutie beelden van 9131 personen, met variatie in pose, leeftijd, belichting en etniciteit. Hoewel de dataset geen expliciete leeftijdslabels heeft, wordt zij vaak gebruikt om robuuste feature-extractors (zoals ResNet en SEnet) te trainen die later ook toegepast worden voor leeftijdsschatting.

Samenvattend: IMDB-WIKI en UTKFace behoren tot de meest gebruikte datasets voor leeftijdsschatting. Omdat IMDB-WIKI echter veel ruis bevat, wordt in dit onderzoek gebruikgemaakt van de opgeschoonde **IMDB-Clean** [10]. Een belangrijk voordeel van IMDB-Clean is dat deze dataset beelden bevat met een hoge resolutie (tot 1024px) en betrouwbare leeftijdslabellen. Hoewel UTKFace ook breed wordt toegepast en een evenwichtige spreiding in leeftijden heeft, bestaat de dataset grotendeels uit beelden van lage resolutie (200×200 px). Daardoor is UTKFace minder geschikt om de invloed van invoerresolutie te analyseren, omdat het geen voldoende hoge-resolutiebeelden bevat ter vergelijking. Daarom sluit IMDB-Clean het beste aan bij de doelstelling van dit onderzoek.

TABLE II
OVERZICHT VAN VEELGEBRUIKTE DATASETS VOOR LEEFTIJDSSCHATTING

Dataset	Aantal beelden	Resolutie / kwaliteit	Leeftijdssrange
UTKFace	20.000+	~ 200 × 200 px (laag)	0–116
FairFace	100.000+	~ 300 × 300 px (medium)	Categorieën (0–2 tot 70+)
IMDB-WIKI	500.000+	vaak >400 px (gemengd)	0–100+
IMDB-Clean	Subset van IMDB-WIKI	tot 1024 px, opgeschoond (hoog)	0–100+
VGGFace2	3.31M	hoge resolutie web images (gemengd)	Geen leeftijdslabellen

1) *Keuze voor bestaande frameworks:* Voor dit onderzoek is gekozen voor *DeepFace* [1] en *InsightFace* [2], omdat deze frameworks voorgetrainde modellen bieden voor leeftijdsvorspelling, die direct inzetbaar zijn zonder extra training. Beide systemen combineren detectie, alignering, feature-extractie en regressie in een gestandaardiseerde pipeline, wat reproduceerbare vergelijking op verschillende resoluties mogelijk maakt.

a) *DeepFace:* DeepFace is een Python-framework voor gezichtsherkenning en leeftijdsvorspelling, dat meerdere state-of-the-art modellen ('backbones') ondersteunt, zoals VGG-Face, ArcFace, FaceNet, Dlib en DeepID [1]. De standaard pipeline bestaat uit detectie, alignering, normalisatie en embedding-extractie. Specifiek voor leeftijdsschatting worden gezichten na cropping en resizen (meestal 224 × 224 px voor VGG-Face of 112 × 112 px voor ArcFace) doorgestuurd naar een regressiemodel dat een leeftijd voorspelt [3]. Volgens de officiële documentatie behaalt dit model een gemiddelde fout van circa ±4.65 jaar (MAE) [13].



Fig. 3. Voorbeeldoutput van DeepFace: leeftijd, emotie, geslacht en ras worden automatisch voorspeld uit een gezichtsafbeelding [1].

b) *InsightFace:* InsightFace is een framework dat oorspronkelijk ontwikkeld is voor gezichtsherkenning, gebaseerd op de ArcFace-architectuur [4]. Het ondersteunt detectie met

RetinaFace en SCRFD, beide inbegrepen in de officiële implementatie [2]. Invoerbeelden worden standaard geschaald naar 112 × 112 px voor verwerking in de ArcFace-backbone. Voor leeftijdsvorspelling biedt InsightFace specifieke modellen met een 'genderage'-head, waarbij eerst gezichts-embeddings worden geëxtraheerd en vervolgens een regressie- of classificatielaag wordt toegepast.

TABLE III
VERGELIJKING VAN DEEPFACE EN INSIGHTFACE (2025)

Eigenschap	DeepFace [1]	InsightFace [2]
Open source	MIT-licentie	MIT-licentie
Backbones	VGG-Face, ArcFace, FaceNet, Dlib, DeepID	ArcFace, Subcenter ArcFace
Detectiemodellen	OpenCV, SSD, Dlib, MTCNN, RetinaFace, etc.	RetinaFace, SCRFD
Input formaat	224×224 (VGG-Face), 112×112 (ArcFace)	112×112
Leeftijdsmodel	Pretrained + regressie	Embedding + regressie/klassificatie (incl. genderage-head)
Updates	Actief (2025, v0.9.5)	Actief (2025, v0.7.3+)

c) *Keuzeoverweging:* Er is niet gekozen voor alternatieven, zoals zelfgetrainde CNN's, OpenCV of YOLO-Age modellen vereisen vaak aangepaste trainingsdata en GPU-resources, wat buiten de haalbaarheid van dit project valt.

DeepFace en InsightFace zijn daarentegen open-source, goed gedocumenteerd en maken gebruik van wetenschappelijk gevalideerde pretrained modellen, waardoor ze beter aansluiten bij het doel van dit onderzoek.

2) *Architecturen voor leeftijdsschatting:* Veel leeftijdsschattingmodellen maken gebruik van convolutionele neurale netwerken (CNN's). Twee veelgebruikte backbone-architecturen in bestaande frameworks zijn **VGG-Face** en **ArcFace**.

Het **VGG-Face**-model [3] is gebaseerd op de VGG-16 architectuur en gebruikt invoerbeelden van 224 × 224 pixels. Het netwerk bevat circa 134 miljoen parameters en produceert embeddings van 4096 dimensies. In frameworks zoals DeepFace wordt VGG-Face hergebruikt als feature extractor, waarbij een regressielaag wordt toegevoegd om de leeftijd te voorspellen [20]. VGG-Face geldt als een klassieke, sequentiële CNN zonder residual connections, met relatief hoge rekenkosten.

ArcFace [4] daarentegen is gebaseerd op een ResNet-architectuur met residual connections en gebruikt invoerbeelden van 112 × 112 pixels. Het model telt circa 34 miljoen parameters, genereert embeddings van 512 dimensies en introduceert de *Additive Angular Margin Loss* voor robuustere representaties. Binnen InsightFace is ArcFace uitgebreid met een genderage-head waarmee leeftijd en geslacht direct voorspeld worden.

Samenvattend vertegenwoordigen VGG-Face en ArcFace twee generaties CNN-architecturen: een klassieke, zware CNN (VGG-Face) versus een moderne, efficiëntere ResNet-variant (ArcFace). Dit verschil in architectuur verklaart mede de betere prestaties en hogere snelheid van InsightFace in de experimenten van dit onderzoek. Er bestaan ook recentere modellen, zoals FusionNet [16] of attention-gebaseerde varianten [17], maar deze vallen buiten de scope van dit onderzoek.

3) *De impact van beeldresolutie*: Factoren zoals resolutie en beeldscherpte spelen een grote rol, omdat subtiele details (zoals rimpels en huidtextuur) snel verloren gaan bij lage resoluties. Uit de studie “*Effects of Varying Resolution on Performance of CNN based Image Classification*” [5] blijkt dat de resolutie van beelden een significante invloed heeft op de prestaties van CNN-modellen. Zo neemt de nauwkeurigheid merkbaar af wanneer een model wordt getraind op hogeresolutiebeelden maar getest wordt op datasets met lagere resoluties.

Table 3 Performance result of experimental study on CIFAR10 dataset

CIFAR10 Dataset Resolution	TOTV			TTTV		
	Trained on original resolution dataset (32x32) and tested on varying resolution dataset (32x32, 24x24, 16x16, 8x8)			Trained and tested on each varying resolution dataset separately (32x32, 24x24, 16x16, 8x8)		
	Accuracy	Precision	F1 Score	Accuracy	Precision	F1 Score
32x32	0.8752	0.87652	0.87548	0.8752	0.87652	0.87548
24x24	0.6409	0.72365	0.65320	0.6204	0.70501	0.63220
16x16	0.3166	0.48415	0.29897	0.4233	0.62030	0.40654
8x8	0.1855	0.27090	0.13986	0.3020	0.54599	0.24262

Fig. 4. Effect van variërende resolutie op de prestaties van een CNN op CIFAR-10. Te zien is dat nauwkeurigheid, precisie en F1-score sterk afnemen bij lagere resoluties [5].

Figuur 4 bevestigt dat beeldresolutie een essentieel aspect is om rekening mee te houden bij het gebruik van bestaande frameworks voor leeftijdsdetectie, zoals *DeepFace* [1] en *InsightFace* [2]. Ook in andere domeinen is dit effect zichtbaar: in medische beeldanalyse (endoscopie) leidt lagere resolutie eveneens tot een duidelijke afname van de prestaties van CNN-modellen [18].

Daarom onderzoekt deze studie expliciet de invloed van verschillende beeldresoluties op de nauwkeurigheid van leeftijdsvoorspellingen met *DeepFace* en *InsightFace*.

III. RESEARCH METHODOLOGY

A. Dataset en preprocessing

Voor dit onderzoek is een aangepaste subset gebruikt van de IMDB-Clean dataset.¹ De selectie bestaat uit 1000 afbeeldingen die voldoen aan de volgende voorwaarden:

- genomen vanaf 2010 (volgens bestandsmetadata),
- minimaal 1024 pixels breed of hoog,
- precies één gezicht (gedetecteerd met het standaard InsightFace-model:

`FaceAnalysis(allowed_modules=["detection"])`).

De filtering is gedaan met een Python-script dat automatisch beelden zonder gezicht of met meerdere gezichten verwijdert. De uiteindelijke set is openbaar beschikbaar via Hugging Face.²

Van elke afbeelding zijn zeven versies gemaakt in de resoluties: 64×64 , 112×112 , 224×224 , 256×256 , 512×512 , 720×720 en 1080×1080 pixels.

Hiervoor is `Pillow` gebruikt om de beelden te resizen. Omdat ze vierkant zijn, was letterboxing (padding) niet nodig.

¹<https://www.kaggle.com/datasets/yuulind/imdb-clean>

²<https://huggingface.co/datasets/codershiyar/imdb-clean-2010plus-singleface>

Alle versies zijn opgeslagen in één map, met de resolutie in de bestandsnaam (bijv. `_64.jpg`, `_512.jpg`).



Fig. 5. Voorbeeld van één afbeelding in zeven verschillende resoluties: 64×64 t/m 1080×1080 .

B. Environment setup

Alle experimenten zijn uitgevoerd in Google Colab met een NVIDIA T4 GPU en Python 3 als runtime (zie Figuur 6). De frameworks *DeepFace* (v0.0.95) en *InsightFace* (v0.7.3) zijn geïnstalleerd via `pip`:

```
!pip install deepface==0.0.95 insightface==0.7.3
onnxruntime-gpu-1.23.0
```

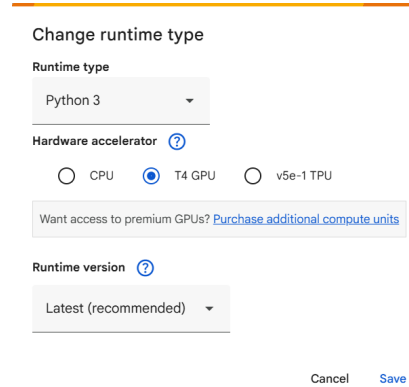


Fig. 6. Screenshot van de gebruikte Colab-runtime (T4 GPU, Python 3).

Voor *DeepFace* is de analyse uitgevoerd met:

```
DeepFace.analyze(
    img_path=img_path, actions=["age"], enforce_detection=
    False
)
```

zodat uitsluitend leeftijd werd voorspeld.

Voor *InsightFace* is gebruikgemaakt van:

```
FaceAnalysis(allowed_modules=["detection", "genderage"])
```

Hierbij wordt gezichtsdetectie uitgevoerd met SCRFD, en leeftijd geschat via de ingebouwde `genderage-head` op de ArcFace-backbone.

De volledige code en verwerkingspipeline zijn publiek beschikbaar [15].

C. Experiment design

In elk experiment zijn dezelfde 1000 afbeeldingen gebruikt (de geselecteerde subset uit IMDB-Clean). Voor elke afbeelding zijn zeven resoluties getest (64×64 t/m 1080×1080). Elke resolutie is getest met beide frameworks:

- **DeepFace (v0.0.95)** – standaard detectie en leeftijdsanalyse met pretrained regressiemodellen.
- **InsightFace (v0.7.3)** – standaard detectie (SCRFD) en leeftijdsschatting via de ingebouwde `genderage-head`.

Door beide frameworks systematisch toe te passen op alle resoluties ontstaat een eerlijke vergelijking. De uitkomsten worden geëvalueerd met MAE, SD en MedAE (zie Sectie III-D).

D. Evaluatiemetrieken

Om de prestaties van DeepFace en InsightFace onder verschillende resoluties te beoordelen, is per combinatie van model en resolutie de **Gemiddelde absolute fout (MAE)** berekend. Dit cijfer geeft de gemiddelde afwijking weer tussen de voorspelde leeftijd en de werkelijke leeftijd in jaren.

Daarnaast zijn twee aanvullende statistieken opgenomen:

- **Standaarddeviatie (SD)** – meet de spreiding van de absolute fouten. Een hoge SD wijst op onregelmatige voorspellingen, een lage SD op consistente prestaties.
- **Mediaan absolute fout (MedAE)** – de middelste foutwaarde. Deze maat is robuuster dan het gemiddelde en laat zien of de MAE wordt beïnvloed door uitschieters.

Samen geven MAE, SD en MedAE een volledig beeld van zowel de nauwkeurigheid als de betrouwbaarheid van de leeftijdsvoorspellingen per resolutie.

Alle berekeningen zijn uitgevoerd op exact dezelfde 1000 gezichten per resolutie, zodat de uitkomsten eerlijk vergeleken kunnen worden tussen modellen en resoluties. De resultaten worden samengevat in Tabel IV.

a) *Gebruikte formules:* Voor elke resolutie en elk model zijn de volgende formules gebruikt, waarbij y_i de voorspelde leeftijd is, \hat{y}_i de werkelijke leeftijd en n het aantal afbeeldingen:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i| - MAE)^2}$$

$$MedAE = \text{mediaan}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$$

IV. RESULTATEN EN DISCUSSIE

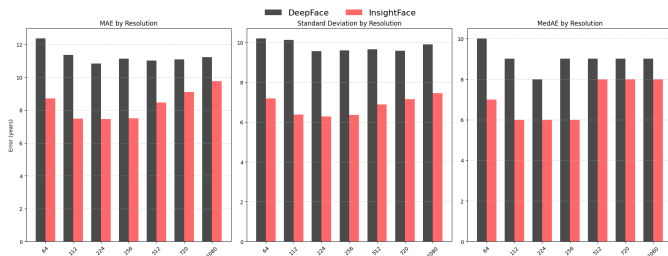


Fig. 7. MAE, standaarddeviatie en MedAE per resolutie voor DeepFace en InsightFace.

Tabel IV en Figuur 7 tonen de foutmaten (MAE, SD en MedAE) per resolutie voor DeepFace en InsightFace. Beide frameworks behalen de laagste MAE bij 224×224 pixels: 10.83 jaar voor DeepFace en 7.46 jaar voor InsightFace.

Bij 64×64 stijgt de fout naar 12.36 (DeepFace, +14.1%) en 8.72 (InsightFace, +16.9%), ten opzichte van 224×224 . Eerder werk suggereert dat dit kan samenhangen met verlies van leeftijdsrelevante details, zoals rimpels en huidtextuur, bij lage resoluties [5]. Bij hoge resoluties (1080×1080)

TABLE IV
GEMIDDELDE ABSOLUTE FOUT (MAE), STANDAARDDEVIAATIE (SD) EN MEDIAAN ABSOLUTE FOUT (MedAE) PER RESOLUTIE VOOR DEEPFACE EN INSIGHTFACE.

Resolutie	DeepFace			InsightFace		
	MAE ↓	SD ↓	MedAE ↓	MAE ↓	SD ↓	MedAE ↓
64 × 64	12.36	10.19	10.0	8.72	7.19	7.0
112 × 112	11.36	10.12	9.0	7.48	6.38	6.0
224 × 224	10.83	9.57	8.0	7.46	6.28	6.0
256 × 256	11.13	9.60	9.0	7.51	6.36	6.0
512 × 512	11.02	9.65	9.0	8.45	6.89	8.0
720 × 720	11.08	9.58	9.0	9.09	7.14	8.0
1080 × 1080	11.22	9.90	9.0	9.76	7.45	8.0

blijft DeepFace stabiel (11.22 jaar; +3.6%), terwijl InsightFace verslechtert tot 9.76 jaar (+30.8%). DeepFace blijft relatief stabiel rond ~ 11 jaar MAE, terwijl de fout bij InsightFace gradueel toeneemt vanaf 224×224 . Dit kan erop wijzen dat het interne resizingproces bij InsightFace minder goed omgaat met extreme downscaling van hogeresolutiebeelden naar 112×112 pixels, waardoor details verloren gaan.

Een opvallende bevinding is dat InsightFace bij externe invoer van 224×224 pixels een iets lagere MAE behaalt (7.46 jaar) dan bij 112×112 pixels (7.48 jaar), ondanks interne resize naar 112×112 . Dit wijst erop dat hogere invoeresoluties mogelijk bijdragen aan betere gezichtsdetectie (SCRFD) en alignment vóórdat de interne resizing plaatsvindt.

TABLE V
GEMIDDELDE FOUTMATEN (MAE, SD EN MEDAE) OVER ALLE RESOLUTIES.

Framework	MAE ↓	SD ↓	MedAE ↓
DeepFace	11.28	9.80	9.00
InsightFace	8.35	6.81	7.00

Gemiddeld over alle resoluties presteert InsightFace beter dan DeepFace (Tabel V). De gemiddelde MAE bedraagt 8.35 jaar voor InsightFace en 11.28 jaar voor DeepFace.

Naast de gemiddelde absolute fout (MAE) geven de standaarddeviatie (SD) en de mediaan absolute fout (MedAE) extra inzicht in de prestaties van de modellen. Voor InsightFace bij 224×224 pixels is de SD 6,28 jaar en de MedAE 6,0 jaar; dat betekent dat de fouten redelijk consistent zijn en dat de helft van de voorspellingen maximaal 6 jaar afwijkt van de werkelijke leeftijd. DeepFace heeft een hogere SD en MedAE, wat wijst op grotere variatie en minder typische nauwkeurigheid. Dit verschil hangt samen met de gebruikte architecturen: InsightFace maakt gebruik van ArcFace met ResNet-backbone en Additive Angular Margin Loss, terwijl DeepFace is gebaseerd op VGG-Face met softmax loss [3], [4].

A. Verwerkingstijd

Tabel VI laat zien dat de verwerkingstijd sterk verschilt tussen de frameworks. InsightFace verwerkt consequent sneller (46-66 beelden per seconde) en blijft relatief stabiel bij hogere resoluties. DeepFace daarentegen vertraagt exponentieel: van 15.6 beelden per seconde bij 64×64 naar slechts 1.3 beelden per seconde bij 1080×1080 . Gezien de combinatie van lagere

fouten (Tabel IV) en hogere verwerkingssnelheid kan worden vastgesteld dat InsightFace in deze experimenten zowel nauwkeuriger als efficiënter presteert dan DeepFace.

TABLE VI
VERWERKINGSTIJD PER 1000 AFBEELDINGEN BIJ VERSCHILLENDE
RESOLUTIES.

Resolutie	DeepFace	InsightFace
64 × 64	1:03 (15.6/s)	0:16 (58.9/s)
112 × 112	1:00 (16.5/s)	0:15 (65.6/s)
224 × 224	1:29 (11.1/s)	0:15 (64.0/s)
256 × 256	1:41 (9.9/s)	0:16 (61.8/s)
512 × 512	4:02 (4.1/s)	0:17 (58.7/s)
720 × 720	6:34 (2.5/s)	0:18 (52.7/s)
1080 × 1080	12:29 (1.3/s)	0:21 (46.1/s)

V. CONCLUSIE

Dit onderzoek toont aan dat invoerresolutie significant invloed heeft op leeftijdsschatting met DeepFace en InsightFace, ondanks hun interne preprocessing. De optimale resolutie voor beide frameworks is 224×224 pixels. Afwijkingen hiervan zowel lager als hoger leiden tot nauwkeurigheidsverlies, waarbij InsightFace gevoeliger is voor zeer hoge resoluties (+31% bij 1080×1080) dan DeepFace (+3,6%). Een opvallende bevinding is dat InsightFace bij externe invoer van 224×224 pixels nauwkeuriger voorspelt (MAE: 7,46 jaar) dan bij 112×112 pixels (MAE: 7,48 jaar), ondanks interne resize naar 112×112. Dit bevestigt dat de externe invoerresolutie van invloed blijft op de eindprestatie. Dit onderzoek is gebaseerd op 1000 afbeeldingen uit IMDB-Clean, een dataset met foto's van bekende personen. Toekomstig onderzoek kan zich richten op datasets met gevarieerde opnamehoeken, beeldkwaliteiten en demografische groepen.

REFERENCES

- [1] S. Serengil and A. Ozpinar, "A benchmark of facial recognition pipelines and co-usability performances of modules," *Journal of Information Technologies*, vol. 17, no. 2, pp. 95–107, 2024. doi: 10.17671/gazibtd.1399077. [Online]. Available: <https://dergipark.org.tr/en/download/article-file/3573195>. Accessed: Oct. 14, 2025.
- [2] J. Guo, J. Deng, and S. Zafeiriou, "InsightFace: 2D and 3D Face Analysis Project," GitHub repository, 2025. [Online]. Available: <https://github.com/deepinsight/insightface>. [Accessed: Oct. 5, 2025]
- [3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. British Machine Vision Conf. (BMVC)*, 2015. [Online]. Available: <https://ora.ox.ac.uk/objects/uuid:a5f2e93f-2768-45bb-8508-74747f85cad1>. Accessed: Oct. 14, 2025.
- [4] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *arXiv preprint arXiv:1801.07698*, Jan. 2018. [Online]. Available: <https://arxiv.org/abs/1801.07698>. Accessed: Oct. 14, 2025.
- [5] S. P. Kannoja and G. Jaiswal, "Effects of varying resolution on performance of CNN based image classification: An experimental study," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 9, pp. 451–456, Sept. 2018. doi: 10.26438/ijcse/v6i9.451456. [Online]. Available: https://www.ijcseonline.org/full_paper_view.php?paper_id=2890, https://www.researchgate.net/publication/328960034_Effects_of_Varying_Resolution_on_Performance_of_CNN_based_Image_Classification_An_Experimental_Study. Accessed: Oct. 14, 2025.
- [6] Y. Zhu, Y. Li, G. Mu, and G. Guo, "A study on apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Santiago, Chile, Dec. 2015, pp. 25–31. doi: 10.1109/ICCVW.2015.43. [Online]. Available: <https://ieeexplore.ieee.org/document/7406392>. Accessed: Oct. 14, 2025.
- [7] A. Garain, B. Ray, P. K. Singh, and A. Ahmadian, "GRANet: A deep learning model for classification of age and gender from facial images," *IEEE Access*, vol. 9, pp. 85672–85689, June 2021. doi: 10.1109/ACCESS.2021.3085971.
- [8] J. Chen, L. Cheng, X. Yang, J. Liang, B. Quan, and S. Li, "Joint learning with both classification and regression models for age prediction," *J. Phys.: Conf. Ser.*, vol. 1168, 032016, 2019. doi: 10.1088/1742-6596/1168/3/032016.
- [9] R. Rothe, R. Timofte, and L. Van Gool, "DEX: Deep EXpectation of apparent age from a single image," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Dec. 2015. [Online]. Available: <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>. Accessed: Oct. 14, 2025.
- [10] Y. Lin, J. Shen, Y. Wang, and M. Pantic, "FP-Age: Leveraging face parsing attention for facial age estimation in the wild," *arXiv preprint arXiv:2106.11145*, Jun. 2021. doi: 10.48550/arXiv.2106.11145. [Online]. Available: <https://arxiv.org/abs/2106.11145>. Accessed: Oct. 14, 2025.
- [11] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, pp. 144–157, 2018. doi: 10.1007/s11263-016-0940-3.
- [12] K. Karkkainen and J. Joo, "FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1548–1558. doi: 10.1109/WACV48630.2021.00159. [Online]. Available: <https://ieeexplore.ieee.org/document/9423296>. Accessed: Oct. 14, 2025.
- [13] S. Serengil, "DeepFace: A lightweight face recognition and facial attribute analysis framework," GitHub repository, 2025. [Online]. Available: <https://github.com/serengil/deepface>. Accessed: Oct. 14, 2025.
- [14] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 2018 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG 2018)*, Xi'an, China, May 2018, pp. 67–74. doi: 10.1109/FG.2018.00020. [Online]. Available: <https://ieeexplore.ieee.org/document/8373813>. Accessed: Oct. 14, 2025.
- [15] S. Jamo, "Experimentcode: Effect van beeldresolutie op leeftijdsschatting," Colab notebook, 2025. [Online]. Available: <https://colab.research.google.com/drive/1roSZ2Y4Ne9Yy71DM2mRmY0VetfuydbEi>. Accessed: Oct. 2, 2025.
- [16] H. Wang, X. Wei, V. Sanchez, and C.-T. Li, "Fusion network for face-based age estimation," *arXiv preprint arXiv:1807.10421*, Jul. 2018. doi: 10.48550/arXiv.1807.10421. [Online]. Available: <https://arxiv.org/abs/1807.10421>. Accessed: Oct. 14, 2025.
- [17] H. Wang, V. Sanchez, and C.-T. Li, "Improving face-based age estimation with attention-based dynamic patch fusion," *arXiv preprint arXiv:2112.10167*, Dec. 2021. doi: 10.48550/arXiv.2112.10167. [Online]. Available: <https://arxiv.org/abs/2112.10167>. Accessed: Oct. 14, 2025.
- [18] V. Thambawita, S. A. Hicks, I. Strümke, P. Halvorsen, S. Parasa, and M. A. Riegler, "Impact of image resolution on deep learning performance in endoscopy image classification," *Diagnostics*, vol. 11, no. 12, Art. no. 2183, Dec. 2021. doi: 10.3390/diagnostics11122183. [Online]. Available: <https://www.mdpi.com/2075-4418/11/12/2183>. Accessed: Oct. 5, 2025.
- [19] O. Agbo-Ajala, S. Viriri, M. Oloko-Oba, and O. Ekundayo, "Apparent age prediction from faces: A survey of modern approaches," *Frontiers in Big Data*, vol. 5, Art. no. 1025806, Oct. 2022. doi: 10.3389/fdata.2022.1025806. [Online]. Available: <https://doi.org/10.3389/fdata.2022.1025806>. Accessed: Oct. 7, 2025.
- [20] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, "Deep convolutional neural network for age estimation based on VGG-Face model," *arXiv preprint arXiv:1709.01664*, Sep. 2017. doi: 10.48550/arXiv.1709.01664. [Online]. Available: <https://arxiv.org/abs/1709.01664>. Accessed: Oct. 14, 2025.
- [21] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," *arXiv preprint arXiv:1702.08423*, Mar. 2017. doi: 10.48550/arXiv.1702.08423. [Online]. Available: <https://arxiv.org/abs/1702.08423>. Accessed: Oct. 14, 2025.