

# Computer Vision voor Leeftijdsherkenning bij Zelfscankassa's in Supermarkten

KAAN GÖGCAY, Amsterdam University of Applied Sciences, Nederland

YUSRA EL GHAZI, Amsterdam University of Applied Sciences, Nederland

MERVE HALICI, Amsterdam University of Applied Sciences, Nederland

LEON VAN DIJK, Amsterdam University of Applied Sciences, Nederland

Zelfscankassa's maken het mogelijk om aankopen zelfstandig af te rekenen, maar leveren uitdagingen op bij het controleren van de wettelijke minimumleeftijd voor alcoholverkoop. Dit onderzoek richt zich op de inzet van computer vision voor leeftijdsverificatie bij zelfscankassa's, binnen de geldende wet- en regelgeving en met inachtneming van privacy. Op basis van een juridische, ethische en organisatorische analyse is een prototype ontwikkeld dat leeftijden in real time schat. Klanten met een geschatte leeftijd van 25 jaar of ouder worden automatisch goedgekeurd en in andere gevallen vindt controle plaats door een medewerker. De toepassing werkt op basis van vrijwillige deelname (opt-in) en slaat geen beelden op. Het model is getraind met de UTKFace-dataset en stap voor stap verbeterd. De eindversie heeft een gemiddelde afwijking van 5.06 jaar tussen de geschatte en werkelijke leeftijd, een nalevingspercentage van 99.19% (het aandeel correct geblokkeerde minderjarigen) en een verwerkingstijd van minder dan 500 milliseconden per voorspelling (inferentietijd). De nauwkeurigheid is het hoogst in de leeftijdscategorie 15–25 jaar, die bepalend is voor naleving van de wetgeving. De resultaten tonen aan dat leeftijdsinschatting met computer vision technisch uitvoerbaar is binnen de kaders van de AVG, UAVG en CBL-richtlijnen. In de afronding worden ontwerp- en implementatieadviezen gegeven voor toepassing in supermarkten, met aandacht voor gegevensbescherming, datadiversiteit en het volgen van prestatieverschillen tussen groepen.

---

Authors' Contact Information: Kaan Gögcay, Amsterdam University of Applied Sciences, Amsterdam, Nederland, [kaan@email.com](mailto:kaan@email.com); Yusra El Ghazi, Amsterdam University of Applied Sciences, Amsterdam, Nederland, [yusra@email.com](mailto:yusra@email.com); Merve Halici, Amsterdam University of Applied Sciences, Amsterdam, Nederland, [merve@email.com](mailto:merve@email.com); Leon van Dijk, Amsterdam University of Applied Sciences, Amsterdam, Nederland, [leon@email.com](mailto:leon@email.com).

CONTENTS

Abstract	1
Contents	2
1    Introductie	3
1.1   Deelvragen	3
2    Achtergrond	4
2.1   Huidige situatie	4
2.2   Regelgeving	4
2.3   Gebruikersonderzoek	5
2.4   Gerelateerde Literatuur	8
2.5   Overwegingen voor het inzetten van AI	9
3    Requirements	10
4    Methodologie	12
4.1   Concept	12
4.2   Data	16
4.3   Model	20
5    Resultaten	23
5.1   Resultaten Baseline (ResNet-50)	23
5.2   Resultaten VGG16	24
6    Evaluatie	26
6.1   Evaluatie van Modelresultaten	26
6.2   Onderbouwing van wijzigingen tussen concepten	26
6.3   Volledig geïntegreerde evaluatie van het model binnen het concept	27
7    Discussie	28
8    Conclusie	28
A    Testplan	28
References	30

\*

## 1 Introductie

De invoering van zelfscankassa's in supermarkten neemt de laatste jaren sterk toe en levert voordelen op voor zowel consumenten als supermarkten, zoals kortere wachttijden en lagere personeelskosten. Tegelijkertijd leidt dit systeem tot nieuwe uitdagingen bij het proces van leeftijdsverificatie voor alcoholverkoop. Waar voorheen een kassamedewerker verantwoordelijk was voor één kassa, is een zelfscan kassamedewerker verantwoordelijk voor een zelfscanplein bestaande uit meerdere kassa's. Uit onderzoek onder verschillende stakeholders blijkt dat deze problematiek zich op drie niveaus manifesteert: klanten ervaren wachttijden en frustratie, supermarkten lopen risico's door nalevingsfouten bij drukte, en medewerkers hebben te maken met extra werkdruk en soms agressief gedrag. Hoewel klanten over het algemeen tevreden zijn met het huidige proces, geven zij wachttijden regelmatig als pijnpunt aan. Voor supermarkteigenaren vormen juist de nalevingsproblemen het grootste risico.

Om het proces van de leeftijdsverificatie te verbeteren, wordt gekeken naar hoe computer-vision gebaseerde leeftijdsherkenning ondersteuning kan bieden vanuit de hoofdvraag:

*"Hoe kan een computer-vision model voor leeftijdsverificatie worden ontwikkeld, geëvalueerd en geïmplementeerd bij zelfscankassa's in supermarkten om het verificatieproces bij de verkoop van alcohol te verbeteren?"*

Deze vraag brengt uitdagingen op verschillende aspecten met zich mee. Zo is nog onduidelijk wat in deze context precies een verbetering betekent: gaat het om hogere naleving, snellere doorstroming, lagere werkdruk of meer klanttevredenheid? Dit zal worden vastgesteld middels gebruikersonderzoek en een stakeholderanalyse. Ook is er een groot aantal wetten op het gebied van gegevensverwerking en alcoholverkoop, welke implementatie kunnen belemmeren. Zo is bijvoorbeeld een eerdere poging om een ID-scan in supermarkten te gebruiken wettelijk afgekeurd [26].

Daarnaast zal uitvoerig onderzoek gedaan worden naar geschikte datasets en modellen. Het uiteindelijke computer-vision model zal iteratief worden ontwikkeld, geëvalueerd en getoetst op de vooraf opgestelde technische, functionele, juridische, ethische én organisatorische requirements.

### 1.1 Deelvragen

Om gestructureerd antwoord te geven op de hoofdvraag, zijn de volgende deelvragen vastgesteld:

- (1) Welke tekortkomingen kent het huidige leeftijdsverificatieproces bij zelfscankassa's en in hoeverre biedt computer vision hiervoor een passende oplossing?
- (2) Welke computer-vision technieken en architecturen zijn geschikt om een leeftijdsverificatiemodel te trainen?
- (3) Welke datasets zijn geschikt om een leeftijdsverificatiemodel te ontwikkelen en te testen in de context van supermarkten?
- (4) In hoeverre voldoet het ontwikkelde leeftijdsverificatiemodel aan de vooraf opgestelde technische requirements?
- (5) Op welke manier kunnen het model, de integratie en de gebruikersinteractie in zelfscankassa's binnen supermarkten worden ontworpen zodat het systeem praktisch inzetbaar is bij de verkoop van alcohol.
- (6) In hoeverre voldoet de voorgestelde leeftijdsverificatie-oplossing aan relevante wettelijke en ethische randvoorwaarden (o.a. AVG/UAVG en CBL-richtlijnen) en welke gevolgen heeft dit voor ontwerp en implementatie?

## 2 Achtergrond

### 2.1 Huidige situatie

In supermarkten is het verplicht om bij de verkoop van alcohol de leeftijd van de klant te controleren. Deze verplichting staat beschreven in de CBL Code ‘Verantwoorde alcoholverkoop’ [4]. Bij de zelfscankassa is dit een taak van het aanwezige personeel. Wanneer een klant alcohol afrekenen krijgt de medewerker een melding op de kassa. De medewerker moet dan beoordelen of de klant mogelijk jonger is dan 25 jaar. Als dat zo lijkt vraagt de medewerker om een legitimatie. Daarna voert de medewerker de geboortedatum handmatig in het systeem in zodat de klant de aankoop kan afronden.

Dit proces vraagt veel inzet van het personeel. Zij moeten niet alleen leeftijdscontroles uitvoeren maar ook andere taken doen bij de zelfscankassa’s. Denk aan het helpen van klanten die vastlopen bij het afrekenen en het uitvoeren van steekproeven. Tijdens drukke momenten ontstaat hierdoor vaak hoge werkdruk. Vooral als er weinig medewerkers aanwezig zijn kan het lastig zijn om overal tegelijk bij te helpen. Voor klanten betekent dit dat het afrekenen langer duurt. Zij moeten soms wachten tot een medewerker tijd heeft voor de leeftijdscontrole. Dit kan leiden tot opstoppen bij de zelfscankassa’s en langere wachtrijen. Vooral op piekmomenten zoals in de avond of in het weekend wordt dit duidelijk zichtbaar.

Het doel is om de leeftijdscontrole bij de zelfscankassa sneller en betrouwbaarder te maken. Het personeel blijft betrokken, maar krijgt ondersteuning van een slim systeem. Dit systeem kan automatisch aangeven wanneer een klant jonger lijkt dan 25 jaar. In dat geval wordt er een controle uitgevoerd en kan de medewerker dit proces eenvoudig afronden. Op die manier hoeven medewerkers minder vaak zelf te beoordelen of iemand jong oogt. Het proces verloopt daardoor vlotter en de druk voor medewerkers neemt af. Voor klanten betekent dit dat zij minder lang hoeven te wachten bij de aankoop van alcohol en dat het afrekenen soepeler gaat, ook tijdens drukke momenten.

### 2.2 Regelgeving

De verkoop van alcohol en tabak is streng gereguleerd en supermarkten zijn verplicht de leeftijd van klanten te verifiëren. Bij inzet van technologie zoals computer vision aan de zelfscankassa gelden daarnaast privacyvoorschriften. Dit hoofdstuk geeft een overzicht van de belangrijkste regels rond leeftijdscontrole en mogelijke oplossingsrichtingen.

**2.2.1 Leeftijdsverificatie.** Het beleid “Verantwoorde alcoholverkoop in de supermarkt” van het Centraal Bureau Levensmiddelenhandel (CBL) bevat de richtlijnen die alle supermarkten moeten volgen. De leeftijdsverificatie verloopt daarbij in vaste stappen. Allereerst maakt het personeel een inschatting of de klant 25 jaar of ouder lijkt te zijn. Wanneer dit niet het geval is, moet de klant een geldig identiteitsbewijs laten zien. Is de klant jonger dan 18 jaar, dan mag de verkoop niet doorgaan. Wanneer er meerdere personen bij de aankoop aanwezig zijn, moet van iedereen de leeftijd worden gecontroleerd. De verkoop gaat in dat geval alleen door als alle betrokken personen 18 jaar of ouder zijn [4].

**2.2.2 Privacy.** Wanneer leeftijdsverificatie wordt ondersteund met technologie, bijvoorbeeld door camera’s bij de zelfscankassa, komen ook privacyregels in beeld. Camerabeelden kunnen persoonsgegevens bevatten, omdat gezichten en andere kenmerken direct naar een individu te herleiden zijn. Volgens de Algemene verordening gegevensbescherming (AVG) mogen persoonsgegevens alleen verwerkt of bewaard worden indien daarvoor een wettelijke grondslag bestaat, zoals het voldoen aan een wettelijke verplichting (art. 6) [1]. Daarbij geldt het beginsel van dataminimalisatie en opslagbeperking: er mogen zo min mogelijk gegevens verzameld en niet langer bewaard worden dan noodzakelijk (art. 5), terwijl klanten altijd duidelijk geïnformeerd moeten worden over het gebruik ervan (art. 12–13) [1].

Een terugkerende vraag bij leeftijdsverificatie is of een identiteitsbewijs mag worden gescand, gefotografeerd of gekopieerd. Dit is volgens de AVG in de meeste gevallen niet toegestaan, omdat een ID gevoelige persoonsgegevens bevat, zoals het burgerservicenummer (BSN) en pasfoto (art. 4(2)) [1]. De Uitvoeringswet AVG (UAVG) staat verwerking van het BSN bovendien alleen toe wanneer dit wettelijk verplicht is (art. 46) [2]. Voor supermarkten betekent dit dat enkel visuele inspectie van een ID door een medewerker is toegestaan; het opslaan of kopiëren van een ID is verboden.

Nieuwe vormen van leeftijdsverificatie kunnen gebruikmaken van biometrische gegevens, zoals gezichtsherkenning. Deze vallen onder de bijzondere categorieën persoonsgegevens (art. 4(14)), waarvan verwerking in beginsel verboden is (art. 9(1)) [1]. De UAVG voorziet slechts in een beperkte uitzondering: biometrie mag uitsluitend worden toegepast wanneer dit strikt noodzakelijk is voor authenticatie of beveiliging (art. 29) [2]. Voor supermarkten is biometrische leeftijdscontrole daarom juridisch en technisch slechts beperkt toepasbaar.

### 2.3 Gebruikersonderzoek

Om het probleem beter in kaart te krijgen vanuit het perspectief van de gebruikers, is een gebruikersonderzoek gedaan. Het doel van dit gebruikersonderzoek is om van bij gebruikers de mening te peilen over het huidige proces van leeftijdsverificatie bij de zelfscankassa, als de mening over beeldherkenning voor leeftijdsschatting om dit proces te verbeteren. Onder gebruikers wordt verstaan: een ieder die met de technologie moet gaan werken, bijvoorbeeld supermarktklanten, medewerkers en eigenaren. Voor dit gebruikersonderzoek is gebruik gemaakt van een enquête voor supermarktklanten. Ook is er een interview gehouden met een supermarkteigenaar.

**2.3.1 Enquête.** De enquête telt 35 respondenten. De leeftijdsverdelingen over de respondenten was als volgt:

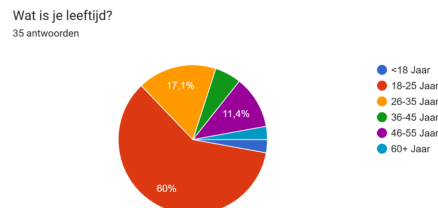


Fig. 1. Leeftijdsverdeling respondenten (N=35).

Uit de enquête bleek dat 61,8% van de klanten 1–3 minuten wacht op hulp bij leeftijdsverificatie. Hoewel 55,9% dit als “geen probleem” ervaart, vindt 35,3% het “een beetje irritant” en 8,8% “erg vervelend”. Het wachten wordt dus als pijnpunt ervaren. De algemene tevredenheid over alcoholaankopen bij zelfscankassa's is echter goed: 26,5% waardeert dit met 5/5, 47,1% met 4/5 en 26,5% met 3/5.



Fig. 2. Tevredenheid alcoholaankopen zelfscankassa (N=34).

Over automatische leeftijdsschatting via camera's zijn de meningen verdeeld. Van de respondenten vindt 48,6% het een goed tot uitstekend idee, terwijl 31,4% het (zeer) slecht vindt en 20% neutraal staat. Er bestaat geen duidelijke consensus onder klanten.

Stel: Een camera kan automatisch inschatten of je 18+ bent. Mensen die er duidelijk volwassen uitzien worden direct goedgekeurd. Wat vind je van dit idee?

35 antwoorden

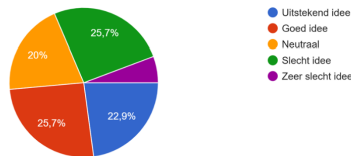


Fig. 3. Mening over automatische leeftijdsschatting via camera (N=35).

De belangrijkste voordelen die respondenten zien zijn: niet hoeven wachten op medewerkers (74,3%), sneller afrekenen (65,7%), minder frustratie (31,4%) en geen ID tonen (31,4%). Echter, 17,1% ziet geen voordelen.

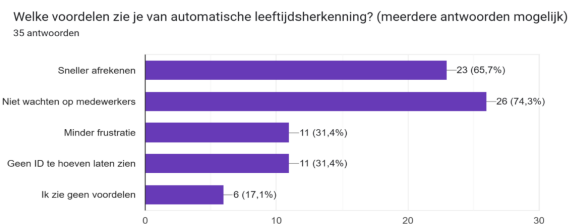


Fig. 4. Voordelen automatische leeftijdsherkenning (N=35, meerdere antwoorden mogelijk).

Voor acceptatie van beeldherkenning stellen klanten verschillende voorwaarden. Vrijwillig gebruik is het belangrijkste (70,6%), gevolgd door directe verwijdering van beelden (55,9%). Ook transparantie over nauwkeurigheid (32,4%), onafhankelijke privacycontrole (29,4%) en duidelijke uitleg (26,5%) zijn belangrijke voorwaarden.



Fig. 5. Voorwaarden voor acceptatie beeldherkenning (N=34, meerdere antwoorden mogelijk).

**2.3.2 Stakeholder Interview.** Andere cruciale stakeholders in dit project zijn de supermarkteigenaren. Zij beslissen immers over investeringen in nieuwe technologie en apparatuur. Om hun perspectief te belichten, is een interview gehouden met de eigenaar van een PLUS supermarkt in Vollenhove, een dorp met circa 4500 inwoners. Sinds 2022 beschikt de winkel over vier zelfscankassa's.

Volgens de eigenaar verloopt leeftijdsverificatie bij zelfscan eenvoudig: bij het scannen van een leeftijdsgelabeld product verschijnt een melding op de handscanner en gaat een lampje boven de kassa branden. Een medewerker voert de geboortedatum in en keurt zo de aankoop goed. Een eerdere bug maakte automatische doorgang na time-out mogelijk, maar dit is verholpen.

Alle caissières volgen een verplichte e-learning over leeftijdsverificatie, waarin slechts één fout is toegestaan. Het zelfscanpercentage ligt rond 63%, waardoor het vaak druk is. Er is altijd een medewerker aanwezig; de verificatie kost weinig tijd, maar zorgt bij piekmomenten toch voor oponthoud. Klachten daarover zijn zeldzaam, omdat klanten de controles accepteren.

Toch zijn er bij steekproeven alcohol en andere producten aan minderjarigen verkocht. De supermarkt kreeg hiervoor twee officiële waarschuwingen. Dit gebeurt vooral bij drukte, wanneer medewerkers slordiger worden. Dit wordt als groot pijnpunt gezien vanwege de mogelijke consequenties.

De eigenaar staat positief tegenover computer vision voor leeftijdsschatting. Technologie die medewerkers ontlast en de doorstroming bevordert, is welkom. In stedelijke gebieden kan dit bovendien agressie tegen medewerkers verminderen en zo de veiligheid verhogen.

Wel ziet hij voorwaarden: sommige klanten zullen bezwaar maken, daarom is duidelijke communicatie essentieel en moet keuzevrijheid blijven bestaan. Beelden mogen niet worden opgeslagen en klanten moeten desgewenst de camera kunnen afschermen. Als hieraan wordt voldaan, ziet hij vooral voordelen in AI-gestuurde leeftijdsverificatie.

**2.3.3 Conclusie.** Uit het gebruikersonderzoek komen drie hoofd-tekortkomingen naar voren. Ten eerste ervaren klanten wachttijden en frustratie, waarbij 44% van de klanten de gemiddelde wachttijd van 1-3 minuten als irritant beschouwt, wat de klanttevredenheid beïnvloedt ondanks een algemeen positieve beoordeling van het proces. Ten tweede ontstaan er nalevingsproblemen bij drukte, waarbij fouten leiden tot alcoholverkoop aan minderjarigen en officiële waarschuwingen voor supermarkten. Dit vormt het grootste pijnpunt voor eigenaren. Ten derde ervaren medewerkers werkdruk en soms agressief gedrag van klanten, wat de effectiviteit van het controleproces ondermijnt.

Voor succesvolle implementatie van computer vision moet een oplossing aan specifieke acceptatiecriteria voldoen. Het belangrijkste is vrijwillige deelname, wat 70,6% van de klanten eist, gevolgd door directe verwijdering van beeldmateriaal (55,9%). Daarnaast zijn transparantie over nauwkeurigheid en werking, behoud

van keuzemogelijkheden voor klanten, en duidelijke communicatie over privacy en functionaliteit essentiële voorwaarden.

## 2.4 Gerelateerde Literatuur

Safa [22] onderzocht hoe convolutionele neurale networks (CNN's) kunnen worden ingezet om de leeftijd van personen automatisch te voorspellen op basis van gezichtsafbeeldingen. In de studie werden verschillende deep learning-architecturen vergeleken, waaronder recurrente neurale networks (RNN's) en generative adversarial networks (GAN's). CNN's bleken echter het meest geschikt om leeftijdspatronen in gezichten te herkennen. Het model werd getraind met datasets waarin de gezichten waren ingedeeld in leeftijdsklassen, zodat zowel classificatie in leeftijdsgroepen als benadering van de werkelijke leeftijd mogelijk was. Uit de resultaten bleek dat CNN's aanzienlijk beter presteren dan de andere onderzochte methoden, vooral wanneer er voldoende variatie aanwezig is in de trainingsdata. Het onderzoek onderstreept daarmee dat deep learning-modellen, mits getraind op representatieve datasets, een effectieve basis vormen voor automatische leeftijdsschatting.

Rothe, Timofte en Van Gool [21] onderzochten hoe de apparente leeftijd (de leeftijd die iemand lijkt te hebben) automatisch kan worden geschat op basis van één gezichtsfoto. Hiervoor ontwikkelden zij de grootschalige IMDB-WIKI dataset met ruim 260.000 afbeeldingen voorzien van biologische leeftijden en gebruikten zij de LAP dataset met ongeveer 4.700 beelden die door beoordelaars waren gelabeld met apparente leeftijden. De auteurs pasten een VGG-16 CNN toe, voorgetraind op ImageNet en vervolgens gefinetuned op beide datasets. In plaats van directe regressie werd leeftijd geformuleerd als een classificatieprobleem (0–100 jaar), waarbij de uiteindelijke voorspelling werd berekend als de verwachte waarde van de softmax-output. Deze aanpak, gecombineerd met data-augmentatie en een ensemble van twintig netwerken, leverde state-of-the-art resultaten op: een gemiddelde fout van 3,2 jaar en een  $\epsilon$ -error van 0,2649, waarmee het model de eerste plaats behaalde in de ChaLearn LAP 2015 challenge en zelfs beter presteerde dan menselijke beoordelaars. Dit onderzoek toont aan dat deep learning in combinatie met grootschalige en diverse trainingsdata zeer effectief is voor leeftijdsschatting.

Singh et al. [25] richtten hun onderzoek specifiek op binaire leeftijdsclassificatie met als doel te bepalen of een persoon meerderjarig is of niet. Voor dit onderzoek maakten zij gebruik van de Multi-Ethnicity dataset, waarin personen zijn gelabeld als jonger dan 18 jaar of ouder. De auteurs ontwikkelden hiervoor de Class Specific Mean Autoencoder, een methode die gezichtskenmerken leert en deze dichter bij het gemiddelde van de juiste klasse (minderjarig of volwassen) positioneert. Deze aanpak bleek effectiever dan zowel traditionele algoritmen als commerciële systemen. Op de Multi-Ethnicity dataset behaalde het model een nauwkeurigheid van 92,09%. Dit onderzoek laat zien dat binaire leeftijdsclassificatie met behulp van deep learning een betrouwbare methode kan zijn voor toepassingen zoals leeftijdsverificatie bij leeftijdsgebonden producten.



## 2.5 Overwegingen voor het inzetten van AI

Voor de toepassing van AI bij leeftijdsschatting aan de zelfscankassa is het noodzakelijk om de gevolgen vanuit verschillende invalshoeken te beschouwen. Dit betreft technische haalbaarheid, maatschappelijke impact en ethische implicaties.

**2.5.1 Technische overwegingen.** AI-modellen voor leeftijdsschatting kunnen in veel gevallen een redelijke nauwkeurigheid behalen, maar foutmarges blijven aanwezig. De prestaties hangen sterk af van factoren zoals camerakwaliteit, belichting, gezichtspositie en de diversiteit van de gebruikte trainingsdata. Uit het NIST FATE AEV-rapport blijkt bijvoorbeeld dat onscherpe of onderbelichte beelden leiden tot slechtere leeftijdsschattingen [18].

Modellen die onvoldoende getraind zijn op verschillende leeftijden, huidskleuren of gezichtsstructuren kunnen bovendien systematische fouten vertonen. Een verkeerde inschatting kan resulteren in zowel onterechte weigering van legitieme klanten als in het per ongeluk goedkeuren van de verkoop aan minderjarigen. Omdat de inschatting volledig gebeurt op basis van visuele kenmerken, is het hanteren van een ruime veiligheidsmarge vaak noodzakelijk om risico's te beperken. Daarbij geldt dat een hogere MAE bij oudere leeftijden (bijvoorbeeld boven de 80 jaar) acceptabeler is dan bij jongere leeftijdsgroepen: een fout van 15 jaar bij een 85-jarige heeft in de praktijk geen consequenties voor alcoholverkoop, terwijl een afwijking van 5 jaar bij een 20-jarige wel cruciaal is.

**2.5.2 Maatschappelijke overwegingen.** Uit een interview met een Jumbo-ondernemer kwam naar voren dat een dergelijk systeem vooral in stedelijke context voordelen kan hebben. Hij gaf aan dat medewerkers daar regelmatig te maken krijgen met agressie van klanten bij leeftijdscontroles, en dat automatische goedkeuring veel druk zou kunnen wegnemen. Tegelijkertijd zijn er ook andere maatschappelijke aandachtspunten. Wanneer de AI faalt, kunnen medewerkers alsnog extra belast worden met controles of correcties. Daarnaast bestaat het risico dat jongeren proberen de systemen te omzeilen. Ook kan de maatschappelijke acceptatie variëren: sommige klanten zien het als een nuttige innovatie, terwijl anderen het ervaren als een inbreuk op hun privacy, al blijkt uit onderzoek dat de meeste gebruikers zich juist veilig voelen bij het gebruik van gezichtsherkenning [15].

**2.5.3 Ethische overwegingen.** De inzet van AI voor leeftijdscontrole roept vragen op over privacy, transparantie en verantwoordelijkheid. Het is belangrijk dat klanten erop kunnen vertrouwen dat hun biometrische gegevens niet structureel worden opgeslagen of misbruikt. Daarnaast moet nagedacht worden over de mate van automatisering: een volledig geautomatiseerd systeem zonder menselijke tussenkomst kan leiden tot problemen bij foutieve beslissingen. Hierbij speelt ook de vraag wie verantwoordelijk is als een AI-systeem een verkeerde inschatting maakt: de ontwikkelaar, de supermarkt of de leverancier van de technologie. Vanuit ethisch oogpunt is het wenselijk dat er altijd een mogelijkheid tot menselijke tussenkomst blijft bestaan, zodat fouten hersteld kunnen worden en de rechten van de consument gewaarborgd blijven. Dit sluit aan bij de interpretatie van artikel 22 van de GDPR, waarin het recht op menselijke beoordeling bij geautomatiseerde besluitvorming wordt benadrukt [14].

### 3 Requirements

In dit hoofdstuk worden de eisen beschreven waaraan de AI-gebaseerde leeftijdsverificatie moet voldoen. De requirements zijn opgesteld op basis van wetgeving, ethische richtlijnen en input van betrokken stakeholders. Tabel 1 geeft een overzicht van de gestructureerde MoSCoW-requirements, ingedeeld naar juridische, ethische, organisatorische, functionele en technische categorieën.

Table 1. Gesorteerde MoSCoW Requirements voor AI-gebaseerde Leeftijdsverificatie.

Categorie	Requirement	MoSCoW	Stakeholder
Juridisch	Als de klant jonger dan 25 jaar is, moet de leeftijd vastgesteld worden middels een ID-check	Must	Wetgever
Juridisch	Als er meerdere klanten aan dezelfde kassa staan, moet van elke klant de leeftijd worden vastgesteld	Must	Wetgever
Juridisch	ID's mogen niet gescand worden	Must	Wetgever
Juridisch	Het ontwerp moet voldoen aan de AVG artikel 4 lid 14 en artikel 9 lid 1	Must	Wetgever
Juridisch	Het ontwerp moet voldoen aan de Uitvoeringswet AVG artikel 29	Must	Wetgever
Juridisch	De gegevens mogen niet verwerkt worden, volgens AVG artikel 4 lid 2	Must	Wetgever
Juridisch	Alvorens gegevens van de klant worden afgenomen, wordt de klant de benodigde informatie verstrekt, volgens AVG artikel 13	Must	Wetgever
Juridisch	De klant wordt duidelijk geïnformeerd over de verwerking van zijn gegevens, naar AVG artikel 12 lid 1	Must	Wetgever
Ethisch	Het systeem moet de privacy en de waardigheid van de klant respecteren	Must	Klant
Ethisch	Het gebruik van AI moet proportioneel zijn: alleen toegepast bij leeftijdscontrole voor alcohol	Must	Klant
Organisatorisch	Medewerkers moeten te allen tijde de controle-functie behouden (Human in the loop)	Must	Wetgever
Functioneel	Het gebruik van AI voor leeftijdsverificatie moet op vrijwillige basis zijn	Must	Klant
Functioneel	De camera wordt afgedekt totdat er gekozen is voor het gebruik van AI voor leeftijdsverificatie	Must	Supermarkteigenaar
Functioneel	De AI oplossing vermindert het aantal interacties tussen klanten en medewerkers	Must	Medewerker/Supermarkteigenaar
Functioneel	Het model moet vaststellen of de persoon jonger dan 25 jaar is, of 25 jaar of ouder	Must	Ontwikkelaar
Technisch	Camerabeelden worden niet langer opgeslagen dan nodig is om de voorspelling te maken	Must	Ontwikkelaar
Technisch	Het model scoort op de testset een nalevingspercentage van ten minste 99%	Must	Ontwikkelaar

Vervolg van Tabel 1

Categorie	Requirement	MoSCoW	Stakeholder
Technisch	Het model scoort op de testset een Mean Absolute Error lager dan 5.0 jaar in de leeftijdsgroepen 15–20, 20–25 en 25–30 jaar	Must	Ontwikkelaar
Technisch	Het model scoort op de testset een Mean Absolute Error lager dan 10.0 jaar in de leeftijdsgroepen 0–5, 5–10, 10–15 en 35–40 jaar	Must	Ontwikkelaar
Technisch	Voor de overige leeftijdsgroepen mag de MAE-waarde, afgetrokken van de ondergrens van de groep, niet onder 25 uitkomen	Must	Ontwikkelaar
Technisch	Het model heeft op de MAE fairness metrics een	Must	Ontwikkelaar
Technisch	Het model heeft een inference time van minder dan 500 ms	Must	Ontwikkelaar
Ethisch	Het model moet getraind worden op een diverse dataset (leeftijd, geslacht en etniciteit) om bias in de prestaties te vermijden	Should	Ontwikkelaar
Organisatorisch	Medewerkers moeten duidelijke instructies krijgen hoe om te gaan met foutmeldingen of twijfelgevallen	Should	Medewerkers

## 4 Methodologie

### 4.1 Concept

Het concept richt zich op het onderzoeken en ontwikkelen van een AI-ondersteunde leeftijdscontrole aan de zelfscankassa. Hierbij wordt eerst gekeken naar de waarde voor verschillende belanghebbenden, gevolgd door een analyse van het huidige handmatige proces. Vervolgens wordt de conceptuele uitbreiding toegelicht, inclusief de toegepaste ontwerpprincipes, het niveau van automatisering en de evaluatie via paper prototyping.

*4.1.1 Value Proposition.* Een geautomatiseerde leeftijdscheck aan de zelfscankassa biedt voordelen voor meerdere belanghebbenden: klanten, zelfscanmedewerkers en supermarkteigenaren. Voor klanten die bij een reguliere kassa doorgaans niet op leeftijd worden gecontroleerd, maar dit bij een zelfscankassa wel ervaren, vormt de AI-controle een efficiënte oplossing. Zo hebben zij baat bij de kortere wachttijden die het met zich mee brengt. Voor medewerkers verlaagt de automatische controle de werkdruk en vermindert het aantal interacties met klanten. In stedelijke gebieden waar agressie tegen medewerkers veel voorkomt, biedt dat een uitkomst. Supermarkteigenaren profiteren daarnaast van een kleinere kans op boetes door foutieve controles.

Maar dit concept brengt niet alleen voordelen. Een groot ethisch nadeel is dat de AI, net als de mens, niet perfect is. Dezelfde fouten die gemaakt worden door de mens, kunnen ook worden gemaakt door een AI. In het ergste geval schat de AI iemand ver boven de alcohol grens met hoge zekerheid, waardoor een minderjarige aan alcohol kan komen. Ook is er een merkwaardig maatschappelijk nadeel. Als er op elke zelfscankassa en op de terminal van de medewerkers een AI model zal draaien, dan verbruikt het meer stroom dan dat het momenteel doet. Dit kan zorgen voor extra CO2 uitstoot.

*4.1.2 Huidige Staat.* Uit een interview met medewerkers van de Plus-supermarkt is een flowchart opgesteld van het huidige proces van de leeftijdscontrole bij de zelfscankassa, zie Figuur 6. De flowchart start bij de groene afgeronde rechthoek bovenaan en eindigt in de rode afgeronde rechthoeken. Momenteel verloopt het volledige proces handmatig, wat mogelijkheden biedt voor verbetering en automatisering.

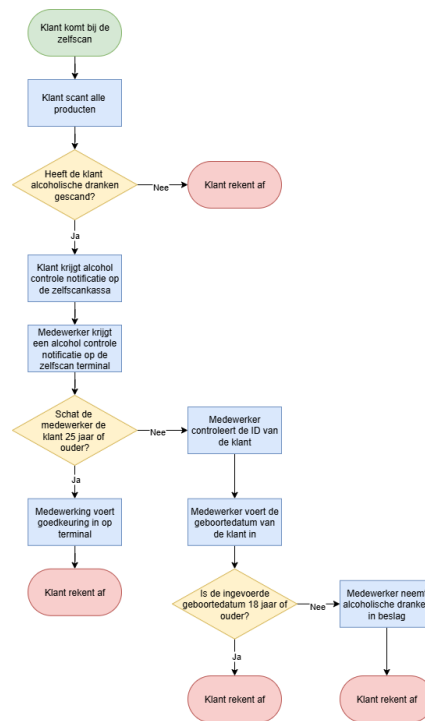


Fig. 6. Flowchart van het huidige proces aan de zelfscankassa's in supermarkten

**4.1.3 Conceptuele Uitbreiding.** Zoals eerder aangegeven, willen we proberen een gedeelte van het proces te automatiseren. Dit kan worden gerealiseerd met een kleine uitbreiding in de eerder getoonde flowchart (Figuur 7). Deze uitbreiding zorgt ervoor dat klanten die op hoge leeftijd worden ingeschat met een hoge mate van zekerheid geen hulp van een zelfscanmedewerker nodig hebben. Dit biedt, zoals eerder genoemd, meerdere voordelen. Een ander voordeel van deze uitbreiding is dat het bestaande proces vrijwel volledig intact blijft, waardoor de implementatie eenvoudiger is dan wanneer het hele proces zou worden aangepast.

Er is ook rekening gehouden met het feit dat de mens als eindverantwoordelijke betrokken blijft (human-in-the-loop). Stel dat het AI-model niet een bepaalde mate van zekerheid bereikt, dan voert de zelfscanmedewerker alsnog een controle uit.

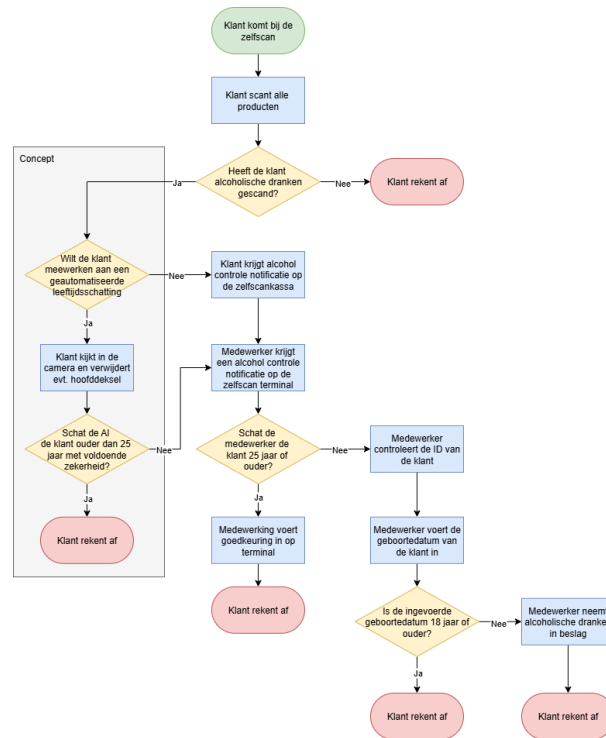


Fig. 7. Conceptuele uitbreiding aan het huidige proces van de zelfscan kassa's

**4.1.4 AI-Breakdown.** De AI speelt in dit proces de rol van eerste controlemechanisme (zie het blokje "Concept" in Figuur 7). Het systeem gebruikt gezichtsherkenning om de leeftijd van de klant in te schatten. De klant mag direct afrekenen zodra het model 25 jaar of ouder inschat met voldoende zekerheid. In alle andere gevallen grijpt de menselijke controle in. Dit principe valt onder augmentation: de AI ondersteunt en versnelt het proces, maar de uiteindelijke betrouwbaarheid wordt gewaarborgd doordat een medewerker ingrijpt bij twijfel. Zo ontstaat een hybride systeem waarbij technologie en mens elkaar aanvullen.

De AI is ontworpen met oog voor privacy. Beelden worden uitsluitend real-time gebruikt en niet opgeslagen. Alleen het resultaat van de beoordeling wordt gelogd, bijvoorbeeld of de AI de klant heeft doorgelaten of heeft doorverwezen naar de medewerker. Daarnaast is de fallback, die in het flow diagram zichtbaar is als de overgang naar de medewerker, cruciaal om bias of fouten te voorkomen: zodra de AI geen betrouwbaar oordeel kan vellen, wordt de beslissing veilig bij een mens gelegd.

**4.1.5 Gebruikte Design Patterns.** Binnen dit proces zijn meerdere ontwerpprincipes te herkennen die te zien zijn in de processtroom. Het belangrijkste is het Chain of Responsibility-pattern: de aanvraag tot alcoholverkoop gaat eerst langs de AI en wordt bij onvoldoende resultaat doorgeschakeld naar een medewerker. Daarnaast zien we het Strategy-pattern, waarbij er twee verschillende strategieën voor leeftijdscontrole bestaan (AI-schatting en handmatige controle), die afhankelijk van de situatie worden toegepast. Ook het idee van graceful degradation is zichtbaar in het flowdiagram: als de AI niet toepasbaar is (bijvoorbeeld opt-out, meerdere gezichten, afgedekt gezicht of technische storing), valt het systeem automatisch terug op de menselijke controle. Dit laat goed zien hoe augmentation in de praktijk werkt: het systeem wordt sneller en efficiënter door AI, maar blijft robuust

doordat de menselijke factor altijd beschikbaar blijft. De besluitlogica is deterministisch: bij een voorspelde leeftijd van  $\geq 25$  gaat de aankoop door; bij  $< 25$  volgt overdracht naar een medewerker.

Qua gebruikerservaring zijn er duidelijke aanwijzingen die passen bij error prevention en progressive disclosure. In het diagram wordt dit bijvoorbeeld zichtbaar bij de instructie om het hoofddeksel te verwijderen: de klant krijgt alleen aanvullende stappen wanneer dat nodig is. Tot slot sluit de inrichting aan bij privacy by design, omdat er zo min mogelijk gegevens worden verwerkt en klanten altijd de keuze hebben om de AI te weigeren zonder dat het koopproces volledig stopt.

Concretisering met paper prototyping (Iteratie 1  $\rightarrow$  Iteratie 2). In de eerste papieren prototypeversie, getest met twee CMD-studenten, is de hoofdflow gevalideerd: na expliciete opt-in (privacy by design) beoordeelt de AI de leeftijd en wordt de klant bij een voorspelde leeftijd van 25 jaar of ouder automatisch doorgelaten; in alle andere gevallen vindt overdracht plaats naar de medewerker. Daarmee worden het Chain of Responsibility-patroon (eerst AI, dan medewerker) en het Strategy-patroon (AI-route versus handmatige ID-controle) concreet zichtbaar. Graceful degradation komt naar voren zodra AI niet toepasbaar is (bijvoorbeeld bij opt-out, meerdere gezichten of afgedekte gezichten): het proces loopt dan via de menselijke route door. De interface ondersteunt progressive disclosure en error prevention door contextuele instructies zoals “kijk naar de camera / verwijder hoofddeksel” alleen te tonen wanneer dat nodig is. Zo fungeert de AI als versneller van routinegevallen, terwijl met human-in-the-loop de eindverantwoordelijkheid bij de medewerker blijft.

Op basis van de CMD-feedback is in de tweede iteratie de uitwerking aangescherpt en direct aan de design patterns gekoppeld. Ter versterking van privacy by design en transparantie is op het opt-in-scherm een informatieknoop toegevoegd met de boodschap dat tijdens de AI-scan geen beelden of persoonsgegevens worden opgeslagen en dat verwerking uitsluitend realtime plaatsvindt. Vóór de AI-scan wordt nu expliciet gecommuniceerd dat bij een voorspelde leeftijd onder 25 jaar een medewerker de controle uitvoert, wat de overdracht binnen Chain of Responsibility voorspelbaar maakt en progressive disclosure concreet invult. Voor error prevention is de invoer kwaliteit verbeterd: wanneer een hoofddeksel wordt gedetecteerd, vraagt de UI om dit te verwijderen; bij een hoofddoek volstaat dat het gezicht zichtbaar is. De besturing is uniform gemaakt met Terug- en Annuleer-knoppen op alle relevante schermen (uitgezonderd tijdens de feitelijke leeftijdsmeting), waarbij Annuleer de flow direct naar de medewerkeroute leidt een robuuste invulling van graceful degradation. Tot slot communiceert het eindscherm de uitkomst ondubbelzinnig met een ✓/doorgaan ×weigering, wat past bij error prevention en misinterpretaties voorkomt. Gezamenlijk concretiseren deze aanpassingen de eerder benoemde patronen in de feitelijke interactiestroom: routinegevallen worden autonoom afgehandeld (Strategy: AI-route), uitzonderingen en randgevallen worden betrouwbaar opgevangen door de medewerker (Strategy: handmatige route), en het systeem degradeert veilig zonder dat het koopproces stilvalt (graceful degradation), met blijvende waarborgen voor privacy by design en human-in-the-loop.

Transparantie. De interface bevat een informatieknoop waarin wordt toegelicht dat verwerking uitsluitend realtime plaatsvindt en er geen beelden of persoonsgegevens worden opgeslagen. Handover. Vóór de AI-scan wordt expliciet gecommuniceerd dat bij een voorspelde leeftijd onder 25 jaar een medewerker de controle uitvoert. Fallback. Wanneer de AI niet toepasbaar is, degradeert het systeem veilig naar handmatige controle zonder het afrekenproces te blokkeren.

**4.1.6 Niveau van Automatisering.** Volgens de schaal van Sheridan & Verplank (1978) valt dit ontwerp onder Level 9. Het systeem handelt autonoom: de AI schat de leeftijd en laat klanten automatisch passeren wanneer de voorspelde leeftijd  $\geq 25$  jaar is, zonder voorafgaande menselijke goedkeuring. De mens wordt alleen geïnformeerd en ingeschakeld wanneer het systeem dat nodig acht, bijvoorbeeld wanneer de voorspelde leeftijd  $< 25$  is, wanneer de klant voor opt-out kiest, wanneer meerdere gezichten of afdekking worden gedetecteerd, of bij technische storingen. In die gevallen vindt directe overdracht plaats naar de medewerker (human-in-the-loop), die de controle afrondt (bijv. visuele beoordeling/ID-check).

De grenswaarde van 25 jaar is, zoals vastgelegd in de juridische requirements (zie hoofdstuk 3), gebaseerd op de wetgeving die voorschrijft dat klanten die jonger lijken dan 25 jaar altijd gecontroleerd moeten worden. Deze grens vormt het beslismoment binnen het systeem: boven de 25 jaar verloopt de verificatie automatisch, daaronder wordt direct een medewerker ingeschakeld. Zo blijft het proces zowel wettelijk correct als betrouwbaar.

Deze inrichting combineert snelheid in routinetaken met betrouwbare afhandeling van uitzonderingen, sluit aan bij privacy by design (opt-in, geen opslag), en maakt het besluitvormingsproces duidelijk uitlegbaar binnen de winkeloperatie.

Tegelijkertijd blijft de inzet van AI een afweging tussen technische efficiëntie en maatschappelijke aanvaardbaarheid. Hoewel het systeem leeftijdsverificatie sneller en consistentere maakt, kan het nooit volledig losstaan van menselijke beoordeling. Door AI in te zetten als ondersteunend hulpmiddel in plaats van vervanging van menselijk toezicht, ontstaat een evenwichtige oplossing die technisch uitvoerbaar, maatschappelijk verantwoord en ethisch verdedigbaar is.

*4.1.7 Testopzet & Evaluatie.* Voor de evaluatie van het concept is gebruikgemaakt van paper prototyping en een conceptuele applicatie. In de eerste iteratie zijn ontwerpschetsen getest met CMD-studenten, gevolgd door een eenvoudige functionele applicatie die is getest met een zelfscanmedewerker en klanten. In de derde iteratie is een YOLO-model toegevoegd om meerdere gezichten te detecteren en zo te voldoen aan de juridische eis dat alle ID's aan dezelfde kassa worden gecontroleerd bij alcoholverkoop.

*Eerste Iteratie.* De paper prototyping toonde aan dat de flow in functioneel opzicht volledig voldoet, met uitzondering van de mogelijkheid voor klanten om de automatische leeftijdscontrole te annuleren. Daarnaast lieten de prototyping-sessies zien dat situaties met meerdere klanten tegelijk aan de zelfscankassa extra aandacht voor het procesontwerp vereisen.

*Tweede Iteratie.* De tweede iteratie hebben we de testen gestructureerd voorbereid en getest. Dit hebben we gedaan met behulp van een testplan (A). In dit testplan staat exact beschreven wat er in de test behandeld moet worden: denk aan de scope, doelen, rollen, eventuele resources en vragen die we na afloop wilden stellen. Ook kregen de gebruikers de mogelijkheid om zelf open feedback te geven. De belangrijkste observaties tijdens het testen zijn dat de applicatie en de flow functioneel duidelijk zijn, en dat het info-knopje dat we hebben toegevoegd op basis van de feedback op de eerste iteratie, tijdens de test niet gezien werd en te onduidelijk was.

*Derde Iteratie.* In de derde iteratie is een YOLO-model geïntegreerd om te testen of meerdere gezichten tegelijk kunnen worden gedetecteerd. Wanneer het model meerdere bounding boxes herkent, betekent dit dat er meerdere personen bij de zelfscankassa staan. In dat geval ontvangt de zelfscankassamedewerker automatisch een melding dat er meerdere mensen zijn geïdentificeerd, terwijl op het klantscherm een bericht verschijnt dat er een medewerker onderweg is.

## 4.2 Data

Dit hoofdstuk beschrijft de datakeuze en de verkenning van drie gezichtsdatasets voor leeftijdsschatting. We voeren een EDA uit om de geschiktheid vast te stellen. Per dataset beoordelen we de labelkwaliteit en de leeftijdsdekking. We kijken naar balans in geslacht en populatiegroepen. We controleren resolutie en beeldkwaliteit. We inspecteren pose, belichting en oclusies. Ook nemen we omvang en licentie mee. Het hoofdstuk sluit af met een gemotiveerde keuze voor de definitieve dataset en de gevolgen voor de modelarchitectuur en evaluatie.

*4.2.1 Overwogen Datasets.* Voor de datasetselectie zijn FG-NET, MORPH-II, IMDB-Clean en UTKFace geanalyseerd. FG-NET is met slechts 1.002 afbeeldingen te klein en bevat onvolledige metadata (zoals ontbrekende gender- en etniciteitsvelden), waardoor deze ongeschikt is. MORPH-II bevat wel 50.015 afbeeldingen en geslachtslabels,



maar kent een sterke genderimbalance (86,1% man) en mist etniciteitsdata, wat bias kan veroorzaken. Daarom is ook deze set niet gebruikt.

De UTKFace-dataset is geselecteerd voor de modelontwikkeling vanwege de duidelijke structuur, gebalanceerde verdeling van leeftijd en geslacht, en betrouwbare annotaties [5]. De IMDB-Clean-dataset is wel gecontroleerd, maar uiteindelijk uitgesloten omdat de leeftijdslabels vaak onnauwkeurig bleken door foutieve geboortedata en onjuiste extractie uit metadata, een probleem dat ook door eerdere studies is bevestigd (Rothe et al., 2015). Hierdoor zou opname van IMDB-Clean leiden tot inconsistente training en verminderde modelbetrouwbaarheid.

Daarom is uitsluitend UTKFace gebruikt als primaire bron voor het leeftijdsverificatiemodel.

**4.2.2 UTKFace.** De UTKFace-dataset is een openbaar beschikbare set van ruim 20.000 gezichtsafbeeldingen die veel wordt gebruikt binnen computer vision. Elke afbeelding is gelabeld via de bestandsnaam volgens het patroon `age_gender_race_timestamp.jpg`, met leeftijd (0–116 jaar), geslacht (0 = man, 1 = vrouw) en vijf etnische categorieën. De leeftijdslabels zijn deels automatisch geschat met een leeftijdsschattingalgoritme en vervolgens handmatig gecontroleerd door een annotator [5]. Door de variatie in pose, expressie en belichting is de dataset geschikt voor taken als leeftijdsschatting, geslachtsclassificatie en populatieherkenning. Tijdens de exploratieve data-analyse zijn bestanden met afwijkende naamstructuren verwijderd om labelruis te beperken. De leeftijdsgroep 20–50 jaar is het best vertegenwoordigd, terwijl het aantal voorbeelden boven de 70 jaar beperkt is.

Na herstructurering in leeftijdscategorieën blijkt dat de verdeling tot 17 jaar ongeveer gelijk is tussen mannen en vrouwen. Bij jongvolwassenen (18–30 jaar) zijn vrouwen oververtegenwoordigd, terwijl mannen domineren vanaf 31 jaar. Voor de etnische verdeling geldt dat de categorie ‘White’ in alle leeftijdsgroepen het meest voorkomt, gevolgd door Black en Asian. Indian en Other zijn het minst vertegenwoordigd.

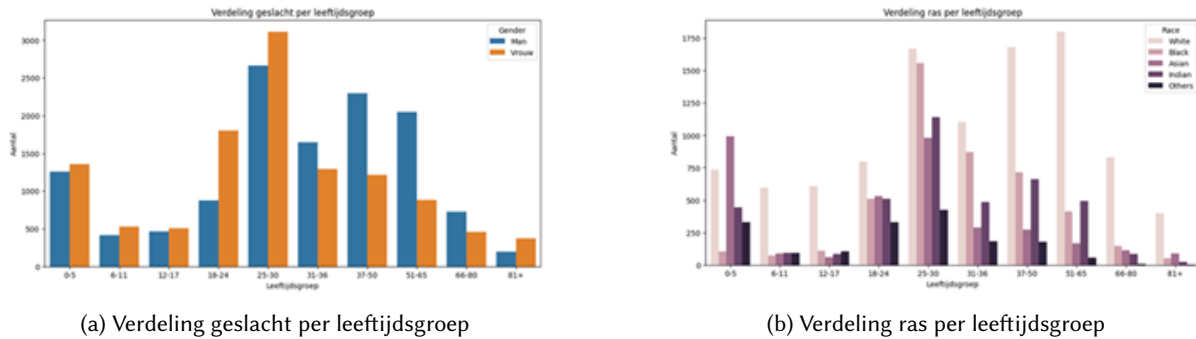


Fig. 8. Exploratieve data-analyse van de UTKFace-dataset

Een resolutieanalyse toonde aan dat de afbeeldingen gemiddeld  $638 \times 662$  pixels groot zijn, met aanzienlijke spreiding. Afbeeldingen kleiner dan  $200 \times 200$  pixels of extreem groot ( $> 2000 \times 2000$  pixels) zijn handmatig gecontroleerd; 610 lage-kwaliteitbeelden zijn verwijderd om ruis in het model te voorkomen.

Hoewel UTKFace een solide basis vormt dankzij de diversiteit in leeftijd, geslacht en etniciteit, blijft de totale omvang beperkt en zijn sommige subgroepen ondervertegenwoordigd. Daarom is de dataset uitgebreid met IMDB-clean.

**4.2.3 Data Preprocessing.** Voor de voorbereiding van de dataset is een pipeline ontwikkeld om gezichten uit de UTKFace-dataset te herkennen en extraheren. De aanpak bestond uit drie stappen: detectie, selectie en normalisatie, aangevuld met logging voor kwaliteitscontrole.



Fig. 9. Afbeeldingen kleiner dan 200 x 200

**4.2.4 YOLO-model.** Als detector is YOLOv8n-face gebruikt. Dit model is een lichte variant van de YOLOv8-architectuur, geoptimaliseerd voor gezichtsdetectie. Dit is ontworpen voor lagere rekencapaciteit met behoud van redelijke detectieprestaties, waardoor ze geschikt zijn voor real-time toepassingen in visuele systemen [31]. De confidence-drempel is op 0.25 gezet, gelijk aan de Ultralytics-standaard voor YOLO-inference, om recall te prioriteren tijdens het croppen[28].

Na de detectie kreeg elke afbeelding een status: no\_face bij geen gezichten en multi\_face bij meerdere gezichten. Deze categorieën zijn uitgesloten om de dataset te beperken tot afbeeldingen met één gezicht. Alleen dan werd de bounding box gebruikt. Een steekproef liet zien dat zwart-witafbeeldingen niet werd herkend en dat het model soms personen op de achtergrond als extra gezichten detecteerde.



Fig. 10. Voorbeelden waarin geen gezichten zijn getecteerd



Fig. 11. Voorbeelden waarin meerdere gezichten gedetecteerd zijn

Met een crop- en resize-functie zijn de gezichten uitgesneden en met 12% padding vergroot om verlies van randen te voorkomen. De uitsnedes zijn genormaliseerd naar 200×200 pixels en omgezet van BGR naar RGB, omdat OpenCV standaard BGR gebruikt terwijl deep learning modellen doorgaans RGB verwachten. Aangezien veel modellen zijn getraind op afbeeldingen met drie RGB-kanalen is dit formaat nodig om correcte representaties en prestaties te waarborgen [13].

**4.2.5 Data Splitting.** Uit de bestandsnamen van *UTKFace* zijn de labels leeftijd, gender en populatiecategorie geëxtraheerd. Leeftijd is vervolgens ingedeeld in tien klassen (*age\_class*). Voor een evenwichtige verdeling is een gecombineerd label opgebouwd, en zeldzame combinaties met minder dan zes voorbeelden zijn vooraf gefilterd om een stabiele stratificatie mogelijk te maken.

De dataset is daarna gestratificeerd gesplitst in 70/15/15 met *random\_state=42* voor reproduceerbaarheid. Om datalekken te voorkomen en herhaalbaarheid te waarborgen, zijn de bestanden fysiek naar aparte mappen gekopieerd. Deze procedure behoudt zowel de marginale als de gezamenlijke verdelingen van leeftijd, gender en populatiecategorie in alle subsets, en beperkt zo sampling bias. De verdeling per set is weergegeven in de onderstaande tabellen.

Table 2. Aantal per set.

Set	Aantal	% van totaal
Train	15 845	70%
Val	3 396	15%
Test	3 396	15%

Table 3. Verdeling naar geslacht.

Set	Man	Vrouw	% Man	% Vrouw
Train	8 289	7 556	52.3%	47.7%
Val	1 780	1 616	52.4%	47.6%
Test	1 782	1 614	52.5%	47.5%

Table 4. Verdeling naar etniciteit.

Set	White	Black	Asian	Other
Train	6 700	2 981	2 238	2 786
Val	1 431	636	478	597

### 4.3 Model

Deze sectie behandelt het volledige modelleringsproces. Te beginnen met de afweging tussen regressie of classificatie, waarbij de keuze is gevallen op regressie. Om een benchmark te stellen waarmee de prestaties van het uiteindelijke model vergeleken kan worden, is ook een baseline-model opgesteld (ResNet50). Vervolgens wordt de keuze voor het uiteindelijke model-architectuur (VGG16) gerechtvaardigd op basis van literatuur. Hierna worden per iteratie van het model alle aanpassingen besproken. Afsluitend worden een aantal prestatie- en fairness-metrics opgesteld om de prestaties van de modellen vast te leggen.

**4.3.1 Regressie vs Classificatie.** In dit onderzoek wordt leeftijdsschatting geformuleerd als een regressieprobleem. Leeftijd is een continue getal, en dat sluit goed aan bij de output van een regressiemodel. Daarmee wordt rekening gehouden met de grootte van de afwijking, wat bij training van classificatiemodellen niet wordt meegenomen. Daar geldt dat een foute voorspelling gewoon fout is, ongeacht hoe ver de voorspelling van de ground truth verwijderd is. Naast regressie, zal ook de hybride aanpak van Rothe et al [21] in overweging worden genomen [20]. Dit model wordt getraind als classificatie met cross-entropy lossfunctie, maar voorspelt leeftijd door de verwachtingswaarde over de softmax-verdeling te nemen. Hierdoor kan ook makkelijk MAE worden gebruikt om de voorspellingsfout uit te drukken. Deze aanpak behaalde in het paper van Rothe et al. [20] de laagste voorspellingsfout (MAE 3.21). Door altijd een continue waarde te houden als output van het model, komt er nog een extra voordeel bij. Wanneer het model gebruikt wordt in een andere context waarbij een andere leeftijd als grenswaarde geldt, hoeft het model niet opnieuw te worden getraind.

**4.3.2 Baseline: ResNet-50.** Vroege studies, zoals Niu et al. [17], gebruikten ondiepe, aangepaste CNN's als regressiebaseline. Recente literatuur toont echter een duidelijke verschuiving naar ResNet-50 als standaardbaseline vanwege de bewezen stabiliteit en generaliseerbaarheid [9, 16]. ResNet-50 is een 50-laags convolutioneel neurale netwerk opgebouwd uit bottleneck-residual blocks ( $1 \times 1-3 \times 3-1 \times 1$ ) en identity shortcuts die het degradatieprobleem bij diepe netwerken voorkomen. Het model is getraind op ImageNet en wordt veel gebruikt als backbone voor transfer learning [6].

In dit project dient ResNet-50, geïnitieerd met ImageNet-gewichten, als backbone. De convolutionele lagen blijven in de eerste fase bevroren om generieke visuele kenmerken te behouden [30]. Daarbovenop volgen een Global Average Pooling-laag, Dropout (0.2) en een Dense(1, lineaire) laag voor leeftijdsregressie [20]. Het model wordt gecompileerd met de Adam-optimizer ( $\alpha = 1 \times 10^{-3}$ ), vanwege de adaptieve leersnelheden en snelle convergentie [12].

Als verliesfunctie is Huber loss gekozen, omdat deze de precisie van MSE bij kleine fouten combineert met de robuustheid van MAE bij uitschieters [10]. Evaluatie gebeurt met MAE, dat direct de gemiddelde afwijking in jaren weergeeft. Tijdens de training worden de train-MAE en val-MAE per epoch gemonitord. Daarnaast worden *ModelCheckpoint*, *ReduceLROnPlateau* en *EarlyStopping* gebruikt om overfitting te beperken en de generaliseerbaarheid te waarborgen.

**4.3.3 Iteratie 1: VGG16.** Het doel van het modelleren is het behalen van een zo laag mogelijke mean absolute error (MAE). Om dit te bereiken is in de literatuur gezocht naar modellen die hierin het best presteren. Het paper van Rothe et al. [20] liet een implementatie zien gebaseerd op VGG16 die de laagst vindbare MAE behaalde (3.21 jaar). Dit paper rechtvaardigt de keuze om verder te modelleren met de VGG16-backbone die geïnitieerd is met de gewichten van ImageNet. Deze methode van transfer learning wordt vaak gebruikt omdat de geleerde feature-extracties goed generaliseren naar nieuwe data [11].

Om te werken met regressie is de basis architectuur van VGG16 aangepast. Zo is de output laag veranderd, van een softmax laag voor classificatie naar een lineaire laag met een enkele continue output. Zowel het model als de trainingssetup is geïmplementeerd in PyTorch [19]. De optimizer is Adam met een learning rate  $\alpha = 1 \times 10^{-4}$ . Als

loss functie is MAE gekozen, omdat deze direct de afwijking in jaren representeert en eenvoudig te interpreteren is.

**4.3.4 Iteratie 2: VGG16 - Aangepaste Regressie-head.** In de tweede iteratie is de regressie-head van het VGG16-model volledig herzien. Waar in de eerste iteratie alleen de laatste lineaire laag was vervangen door één uitgangsgeneuron ( $4096 \rightarrow 1$ ), is nu de volledige classifier van VGG16 verwijderd en vervangen door een eigen regressie-architectuur. Deze bestaat uit meerdere volledig verbonden lagen die in meer stappen toe werkt naar de enkele output ( $25088 \rightarrow 1024 \rightarrow 256 \rightarrow 1$ ). Deze aanpassing komt voort uit de observatie dat de prestaties van de eerste iteratie (met slechts één lineaire laag) onvoldoende waren. Met de aangepaste structuur werd onderzocht of extra capaciteit en regularisatie in de regressie-head konden bijdragen aan een verbeterde voorspelling.

**4.3.5 Iteratie 3: VGG16 - Classificatie met verwachtingswaarde.** De tweede iteratie gaf al grote vooruitgang in de MAE van het model. Om te kijken of deze prestatie nog verder verbeterd kon worden, is in de derde iteratie is getest met de classificatie met verwachtingswaarde, zoals geïmplementeerd door [20]. Hievoor is de laatste laag van het VGG16-model aangepast, zodat het 101 klassen als output heeft. Elke klasse fungeert als een leeftijd als heel getal. Om toch een continue waarde als output te hebben, wordt bij een voorspelling de verwachtingswaarde berekend over de kansverdeling over de klassen. Het model is getraind met een Cross-Entropy loss functie.

**4.3.6 Iteratie 4: VGG16 - DLDL.** In de vierde iteratie is het model uitgebreid met de *Deep Label Distribution Learning* (DLDL) methode, zoals voorgesteld door Gao et al. [7]. Bij deze aanpak wordt de leeftijd niet langer gemodelleerd als één enkel getal of als één klasse, maar als een labeledistributie die de waarschijnlijkheid van nabijgelegen leeftijden meeneemt. Elke trainingssample wordt hierbij gerepresenteerd als een Gaussische verdeling rond de werkelijke leeftijd, waardoor het model leert dat leeftijden in elkaars buurt verwant zijn aan elkaar. Dit maakt het mogelijk om te leren over nabijgelegen leeftijden die minder vaak voorkomen in de dataset. Tijdens het trainen wordt hiervoor de Kullback–Leibler-divergentie gebruikt als loss functie, wat gebruikelijk is bij DLDL.

4.3.7 *Evaluatiemetrics voor het model*. Als evaluatiemetric gebruiken we de Mean Absolute Error (MAE), omdat deze in jaren aangeeft hoe ver een voorspelling gemiddeld van de echte leeftijd afwijkt. De MAE is gedefinieerd als de volgende formule.

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

Dit wordt breed toegepast voor regressie-evaluatie. Een lagere MAE betekent een beter model (op dezelfde dataset/split). Deze maat is bovendien minder gevoelig voor uitschieters dan MSE/RMSE, omdat bij RMSE fouten worden gekwadrateerd en grote fouten daardoor zwaarder meetellen, terwijl MAE absolute fouten optelt [23, 29].

Naast de MAE, is er een andere metric die van belang is in de context van alcoholverkoop bij supermarkten. Namelijk het nalevingspercentage. Dit percentage is voor het model gedefinieerd als het percentage foutieve verkopen. Dat wil zeggen: het model voorspelde een leeftijd van 25 jaar of ouder, maar de werkelijke leeftijd van de klant was 17 of jonger. Het nalevingspercentage kan middels de volgende formule worden berekend.

$$C = 1 - \frac{|\{i \mid \hat{y}_i \geq 25 \wedge y_i \leq 17\}|}{N}$$

waarbij  $C$  het nalevingspercentage is,  $N$  het totaal aantal gecontroleerde personen,  $y_i$  de werkelijke leeftijd van persoon  $i$  en  $\hat{y}_i$  de door het model voorspelde leeftijd. De teller geeft het aantal personen dat in werkelijkheid 17 jaar of jonger is maar toch door het model als 25 jaar of ouder wordt voorspeld.

Om fairness inzichtelijk te maken wordt gebruikgemaakt van *MAE parity*, waarbij de gemiddelde absolute fout (MAE) per groep wordt vergeleken met de totale gemiddelde fout over alle voorbeelden.

Voor gender definiëren we de relatieve afwijking als

$$p_{\text{MAE}}^{\text{gender},g} = \frac{\text{MAE}_g}{\text{MAE}_{\text{overall}}} \times 100\%,$$

waarbij  $\text{MAE}_g = \frac{1}{|S_g|} \sum_{i \in S_g} |\hat{y}_i - y_i|$  de gemiddelde fout binnen gendergroep  $g$  voorstelt, en  $\text{MAE}_{\text{overall}} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$  de gemiddelde fout over de volledige dataset is. Een waarde van 100% betekent dat de fout gelijk is aan het modelgemiddelde, terwijl hogere of lagere waarden respectievelijk een slechtere of betere prestatie voor die groep aangeven.

Voor etniciteit geldt de definitie:

$$p_{\text{MAE}}^{\text{ethnicity},g} = \frac{\text{MAE}_g}{\text{MAE}_{\text{overall}}} \times 100\%,$$

met  $g \in G_{\text{ethnicity}}$  de etnische categorieën in de dataset. Ook hier geeft een waarde rond 100% aan dat de fout vergelijkbaar is met de totale gemiddelde fout, terwijl hogere waarden wijzen op relatief grotere voorspellingsfouten voor die groep.

## 5 Resultaten

In dit hoofdstuk worden de prestaties van de ontwikkelde modellen op de validatieset gepresenteerd, met de gemiddelde absolute fout (MAE) als hoofdmetriek. Eerst wordt de baseline met ResNet-50 besproken, gevolgd door de resultaten van de vier iteraties van het VGG16-model. Daarbij wordt aandacht besteed aan zowel de nauwkeurigheid als de fairness van de voorspellingen. Tot slot worden de belangrijkste kwaliteitsaspecten, zoals robuustheid, schaalbaarheid en uitlegbaarheid, kort geëvalueerd.

### 5.1 Resultaten Baseline (ResNet-50)

Het ResNet-50 model dat als baseline dient, haalde op de validatieset een MAE van 7.87. In Figuur 12 is de MAE per leeftijd te zien.

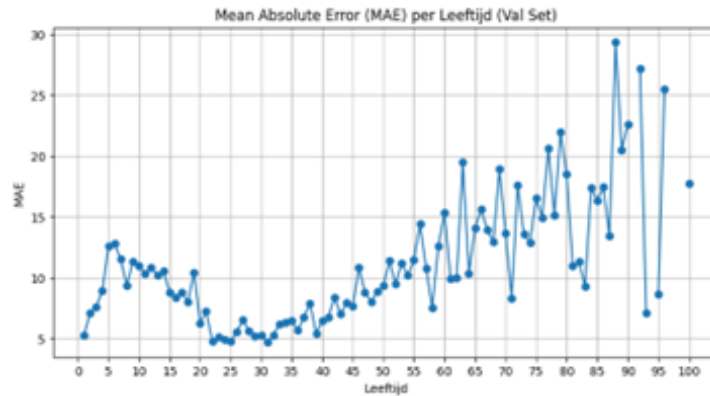


Fig. 12. Training results for ResNet-50.

Deze laat zien dat de fout het laagst is rond 25–40 jaar. In dit bereik voorspelt het model het nauwkeurigst. Dit sluit aan bij de EDA van de UTKFace dataset, waar de datadichtheid het hoogst was in deze groep. Vanaf ongeveer 50 jaar neemt de MAE duidelijk toe en tussen 80–95 jaar zijn scherpe pieken te zien.

## 5.2 Resultaten VGG16

Tijdens het modelleren van de verdere iteraties gebaseerd op de VGG16-backbone, zijn per iteratie de prestatie metrics en fairness metrics bijgehouden. Tijdens het trainen is telkens het model met de laagste MAE op de validatieset gekozen. Tabel 5 geeft een duidelijk overzicht van de progressie door de iteraties heen, inclusief de MAE-parity per geslacht en etnische groep.

Table 5. Modelprestaties en MAE-parity per groep.

Model	Prestaties		Parity (fairness)						
	Val MAE	Nalevingspercentage	Male	Female	White	Black	Asian	Indian	Other
Iteratie 1	10.47	97.50	97.16%	100.66%	120.56%	81.15%	89.94%	85.59%	59.46%
Iteratie 2	6.76	98.21	100.91%	97.48%	110.05%	104.79%	79.75%	88.24%	77.21%
Iteratie 3	5.43	98.79	104.85%	93.48%	108.61%	109.24%	72.87%	94.06%	79.69%
Iteratie 4	<b>5.18</b>	<b>98.99</b>	104.01%	93.99%	107.05%	114.57%	72.84%	97.77%	68.03%

Het is duidelijk te zien dat iteratie 4 het hoogste scoort op beide prestatie-metrics, met een validatie MAE van 5.18 en een nalevingspercentage van 98.99%. Als we kijken naar de fairness metrics is er niet meteen duidelijk een verband óf verbetering te zien door de iteraties heen. Wel is te zien dat afbeeldingen uit de ethniciteiten Asian en Other over het algemeen beter scoren vergeleken met de algehele MAE. Als we kijken naar Iteratie 4, zien we dat de grootste negatieve afwijking ten opzichte van de algehele MAE bij de ethniciteit Black zit. Deze afwijking van 114.57% kan vertaald worden naar een verschil in MAE van 0.72 jaar. Positieve afwijkingen (beter dan de algehele MAE) worden niet als oneerlijk gezien, en worden niet meegenomen in de evaluatie.

Om wat meer informatie te geven over de leeftijds groepen waar de fouten zich concentreren, kan specifiek gekeken worden naar de MAE per leeftijd. Voor elk van de iteraties is dit geplot in Figuur 13.

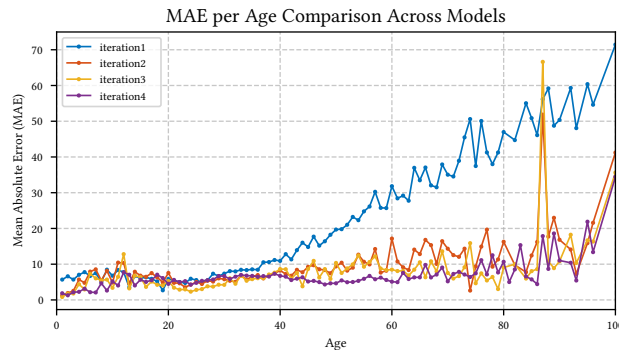


Fig. 13. MAE per leeftijd op de validatieset voor opeenvolgende iteraties.

De grafiek toont dat de fout bij jonge leeftijden het hoogst is en vervolgens afneemt richting de middelbare leeftijden. Tussen de leeftijden van 20 en 40 jaar presteert iteratie 3 het beste, maar onder de 20 jaar behaalt iteratie 4 duidelijk lagere fouten. Dit verklaart mogelijk het hogere nalevingspercentage van iteratie 4, aangezien leeftijdsvoorspellingen in de risico-categorie accurater zijn.



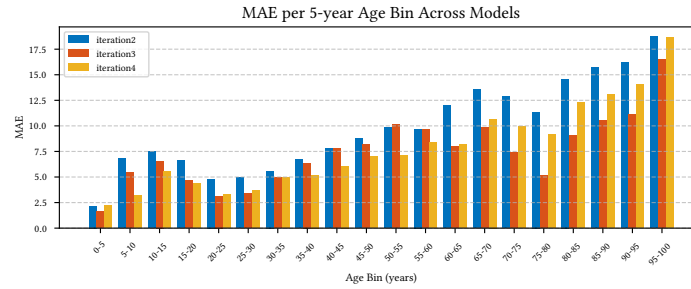


Fig. 14. MAE per leeftijdsbin op de validatieset.

Dit wordt ook weerspiegeld in Figuur 14. Deze grafiek geeft overzichtelijk aan wat de MAE per leeftijdsgroep (in bins van 5 jaar) is. Iteratie 1 is hier buiten beschouwing gelaten omdat deze zodanig veel slechter presteert dan verdere iteraties, dat het niet nodig is om dit nogmaals te vergelijken.

**5.2.1 Kwaliteitscriteria.** Robuustheid betekent dat een model stabiel blijft bij kleine verstoringen [8, 32]. Een gangbare test is het toevoegen van Gaussiaanse ruis aan invoerbeelden, omdat dit in de praktijk ook voorkomt, bijvoorbeeld door compressie of belichting (Zhou et al., 2021). In dit experiment is deze ruis toegevoegd aan de validatieset. De MAE steeg hierdoor naar 17.89. Dit kan komen doordat het model alleen getraind is op schone beelden en patronen daardoor verstoord raken bij ruis. In de praktijk kan dit leiden tot minder betrouwbare voorspellingen.

Scalability betreft opschalen naar grotere batches en datasets [3]. TensorFlow ondersteunt dit via efficiënte batching en GPU-distributie (Abadi et al., 2016). Op de validatieset bleef de MAE 7.8774 bij  $bs=32$ , 64 & 128. De inferentietijd was 28.69s ( $bs=32$ ), 27.23s ( $bs=64$ ) en 31.00s ( $bs=128$ ). Dit duidt op stabiele prestaties bij opschalen. Met de snelste doorvoersnelheid bij  $bs=64$ .

Explainability beschrijft hoe uitlegbaar de voorspellingen zijn. Voor beeldmodellen wordt vaak Grad-CAM gebruikt, waarmee duidelijk wordt welke beeldregio's bijdragen aan de voorspelling [24]. Dit vergroot de transparantie en het vertrouwen in het model.

Model complexity gaat over de omvang en het aantal parameters. Een complex model kan meer patronen leren, maar vergroot ook het risico op overfitting en vraagt meer rekenkracht. Het getrainde model telt 23.589.761 parameters, wat duidt op een groot leervermogen maar ook hoge eisen aan data en resources.

Resource demand gaat over rekenkracht en is belangrijk voor praktische inzetbaarheid [27]. In dit project duurde de training  $\pm 2$  uur voor 50 epochs met batch size 32 op GPU-acceleratie. Dit betekent dat het model rekentechnisch veeleisend is maar operationeel inzetbaar op één GPU.

## 6 Evaluatie

De ontwikkelde oplossing functioneert als hybride systeem: het AI-model schat realtime de leeftijd van klanten in en laat alleen personen waarvan de voorspelde leeftijd 25 jaar is zelfstandig passeren. In alle andere gevallen schakelt het systeem automatisch een medewerker in (human-in-the-loop). Hierdoor blijft het proces betrouwbaar, neemt de werkdruk af en verloopt de doorstroming soepeler. Er wordt een vaste drempel van 25 jaar voor automatische goedkeuring gehanteerd, omdat de MAE in de aangrenzende leeftijdsgroepen (15–25 jaar) laag genoeg is om het risico op false pass (<18 toch doorlaten) te minimaliseren.

### 6.1 Evaluatie van Modelresultaten

De uiteindelijke testresultaten (Tabel 6) bevestigen dit: Test MAE = 5,06 met een nalevingspercentage = 99,19%; de MAE per 5-jaarsbin ligt in 15–25 jaar 3–5, wat de keuze voor de drempel 25 onderbouwt. Fairness vraagt blijvende aandacht: MAE-parity varieert per groep (Male 103,61%, Female 96,43%; White 110,49%, Black 112,78%, Asian 76,76%, Indian 81,98%, Other 77,74%). Operationeel moeten daarom fouten per groep gemonitord blijven en vallen twijfelgevallen standaard onder handmatige controle.

Table 6. Modelprestaties met MAE-parity per groep op de testset.

Model	Prestaties		Parity (fairness)						
	Test MAE	Nalevingspercentage	Male	Female	White	Black	Asian	Indian	Other
Final model	5.06	99.19	103.61%	96.43%	110.49%	112.78%	76.76%	81.98%	77.74%

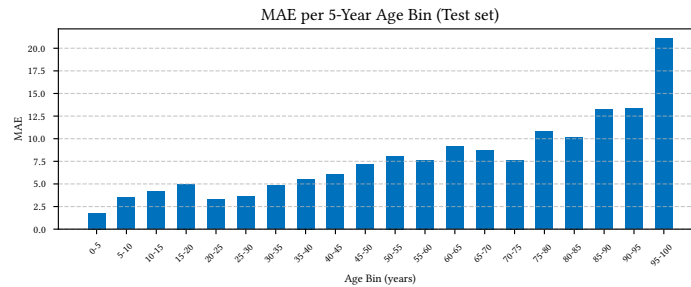


Fig. 15. MAE per leeftijdsgroep op de testset.

### 6.2 Onderbouwing van wijzigingen tussen concepten

Gedurende de ontwikkeling zijn de design patterns stapsgewijs verbeterd. De eerste versie is bijvoorbeeld uitgebreid met een fallback-mechanisme naar medewerkers (Chain of Responsibility). Ook is bewust gekozen voor graceful degradation: wanneer de voorspelde leeftijd < 25 jaar is of technisch faalt, schakelt het altijd over op handmatige controle. Daarnaast is progressive disclosure toegepast, zoals bij instructies om hoofddekseksels te verwijderen, deze verschijnen pas wanneer dit nodig is. Door deze aanpassingen blijft het ontwerp zowel gebruiksvriendelijk als robuust.

### 6.3 Volledig geïntegreerde evaluatie van het model binnen het concept

Technische resultaten zijn telkens vertaald naar functionele en ethische overwegingen: een hogere nauwkeurigheid maakte automatische goedkeuring mogelijk bij meer gevallen, terwijl privacy by design (zonder opslag van beelden) en menselijke eindverantwoordelijkheid gewaarborgd bleven. Deze keuzes zijn getoetst in gesprekken met vakgenoten, studenten en een supermarkteigenaar. Hieruit kwamen verschillende randvoorwaarden naar voren: vrijwillige deelname, directe verwijdering van beeldmateriaal, heldere communicatie naar klanten en in alle gevallen de mogelijkheid tot menselijke herbeoordeling. Zo is de oplossing zowel technisch solide als maatschappelijk verantwoord vormgegeven. Samengevat: de combinatie van drempel 25 en de lage MAE in de kritieke band 15–25 resulteert in hoge naleving; fairnessverschillen worden gemonitord en procedureel afgevangen (opt-in, geen opslag, human-in-the-loop), zodat ongewenste consequenties van AI-gebruik beperkt blijven.

## 7 Discussie

De resultaten tonen aan dat het uiteindelijke model zowel op nauwkeurigheid als naleving sterk presteert. Met een test-MAE van 5.06 jaar en een nalevingspercentage van 99.19% voldoet het model aan de vooraf vastgestelde eisen. De fout ligt op een vergelijkbaar niveau met wat eerder in literatuur werd bereikt voor leeftijdsschatting met VGG-achtige modellen.

De fairness-metrics laten zien dat de prestaties per demografische groep enigszins verschillen. Zo is de MAE lager voor vrouwen (96.43%) dan voor mannen (103.61%), vergeleken met de overall MAE van 5.06 jaar. Tussen etnische groepen zijn de verschillen groter: de groepen *Black* (112.78%) en *White* (110.49%) hebben hogere fouten, terwijl *Asian* (76.76%) en *Other* (77.74%) beter presteren. Om het model ethisch verantwoord te achten, zou het model niet veel slechter moeten presteren op minderheden. De enige etnische minderheid die slechter presteert dan *White* is *Black*. Deze groep heeft een MAE op de test-set van 5.71 jaar. Ondanks dat dit verschilt met de overall MAE, is het verschil niet te groot om voor significant afwijkende prestaties te zorgen in de implementatie.

De foutverdeling over leeftijden bevestigt dat het model het best presteert bij middelbare leeftijden en hogere fouten vertoont bij jonge en oudere gezichten. Iteratie 4, waarin Deep Label Distribution Learning (DLDL) werd toegevoegd, heeft geleid tot stabielere prestaties over de gehele leeftijdsrange, vooral bij lagere leeftijden. Het nalevingspercentage van 99.19% voldoet aan de vooraf gesteld eis en geeft aan dat het model effectief kan inschatten of een klant oud genoeg is.

## 8 Conclusie

Het ontwikkelde systeem voor AI-gebaseerde leeftijdsverificatie voldoet aan de in Tabel 1 vastgestelde *must-requirements*. De evaluatie op de testset toont aan dat het model een Mean Absolute Error (MAE) van 5.06 behaalt en een nalevingspercentage van 99.19%. Hiermee voldoet het aan de technische eisen voor nauwkeurigheid en naleving. Binnen de leeftijdsgroepen 15–30 jaar blijft de fout ruim onder de grens van 5 jaar, terwijl in overige leeftijdsgroepen de MAE niet onder 25 jaar uitkomt, conform de specificatie. De inference time bleef binnen de limiet van 500 ms, waarmee ook de prestatie-eis is behaald.

Het systeem verwerkt geen identificeerbare gegevens, slaat geen beelden op en functioneert uitsluitend op het moment van verificatie. Daarmee wordt voldaan aan de AVG-artikelen 4, 9, 12 en 13, en de relevante bepalingen uit de Uitvoeringswet AVG. Het gebruik van AI is vrijwillig en transparant: klanten worden vooraf geïnformeerd, en de camera is standaard afgedekt totdat expliciet toestemming wordt gegeven. De AI-oplossing ondersteunt de medewerker in plaats van deze te vervangen, wat de vereiste *human-in-the-loop*-structuur waarborgt.

Uit de fairness metrics blijkt dat de prestaties per geslacht en etniciteit verschillen, maar binnen acceptabele marges blijven. De grootste afwijking (112.21%) komt voor bij de etnische groep *Black*, terwijl andere groepen lager scoren. Hoewel hiermee aan de functionele eisen wordt voldaan, blijft voortdurende monitoring van bias noodzakelijk om oneerlijke behandeling te voorkomen. De requirement over training op een diverse dataset is ook gerealiseerd; de dataset bevat variatie in leeftijd, geslacht en etniciteit, maar is niet volledig uitgebalanceerd. Wel is gebruik gemaakt van de enige dataset die überhaupt inzicht geeft in verdeling van etniciteit. Een diversere dataset had niet gekozen kunnen worden.

De impact van het systeem op individu en maatschappij is positief te beoordelen. De oplossing respecteert de privacy van klanten, voorkomt overmatige gegevensverwerking en ondersteunt naleving van de alcoholwet zonder buitenproportioneel discriminerende besluitvorming. .

Samenvattend voldoet de ontwikkelde oplossing aan alle kernvereisten op technisch, juridisch en ethisch vlak. De toepassing is uitvoerbaar, privacyvriendelijk en betrouwbaar.

## A Testplan

# Testplannen

---

In beide testplannen streven we ernaar om te testen of de klant de applicatie zelfstandig kan gebruiken. Dit doen we door simpele en korte scenario's / taken te geven zonder enige context.

## Testplan Klantpaneel

**Scope:** Elk mogelijke scenario testen van het klantpaneel

**Objective:** Erachter komen of dat de user zelfstandig door de applicatie kan navigeren zonder externe hulp.

**Test Strategy:** Task based testing. Ik geef de user 2 taken zonder context.

- Task 1: Allow age verification
- Task 2: Deny age verification

**Roles:** De user

### Resources

We hebben 2 papiertjes waarop elk een taak op staat. De taken zijn expres zo vaag mogelijk gemaakt om te testen of de user er zelf uit komt met enkel het gebruik van de applicatie zonder extern hulp. Zo kunnen we testen of de applicatie duidelijk genoeg communiceert richting de user. De user krijgt de eerste keer de eerste task, vervolgens herhalen we de test met de tweede task.

- Digitale Prototype versie 2
- Papiertje met Task 1
- Papiertje met Task 2

**Post Interview Vragen** Na afloop van de tests stel ik de volgende vragen:

- Liep je ergens tegenaan in het process?
- Was alles duidelijk? zo niet wat?
- zag je het informatie knopje? zo wel, waarom klikte je niet?
- Wil je zelf nog iets inbrengen/opmerken?

## Testplan Medewerkers Paneel

**Scope:** Elk mogelijke scenario van het medewerkerspaneel testen

**Objective:** Erachter komen of de user zelfstandig door de applicatie kan navigeren zonder externe hulp. Bij deze test wordt wel uitgelegd dat er een tweede personage meedoet aan de test en dat er interactie zal plaats vinden tussen de twee personen. Verder wordt er geen context meegegeven.

**Test Strategy:** Task Based Testing. Ik geef de user taken die uitgevoerd moeten worden.

- Task 1: Beoordeel een klant ouder dan 25 jaar
- Task 2: Beoordeel een klant jonger dan 25 jaar en identificeer de klant (ID 1)
- Task 3: Beoordeel een klant jonger dan 25 jaar en identificeer de klant (ID 2)

**Roles:** De User en de klant met de neppe ID's

**Resources:** We hebben 2 neppe papieren ID's. Deze worden bijgehouden door de klant. De eerste keer krijgt de klant Nep ID 1 en de tweede keer krijgt de klant Nep ID 2. De verschillende ID's lokken beide een ander scenario uit binnen de applicatie. Nu testen we of de user de informatie vindt en correct leest en de juiste knoppen kan vinden en drukken.

- Digitale Prototype versie 2
- Nep ID 1 (jonger dan 18)
- Nep ID 2 (ouder dan 18)

### Post Interview Vragen

- Liep je ergens tegenaan in het proces?
- Was alles duidelijk? Zo niet, wat?
- Wil je zelf nog iets inbrengen/opmerken?

## References

- [1] [n. d.]. Algemene verordening gegevensbescherming (AVG). <https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=CELEX:32016R0679>. Verordening (EU) 2016/679 van het Europees Parlement en de Raad van 27 april 2016. Geraadpleegd: 2025-10-03.
- [2] [n. d.]. Uitvoeringswet Algemene verordening gegevensbescherming (UAVG). <https://wetten.overheid.nl/BWBR0040940/2018-05-25>. Wet van 16 mei 2018, houdende regels ter uitvoering van de AVG. Geraadpleegd: 2025-10-03.
- [3] Martin Abadi et al. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *arXiv preprint* (2016). arXiv:1603.04467 <https://arxiv.org/abs/1603.04467>
- [4] Centraal Bureau Levensmiddelenhandel. 2024. Beleid verantwoorde alcoholverkoop in de supermarkt. [https://www.cbl.nl/app/uploads/2019/03/Beleid-verantwoorde-alcoholverkoop-in-de-supermarkt\\_maart-2024.pdf](https://www.cbl.nl/app/uploads/2019/03/Beleid-verantwoorde-alcoholverkoop-in-de-supermarkt_maart-2024.pdf). Accessed: 2025-09-16.
- [5] Yiqing Chen. 2019. Comment on the work of Zhang et al. (2017, Journal of Inequalities and Applications). *Journal of Inequalities and Applications* 2019 (12 2019). doi:10.1186/s13660-019-2142-3
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2009), 248–255.
- [7] Bin Gao, Chao Chen, Shuai Zhang, Yuting Zhao, and Xing Zhang. 2017. Deep Label Distribution Learning for Apparent Age Estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5532–5541. doi:10.1109/CVPR.2017.587
- [8] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2015. Explaining and Harnessing Adversarial Examples. *International Conference on Learning Representations (ICLR)* (2015). <https://arxiv.org/abs/1412.6572>
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
- [10] Peter J. Huber. 1964. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35, 1 (1964), 73–101.
- [11] Anders Kensert, Peter J. Harrison, and Ola Spjuth. 2019. Transfer Learning with Deep Convolutional Neural Networks for Classifying Cellular Morphological Changes. *SLAS Discovery* 24, 4 (2019), 365–377. doi:10.1177/2472555218818756
- [12] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)* (2015). <https://arxiv.org/abs/1412.6980>
- [13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, Vol. 25. Curran Associates, Inc., 1097–1105. [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
- [14] Anna Levitina. 2025. *Automated Decision-Making and the Right to Human Intervention under GDPR*. [https://www.cidob.org/sites/default/files/2025-02/121-143\\_ANNA%20LEVITINA\\_ANG.pdf](https://www.cidob.org/sites/default/files/2025-02/121-143_ANNA%20LEVITINA_ANG.pdf) Accessed: 2025-09-29.
- [15] Per Lindmark. 2021. *User Perceptions of Facial Recognition in Retail Environments*. <https://www.diva-portal.org/smash/get/diva2:1627145/FULLTEXT01.pdf> Accessed: 2025-09-29.
- [16] Zhihao Liu, Chao Zhang, and Xian Wu. 2023. A Survey on Deep Learning-Based Age Estimation from Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
- [17] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. 2016. Ordinal Regression with Multiple Output CNN for Age Estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4920–4928.
- [18] National Institute of Standards and Technology (NIST). 2024. *Face Recognition Vendor Test (FRVT) – Age Estimation Accuracy Evaluation Report (AEV)*. Technical Report NIST IR 8450. NIST. <https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8450.pdf>
- [19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 32. Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf)
- [20] Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2018. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. *International Journal of Computer Vision* 126, 2-4 (2018), 144–157. doi:10.1007/s11263-016-0940-3
- [21] Rasmus Rothe, Radu Timofte, and Luc Van Gool. 2015. DEX: Deep EXpectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*. [https://openaccess.thecvf.com/content\\_iccv\\_2015\\_workshops/w11/papers/Rothe\\_DEX\\_Deep\\_EXpectation\\_ICCV\\_2015\\_paper.pdf](https://openaccess.thecvf.com/content_iccv_2015_workshops/w11/papers/Rothe_DEX_Deep_EXpectation_ICCV_2015_paper.pdf)
- [22] R. Safa, N. Petrovic, V. Nozick, and L. Cevallos-Torres. 2024. Human age prediction based on facial patterns using CNN. *CETI Journal* (2024). <https://www.ceti.reapress.com/journal/article/view/30/50>
- [23] scikit-learn developers. n.d.. *sklearn.metrics.mean\_absolute\_error*. [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean\\_absolute\\_error.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html) Accessed: 2025-09-29.
- [24] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 618–626. doi:10.1109/ICCV.2017.74

- [25] A. Singh, S. Nagpal, R. Singh, and M. Vatsa. 2018. Class specific mean autoencoder for age-invariant face verification. *arXiv preprint* (2018). <https://arxiv.org/pdf/1803.07385>
- [26] Staatscourant. 2012. Richtsnoeren identificatie en verificatie van persoonsgegevens (Gebruik van 'kopietje paspoort' in de private sector). <https://zoek.officielebekendmakingen.nl/stcrt-2012-14741.html>. Accessed: 2025-09-23.
- [27] Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the International Conference on Machine Learning (ICML)*. <https://arxiv.org/abs/1905.11946>
- [28] Ultralytics. 2025. Model prediction with Ultralytics YOLO — Predict: Inference arguments. <https://docs.ultralytics.com/modes/predict/#inference-arguments>.
- [29] Cort J. Willmott and Kenji Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, 1 (2005), 79–82. doi:10.3354/cr030079
- [30] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How Transferable Are Features in Deep Neural Networks?. In *Advances in Neural Information Processing Systems*, Vol. 27.
- [31] X. Zhou and Y. Li. 2024. Masked face detection algorithm based on YOLOv8. In *Proceedings of SPIE 13690, Sixth International Conference on Computer Vision, Image and Deep Learning*. SPIE. doi:10.1117/12.3068858
- [32] Yinpeng Zhou, Nan Ye, Jie Feng, Zhiyong He, Shuo Yang, Hao Zhang, and Shirui Pan. 2021. A Survey of Robustness in Deep Learning. *arXiv preprint* (2021). arXiv:2011.01529 <https://arxiv.org/abs/2011.01529>