

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

Dit is waar de samenvatting komt

Authors' Contact Information: Kaan Gögcay, Amsterdam University of Applied Sciences, Amsterdam, Nederland, kaan@email.com; Yusra El Ghazi, Amsterdam University of Applied Sciences, Amsterdam, Nederland, yusra@email.com; Merve Halici, Amsterdam University of Applied Sciences, Amsterdam, Nederland, merve@email.com; Leon van Dijk, Amsterdam University of Applied Sciences, Amsterdam, Nederland, 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 Gerelateerde Literatuur	7
2.4 Overwegingen voor het inzetten van AI	9
3 Requirements	10
4 Methodologie	11
4.1 Concept	11
4.2 Data	13
4.3 Model	15
5 Resultaten	18
5.1 Resultaten ResNet-50	18
5.2 Resultaten VGG16	19
6 Evaluatie	21
6.1 Evaluatie van Modelresultaten	21
6.2 Onderbouwing van wijzigingen tussen concepten	21
6.3 Volledig geïntegreerde evaluatie van het model binnen het concept	21
7 Discussie	22
8 Conclusie	22
References	22

*

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 zelfscancaplein 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 leeftijdsherkenning worden ontwikkeld, geëvalueerd en geïmplementeerd bij zelfscankassa's in supermarkten om het proces van leeftijdsverificatie 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 [23].

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 leeftijdscontroleproces bij zelfscankassa's, en welke criteria bepalen of een computer-vision oplossing dit proces kan verbeteren?
- (2) Welke computer-vision technieken en architecturen zijn geschikt om een leeftijdsherkenningsmodel te trainen?
- (3) Welke datasets zijn geschikt om een leeftijdsherkenningsmodel te ontwikkelen en te testen in de context van supermarkten?
- (4) Hoe presteert het ontwikkelde leeftijdsherkenningsmodel op de gekozen dataset, gemeten aan de hand van relevante evaluatiemetrics en de vooraf opgestelde requirements?
- (5) Op welke manier kan het model worden geïntegreerd en toegepast in zelfscankassa's binnen supermarkten zodat het praktisch inzetbaar is in de verkoop van alcohol?

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 afrekent 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 opstopingen 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 Regeling

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 bovenindien 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.

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

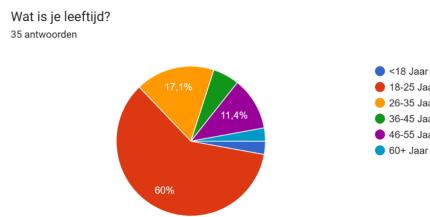


Fig. 1. Leeftijdsverdeling respondenten (N=35).

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.

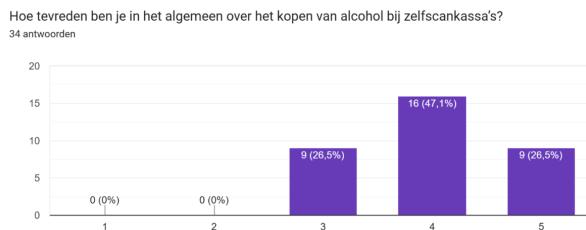


Fig. 2. Tevredenheid alcoholaankopen zelfscankassa (N=34).

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.

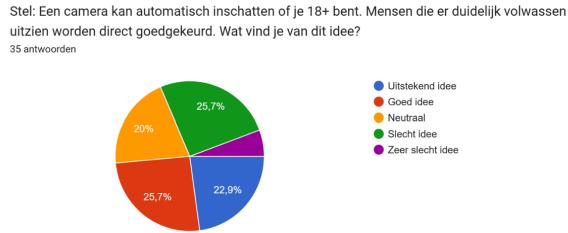


Fig. 3. Mening over automatische leeftijdsschatting via camera (N=35).

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.

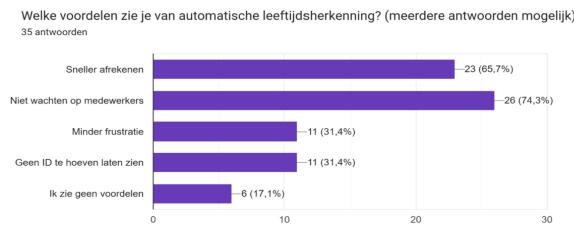


Fig. 4. Voordelen automatische leeftijdsherkenning (N=35, meerdere antwoorden mogelijk).

Voor acceptatie van beeldherkenning stellen klanten verschillende voorwaarden. Vrijwillig gebruik is het belangrijkst (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.

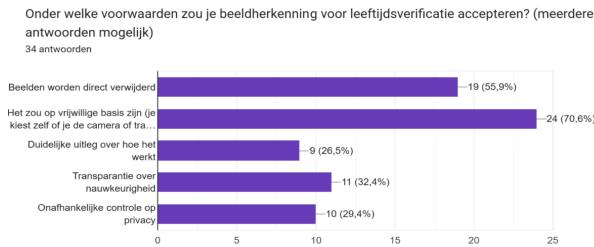


Fig. 5. Voorwaarden voor acceptatie beeldherkenning (N=34, meerdere antwoorden mogelijk).

2.2.4 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 supermarket 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 leeftijdsgebonden 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.

2.2.5 Conclusie. Uit het gebruikersonderzoek komen drie hoofdtekortkomingen 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 belangrijkst 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.

2.3 Gerelateerde Literatuur

Safa [18] 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 [17] 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 ϵ -fout 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. [22] 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.

2.4 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.

2.4.1 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 [14].

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.

2.4.2 Maatschappelijke overwegingen. Uit een interview met een Jumbo-ondernemer kwam naar voren dat een dergelijk systeem vooral in stedelijke context voordeelen 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 [13].

2.4.3 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 [12].

3 Requirements

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 afgenoem, 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 controlefunctie 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
Technisch	Het model scoort op de testset een Mean Absolute Error lager dan 5.0 jaar	Must	Ontwikkelaar
Technisch	Het model heeft een inference time van minder dan 500ms	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 twijfel gevallen	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 ontwerprincipes, 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, zelfscankassemedewerkers 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 verminderd 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.

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.

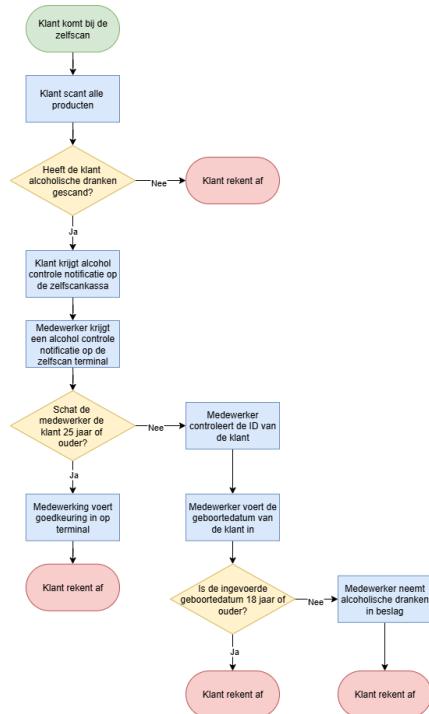


Fig. 6. Flowchart van het huidige proces aan de zelfscankassa's in supermarkten

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, het resultaat is te vinden in Figuur 7. Deze uitbreiding zorgt ervoor dat klanten die op hoge leeftijd worden ingeschat met een hoge mate van zekerheid geen hulp van een zelfscannedewerker nodig hebben. Dit biedt meerdere voordelen, zoals eerder genoemd. 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 zelfscannedewerker alsnog een controle uit.

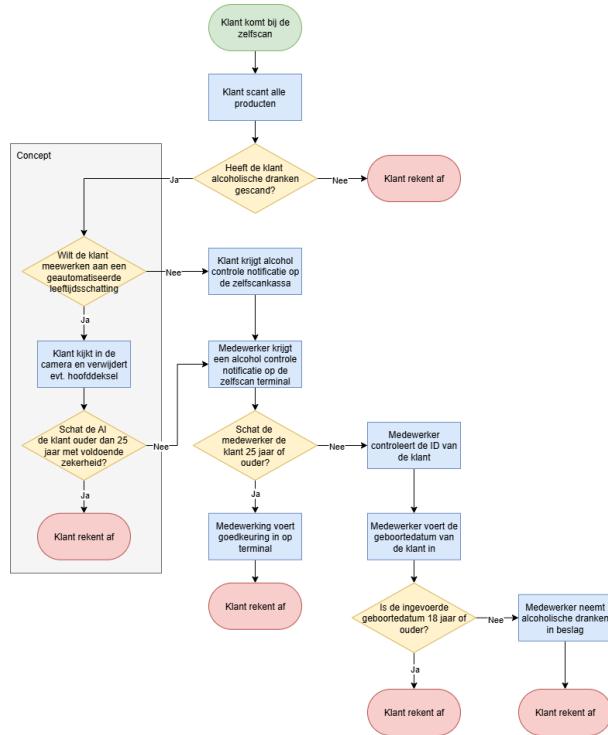


Fig. 7. Conceptuele uitbreiding aan het huidige proces van de zelfscankassa's

4.1.4 AI-Breakdown. De AI speelt in dit proces de rol van eerste controlesysteem (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 ontwerprincipes 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 flow diagram: als de AI niet functioneert of geen zekerheid kan bieden, 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.

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.

4.1.6 Niveau van Automatisering. Binnen het model van levels of automation komt dit proces overeen met Level 5. De AI neemt beslissingen wanneer de uitkomst voldoende betrouwbaar is, bijvoorbeeld door zelfstandig vast te stellen dat een klant naar schatting 25 jaar of ouder is. In twijfelgevallen of bij weigering wordt automatisch overgeschakeld naar een medewerker. Dit weerspiegelt het principe van augmentation met human in the loop: routinetaken worden versneld door AI, terwijl de mens verantwoordelijk blijft voor uitzonderingen en fouten, wat leidt tot een balans tussen efficiëntie en betrouwbaarheid.

4.1.7 Testopzet & Evaluatie. Voor de evaluatie van het concept is gebruikgemaakt van paper prototyping, uitgevoerd in twee iteraties met CMD-studenten. Zij kregen de flow in papieren vorm voorgelegd en werden geconfronteerd met praktijkscenario's, zoals het dragen van een pet of hoofddoek en de aanwezigheid van meerdere personen bij de kassa. Tijdens de tests werd geobserveerd hoe de studenten reageerden en welke vragen of opmerkingen zij formuleerden.

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.

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 leeftijds-dekking. We kijken naar balans in geslacht en populatiegroepen. We controleren resolutie en beeldkwaliteit. We inspecteren pose, belichting en occlusies. Ook nemen we omvang en licentie mee. Het hoofdstuk sluit af met een gemotiveerde keuze voor de definitieve dataset en de gevonden voordeelen voor de modelarchitectuur en evaluatie.

4.2.1 FGNET. De FGNET dataset is een kleine dataset die zich specifiek richt op leeftijdsprogressie. Dat houdt in dat een gros van de afbeeldingen uit de dataset van dezelfde personages komen. Deze eigenschap maakt de dataset erg uniek, maar de dataset heeft ook meerdere valkuilen. Zo heeft de dataset maar een totaal van 1002 afbeeldingen én ontbreken de velden gender en etniciteit. Door deze redenen lijkt het ons geen handige keuze om FGNET mee te nemen in ons verdere onderzoek aangezien het meer nadelen heeft dan voordeelen

4.2.2 UTKFace. Voor dit onderzoek is UTKFace gebruikt: meer dan 20.000 gezichten (leeftijd 0–116 jaar) met labels voor leeftijd, geslacht en etniciteit (vijf categorieën). Deze keuze sluit aan bij de requirement: “het model moet getraind worden op een diverse dataset (leeftijd, geslacht en etniciteit) om bias in de prestaties te vermijden.”

Een nadeel is dat de totale hoeveelheid data nog beperkt is en sommige leeftijds- of etnische groepen ondervertegenwoordigd zijn. Dit kan tot bias en een hogere fout in die subgroepen leiden; bovendien kunnen de labels ruis bevatten.

Het plan is om later extra data toe te voegen en de balans per groep te verbeteren. Daarmee wordt het model robuuster en voldoet het beter aan de requirement.

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.

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 [28]. De confidencedrempel is op 0.25 gezet, gelijk aan de Ultralytics-standaard voor YOLO-inference, om recall te prioriteren tijdens het croppen[25].

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 werden herkend en dat het model soms personen op de achtergrond als extra gezichten detecteerde.



Fig. 8. Voorbeelden waarin geen gezichten zijn getecteerd

Met een crop- en resize-functie zijn de gezichten uitgesneden en met 12% padding vergroot om verlies van randen te voorkomen. De uitsneden 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 [11].



Fig. 9. Voorbeelden waarin meerdere gezichten gedetecteerd zijn

4.3 Model

Deze sectie behandelt de afweging tussen classificatie en regressie voor leeftijdsschatting, de gemaakte keuzes in netwerkarchitectuur, en de iteratieve stappen in de modelontwikkeling. Zo wordt beargumenteerd welke baseline gebruikt wordt, en voor elke iteratie in het ontwikkelproces worden de aanpassingen weergegeven. Daarnaast worden de evaluatiemetrics uiteengezet waarmee de prestaties en de fairness van de modellen worden beoordeeld.

4.3.1 Regressie vs Classificatie. In dit onderzoek wordt leeftijdsschatting geformuleerd als een regressieprobleem. Regressie sluit beter aan bij het continue karakter van leeftijd en maakt het mogelijk om de fout direct te optimaliseren op een interpreteerbare maat zoals de Mean Absolute Error (MAE). Daarmee wordt rekening gehouden met de grootte van de afwijking, wat bij training van classificatiemodellen niet wordt meegenomen. Daarnaast zal ook de hybride aanpak van Rothe et al. (DEX) in overweging worden genomen [16]. 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. [16] de laagste voorspellingsfout (MAE 3.21).

4.3.2 Baseline: ResNet-50. Voor face age estimation varieerden baselines. Vroege studies, zoals Niu et al. (2016), gebruikten ondiepe, custom CNN's als regressiebaseline. Recente literatuur laat een duidelijke verschuiving zien naar ResNet-50 als standaardbaseline vanwege de bewezen stabiliteit en generaliseerbaarheid [21, 27]. ResNet-50 wordt explicet ingezet als uitgangspunt voor transfer learning bij leeftijdsvoorspellingen [21] en als referentiemodel voor portret-gebaseerde leeftijdsherkenning [27]. ResNet-50 is een 50-laags CNN met bottleneck-residual blocks ($1 \times 1 - 3 \times 3 - 1 \times 1$) en identity shortcuts. Deze residual-koppelingen verhelpen het degradatieprobleem bij diepe netten en vergemakkelijken de optimalisatie. Het model is op ImageNet getraind en wordt daardoor breed ingezet als generieke backbone voor transfer learning [7]. In dit project dient ResNet-50 (ImageNet-gewichten) als backbone. De convolutionele lagen blijven in de eerste fase bevroren om generieke visuele kenmerken te behouden [6]. Bovenop komt Global Average Pooling, gevolgd door Dropout (0.2) en een Dense(1, lineair) voor leeftijdsregressie [8].

Het model wordt gecompiled met Adam ($\text{lr} = 1\text{e-}3$) vanwege adaptieve leersnelheden en snelle convergentie [10]. Als verliesfunctie kiezen we Huber loss. Dit combineert de precisie van MSE bij kleine fouten met de robuustheid van MAE bij grotere afwijkingen en dempt zo de impact van uitschieters in leeftijdsannotaties [9].

Voor de evaluatie gebruiken we MAE, omdat deze metriek direct de gemiddelde afwijking in jaren weergeeft. Train-MAE en val-MAE worden per epoch gemonitord, waarbij de training is ingericht op het optimaliseren van val-MAE. Met ModelCheckpoint wordt het beste model opgeslagen. ReduceLROnPlateau verlaagt de learning

rate bij stagnatie en EarlyStopping stopt de training tijdig en herstelt de beste gewichten. Zo wordt overfitting beperkt en blijft het model generaliseerbaar.

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. [16] 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 geïnitialiseerd is met de gewichten van ImageNet. Deze methode van transfer learning wordt vaak gebruikt omdat de geleerde feature-extracties goed generaliseren naar nieuwe data [?].

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 [15]. 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 represeneert en eenvoudig te interpreteren is. Tijdens het trainen wordt na elke epoch zowel de trainings- als validatiefout berekend. Het model wordt opgeslagen zodra de validatiefout verbeterd, terwijl early stopping voorkomt dat er onnodig doorgetraind wordt wanneer er gedurende meerdere epochs geen verbetering optreedt.

4.3.4 Iteratie 2: VGG16 - Aangepaste Regressie-head. In de tweede iteratie is de regressie-head van het VGG16-model aangepast. Waar in de eerste opzet een enkele lineaire laag direct de leeftijd voorspelde, is dit vervangen door een diepere structuur bestaande uit meerdere volledig verbonden lagen, batch-normalisatielagen, ReLU-activaties en dropout. Deze architectuur eindigt nog steeds in één lineaire outputlaag, zodat de continue leeftijdswaarde wordt voorspeld. BETER WEERGEVEN WAT DE AANPASSING IS De keuze voor deze uitbreiding kwam 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 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 [19, 26].

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 bias inzichtelijk te maken wordt gebruikgemaakt van *MAE parity*, waarbij de relatieve verschillen in gemiddelde absolute fout (MAE) tussen groepen worden gemeten. Voor gender definiëren we het percentageverschil als

$$\Delta_{\text{MAE}}^{\text{gender}} = \frac{|\text{MAE}_{\text{male}} - \text{MAE}_{\text{female}}|}{\frac{1}{2}(\text{MAE}_{\text{male}} + \text{MAE}_{\text{female}})} \times 100\%,$$

met $\text{MAE}_g = \frac{1}{|S_g|} \sum_{i \in S_g} |\hat{y}_i - y_i|$, waarbij S_g de verzameling personen met gender g is.

Voor etniciteit generaliseren we dit naar meerdere groepen G door te kijken naar het maximale relatieve verschil:

$$\Delta_{\text{MAE}}^{\text{ethnicity}} = \max_{g,h \in G} \frac{|\text{MAE}_g - \text{MAE}_h|}{\frac{1}{2}(\text{MAE}_g + \text{MAE}_h)} \times 100\%.$$

Een lage waarde van deze metrics duidt op een eerlijker verdeling van fouten tussen groepen.

5 Resultaten

Dit hoofdstuk geeft een overzicht van de modelprestaties op de trainings- en validatieset, met MAE als hoofdmetriek. Eerst wordt de ontwikkeling over de epochs kort samengevat. Vervolgens volgt een beknopte uitsplitsing van de fout per leeftijd en de relatie met de dataverdeling. Tot slot worden de modellen beoordeeld op de belangrijkste kwaliteitscriteria.

5.1 Resultaten ResNet-50

De training is uitgevoerd voor maximaal 50 epochs. EarlyStopping herstelde na afloop de gewichten van de beste epoch (47). Op dat punt behaalde het model een train-MAE van 7.92 en een val-MAE van 7.88.

De learning curves (Figuur 10) tonen een snelle daling in de eerste 10 epochs, gevolgd door geleidelijke verbetering. Kleine stapjes winst vallen samen met learning-rate verlagingen. De kloof tussen train-MAE en val-MAE blijft klein (0.05), wat wijst op goede generalisatie en beperkt risico op overfitting.

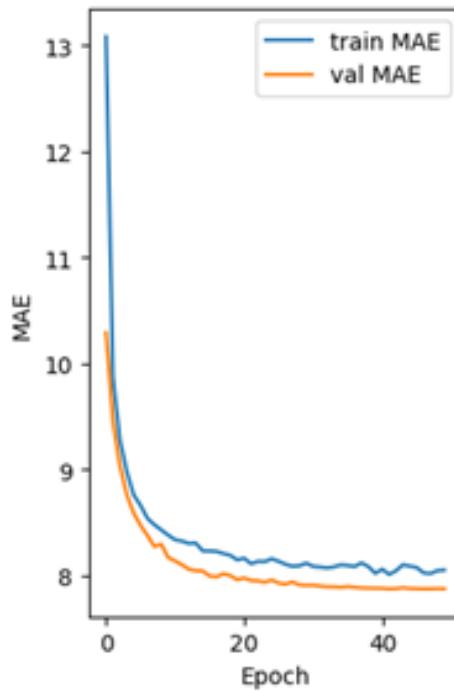


Fig. 10. Training results for ResNet-50.

De MAE per leeftijd op de validatieset (Figuur 11) 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, waar de dadadichtheid 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. Dit kan worden verbeterd door extra data toe te voegen en zo de balans per leeftijd te vergroten.

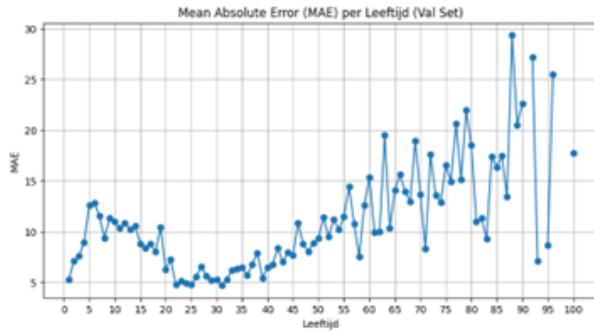


Fig. 11. Training results per Age for ResNet-50.

5.2 Resultaten VGG16

De resultaten van het modelleerproces worden vastgelegd in een tabel, waarin per iteratie de laagst behaalde MAE wordt weergegeven. Omdat het modelleerproces nog gaande is, worden de keuzes voor verdere aanpassingen telkens gebaseerd op de tot dan toe behaalde MAE-waarden. Op deze manier fungeert de MAE als leidende maatstaf voor het evalueren en bijsturen van het model.

Table 2. Overzicht van de behaalde MAE per iteratie (lager is beter).

Iteratie	Train MAE	Validatie MAE
Iteratie 1 – Enkelvoudige regressielag	10.26	9.98
Iteratie 2 – Aangepaste regressie-head	6.62	6.86
Iteratie 3 – Classificatie + verwachtingswaarde	3	5.28

5.2.1 Iteration 2. De grafiek in Figuur 12. toont de MAE per leeftijd op de validatieset. Te zien is dat het model relatief goed presteert bij jongere leeftijden, met lage MAE-waarden tussen 5 en 10 jaar. Naarmate de leeftijd toeneemt, neemt de fout geleidelijk toe. Vanaf ongeveer 50 jaar stijgt de MAE sterk, met uitschieters boven de 50 jaar. Dit wijst erop dat het model moeite heeft met het accuraat voorspellen van hogere leeftijden. Dit komt door een kleinere representatie van deze leeftijdsgroep in de dataset.

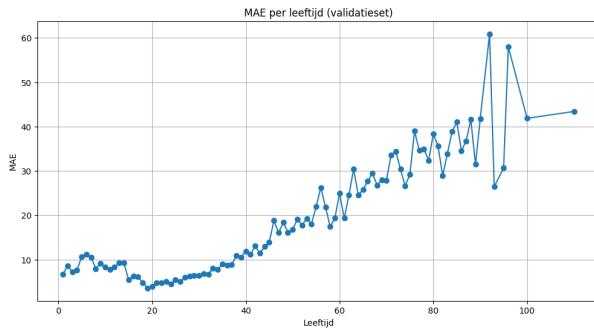


Fig. 12. MAE per Leeftijd voor VGG16 - Iteratie 2

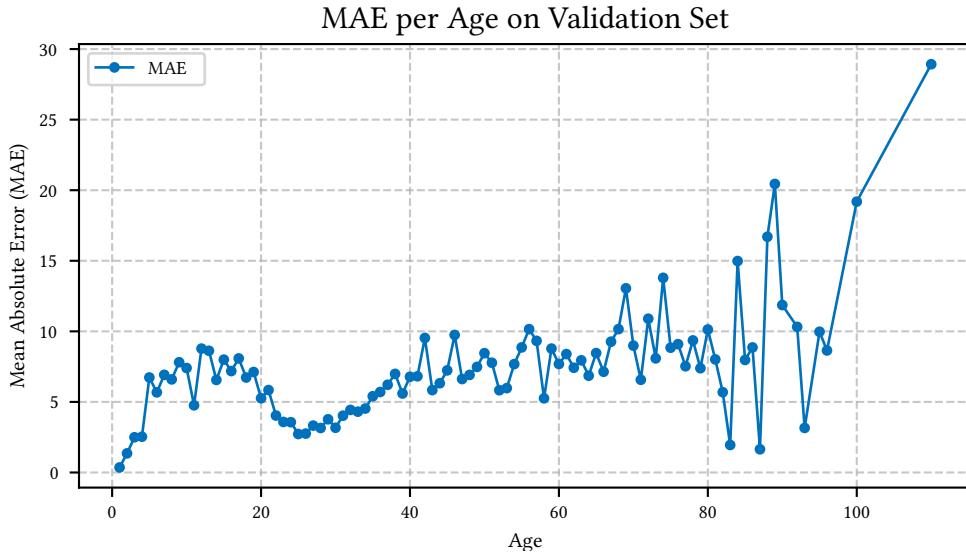


Fig. 13. MAE per age on the validation set.

5.2.2 Iteration 3.

5.2.3 Kwaliteitscriteria. Robuustheid betekent dat een model stabiel blijft bij kleine verstoringen [5, 29]. 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 [20]. 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 leer vermogen maar ook hoge eisen aan data en resources.

Resource demand gaat over rekenkracht en is belangrijk voor praktische inzetbaarheid [24]. In dit project duurde de training ± 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 met een hoge zekerheid van 25 jaar of ouder 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.

6.1 Evaluatie van Modelresultaten

De experimenten tonen een foutmarge (MAE) tussen ongeveer 5,3 en 7,9 jaar, afhankelijk van het gekozen model en de iteratie. Belangrijk hierbij is dat fouten bij hogere leeftijden, zoals een 60-jarige die wordt ingeschat op 40 jaar, in de praktijk nauwelijks gevlogen hebben. Afwijkingen rondom de grens van 18-25 jaar zijn daarentegen wel cruciaal. Daarom tellen fouten onder de 40 jaar zwaarder mee in de evaluatie dan die boven de 60 jaar. Deze nuancing verklart de relatief hoge foutmarges bij oudere leeftijden en toont dat het model binnen de relevante leeftijdsgroep nauwkeurig genoeg presteert.

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 het systeem onvoldoende zekerheid heeft of technisch faalt, schakelt het altijd over op handmatige controle. Daarnaast is progressive disclosure toegepast, zoals bij instructies om hoofddeksels 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.

7 Discussie

8 Conclusie

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 arXiv:1603.04467* (2016). <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-verantwoerde-alcoholverkoop-in-de-supermarkt_maart-2024.pdf. Accessed: 2025-09-16.
- [5] 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>
- [6] Y. Gu and H. Lee. 2024. Transfer Learning Strategies for Face Age Estimation. *IEEE Transactions on Biometrics* (2024).
- [7] 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 (CVPR)*. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf
- [8] X. Hu and Y. Xia. 2025. Improved Regression Architectures for Face Age Estimation. *arXiv preprint arXiv:2501.01234* (2025).
- [9] Katarzyna Janocha and Wojciech Marian Czarnecki. 2017. On Loss Functions for Deep Neural Networks in Classification. *arXiv preprint arXiv:1702.05659* (2017). <https://arxiv.org/abs/1702.05659>
- [10] 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>
- [11] 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
- [12] 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 Accessed: 2025-09-29.
- [13] 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.
- [14] 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>
- [15] 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
- [16] 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
- [17] 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
- [18] 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>
- [19] scikit-learn developers. n.d.. *sklearn.metrics.mean_absolute_error*. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html Accessed: 2025-09-29.
- [20] 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
- [21] A. Sheoran et al. 2021. Face Age Estimation using Deep Residual Networks. *Journal of Computer Vision Research* (2021).
- [22] 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>
- [23] 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.

- [24] 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>
- [25] Ultralytics. 2025. Model prediction with Ultralytics YOLO — Predict: Inference arguments. <https://docs.ultralytics.com/modes/predict/#inference-arguments>.
- [26] 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
- [27] L. Xi et al. 2023. Portrait-Based Age Estimation with Residual Networks. *Pattern Recognition Letters* (2023).
- [28] 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
- [29] 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>