

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.

, Vol. 1, No. 1, Article . Publication date: September 2025.

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	4
2.4 Gerelateerde Literatuur	7
2.5 Overwegingen voor het inzetten van AI	8
3 Requirements	9
4 Methodologie	10
4.1 Concept	10
4.2 Data	13
4.3 Model	14
5 Resultaten	15
5.1 Resultaten ResNet-50	16
5.2 Resultaten VGG16	17
6 Evaluatie	18
7 Discussie	18
8 Conclusie	18
References	18

*

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 zelfscankassaplein bestaande uit meerdere kassa's. Uit nalevingsonderzoek blijkt ook dat de handhaving van de 18-jaarsgrens in supermarkten tekortschiet. In 2024 werd in bijna een derde van de gevallen de leeftijds grens niet correct nageleefd [2].

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 gebruiksonderzoek 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 [22].

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’ [3]. 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 Regelgeving

Wachten op aanpassingen Merve

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:

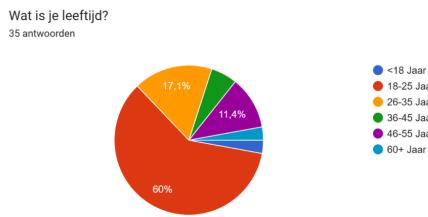


Fig. 1. Leeftijdsverdeling respondenten (N=35).

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.

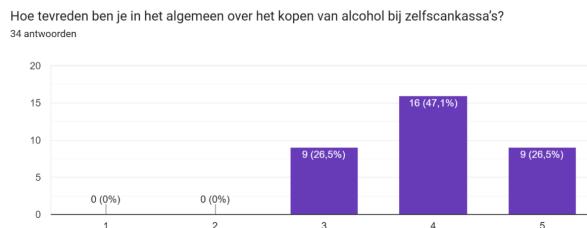


Fig. 2. Tevredenheid alcoholaankopen zelfscankassa (N=34).

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.

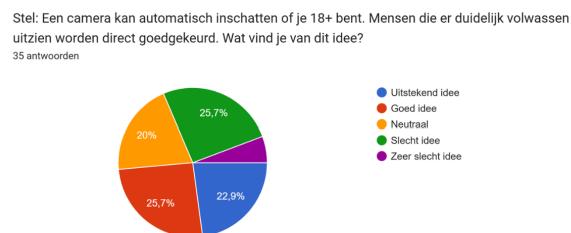


Fig. 3. Mening over automatische leeftijdsschatting via camera (N=35).

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.

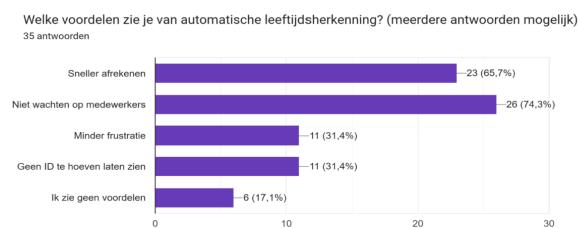


Fig. 4. Voordelen automatische leeftijdsherkenning (N=35, meerdere antwoorden mogelijk).

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).

2.3.2 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 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.

2.3.3 Conclusie. Uit het gebruikeronderzoek 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.

2.4 Gerelateerde Literatuur

Safa [17] onderzocht hoe convolutionele neurale networks (CNN's) kunnen worden ingezet om de leeftijd van personen automatisch te voorspellen op basis van gezichtsaafbeeldingen. In de studie werden verschillende deep learning-architecturen vergeleken, waaronder recurrente neurale networks (RNN's) en generatieve 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 [16] 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. [21] 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 gezichtskennmerken 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 [13].

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.

2.5.2 Maatschappelijke overwegingen. Uit een interview met een Jumbo-ondernehmer 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 leeftidscontroles, 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 [12].

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 [11].

3 Requirements

Table 1. Gesorteerde MoSCoW Requirements voor AI-gebaseerde Leeftijdsverificatie

Categorie	Requirement	MoSCoW
Juridisch	Als de klant jonger dan 25 jaar is, moet de leeftijd vastgesteld worden middels een ID-check	Must
Juridisch	Als er meerdere klanten aan dezelfde kassa staan, moet van elke klant de leeftijd worden vastgesteld	Must
Juridisch	ID's mogen niet gescand worden	Must
Juridisch	Het ontwerp moet voldoen aan de AVG artikel 4 lid 14 en artikel 9 lid 1	Must
Juridisch	Het ontwerp moet voldoen aan de Uitvoeringswet AVG artikel 29	Must
Juridisch	De gegevens mogen niet verwerkt worden, volgens AVG artikel 4 lid 2	Must
Juridisch	Alvorens gegevens van de klant worden afgenoemt, wordt de klant de benodigde informatie verstrekt, volgens AVG artikel 13	Must
Juridisch	De klant wordt duidelijk geïnformeerd over de verwerking van zijn gegevens, naar AVG artikel 12 lid 1	Must
Ethisch	Het systeem moet de privacy en de waardigheid van de klant respecteren	Must
Ethisch	Het gebruik van AI moet proportioneel zijn: alleen toegepast bij leeftijdscontrole voor alcohol	Must
Organisatorisch	Medewerkers moeten te allen tijde de controlefunctie behouden (Human in the loop)	Must
Functioneel	Het gebruik van AI voor leeftijdsverificatie moet op vrijwillige basis zijn	Must
Functioneel	De camera wordt afgedekt totdat er gekozen is voor het gebruik van AI voor leeftijdsverificatie	Must
Functioneel	Het model moet vaststellen of de persoon jonger dan 25 jaar is, of 25 jaar of ouder	Must
Technisch	Camerabeelden moeten real-time verwerkt worden	Must
Technisch	Er wordt een confidence waarde gegeven bij de voorspelling	Must
Technisch	Voorspellingen met een te lage confidence worden handmatig geverified	Must
Ethisch	Het model moet getraind worden op een diverse dataset (leeftijd, geslacht en etniciteit) om bias in de prestaties te vermijden	Should
Organisatorisch	Medewerkers moeten duidelijke instructies krijgen hoe om te gaan met foutmeldingen of twijfelgevallen	Should

4 Methodologie

4.1 Concept

In de eerste versie van het concept (zie Figuur 6) lag de nadruk voornamelijk op de scheiding tussen automatische leeftijdscontrole en handmatige controle door een medewerker. Tijdens de eerste paper-prototypingssessie met CMD-studenten werd echter waardevolle feedback verkregen met betrekking tot zowel de interactie als de gebruikerservaring. Uit deze sessie bleek dat er meer aandacht nodig was voor scenario's waarin meerdere personen gelijktijdig bij de kassa aanwezig zijn, evenals voor situaties waarin klanten hoofddeksels dragen, zoals een pet of hoofddoek. Daarnaast adviseerden de studenten om expliciter te communiceren dat een medewerker wordt ingeschakeld zodra een klant jonger lijkt dan 25 jaar. Verder werd het belang benadrukt van het toevoegen van een terugknop en een annuleerknop, en werd geadviseerd om de laatste schermen visueel duidelijker te maken door het gebruik van herkenbare symbolen, zoals een vink of kruis.



Fig. 6. Flowchart 1e Iteratie

Op basis van deze input werd een tweede versie van de flowchart ontwikkeld (zie tweede flowchart). In deze aangepaste versie is de rol van de medewerker nadrukkelijker opgenomen, zijn er extra opties toegevoegd voor de gebruiker en is de visuele feedback verbeterd. Tijdens de tweede iteratie met CMD-studenten werd deze nieuwe flow getest. Daarbij werd bevestigd dat de wijzigingen de helderheid en gebruiksvriendelijkheid aanzienlijk hebben versterkt.

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 verminderd het aantal interacties met klanten. In stedelijke gebieden waar agressie tegen medewerkers veel voor komt, biedt dat een uitkomst. Supermarkteigenaren profiteren daarnaast van een kleinere kans op boetes door foutieve controles.

4.1.2 Flow Diagram. Het proces van alcoholverkoop (zie Figuur 7) start met de vraag of de klant wil meewerken aan de AI-leeftijdscontrole. Indien de klant instemt, kijkt hij of zij in de camera en verwijderd zo nodig een hoofddeksel, waarna de AI een leeftijdsinschatting maakt. Wanneer de klant naar schatting 25 jaar of ouder is, kan de aankoop direct worden afgerond. Dit vormt de happy flow, waarbij de verkoop soepel en zonder tussenkomst van een medewerker verloopt. In alle gevallen waarbij de klant niet wil meewerken of jonger dan 25 jaar wordt geschat, treedt het principe van human-in-the-loop in werking: het systeem schakelt over naar een medewerker die de leeftijd handmatig controleert. Indien de medewerker bevestigt dat de klant minstens 18 jaar oud is, mag de verkoop alsnog doorgaan. Blijkt de klant jonger dan 18 jaar, of weigert hij of zij volledig mee te werken aan de controle, dan eindigt het proces met een weigering van de verkoop. Het flowdiagram maakt hiermee een duidelijk onderscheid tussen de positieve route, waarin de verkoop doorgaat, en de negatieve route, waarin de transactie wordt beëindigd.

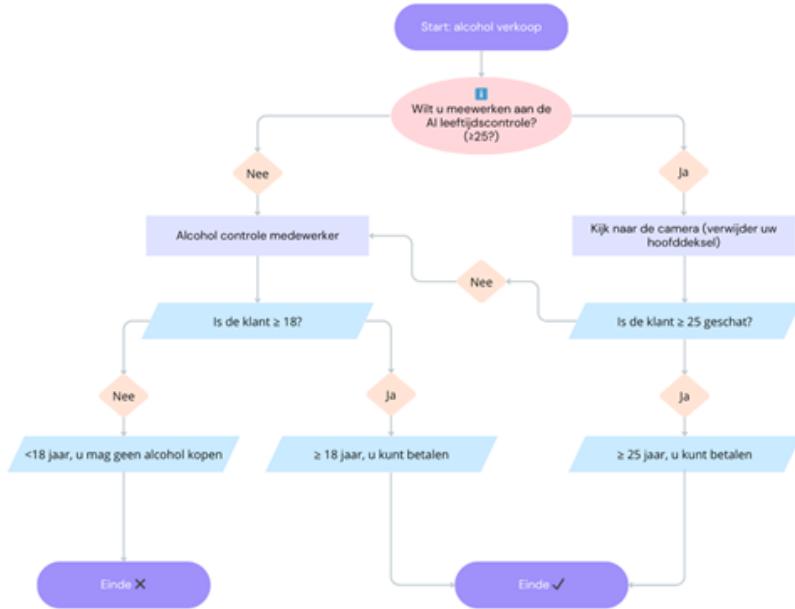


Fig. 7. Caption

In het flow diagram wordt duidelijk dat human-in-the-loop wordt toegepast wanneer de AI-leeftijdscontrole geen besluit kan nemen: als de klant niet wil meewerken aan de AI-leeftijdscontrole of jonger dan 25 jaar wordt geschat. Op dat moment schakelt het proces over naar een medewerker die handmatig de leeftijd controleert. Zo combineert het systeem snelheid door inzet van AI met betrouwbaarheid dankzij menselijke beoordeling. Wanneer de medewerker bepaalt dat de klant minstens 18 jaar oud is, kan de verkoop alsnog doorgaan.

Het diagram laat ook zien hoe de error flows verlopen. Een error flow ontstaat wanneer een klant jonger blijkt dan 18 jaar of helemaal niet wil meewerken aan leeftijdscontrole. In die gevallen eindigt het proces met een weigering en mag er geen alcohol worden verkocht. Op die manier wordt in de processtroom duidelijk onderscheid gemaakt tussen de positieve route, waarin de verkoop doorgaat, en de negatieve route, waarin de verkoop wordt stopgezet.

4.1.3 AI-Breakdown. De AI speelt in dit proces de rol van eerste controlemechanisme (zie rechterzijde van het flow diagram). Het systeem gebruikt gezichtsherkenning om de leeftijd van de klant in te schatten. Alleen wanneer de schatting op of boven de grens van 25 jaar ligt en het model voldoende zeker is van die voorspelling, krijgt de klant toestemming om direct af te rekenen. 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.

4.1.4 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 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.

4.1.5 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.

Het flowdiagram toont daarbij de volledige processtroom: de happy flow treedt op wanneer de klant als ouder dan 25 jaar wordt geschat en direct kan betalen, terwijl error flows ontstaan bij weigering of een lagere leeftijdsinschatting. Indien de klant jonger blijkt dan 18 jaar, eindigt het proces met een weigering van de verkoop.

4.1.6 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 focus lag op gebruiksvriendelijkheid, communicatie en uitzonderingssituaties. Deze aanpak maakte het mogelijk om al vroeg in het ontwerpproces verbeterpunten te signaleren en direct te verwerken in een herziene flowchart. De sessies bewezen daarmee een laagdrempelige en effectieve methode om de interactie tussen klant, AI en medewerker te evalueren.

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 occlusies. 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 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 voordelen

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 [27]. De confidence-drempel is op 0.25 gezet, gelijk aan de Ultralytics-standaard voor YOLO-inference, om recall te prioriteren tijdens het croppen[24].

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.

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 [10].



Fig. 8. Caption



Fig. 9. Caption

4.3 Model

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 [15]. 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. [15] 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 [20, 26]. ResNet-50 wordt explicet ingezet als uitgangspunt voor transfer learning bij leeftijdsvoorspellingen [20] en als referentiemodel voor portret-gebaseerde leeftijdsherkenning [26]. 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 [6]. 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 [5]. Bovenop komt Global Average Pooling, gevolgd door Dropout (0.2) en een Dense(1, lineair) voor leeftijdsregressie [7].

Het model wordt gecompiled met Adam ($\text{lr} = 1e-3$) vanwege adaptieve leersnelheden en snelle convergentie [9]. 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 [8].

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. [15] 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 [14]. 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 verbetert, 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.

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 evaluatiemeter 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 [18, 25].

5 Resultaten

Dit hoofdstuk geeft een overzicht van de modelprestaties op de trainings- en validatieset, met MAE als hoofdmeter. 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.

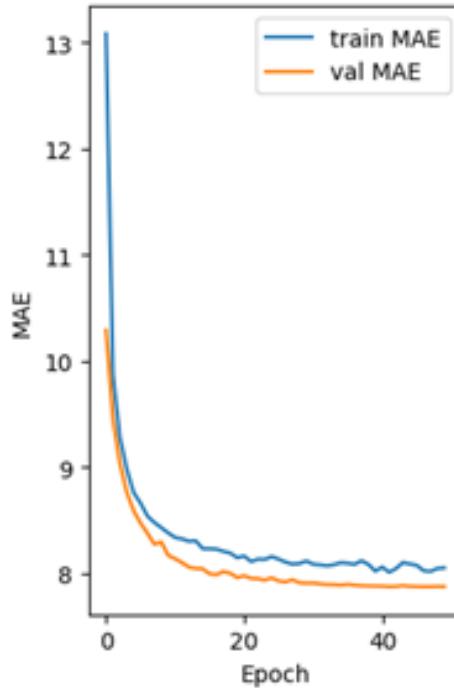


Fig. 10. Training results for ResNet-50.

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.

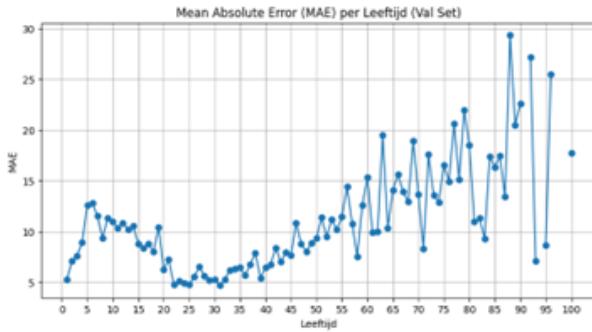


Fig. 11. Training results per Age for ResNet-50.

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 regressielaa	10.26	9.98
Iteratie 2 – Aangepaste regressie-head	6.62	6.86
Iteratie 3 – –	–	–

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.

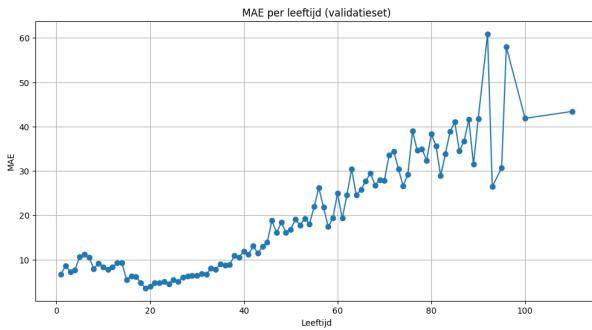


Fig. 12. MAE per Leeftijd voor VGG16 - Iteratie 2

5.2.1 Kwaliteitscriteria. Robuustheid betekent dat een model stabiel blijft bij kleine verstoringen [4, 28]. 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 [1]. 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 [19]. 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 [23]. 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

7 Discussie

8 Conclusie

References

- [1] Martín Abadi et al. 2016. TensorFlow: Large-scale machine learning on heterogeneous systems. *arXiv preprint arXiv:1603.04467* (2016). <https://arxiv.org/abs/1603.04467>
- [2] Bureau Objectief. 2024. Nalevingscijfer leeftijdsgrafs 2024. <https://allesovernix18indesupermarkt.bureauobjectief.nl/nalevingscijfer-leeftijdsgrafs-2024/>. Accessed: 2025-09-23.
- [3] 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.
- [4] 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>
- [5] Y. Gu and H. Lee. 2024. Transfer Learning Strategies for Face Age Estimation. *IEEE Transactions on Biometrics* (2024).
- [6] 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
- [7] X. Hu and Y. Xia. 2025. Improved Regression Architectures for Face Age Estimation. *arXiv preprint arXiv:2501.01234* (2025).
- [8] 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>
- [9] 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>
- [10] 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
- [11] 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.
- [12] 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.

- [13] 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>
- [14] 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/bdbca288fee7f92fbfa9f7012727740-Paper.pdf
- [15] 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
- [16] 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
- [17] 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>
- [18] 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.
- [19] 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
- [20] A. Sheoran et al. 2021. Face Age Estimation using Deep Residual Networks. *Journal of Computer Vision Research* (2021).
- [21] 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>
- [22] 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.
- [23] 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>
- [24] Ultralytics. 2025. Model prediction with Ultralytics YOLO — Predict: Inference arguments. <https://docs.ultralytics.com/modes/predict/#inference-arguments>.
- [25] 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
- [26] L. Xi et al. 2023. Portrait-Based Age Estimation with Residual Networks. *Pattern Recognition Letters* (2023).
- [27] 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
- [28] 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>