

ONDERZOEKSVOORSTEL

Machine Learning pipeline voor Facial Image Analysis in camera-gebaseerde gezondheidsmetingen: Schatten van leeftijd en geslacht voor het beoordelen van geestelijke gezondheid.

Bachelorproef, 2023-2024

Alexandra Stalmans

E-mail: alexandra.stalmans@student.hogent.be

Co-promotor: T.Sanglet (IntelliProve, tanguy.sanglet@intelliprove.com)

Samenvatting

Gezichtsanalyse heeft de laatste jaren veel aandacht gekregen vanwege de brede toepassingen op verschillende gebieden, zoals gezondheidszorg, beveiliging en marketing.

De focus van de bachelorproef ligt op het ontwikkelen en implementeren van een robuust systeem voor het schatten van leeftijd en geslacht op basis van gezichtsfoto's, met behulp van machine learning technieken. Dit onderzoek is van bijzonder belang voor het verbeteren van de beoordeling van geestelijke gezondheidszorg door camera-gebaseerde gezondheidsmetingen. Deze inspanningen dragen bij aan de doelstellingen van IntelliProve, een platform dat online gezondheidsoplossingen biedt. Er wordt onderzocht welke machine learning technieken de beste resultaten tonen en uit welke elementen de volledige pipeline zal bestaan.

In de eerste fase van het onderzoek wordt een literatuurstudie uitgevoerd om uit bestaand onderzoek de gebruikte modellen te verkennen en hun sterktes, limitaties en performantie te achterhalen. In de proof-of-concept gebeurt eerst de preprocessing, feature extractie en gezichtsdetectie. Daarna worden de modellen uit de literatuurstudie getest op bestaande datasets en wordt een prestatie-evaluatie opgesteld op basis van metrieken zoals accuracy, precision, recall, F1-score en ROC curve. Het verwachte resultaat is een robuuste machine learning pipeline waarmee IntelliProve de leeftijd en het geslacht kan voorspellen op basis van een gezichtsafbeelding.

Keuzerichting: AI & Data Engineering

Sleutelwoorden: Facial Image Analysis, Machine Learning, Facial Recognition, Predictive Modelling

Inhoudsopgave

1	Introductie	1
2	Stand van zaken	2
	2.1 Gezichtsdetectie	2
	2.1.1 Normalisatie	2
	2.2 Bestaande Machine-Learning technieken	2
	2.2.1 Extreme Gradient Boost	3
	2.2.2 Adaptive Boosting	3
	2.2.3 Random Forest Classifier	3
	2.2.4 Ensemble	4
	2.3 Bestaande uitdagingen	4
3	Methodologie	4
	3.1 Requirements	4
	3.2 Literatuurstudie	4
	3.3 Proof-of-concept	4
	3.4 Conclusie	4
	3.5 Afwerken scriptie	5
4	Verwacht resultaat, conclusie	5
	Referenties	5

1. Introductie

De onderzoeksvraag werd aangeboden door het bedrijf IntelliProve. IntelliProve biedt online

gezondheidsoplossingen, een software die in staat is om binnen enkele seconden nauwkeurig gezondheidsparameters te bepalen, gebaseerd op een optische meting van het gezicht. Het doel van de bachelorproef is het ontwikkelen en implementeren van een robuust systeem voor het schatten van de leeftijd en het geslacht van personen op basis van gezichtsfoto's, met behulp van machine learning-technieken. Dit project is van bijzonder belang voor het verbeteren van de beoordeling van de geestelijke gezondheidszorg door middel van camera-gebaseerde gezondheidsmetingen. Het onderzoek beoogt bij te dragen aan de vooruitgang op dit gebied door gebruik te maken van geavanceerde algoritmen om leeftijd en geslacht nauwkeurig te voorspellen aan de hand van gezichtsbeelden. De literatuurstudie biedt een inzicht in facial analysis, de bestaande machine learning modellen en hun functionaliteiten. De proof-of-concept zal bestaan uit het ontwikkelen van een machine learning pipeline die in staat is om leeftijd en geslacht te voorspellen op basis van bestaande datasets. De pipeline omvat verschillende image preprocessing technieken om de dataset voor te bereiden op de mo-

deltraining. Om betrouwbaarheid en accuracy te garanderen, worden de modellen verfijnd en geoptimaliseerd om de hoogst mogelijke nauwkeurigheid te bereiken bij het schatten van leeftijd en geslacht.

2. Stand van zaken

Gezichtsanalyse bestaat uit het definiëren van menselijke gezichten in real time, met behulp van computeralgoritmen en machine learning technieken. Voor mensen en computersystemen bevat een gezichtsbeeld details zoals leeftijd, geslacht, stemming, afkomst, et cetera. Gezichtsanalyse omvat dan ook het lokaliseren en meten van gezichtskenmerken in een afbeelding. Vervolgens wordt het gezichtsbeeld geanalyseerd voor het extraheren van kenmerken (Sanil e.a., 2023). Gezichtsanalyse speelt dan ook een grote rol in real world applicaties, zoals animatie, identiteitsverificatie, medische diagnose, et cetera. Ondanks het bestaande onderzoekswerk over dit onderwerp, is gezichtsanalyse nog steeds een lastige taak vanwege verschillende factoren zoals veranderingen in hoek, gezichtsuitdrukkingen en achtergrond (Siddiqi e.a., 2022).

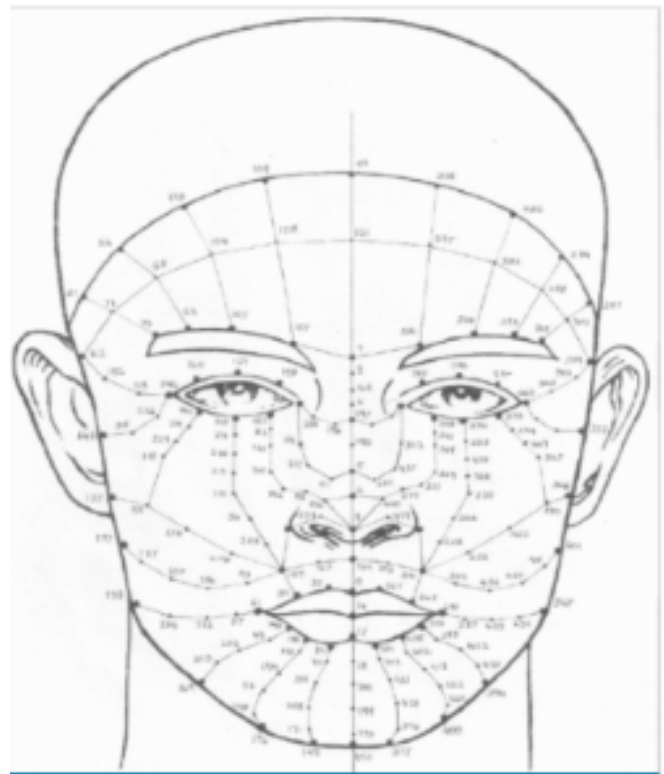
Volgens onderzoek van Basystiuk e.a. (2023) is beeldherkenning een simpele procedure die slechts uit 3 stappen bestaat.

1. Preprocessing: We voegen filters toe aan de afbeelding om het geschikter te maken voor herkenning.
2. Feature Extractie: We identificeren belangrijke data en behouden deze om mee verder te werken. Dit wordt verder besproken in 2.1
3. Classificatie: het analyseren en identificeren van de data na de feature extractie.

2.1. Gezichtsdetectie

Gezichtsdetectie is de eerste stap in het bepalen van de gezichtsfeatures. Deze features zijn de interessante delen van het gezicht, bijvoorbeeld ogen, neus en mond en worden ook wel eens landmarks genoemd (Coppens, 2018). In dit onderzoek raadt de auteur 3 gezichtsdetectieservices aan: Microsoft Face API, Face++ en Kairos. Deze zijn alle 3 ook geschikt om geslacht en leeftijd te voorspellen. Er wordt met 3 services tegelijk gewerkt om elkaars beperkingen op te vangen. In het onderzoek van Sanil e.a. (2023) berekenen ze de landmarks op 2 verschillende methoden: Euclidische afstand en Geodetische afstand.

Sanil e.a. (2023) stelt 468 landmarks visueel voor, in plaats van 68, om de accuracy te verbeteren. Dit door gebruik te maken van de MediaPipe library (Zubair (2021)). Op figuur 1 staan de belangrijkste antropometrische landmarks aangeduid. Het is noodzakelijk om de belangrijkste



Figuur 1: De belangrijkste anthropometrische landmarks (Sanil e.a., 2023).

features te identificeren die kunnen leiden tot het correct voorspellen van de leeftijd of het geslacht. Om de leeftijd te voorspellen kunnen we bijvoorbeeld de rimpels op de gezichtsfoto analyseren (Kwon & da Vitoria Lobo, 1994).

2.1.1. Normalisatie

In onderzoek van (C. Chen, 2011) voor het voorspellen van leeftijd op basis van afbeeldingen werden de gezichtsafbeeldingen genormaliseerd voor de feature extraction plaatsvond. De geometrie van de afbeeldingen kan worden genormaliseerd op basis van de gedetecteerde features, zoals oog of mond coördinatie. De auteur maakt gebruik van een mask om de pixels die zich niet in de ovaal van de typische gezichtsregio bevinden te verwijderen. Dit gaat dan bijvoorbeeld over haar en hemdskragen. Zo blijven enkel de belangrijkste features van het gezicht over. Figuur 2 geeft een voorbeeld van een dataset met gezichtsfoto's. In figuur 3 worden deze afbeeldingen genormaliseerd, waardoor er minder variaties overblijven in de afbeeldingen en we bijvoorbeeld al geen achtergrond overhouden.

2.2. Bestaande Machine-Learning technieken

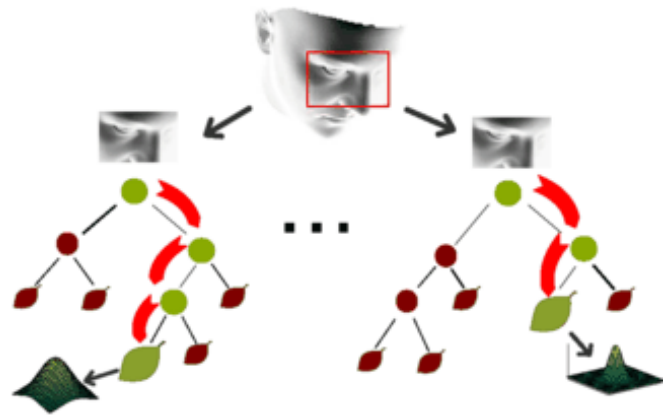
Uit onderzoek naar gezichtsdetectie van Sanil e.a. (2023) bleek dat Extreme gradient boosting de beste resultaten gaf, met een accuracy van 78%, gevolgd door Adaptive Boosting (77%)



Figuur 2: Gezichtsafbeeldingen voor normalisatie (C. Chen, 2011).



Figuur 3: Gezichtsafbeeldingen na het toepassen van normalisatie (C. Chen, 2011).



Figuur 4: Voorbeeld van een random forest voor het schatten van de houding van het hoofd (Fanelli e.a., 2012).

en Random Forest (75%). Onderzoek van (Khan, 2017) gebruikt dan weer een Ensemble systeem.

2.2.1. Extreme Gradient Boost

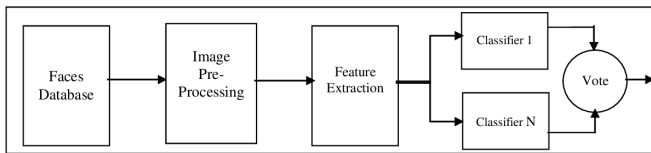
Extreme Gradient Boost, ofwel XGBoost, is een supervised learning model, waar trainingdata, met meerdere features, x_i wordt gebruikt om target waarde y_i te voorspellen (XGBoost, 2023). Uit onderzoek van (X. Chen & Zhang, 2023), waarbij vermoeidheid werd voorspeld op basis van gezichtsafbeeldingen met XGBoost, wordt een boom aangemaakt op basis van de geleverde features. Dit model gaf een hoge accuracy en is minder gevoelig voor noise.

2.2.2. Adaptive Boosting

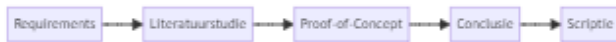
De Adaptive Boost, of AdaBoost, is typisch een classificatie tussen twee klassen. Om het herkeningsprobleem met meerdere klassen op te lossen, kan een majority voting (MV) strategie worden gebruikt om alle paarsgewijze classificatieresultaten te combineren. Hierbij kiezen we de meest voorkomende klasse als voorspelling. AdaBoost is een adaptief algoritme om een reeks classificeerders te boosten, in die zin dat de gewichten dynamisch worden bijgewerkt op basis van de fouten in eerdere leerresultaten (Guo & Zhang, 2001).

2.2.3. Random Forest Classifier

Random forests zijn verzamelingen decision trees die elk getraind zijn op een willekeurig gekozen deelverzameling van de beschikbare data. Deze decision trees worden willekeurig door de trainingsvoorbeelden die aan elke boom worden gegeven, maar ook door een willekeurige subset van tests die beschikbaar zijn voor optimalisatie op elke node (Fanelli e.a., 2012). Een voorbeeld voor het gebruik van Random Forest in gezichtsanalyse is figuur 4. Hierbij wordt een boom getoond voor het schatten van de houding van het hoofd.



Figuur 5: Robuust Gezichtsherkenning systeem(Khan, 2017).



Figuur 6: Methodologie voor het onderzoek.

2.2.4. Ensemble

Een set van 30 Lineaire Discriminant Analyse (LDA) dienen als basis classifiers en dragen bij tot het ensemble in het model van Khan (2017). Het ensemble gebruikt de gecombineerde classificatiecapaciteiten van de base classifiers, zo verbeteren we de algehele prestaties. De volledige pipeline voor de gezichtsanalyse vinden we in figuur 5.

2.3. Bestaande uitdagingen

Ondanks het onderzoekwerk, heersen er nog vele uitdagingen op het vlak van gezichtsanalyse. Deze uitdagingen zijn het gevolg van verschillende factoren zoals gezichtsuitdrukkingen, ruis, belichting, et cetera. Om de nauwkeurigheid van de gezichtsherkenning te verbeteren, is het belangrijk om taken van de gezichtsanalyse met elkaar te correleren. Het is bijvoorbeeld zeer waarschijnlijk dat mannen een baard of snor kunnen hebben, maar vrouwen en kinderen niet (Siddiqi e.a., 2022). Ook is het detecteren van de locatie waar de gezichten zich bevinden op de afbeelding een uitdaging. Dit wordt vaak meegenomen in de preprocessing stap van het analysesysteem (Jiang & Chen, 2008). De problemen situeren zich niet enkel op het vlak van de afbeelding, maar ook bij de machine learning modellen. Eén van de grootste problemen hierbij is overfitting. Dit gebeurt wanneer we het algoritme te veel trainen, waardoor het te hard lijkt op de training-data. Hierdoor zal het systeem falen wanneer we nieuwe data willen classificeren (Coppens, 2018).

3. Methodologie

Figuur 6 geeft een visuele representatie van de opgestelde methodologie.

3.1. Requirements

In de eerste week wordt nagevraagd aan belanghebbenden van IntelliProve aan welke criteria de modellen moeten voldoen. Alle data (gezichtsfoto's) wordt verzameld. Er wordt onder andere nagegaan over welke functionalitei-

ten de modellen moeten beschikken en wat de verwachte prestatievereisten zijn. Als resultaat verwerven we een lijst van alle functionele en niet-functionele requirements, geordend volgens belang.

3.2. Literatuurstudie

De literatuurstudie omvat een diepgaande verkenning van facial analysis technieken en machine learning modellen. Deze fase biedt inzicht in de verschillende methoden voor het extraheeren van gezichtskenmerken en image preprocessing technieken, specifiek met betrekking tot het schatten van leeftijd en geslacht. Het doel is om kennis uit bestaand onderzoek te vergaren om effectieve methodologieën te identificeren in de huidige benaderingen van facial analysis. Het eindresultaat van deze fase, die 3 weken duurt, is een samenvatting van de belangrijkste bevindingen uit de literatuurstudie, die als basis zal dienen voor de proof-of-concept.

3.3. Proof-of-concept

Deze fase start met het verzamelen en analyseren van de datasets. Technieken zoals normalisatie en scaling van de features, feature-extractie en data augmentation worden gebruikt om de dataset voor te bereiden op modeltraining. Vervolgens worden verschillende machine learning algoritmen geselecteerd op basis van de bevindingen uit de literatuurstudie. De modellen worden getraind en geëvalueerd op basis van de vooropgestelde requirements. De pipeline wordt beoordeeld op betrouwbaarheid en precisie op basis van de opgestelde requirements. Metrieken zoals accuracy, precision, recall, F1-score en ROC curves worden gebruikt om de modellen te evalueren. Er worden cross-validatie technieken gebruikt om de robuustheid van de modellen en de generalisatie van nieuwe data te testen. Deze fase vereist dan ook veel tijd en zal 6 weken duren. Het resultaat van deze fase is een proof-of-concept die bestaat uit een machine learning pipeline, die stappen voor het preprocessen van gegevens en getrainde modellen integreert voor het schatten van leeftijd en geslacht op basis van gezichtsfoto's.

3.4. Conclusie

In de conclusiefase worden de resultaten van de evaluatie grondig geanalyseerd. De prestaties van de ontwikkelde modellen worden beoordeeld, waarbij hun sterke punten en beperkingen worden benadrukt. Het belang van een nauwkeurige schatting van leeftijd en geslacht bij de beoordeling van de geestelijke gezondheid en de implicaties voor de gezondheidszorg worden besproken. Er worden aanbevelingen gedaan voor mogelijke verbeteringen of toekomstige onderzoeksrichtingen op basis van de bevindingen en

beperkingen van het project.

3.5. Afwerken scriptie

De laatste fase, die 2 weken duurt, omvat het afwerken van de bachelorproef. Dit is het eindresultaat van het geleverde onderzoek met een proof-of-concept die zal worden ingediend.

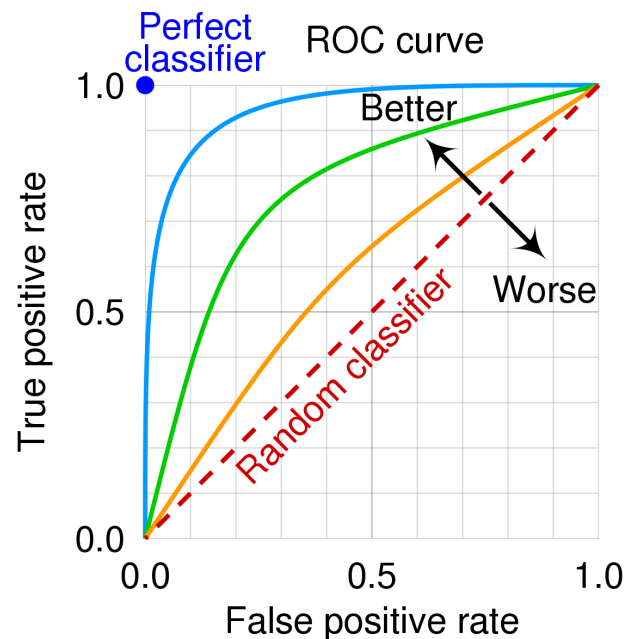
4. Verwacht resultaat, conclusie

Als resultaat van het onderzoek wordt een proof-of-concept opgesteld die een robuuste machine learning pipeline implementeert. We streven naar een accuracy van 75%, wat al haalbaar bleek uit onderzoek van Sanil e.a. (2023). Er wordt verwacht dat de modellen van Boosting, Random Forest en een ensemble de hoogste scores en accuracy opleveren. We geven naast de proof-of-concept ook een prestatie-evaluatie van de gebruikte modellen met uitgebreide analyse en vergelijking. De pipeline kan worden beoordeeld op betrouwbaarheid en precisie op basis van de opgestelde requirements. Metrieken zoals accuracy, precision, recall, F1-score en ROC-curves worden gebruikt om de modellen te evalueren. Onderstaand een mock-up van de scores met op figuur 7 een ROC-curve. Dit onderzoek is van bijzonder belang voor het verbeteren van de beoordeling van de geestelijke gezondheidszorg door middel van camera-gebaseerde gezondheidsmetingen en levert een bijdrage aan de vooruitgang op dit gebied door gebruik te maken van geavanceerde algoritmen om leeftijd en geslacht nauwkeurig te voorspellen aan de hand van gezichtsopnames. Het is een opstap naar verder onderzoek in gezichtsanalyse.

Model	Accuracy	Precision	Recall	F1-Score
Model A	0.72	0.50	0.80	0.82
Model B	0.52	0.25	0.90	0.39
Model C	0.61	0.70	0.40	0.51

Referenties

- Basystiuk, O., Melnykova, N., & Rybchak, Z. (2023). Machine Learning Methods and Tools for Facial Recognition Based on Multimodal Approach. In M. Emmerich, V. Vysotska & V. Lytvynenko (Red.), *Proceedings of the Modern Machine Learning Technologies and Data Science Workshop (MoMLeTDS 2023)*, Lviv, Ukraine, June 3, 2023 (pp. 161–170, Deel 3426). CEUR-WS.org. <https://ceur-ws.org/Vol-3426/paper13.pdf>
- Chen, C. (2011). *Gender Classification from Facial Images* [proefschrift, West Virginia University Libraries]. <https://doi.org/10.33915/etd.3449>



Figuur 7: ROC curve(Gomede, 2023).

- Chen, X., & Zhang, B. (2023). A XGBoost Algorithm-based Fatigue Recognition Model Using Face Detection, 3–4. <https://doi.org/10.21203/rs.3.rs-2814957/v1>
- Coppens, I. (2018). *Persoonlijke aanbevelingen voor video's gebaseerd op gezichtsherkenning* [masterscriptie, Universiteit Gent]. <https://www.scriptieprijs.be/scriptie/2018/persoonlijke-aanbevelingen-voor-videos-gebaseerd-op-gezichtsherkenning>
- Fanelli, G., Dantone, M., Gall, J., Fossati, A., & Van Gool, L. (2012). Random Forests for Real Time 3D Face Analysis. *International Journal of Computer Vision*, 101(3), 437–458. <https://doi.org/10.1007/s11263-012-0549-0>
- Gomede, E. (2023). The ROC Curve: Application and Interpretation in the Health Context. *Medium*. <https://medium.com/@evertongomede/the-roc-curve-application-and-interpretation-in-the-health-context-cc06af05a6ec>
- Guo, G.-D., & Zhang, H.-J. (2001). Boosting for fast face recognition. *Proceedings IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 2. <https://doi.org/10.1109/ratfg.2001.938916>
- Jiang, X., & Chen, Y.-F. (2008). Facial Image Processing. In *Studies in Computational Intelligence* (pp. 29–48). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-76831-9_2
- Khan, H. A. (2017). Feature Fusion and Classifier Ensemble Technique for Robust Face Recognition, 3–8. <https://api.semanticscholar.org/CorpusID:203632225>

- Kwon, Y. H., & da Vitoria Lobo. (1994). Age classification from facial images. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*, 10–12. <https://doi.org/10.1109/cvpr.1994.323894>
- Sanil, G., Prakash, K., Prabhu, S., Nayak, V. C., & Sengupta, S. (2023). 2D-3D Facial Image Analysis for Identification of Facial Features Using Machine Learning Algorithms With Hyper-Parameter Optimization for Forensics Applications. *IEEE Access*, 11, 82521–82538. <https://doi.org/10.1109/access.2023.3298443>
- Siddiqi, M. H., Khan, K., Khan, R. U., & Alsirhani, A. (2022). Face Image Analysis Using Machine Learning: A Survey on Recent Trends and Applications. *Electronics*, 11(8), 1–2. <https://doi.org/10.3390/electronics11081210>
- XGBoost. (2023). *Introduction to Boosted Trees*. <https://xgboost.readthedocs.io/en/stable/tutorials/model.html>
- Zubair, M. (2021). Write a Few Lines of Code and Detect Faces, Draw Landmarks from Complex Images MediaPipe: Detecting face from complex image is easy and fun with MediaPipe. *Towards Data Science*, 3.