

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

Alexandra Stalmans.

Scriptie voorgedragen tot het bekomen van de graad van
Professionele bachelor in de toegepaste informatica

Promotor: Mevr. C. De Leenheer

Co-promotor: Dhr. T. Sanglet

Academiejaar: 2023–2024

Eerste examenperiode

Departement IT en Digitale Innovatie .

**HO
GENT**

Woord vooraf

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Samenvatting

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Inhoudsopgave

Lijst van figuren	vii
1 Inleiding	1
1.1 Probleemstelling	1
1.2 Onderzoeksvraag	2
1.3 Onderzoeksdoelstelling	2
1.4 Opzet van deze bachelorproef	3
2 Stand van zaken	4
2.1 Gezichtsdetectie	4
2.2 Feature extractie	5
2.2.1 Normalisatie	5
2.2.2 AAM	6
2.2.3 LBP	6
2.2.4 HOG features	8
2.3 Feature dimensionaliteitsreductie	9
2.3.1 PCA	9
2.3.2 Truncated SVD	9
2.4 Bestaande uitdagingen	10
2.5 Bestaande Machine-Learning technieken	10
2.5.1 Extreme Gradient Boost	10
2.5.2 Adaptive Boosting	11
2.5.3 Random Forest Classifier	11
2.5.4 Support Vector Machines	12
2.5.5 Logistic Regression	12
2.5.6 Ensemble	13
2.5.7 Multi attribution model	13
2.5.8 Grid Search en Random Search	14
2.6 Dataset	16
2.7 Voorspellen van geslacht en leeftijd	17
2.8 Performantiescores	17
3 Methodologie	20
3.1 Requirements	20

4	Proof of Concept	22
4.1	Dataset	22
4.2	Feature extractie	23
4.3	Feature dimensionaliteitsreductie	28
4.4	Machine learning modellen	31
4.4.1	Random Forest Classifier	32
4.4.2	XGBoost Classifier	34
4.4.3	Support Vector Machines	35
4.4.4	Voting Classifier	36
5	Conclusie	38
A	Onderzoeksvoorstel	40
A.1	Introductie	41
A.2	Stand van zaken	41
A.2.1	Gezichtsdetectie	42
A.2.2	Bestaande Machine-Learning technieken	43
A.2.3	Bestaande uitdagingen	46
A.3	Methodologie	46
A.3.1	Requirements	46
A.3.2	Literatuurstudie	46
A.3.3	Proof-of-concept	47
A.3.4	Conclusie	47
A.3.5	Afwerken scriptie	47
A.4	Verwacht resultaat, conclusie	47
	Bibliografie	50

Lijst van figuren

2.1	Belangrijkste antropometrische landmarks	5
2.2	Gezichtsafbeeldingen voor normalisatie	6
2.3	Gezichtsafbeeldingen na normalisatie	7
2.4	AAM-features	7
2.5	AAM-features	8
2.6	HOG features	8
2.7	Random forest voor houding van het hoofd	11
2.8	SVM	13
2.9	Robuust gezichtsherkenning systeem	14
2.10	Cross validatie voor Grid Search	15
2.11	FairFace, UTKFace en LFWA+ datasets	16
2.12	Confusion matrix	18
3.1	Stappen in de pipeline voor Proof of Concept	21
4.1	5 willekeurige train voorbeelden	23
4.2	5 landmarks op het gezicht	24
4.3	68 landmarks op gezicht	26
4.4	Gezichtsvorm uit afbeelding halen	26
4.5	Normalisatie van gezichtsafbeelding	27
4.6	Geen gezicht gedecteerd op afbeelding	29
4.7	Verklaarde variantie voor de Truncated SVD	30
4.8	Truncated SVD toegepast op genormaliseerde gezichtsafbeelding	31
4.9	Confusion Matrix voor Random Forest Classifier	34
4.10	Confusion Matrix voor XGBoost Classifier	35
4.11	Confusion Matrix voor Support Vector Classifier	36
A.1	Belangrijkste antropometrische landmarks	42
A.2	Gezichtsafbeeldingen voor normalisatie	43
A.3	Gezichtsafbeeldingen na normalisatie	44
A.4	Random forest voor houding van het hoofd	45
A.5	Robuust gezichtsherkenning systeem	45
A.6	Methodologie	46
A.7	ROC curve	49

List of Listings

2.1	GridSearchCV	15
4.1	Functie om FairFace dataset te downloaden en uitpakken (Ramos, 2024)	23
4.2	Functie om mask toe te voegen aan de volledige dataset (Serengil, 2020)	27
4.3	Functie om Truncated SVD toe te passen op de volledige dataset (Scikit-Learn, 2024)	30
4.4	Functie om labels aan te passen voor de MultiOutputClassifier (Roepke, 2024) en (Numpy, 2024)	34

1

Inleiding

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 modeltraining. 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.

1.1. Probleemstelling

Het onderzoek wordt uitgevoerd voor 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. De bachelorproef vormt een opstap naar een applicatie die in de toekomst uitgewerkt zal worden.

1.2. Onderzoeksvraag

Het onderzoek beschrijft de ontwikkeling van een machine learning pipeline voor het schatten van leeftijd en geslacht. Specifiek voor het analyseren van gezichtsafbeeldingen in camera-gebaseerde gezondheidsmetingen. Voor het onderzoek werd volgende onderzoeksvraag opgesteld:

- Hoe kan een efficiënte machine learning-pipeline worden ontwikkeld en geoptimaliseerd voor het analyseren van gezichtsafbeeldingen in camera-gebaseerde gezondheidsmetingen, met als specifieke doelen het schatten van leeftijd en geslacht, toegepast om geestelijke gezondheid te beoordelen?

De onderzoeksvraag kan opgedeeld worden in enkele deelvragen:

1. Wat is de geschikte dataset voor het opgegeven probleem om zo veel mogelijk bias te vermijden?
2. Welke uitdagingen bestaan er al uit voorgaand onderzoek en moet rekening mee gehouden worden in het onderzoek?
3. Uit welke stappen bestaat de machine learning pipeline?
 - (a) Welke feature extractie en/of feature dimensionaliteitsreductie toepassingen zijn het meest geschikt voor het voorspellen van leeftijd en/of geslacht?
4. Welke machine learning modellen zijn er mogelijk voor het voorspellen van leeftijd en geslacht?
 - (a) Wordt er 1 model gemaakt om leeftijd en geslacht te voorspellen of worden er 2 modellen gebruikt die zich elk richten tot een specifieke taak?
 - (b) Welk model, uit een vergelijkende studie, geeft de beste resultaten?
5. Hoe kunnen we de performantie van een model meten?

1.3. Onderzoeksdoelstelling

Het resultaat van de bachelorproef is een proof-of-concept die 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 modeltraining. 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. Er wordt naar een zo hoog mogelijk nauwkeurigheid gestreefd, waarbij de conclusies uit dit onderzoek het belangrijkste zijn. Er kan aangegeven worden waarom bepaalde modellen goed werken of juist niet en of er in de toekomst nog verder onderzoek vereist is.

1.4. Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomein, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

In Hoofdstuk 4 wordt de Proof of Concept uitgewerkt en de code toegelicht.

In Hoofdstuk 5, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

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.

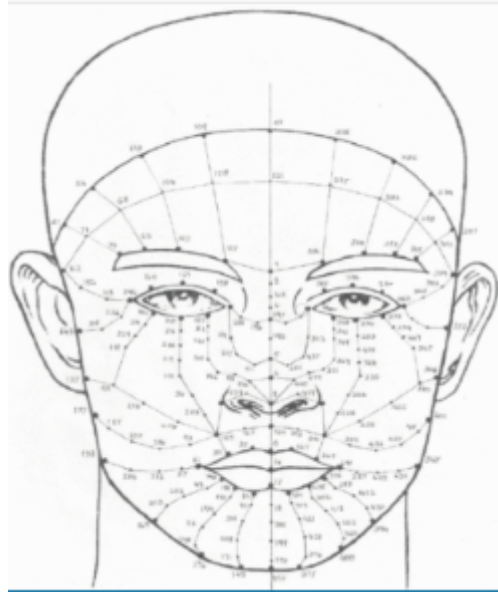
Gezichtsanalyse omvat dan ook het lokaliseren en meten van gezichtskenmerken in een afbeelding. Vervolgens wordt het gezichtsbeeld geanalyseerd voor het extraheren van kenmerken (Sanil e.a., [2023](#)). Gezichtsanalyse speelt dan ook een grote rol in real world applicaties, zoals animatie, identiteitsverificatie, 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 uit slechts 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

**Figuur (2.1)**

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

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 2.1 staan de belangrijkste antropometrische landmarks aangeduid. Het is noodzakelijk om de belangrijkste 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).

2.2. Feature extractie

2.2.1. Normalisatie

In onderzoek van (C. Chen, 2011) voor het voorspellen van leeftijd op basis van afbeeldingen werden de gezichtsafbeeldingen genormaliseerd voor de feature extractie 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 (2.2)**

Gezichtsafbeeldingen voor normalisatie (C. Chen, 2011).

Figuur 2.2 geeft een voorbeeld van een dataset met gezichtsfoto's. In figuur 2.3 worden deze afbeeldingen genormaliseerd, waardoor er minder variaties overblijven in de afbeeldingen en we bijvoorbeeld al geen achtergrond overhouden. De gezichtsafbeeldingen kunnen ook genormaliseerd worden op basis van de coördinaten van de ogen, zodat het centrum van beide ogen van de afbeeldingen op een vaste positie liggen (C. Chen, 2011).

2.2.2. AAM

In het onderzoek van Lakshmiprabha (2016) worden de Active Appearance Model (AAM) features gebruikt in de feature extractie stap. Deze features zijn geschikt om het effect van veroudering weer te geven voor het voorspellen van geslacht. De features die we van AAM krijgen hebben zowel vorm- als textuurinformatie die geschikter is voor veroudering. Figuur 2.4 geeft een representatie van de FG-NET database na het toepassen van de AAM-features. Het resultaat is vergelijkbaar met de normalisatie van C. Chen (2011) op figuur 2.3.

2.2.3. LBP

Local Binary Patterns (LBP) is een textuurdescriptor die gebruikt kan worden om gezichten te representeren, aangezien een gezichtsafbeelding kan worden beschouwd als een samenstelling van microtextuurpatronen (Sánchez López, 2010). De LBP operator kent een label toe aan elke pixel van een afbeelding door een threshold te nemen van de 3x3 burens met de centrale pixelwaarde en het resultaat te beschouwen als een binair getal. Het resultaat na de toepassing van LBP in het onderzoek

**Figuur (2.3)**

Gezichtsafbeeldingen na het toepassen van normalisatie(C. Chen, [2011](#)).

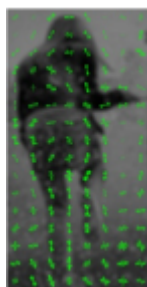
**Figuur (2.4)**

Cohn Kanada database na toepassen van AAM (Lakshmiprabha, [2016](#)).



Figuur (2.5)

Toepassen van LBP op een 16x16 gezichtsafbeelding resulteert in 14x14 label afbeelding (Sánchez López, 2010).



Figuur (2.6)

Toepassen van HOG op een gezichtsafbeelding (Tomasi, 2015).

van Sánchez López (2010) is weergegeven op figuur 2.5.

In het onderzoek van C. Chen (2011) wordt LBP toegepast om de gezichtsfeatures te extraheren uit elk blok die de leeftijdsinformatie bevat. Het resultaat hiervan is een feature dimensie vector van 59. Deze features worden dan samengevoegd in een holistische vector.

2.2.4. HOG features

Histograms of Oriented Gradients (HOG) begint met het tellen van de oriëntatie van de gradiënten in gelokaliseerde delen van een afbeelding. De afbeelding wordt verdeeld in kleine verbonden gebieden, genaamd cellen, en voor elke cel wordt een histogram van gradiëntrichtingen of randoriëntaties voor de pixels binnen de cel gecompileerd (Déniz e.a., 2011).

De HOG features worden exclusief gebruikt voor gezichtsdetectie. De features die uit de omgeving van de landmarks in het gezicht worden gehaald, worden gebruikt voor classificatie en maken gebruik van het nearest neighbour algoritme en de Euclidische afstand.

Tomasi (2015) gebruikt de HOG features om menselijke voetgangers te detecteren. Figuur 2.6 geeft de extractie van de HOG features uit een afbeelding weer.

2.3. Feature dimensionaliteitsreductie

De herkenning van gezicht focust zich voornamelijk op het detecteren van individuele features in het gezicht, zoals ogen, hoofdomtrek, mond of het definiëren van het model van gezicht door positie, grootte of relatie tussen de features (Lin, 2006). Het extraheren van de features speelt een grote rol in de preprocessing fase.

2.3.1. PCA

De Principal Component Analyse (PCA) wordt veelal gebruikt voor gezichtsherkenning. PCA is een unsupervised, statistische techniek die gebruikt wordt voor dimensionaliteitsreductie in machine learning (Salih Hasan & Abdulazeez, 2021). Het reduceert het hoge aantal dimensies in een grote dataset naar een lager aantal dimensies om de opslag en het verwerkingsproces te versnellen. Hierdoor is de data makkelijker te interpreteren en sneller te verwerken. De techniek behoudt het grootste aantal informatie en verwijdert redundante ruis en data. Het grote nadeel van PCA is dat de lineaire transformatie die wordt uitgevoerd om de grootste varianties te behouden, niet altijd goed geschikt is voor het op te lossen probleem, namelijk de classificatie van geslacht (Wang e.a., 2010).

De feature reductie vindt typisch plaats na de feature extractie fase en voor de eigenlijke classificatie (Lakshmiprabha, 2016).

2.3.2. Truncated SVD

De Truncated singular value decomposition, ofwel Truncated SVD, is een benaderingsmethode voor lage-rankmatrices die veelal gebruikt wordt in data science, machine learning en lineaire algebra (Yeh, 2022).

De benadering van de lage-rankmatrix is een dimensionaliteitsreductie methode die van belang is bij de groeiende hoeveelheid data waar traditionele methoden niet mee kunnen omgaan. Deze lage-matrix benaderingsmethoden benaderen een bepaalde matrix als product van meerdere kleinere matrices en wordt gebruikt om de geheugenkost en tijd die nodig is om de modellen te trainen te vermindere.

Truncated SVD is de optimale lage-rank matrix benadering en wordt hierdoor toegepast in machine learning als dimensionaliteitsreductie. PCA is ook een vorm van Truncated SVD, maar hierbij wordt de singulaire waarde, k , bepaald met een statistische interpretatie.

2.4. Bestaande uitdagingen

Ondanks het uitgebreide 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).

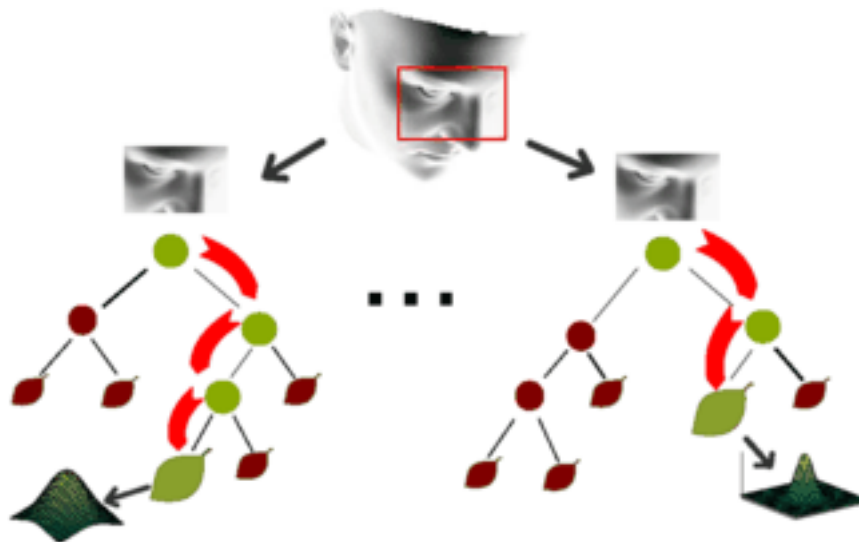
De problemen situeren zich niet enkel op het vlak van de afbeelding, maar ook bij de machine learning modellen. Eén van de grootste problemen hierbij is overfitting. Dit gebeurt wanneer we het algoritme te veel trainen, waardoor het te hard lijkt op de trainingsdata. Hierdoor zal het systeem falen wanneer we nieuwe data willen classificeren (Coppens, 2018).

2.5. Bestaande Machine-Learning technieken

Supervised learning is de methode die input mapt naar een output, gebaseerd op een voorbeeld (Rustam & Ruvita, 2018). Supervised learning is gericht op het voorspellen van de waarde van variabelen of labels, op basis van de input features. Het wordt gewoonlijk gebruikt voor classificatie, benadering, modelleren of identificatie. Het groeperen van mannen of vrouwen op basis van gezichtsherkenning is een classificatieprobleem. Er zijn 2 opties: man of vrouw, ofwel 0 of 1. Dit maakt supervised learning de geschikte methode voor de voorspelling van geslacht. Bij supervised learning bestaat de trainingsdata uit een reeks trainingsvoorbeelden, waarbij elk voorbeeld bestaat uit een input en een verwachte output waarde (Shah, 2012). Het algoritme analyseert de trainingsdata en voorspelt dan de correcte output categorie voor de gegeven data input.

2.5.1. 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. 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%) en Random Forest (75%).

**Figuur (2.7)**

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

2.5.2. Adaptive Boosting

De Adaptive Boost, of AdaBoost, is typisch een classificatie tussen twee klassen (G.-D. Guo & Zhang, 2001). Deze ML-techniek kan dus toegepast worden op de voorspelling van man of vrouw (0 of 1). Om het herkenningsprobleem 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.

2.5.3. 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). Wanneer we een input vector aan het model geven, gaat het model bewegen door de gepaste boom (C. Chen, 2011). De uiteindelijke classificatie is gebaseerd op een majority voting over alle bomen. Uit het onderzoek van Wang e.a. (2010) blijkt dat Random Forest ook geschikt is voor feature selectie.

Een voorbeeld voor het gebruik van Random Forest in gezichtsanalyse is figuur 2.7. Hierbij wordt een boom getoond voor het schatten van de houding van het hoofd.

Het onderzoek van Mady en Hilles (2018) gebruikt random forest classifier voor de herkenning en detectie van gezichten. De classifier wordt toegepast op het gezicht om dit te herkennen uit de databank. Hier wordt aangehaald wanneer we het aantal bomen in de random forest verhogen, zal ook de nauwkeurigheid verhogen, maar is het ook veel tijdrovender. De random forest is dus een trade-off tussen accuracy en tijd. Het toevoegen van Local Binary Pattern (LBP) en Histogram of Oriented Gradient (HOG) verhoogde de algemene accuracy van de classifier.

2.5.4. Support Vector Machines

Een Support Vector Machine (SVM) is een ML-techniek die gebruikt wordt voor patroonclassificatie en regressieanalyse (C. Chen, 2011). Het is gebaseerd op het zoeken naar een lineaire grens tussen twee klassen van patronen. Het SVM mechanisme gaat opzoek naar de optimale classifier die de data in 2 verschillende klassen opdeelt (Rustam & Ruvita, 2018). De optimale classifier wordt de hyperplane genoemd. Dit kunnen we vinden door de marge te maximaliseren en zal zorgen voor zo min mogelijk misclassificatie.

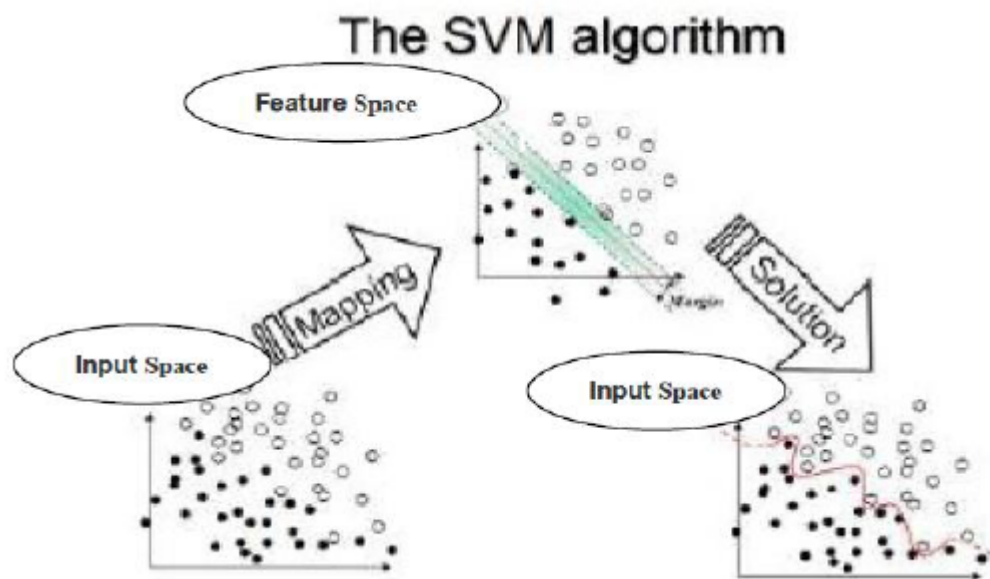
Voor niet-lineair scheidbare data mappen we alle punten naar een feature ruimte door gebruik te maken van een kernel (Shah, 2012). Na het splitsen, kunnen we de punten terug mappen naar de input ruimte met een kromme hyperplane, zoals op figuur 2.8.

Een SVM is zeer effectief wanneer een hoog aantal dimensionele ruimtes zijn. Ze zijn ook veelzijdig, aangezien we meerdere kernels kunnen gebruiken. De keuze van kernel is afhankelijk van de requirements van het model en heeft een grote invloed op de resultaten van een model. Het voordeel van een SVM model is dat het accurate voorspellingen kan maken, onafhankelijk van de grootte (image size) van de afbeeldingen. (S. A. Khan e.a., 2011)

In het onderzoek van Rustam en Ruvita (2018) wordt SVM gebruikt voor de classificatie van leeftijdsgroepen en bereikt de maximale accuracy van 100 %.

2.5.5. Logistic Regression

Logistische regressie is een populair algoritme voor binaire classificatie taken, zoals voorspellen van geslacht (Ramon, 2024). Het modelleert de relatie tussen de input variabelen en de kans dat de afbeelding tot een bepaalde geslachtscategorie behoort. Logistische regressie gebruikt de logistische functie om een lineaire combinatie van de features te maken. Dit wordt uiteindelijk gebruikt om voorspellingen te maken. Logistische regressie wordt gebruikt om de kans dat een instantie behoort tot een bepaalde klasse te voorspellen, zoals bijvoorbeeld de classificatie van spam of niet (Géron, 2019). Als de instantie tot de klasse behoort, voorspelt die 1, anders 0. Dit maakt het dus een binaire classifier.

**Figuur (2.8)**

Support Vector Machine voor niet-lineair scheidbare data

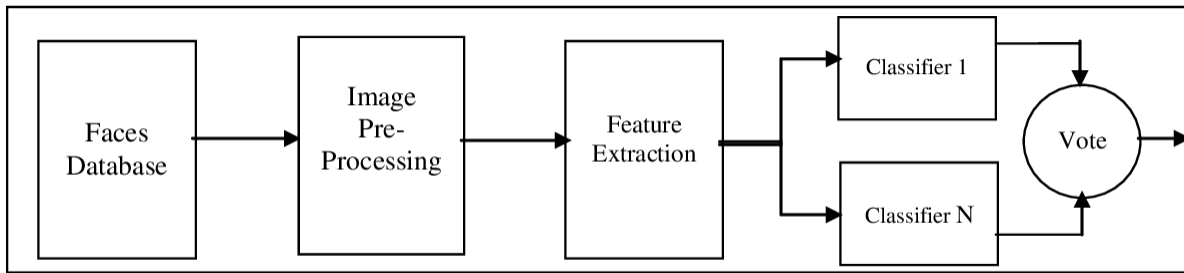
Het onderzoek van Yavuz (2014) gebruikt een logistisch regressiemodel om geslacht te voorspellen. Aan het model werden 2 verschillende optimisatiemethoden toegevoegd, namelijk Stochastic Gradient Following (SGF) en limited memory Broyden-Fletcher-Goldfarb-Shanno algoritme (L-BFGS). SGF is efficiënt en makkelijk te implementeren, wat het één van de meest gebruikte optimalisatie algoritmes maakt. Doordat het onderzoek van Yavuz (2014) slechts een beperkte dataset gebruikt, had het L-BFGS model een grotere error rate dan het SGF model. Dit toont nogmaals aan dat een goede dataset belangrijk is.

2.5.6. Ensemble

Een set van 30 Lineaire Discriminant Analyse (LDA) dienen als basis classifiers en dragen bij tot het ensemble in het model van H. A. 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 2.9.

2.5.7. Multi attribution model

In het onderzoek van (Gupta & Nain, 2022) wordt een multi-attribution model uitgewerkt. Dit voorspelt leeftijd en geslacht door middel van slechts 1 model. Het multi-attribution model voorspelt dus beide: leeftijd en geslacht. Het model van (G. Guo & Mu, 2014), dat gebruik maakt van de Biologically Inspired Feature (BIF), gaf betere resultaten in vergelijking met individuele feature modellen. Een CNN is het meest gebruikte model voor multi-attribution, maar dit valt buiten de scope van deze bachelorproef.

**Figuur (2.9)**

Robuust Gezichtsherkenning systeem(H. A. Khan, 2017).

2.5.8. Grid Search en Random Search

Hyperparameters representeren de specifieke parameters die we meegeven aan een Machine Learning model alvorens de training te beginnen (Ibtisamah, 2023). De hyperparameters kunnen worden gefinetuned op verschillende configuraties om het model beter te doen presteren. Het individueel trainen van deze hyperparameters is tijdrovend en daarom kunnen zoekstrategieën gebruikt worden om meerdere hyperparameters tegelijk uit te testen om de meest geschikte eruit te halen. Grid Search en Random Search zijn de bekendste strategieën en zijn beschikbaar gesteld via Scikit-learn (Scikit-Learn, 2024). Bij Grid Search geven we verschillende hyperparameters mee, die het model gaat onderzoeken. Bij Random Search geven we een range van hyperparameter waarden mee, wat het efficiënter maakt dan Grid Search. Grid Search maakt gebruik van cross validatie, waarbij de dataset in willekeurige groepen wordt opgedeeld (Beheshti, 2022). Een testgroep wordt opzij gehouden en de overige groepen worden getraind. Dit proces wordt herhaald tot elke groep testgroep is geweest. Hiervan wordt dan het gemiddelde genomen. De standaardwaarde voor het aantal herhalingen is 5. Figuur 2.10 geeft hier een visuele representatie van.

	Fold 1	Fold 2	Fold 3	Fold 4
First Iteration	Test			
Second Iteration		Test		
Third Iteration			Test	
Fourth Iteration				Test
Fifth Iteration				

Figuur (2.10)

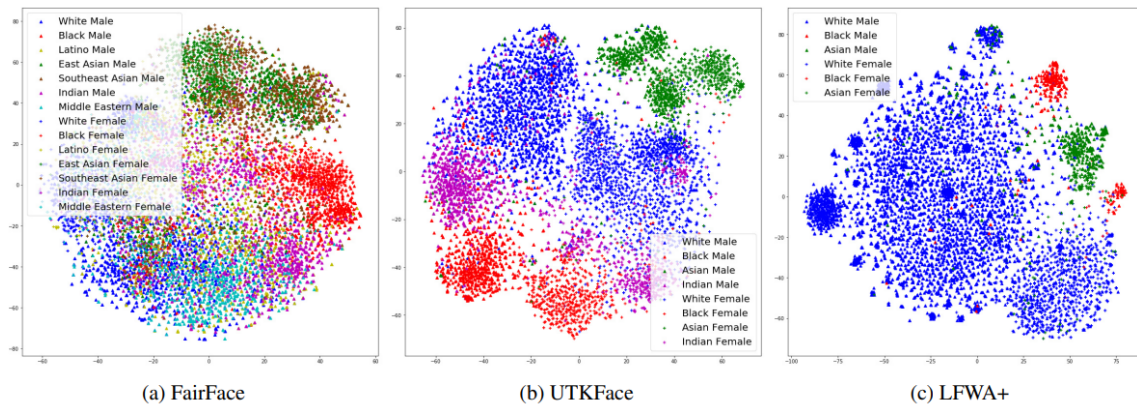
Cross validatie voor Grid Search (Beheshti, [2022](#)).

Hieronder een code voorbeeld van hoe (Educative.io, [2024](#)) met behulp van de Grid Search verschillende hyperparameters definieert.

```

1  from sklearn.model_selection import GridSearchCV, train_test_split
2  from sklearn.ensemble import GradientBoostingRegressor
3  from sklearn.datasets import load_diabetes
4
5  X,y = load_diabetes(return_X_y=True)
6  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33,
7  random_state=42)
8
9  % Creating a Model
10 model = GradientBoostingRegressor()
11
12 % Define the hyperparameter grid
13 param_grid = {'learning_rate': [0.2,0.02,0.02,1],
14               'max_depth'      : [2,4,6,8,10]}
15
16 % Create the GridSearchCV object

```


**Figuur (2.11)**

FairFace, UTKFace en LFWA+ datasets (Karkkainen & Joo, 2021).

```

17 grid_search = GridSearchCV(estimator=model, param_grid=param_grid, cv=5,
18                             n_jobs = -1)
19
20 % Perform the grid search
21 grid_search.fit(X_train, y_train)
22
23 % Access the best hyperparameters and best score
24 best_learning_rate = grid_search.best_params_['learning_rate']
25 best_max_depth = grid_search.best_params_['max_depth']
26 best_score = grid_search.best_score_
27
28 % Print the results
29 print("Best learning rate:", best_learning_rate)
30 print("Best Depth:", best_max_depth)
31 print("Best score (MSE):", best_score)

```

Listing 2.1: GridSearchCV

2.6. Dataset

Uit onderzoek van Karkkainen en Joo (2021) bleek dat bestaande, publieke datasets sterk bevoordeeld zijn naar blanke mensen. Modellen die getraind worden op enkel 1 publieke dataset vertonen dan ook slechte en inconsistente resultaten. Karkkainen gebruikt in zijn onderzoek een zelf samengestelde dataset, met 108501 afbeeldingen die verdeeld zijn over elk ras. De resultaten van het onderzoek op de dataset vertonen een betere accuraatheid over verschillende rassen en leeftijds-groepen. Op figuur 2.11 wordt de verdeling over geslacht en ras van de Fairface dataset die Karkkainen en Joo (2021) gebruikt weergegeven tegenover de UTKFace en LFWA+ dataset. Hieruit is duidelijk dat LFWA+ biased is naar witte mannen en dus niet geschikt is om een accuraat onderzoek uit te voeren.

Ook in het onderzoek van Buolamwini en Gebru (2018) wordt een eigen dataset

aangemaakt. Hier wordt niet enkel gekeken naar ras, maar bijvoorbeeld ook naar dikte en hoeveelheid haar. Het model moet zodoende dezelfde accuraatheid geven over verschillende leeftijdsgroepen, geslachten, rassen, et cetera. Uit het onderzoek bleek ook dat resultaten bij modellen die gebruik maken van Microsoft's Face Detect of Face++ een slechtere precisiescore vertonen, mogelijks door bias in trainingsdata.

2.7. Voorspellen van geslacht en leeftijd

Het eerste baanbrekende onderzoek op het vlak van classificatie op basis van gezichtsafbeeldingen dateert uit 1994. Kwon en da Vitoria Lobo (1994) voorspelde in dit onderzoek slechts 3 leeftijdsgroepen: baby's, tieners en volwassenen. Het onderzoek focuste zich op het vinden van de primaire features in het gezicht, zoals ogen, mond en neus, en combineerde dit met secundaire features, zoals rimpels. Hiermee toonde het onderzoek aan dat de gezichtsfeatures een grote rol spelen in het classificeren van leeftijd. Dit onderzoek vormt dan ook de basis voor vele verdere onderzoeken op het vlak van classificatie.

Menselijke gezichten zijn onderhevig aan veroudering en groei (Gupta & Nain, 2022). Deze verandering in uiterlijk kan verschillen van persoon en is het gevolg van verschillende factoren, zoals gezondheid, levensstijl, ras, roken, et cetera. Vrouwen en mannen verouderen anders, omdat ze een verschillend verouderingspatroon in het gezicht hebben. Dit is afkomstig van het gebruik van makeup, haarstijl, accessoires van vrouwen of snor en baard bij mannen. Vrouwelijke gezichten ogen dan ook vaak jonger dan mannelijke gezichten.

Het voorspellen van leeftijd kan ingedeeld worden in 2 categorieën: leeftijdsgroep classificatie, waarbij we de leeftijd opdelen in ranges (bijvoorbeeld 10-18 jaar) of als regressieprobleem, waarbij we een exact getal gaan voorspellen (Gupta & Nain, 2022). In de proof-of-concept van deze bachelorproef zal een leeftijdsgroep worden voorspeld.

2.8. Performantiescores

Er kunnen verschillende evaluaties uitgevoerd worden om de performantie van het model te testen (Sanil e.a., 2023). Er kan hiervoor een confusion matrix worden opgesteld, zoals weergegeven in figuur 2.12. Deze bevat de True Positives (TP), True Negatives (TN), False Positives (FP) en False Negatives (FN), waarbij TP en TN de correcte voorspellingen weergeven en FP en FN de incorrecte voorspellingen. De confusion matrix werkt voor binaire voorspellingen, ideaal voor de voorspelling van geslacht. Het onderzoek van Sanil e.a. (2023) berekent op basis van de opgestelde

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figuur (2.12)

Confusion matrix(Narkhede, 2018).

matrix 5 scores.

1. De accuracy geeft weer hoe goed het model voorspellingen maakt. Dit wordt voorgesteld door het aantal correcte voorspellingen (Narkhede, 2018).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

2. De precision stelt de hoeveelheid correcte positieve voorspellingen voor.

$$\text{Precision} = \frac{TP}{TP+FP}$$

3. De recall, ook wel sensitivity genoemd, stelt de kans op een positieve klasse voor.

$$\text{Recall} = \frac{TP}{TP+FN}$$

4. De specificity, stelt de kans op een negatieve klasse voor.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

5. De F1-score is het gemiddelde van de precision en recall scores, waarbij 1.0 de maximale score is.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Voor de voorspelling van leeftijd is het ook aangeraden om 2 andere ratios te gebruiken, namelijk True Acceptance Rate (TAR) en False Reject Rate (FRR) (Othman & Adnan, 2014). Hierdoor krijgen we per leeftijdscategorie een overzicht van het aantal juist en fout voorspelde afbeeldingen.

1. De TAR geeft het aantal juist voorspelde afbeeldingen binnen de opgegeven leeftijdscategorie.

$$\text{True Acceptance Rate} = \frac{\text{Juist voorspelde leeftijdscategorie}}{\text{Totale aantal afbeeldingen}} \times 100$$

2. De FRR geeft het aantal fout voorspelde afbeeldingen binnen de opgegeven leeftijdscategorie.

$$\text{False Rejected Rate} = \frac{\text{Fout voorspelde leeftijdscategorie}}{\text{Totale aantal afbeeldingen}} \times 100$$

3

Methodologie

Het onderzoek start met het opstellen van een duidelijke requirementsanalyse. Aan de belanghebbenden van IntelliProve wordt nagevraagd aan welke criteria het model moet voldoen. Deze requirements zijn van belang om de deelvragen voor het onderzoek op te stellen. Deze deelvragen kunnen teruggevonden worden in [1](#).

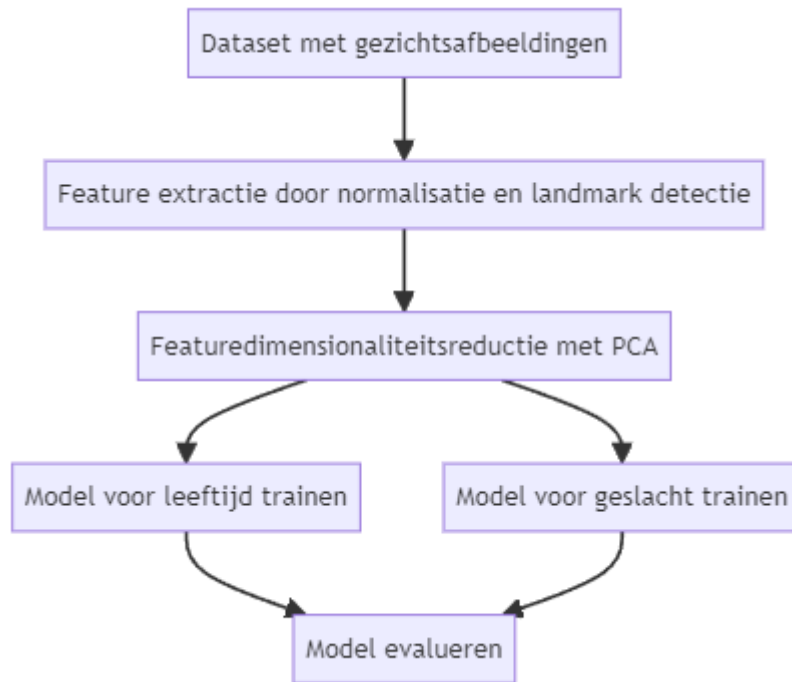
Na het definiëren van de requirements kan een uitgebreide literatuurstudie [2](#) worden uitgevoerd. Hierin wordt allerlei informatie verzameld over technieken en modellen. Na het uitvoeren van de literatuurstudie wordt duidelijk beeld gevormd van welke modellen er bestaan en verwerkt kunnen worden in de Proof of Concept. Er wordt ook een flowchart opgesteld van de pipeline. Hieruit kunnen duidelijk alle stappen gehaald worden om de Proof of Concept stapsgewijs te doen verlopen. Deze flowchart kan gevonden worden op figuur [3.1](#)

Uit de informatie en kennis van de literatuurstudie kan het onderzoek worden uitgevoerd als Proof of Concept [4](#). Alle verschillende modellen en technieken kunnen uitgetest worden en geëvalueerd op de dataset. Het resultaat van de Proof of Concept is een pipeline die op basis van een gezichtsfoto het geslacht en de leeftijd van een persoon gaat voorspellen.

Alle resultaten van het onderzoek en eventueel verder onderzoek worden besproken in de Conclusie [5](#).

3.1. Requirements

- De pipeline moet leeftijd en geslacht voorspellen op basis van een frontale gezichtsfoto. Er kan gewerkt worden met 1 machine learning model die beide voorspelt of 2 verschillende modellen die gecombineerd worden tot 1 voor-

**Figuur (3.1)**

Stappen in de pipeline voor Proof of Concept

spelde waarde.

- De leeftijd wordt in ranges voorspeld, het gaat dus niet om een exact getal.
- Het geslacht wordt voorspeld op basis van het biologische geslacht, man of vrouw.
- De trainings- en test dataset moeten online beschikbaar gesteld zijn. Er is geen interne dataset ter beschikking (omwille van privacy-wetgeving).
- De pipeline bestaat uit preprocessing van de afbeelding en het voorspellen van leeftijd en geslacht.
- Er is geen opgegeven streefdoel voor de accuraatheid van het model. Dit moet zo hoog mogelijk liggen, maar de conclusies die hieruit getrokken kunnen worden zijn van het grootste belang. Andere modellen kunnen voorgesteld worden, de score kan verklaard worden en toekomstig onderzoek kan aangeboden worden.

4

Proof of Concept

4.1. Dataset

Uit de voorgaande literatuurstudie in hoofdstuk 2.6 kan geconcludeerd worden dat de FairFace dataset van Karkkainen en Joo (2021) de meest geschikte is voor gezichtsanalyse. Deze dataset wordt beschikbaar gesteld via Google Drive of Kaggle. De dataset is beschikbaar met een padding (marge) van 0.25 of 1.25. Voor deze Proof of Concept wordt de dataset met een padding van 0.25 gebruikt. Deze is het meest geschikt voor wetenschappelijk onderzoek, omdat deze door de kleinere marge minder opslag verbruikt en het volledige gezicht weergeeft. De dataset is verder opgedeeld in 86.744 train afbeeldingen en 10.954 validatie of test afbeeldingen. Dit maakt de FairFace dataset een bijzonder grote dataset met kleine bias en beperkte opslag, wat ideaal is voor wetenschappelijk onderzoek.

Doordat de zip folder te groot is om direct te downloaden en steeds een waarschuwing genereert via Google Drive, moet de dataset gedownload en uitgepakt worden. Dit kan manueel, of efficiënter in Python via codevoorbeeld 4.1. De dataset kan dus niet in memory opgeslagen worden of rechtstreeks in Google Drive worden aangesproken.

De labels van de trainingsset bevinden zich in een csv bestand (in dezelfde Google Drive). Om uit te testen of de dataset correct werd ingeladen, is het mogelijk om enkele willekeurige afbeeldingen weer te geven met hun leeftijdscategorie en geslacht label. In deze Proof of Concept wordt dan ook een leeftijdscategorie voorspeld en geen leeftijdsgetal. Dit zorgt ervoor dat het voorspellen van de leeftijd een classificatietask wordt, net zoals het voorspellen van geslacht. Wanneer er een leeftijdsgetal wordt voorspeld, is dit een regressieprobleem (Géron, 2019). Een voorbeeld van de uitvoer van deze test wordt weergegeven op figuur 4.1. Voor deze

**Figuur (4.1)**

5 willekeurige train voorbeelden met leeftijd en geslacht label

Proof of Concept worden enkel de labels voor de leeftijd en geslacht gebruikt. Daarnaast bevat de FairFace dataset ook een label voor ras.

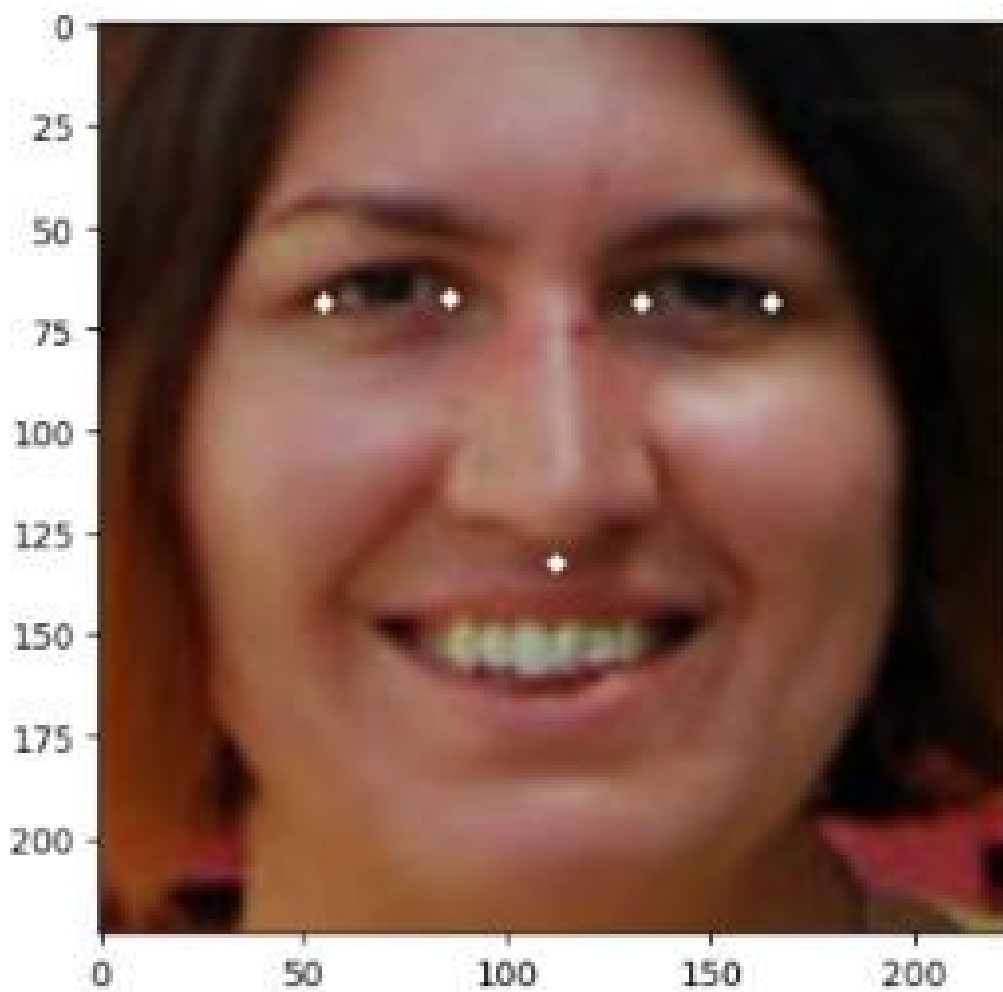
```
1 # Download the dataset from Google Drive and save in the data directory
2 # URL to the Google Drive file
3 url = 'https://drive.google.com/u/0/uc?id=1Z1RqRo0_JiavaZw2yzZG6WETdZQ8qX86&export=download'
4
5 zip_file = gdown.download(url, quiet=False)
6
7 destination_dir = './data'
8 os.makedirs(destination_dir, exist_ok=True) # Create the directory if it doesn't exist
9
10 # Extract the zip file into the specified directory
11 with zipfile.ZipFile(zip_file, 'r') as z:
12 z.extractall(destination_dir)
```

Listing 4.1: Functie om FairFace dataset te downloaden en uitpakken (Ramos, 2024)

4.2. Feature extractie

De feature extractie is de eerste preprocessing stap in de machine learning pipeline. In deze Proof of Concept wordt het getrainde dlib model van King (2024) gebruikt. Dit model detecteert en voorspelt de 68 landmarks uit het gezicht. Het model is beschikbaar via GitHub en is beperkt in opslag. Doordat het model al voorgetraind is, hoeft dit dus niet opnieuw getraind te worden en versnelt het preprocessing proces. Door het detecteren van de landmarks kunnen we deze behouden en de afbeelding normaliseren, zoals in 2.2.1.

King (2024) stelt ook een model ter beschikking waarop slechts 5 landmarks gedetecteerd worden. Dit zijn de 2 uiteinden van de ogen en het midden van de neus. Figuur 4.2 geeft een voorbeeld van een gezichtsafbeelding waarop deze 5 landmarks worden aangeduid. Voor deze use case zijn de 5 landmarks duidelijk te weinig om de volledige vorm van het gezicht weer te geven. Dit model met 5 landmarks is efficiënter bij een simpele gezichtsdetectie, om opslag te besparen en het model te versnellen.



Figuur (4.2)

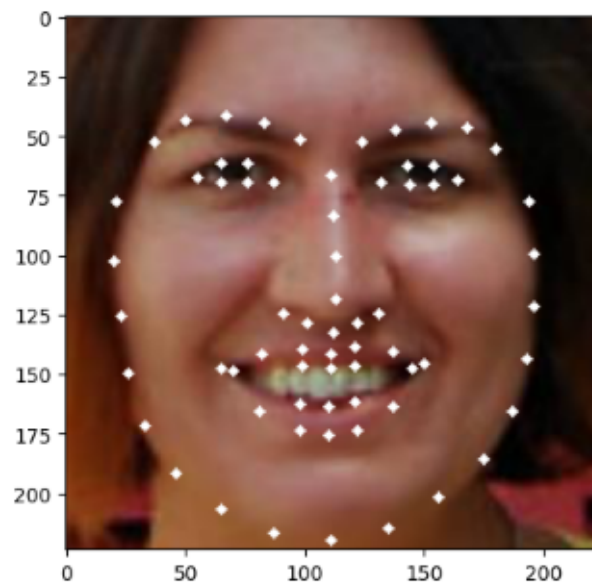
5 landmarks op het gezicht

1. De eerste stap in dit proces omvat het detecteren van de 68 landmarks (Seren-gil, 2020). Deze omvatten de vorm van het gezicht, wenkbrauwen, ogen, neus en mond. Dit wordt uitgevoerd met het *shape_predictor_68_face_landmarks* model van King (2024). Vervolgens wordt de *circle* functie van OpenCV (2024) gebruikt om de landmarks aan te duiden op de afbeelding. Figuur 4.3 geeft het resultaat van deze stap weer.
2. Vervolgens wordt op basis van de gedetecteerde landmarks de vorm van het gezicht uit de afbeelding gehaald. De coördinaten van deze landmarks worden opgehaald en met *line* van OpenCV (2024) omcirkeld. Er wordt een lijn getrokken van het eerste punt naar het volgende in de lijst met coördinaten. Figuur 4.4 geeft het resultaat van deze stap.
3. Ten slotte maken we een mask voor de overgebleven afbeelding. Alles rond

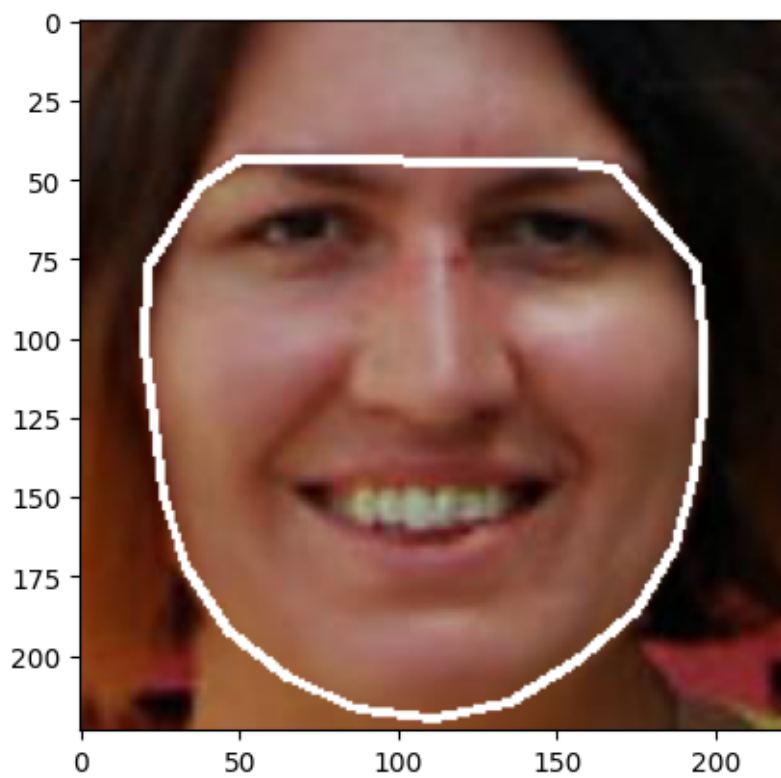
de vorm van het gezicht wordt met *fillConvexPoly* van OpenCV (2024) zwart ingekleurd. Hierdoor wordt de onbelangrijke achtergrond van de afbeelding verwijdert en wordt enkel het gezicht overgehouden. Deze stap voert de normalisatie net zoals C. Chen (2011) uit, gebruikmakend van de HOG features in het detecteren van de landmarks, omschreven in 2.2.4. Figuur 4.5 geeft het resultaat van deze stap weer. Dit is een genormaliseerde afbeelding die gebruikt zal worden voor de feature dimensionaliteitsreductie.

Het normaliseren van de afbeeldingen kan in een functie worden gestoken. Hierdoor kan aan de volledige dataset een mask toegevoegd worden, in plaats van slechts 1 afbeelding voordien. Het toevoegen van een mask aan een afbeelding duurt minder dan een seconde. Dit maakt deze methode geschikt voor de grote dataset. Het toevoegen van de mask wordt in batches gedaan, om het proces te versnellen. Er wordt hierdoor steeds aan een aantal afbeeldingen tegelijk een mask toegevoegd, in plaats van 1 afbeelding per keer. Er wordt gekozen voor een batch size van 128, omdat deze voldoende groot is voor de beschikbare rekenkracht op een standaard laptop. Het volledige proces nam 84 minuten in beslag voor 86.744 afbeeldingen.

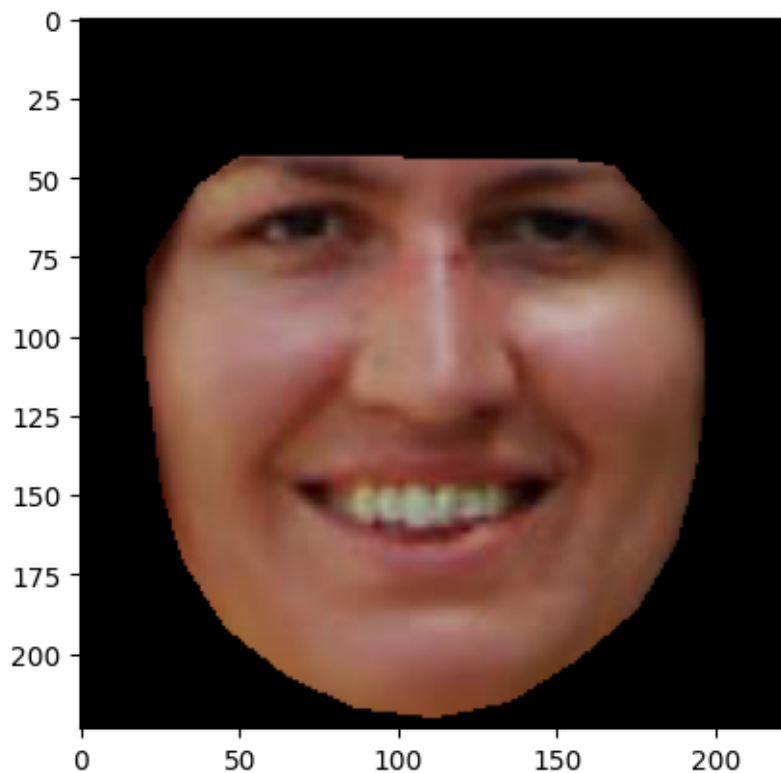
Daarnaast kan deze functie ook valideren dat er geen duidelijke gezichten aanwezig zijn op de afbeeldingen, doordat de 68 landmarks niet gedetecteerd kunnen worden. Deze afbeeldingen zijn onbruikbaar en worden zo ook uit de dataset gehaald. Figuur 4.6 geeft een afbeelding weer waarop de landmarks niet gedetecteerd kunnen worden. Dit komt doordat de foto geen vooraanzicht is en op deze figuur zijn beide ogen niet duidelijk weergegeven. Codevoorbeeld 4.2 geeft deze functie weer (Serengil, 2020). In de genormaliseerde trainingsdataset blijven nog 63.938 afbeeldingen over, in de genormaliseerde testset blijven nog 8.028 over. Beide datasets verliezen na normalisatie zo ongeveer 27% van de gezichtsafbeeldingen, maar zijn nog voldoende groot om mee verder te werken.

**Figuur (4.3)**

68 landmarks op het gezicht

**Figuur (4.4)**

Gezichtsvorm uit afbeelding halen op basis van landmarks

**Figuur (4.5)**

Normalisatie van gezichtsafbeelding

```
1 import os
2 import cv2
3 import dlib
4 import numpy as np
5
6 def get_mask(image_paths, predictor_path, output_folder):
7     # Initialize face detector and landmark detector
8     face_detector = dlib.get_frontal_face_detector()
9     landmark_detector = dlib.shape_predictor(predictor_path)
10
11     for image_path in image_paths:
12         image = dlib.load_rgb_image(image_path)
13
14         # Detect faces in the image -> do not mask images where no face is found
15         faces = face_detector(image, 1)
16         no_faces = []
17         if len(faces) == 0:
18             print("No faces found in the image:", image_path)
19             no_faces.append(image_path)
20             continue
21
22         # Predict facial landmarks
23         face = faces[0] # only one face
24         landmarks = landmark_detector(image, face)
```

```

25 landmarks_tuple = [(landmarks.part(i).x, landmarks.part(i).y) for i in range(68)]
26
27 # Define the route for the mask
28 routes = [i for i in range(16,-1,-1)] + [i for i in range(17,19)] + [i for i in
    range(24,26)] + [16]
29 routes_coordinates = [landmarks_tuple[i] for i in routes]
30
31 # Create a mask
32 mask = np.zeros((image.shape[0], image.shape[1]), dtype=np.uint8)
33 mask = cv2.fillConvexPoly(mask, np.array(routes_coordinates), 255)
34 masked_image = cv2.bitwise_and(image, image, mask=mask)
35
36 # Save the masked image to the output folder with the same filename
37 filename = os.path.splitext(os.path.basename(image_path))[0]
38 output_filename = filename + "_masked.jpg"
39 output_path = os.path.join(output_folder, output_filename)
40 cv2.imwrite(output_path, masked_image)
41
42 print("Masked image saved:", output_path)
43
44 # Configuration
45 predictor_path = "./shape_predictor_68_face_landmarks.dat"
46 input_folder = 'data/train'
47 output_folder = './data/masked_train'
48 batch_size = 128
49
50 # Create the output folder if it doesn't exist
51 if not os.path.exists(output_folder):
52     os.makedirs(output_folder)
53
54 # List all image files in the input folder
55 image_files = sorted([f for f in os.listdir(input_folder) if f.endswith((''.jpg',
    '.jpeg', '.png'))], key=lambda x: int(os.path.splitext(x)[0]))
56
57 # Process images in batches
58 for i in range(0, len(image_files), batch_size):
59     batch_images = image_files[i:i+batch_size]
60     batch_paths = [os.path.join(input_folder, image) for image in batch_images]
61     get_mask(batch_paths, predictor_path, output_folder)

```

Listing 4.2: Functie om mask toe te voegen aan de volledige dataset (Serengil, 2020)

4.3. Feature dimensionaliteitsreductie

Op de genormaliseerde afbeeldingen wordt Truncated SVD toegepast als dimensionaliteitsreductie van de features. Doordat de dataset zeer veel afbeeldingen bevat, wordt geopteerd voor Truncated SVD in plaats van het gekende PCA. PCA vereist een enorme rekenkracht op een grote dataset, omdat deze de covariantie van de matrix van de originele dataset berekent. Truncated SVD berekent de singular value decomposition van de datamatrix, wat veel efficiënter is voor grote en sparse

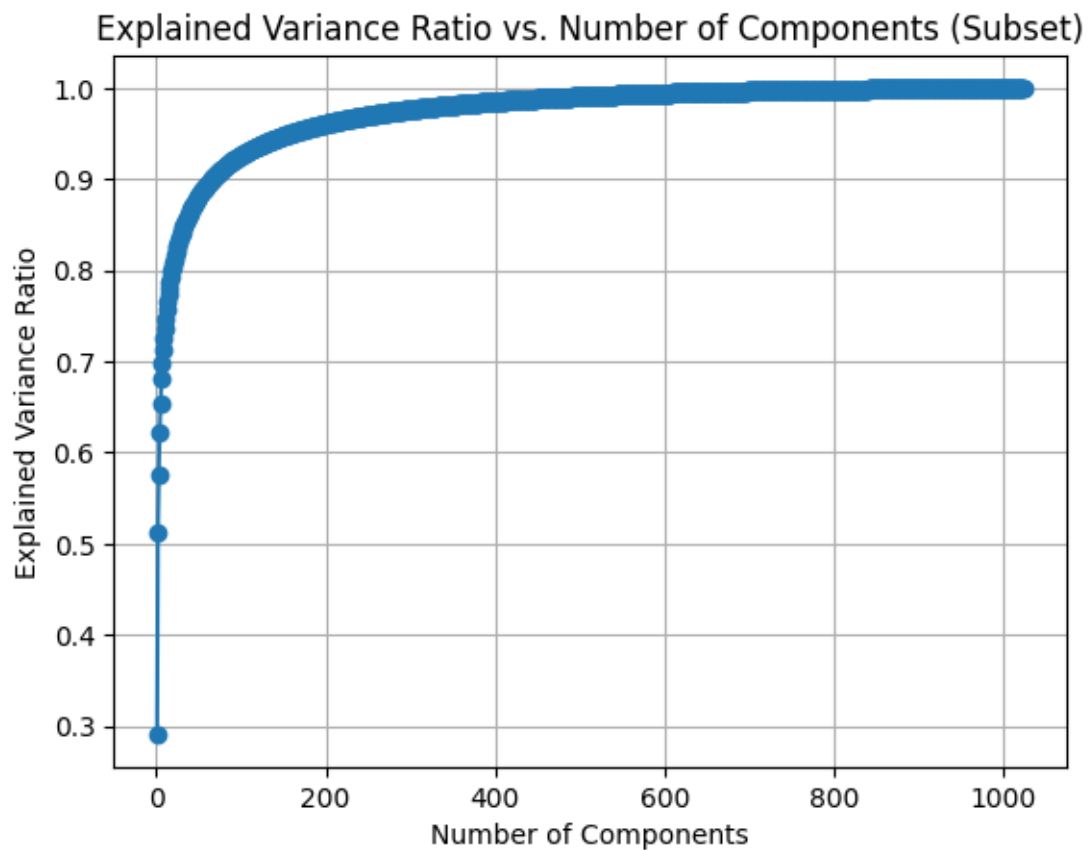
**Figuur (4.6)**

De 68 landmarks werden niet teruggevonden op de gezichtsafbeelding, deze wordt verwijderd uit de dataset.

datasets (Baruah, 2023). De benodigde opslag zal verminderen en de impact van noise op de dataset daalt. Het uitvoeren van PCA op de dataset, of zelfs een subset daarvan, genereert meteen een foutmelding over de benodigde opslag. Een standaard laptop kan deze berekening niet aan.

De Truncated SVD functie wordt beschikbaar gesteld via *sklearn.decomposition*. Aan Truncated SVD wordt het aantal componenten, dat behouden moet worden, als paramater meegegeven. Dit is de gewenste dimensionaliteit van de output data. De dimensionaliteit moet steeds minder zijn dan het aantal features in de data (Scikit-Learn, 2024). Het ideale aantal componenten kan berekend worden aan de hand van de verklaarde variantie. Er wordt een lus gemaakt door de Truncated SVD met een range die gaat tot de shape van de trainingsdata. Voor elk aantal componenten wordt de verklaarde variantie berekend. De grafiek met de verklaarde variantie per aantal componenten kan teruggevonden worden op figuur 4.7.

Er wordt verder gewerkt met 400 componenten, omdat deze een verklaarde variantie van circa 0.99 heeft. Hierdoor daalt het aantal componenten van de data van 1000 naar 400. De features van data worden zo gereduceerd om makkelijker mee verder te werken bij het trainen van de machine learning modellen. Codevoorbeeld 4.3 geeft weer hoe de Truncated SVD wordt toegepast op de volledige trainingsdataset. Het resultaat van de dimensionaliteitsreductie op een gezichtsafbeelding is te vinden op figuur 4.8.

**Figuur (4.7)**

Verklaarde variantie voor de Truncated SVD

```
1 % Apply TruncatedSVD for dimensionality reduction
2 svd = TruncatedSVD(n_components=400)
3 X_train_svd = svd.fit_transform(X_train)
```

Listing 4.3: Functie om Truncated SVD toe te passen op de volledige dataset (Scikit-Learn, [2024](#))

**Figuur (4.8)**

Truncated SVD toegepast op genormaliseerde gezichtsafbeelding

4.4. Machine learning modellen

Om het meest geschikte machine learning model te vinden voor de voorspelling van leeftijd en geslacht op basis van gezichtsafbeeldingen, worden verschillende modellen uit de literatuurstudie in hoofdstuk 2.5 getest. De performantie van 2 aparte modellen die leeftijd en geslacht voorspellen, alsook 1 model die beide voorspelt worden geanalyseerd.

Om optimaal te zoeken naar de beste parameters wordt gebruik gemaakt van Grid-Search 2.5.8. Hieraan worden meerdere parameters meegegeven, zoals aantal estimators of kernel, waarvoor GridSearch de parameter met de beste score na training weergeeft. Aan de parameters van Grid Search wordt 5 als standaardwaarde van de cross validatie meegegeven. Zoals in 2.5.8 aangehaald, worden er 5 herhalingen uitgevoerd op de dataset om het model te optimaliseren. De modellen met hun parameters worden geëvalueerd op basis van de balanced accuracy. Deze score gaat het aantal correct voorspelde instanties na, in een dataset die mogelijks ongelijk verdeeld is (García e.a., 2009). Tabel 4.4 geeft weer hoe de leeftijdscategorieën verdeeld zijn en tabel 4.4 hoe geslacht verdeeld is in de trainingsdata. Hieruit kan afgeleid worden dat de leeftijdscategorieën niet even verdeeld zijn en dat balanced accuracy hier wel degelijk de betere optie is.

De testfase van de GridSearch wordt uitgevoerd op een subset van 1000 willekeurige afbeeldingen uit de dataset om de tijd die nodig is om te trainen te vermindere. Het model met de parameters die de beste score leveren, wordt dan uiteindelijk getraind op de volledige trainingsdataset.

Leeftijdscategorie	Aantal	Percentage van volledige dataset
0-2	1399	2,19%
3-9	8003	12,52%
10-19	6880	10,76%
20-29	18924	29,60%
30 -39	13874	21,70%
40-49	7726	12,08%
50-59	4528	7,08%
60-69	1980	3,10%
more than 70	624	1%

Geslacht	Aantal	Percentage van volledige dataset
Male	32435	50,73%
Female	31503	49,27%

Om een model te trainen die beide (leeftijd en geslacht) gelijktijdig kan voorspellen, wordt gebruik gemaakt van de MultiOutputClassifier van *sklearn.multioutput*. Deze strategie bestaat uit het aanpassen van een classifier per target. Hiermee kunnen classifiers uitgebreid worden die van nature geen multi-target classificatie ondersteunen, zoals Random Forest Classifier (Scikit-Learn, 2024). Het model geeft als output 2 labels, één voor geslacht en één voor de leeftijdscategorie.

Het getrainde model wordt getest op de validatie dataset. Dit gebeurt met de *predict* functie van Scikit-learn. Op basis hiervan wordt de confusion matrix opgesteld en de performantiescores, aangehaald in de literatuurstudie 2, berekent. Scikit-Learn (2024) heeft al enkele ingebouwde functies om de performantiescores, zoals *accuracy_score* en *precision_score*, te berekenen. Deze kunnen uitgevoerd worden op de confusion matrix van het geslacht. Voor de modellen die leeftijd en beide voorspellen dienen de scores handmatig berekend te worden, omdat dit geen 2x2 matrices zijn.

4.4.1. Random Forest Classifier

Het Random Forest Classifier model wordt beschikbaar gesteld via *sklearn.ensemble*. Omdat het model geen tekstuele data kan categoriseren, dienen de labels omgezet te worden naar numerische waarden. Voor het geslacht (Female of Male) wordt

dit 0 of 1. Voor de leeftijdscategorieën wordt OrdinalEncoder gebruikt om deze te mappen naar een getal tussen 0 en 8, waarbij '0-2' categorie 0 wordt en 'more than 70' categorie 8 wordt. Aan deze Ordinal Encoder geven we een mapping mee van de labels. De encoder gaat deze dan fitten op de gegeven labels van de dataset volgens de meegegeven mapping (Scikit-Learn, [2024](#)).

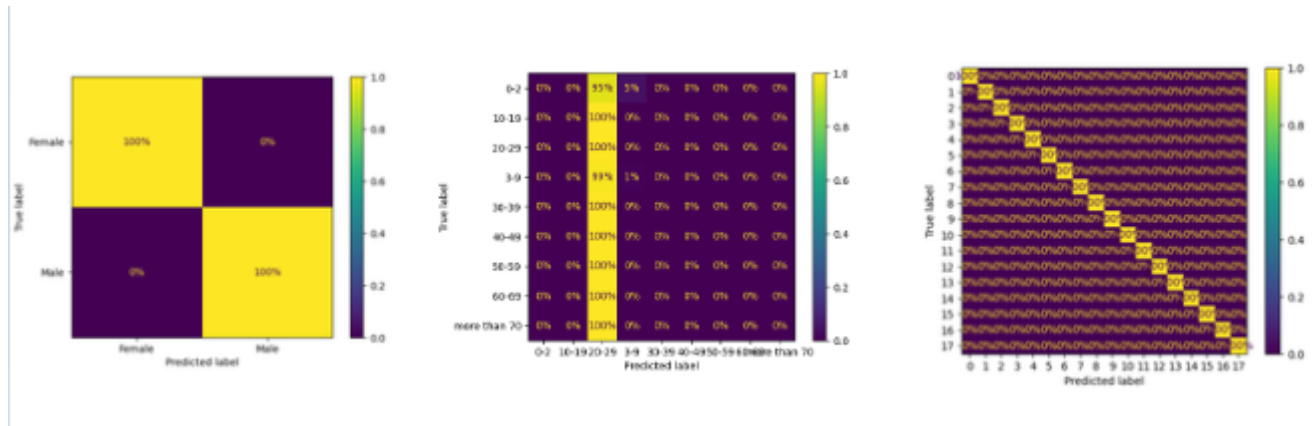
Aan de Grid Search voor de Random Forest Classifier worden volgende parameters meegegeven:

- *n_estimators*: het aantal bomen dat gebruikt wordt (default = 100).
- *max_features*: het aantal features die gebruikt worden om de beste scheiding in de bomen te vinden (default = 'sqrt')
- *min_samples_split*: het minimum aantal instanties die nodig zijn om node te mogen splitsen (default = 2).
- *min_samples_leaf*: minimum aantal instanties die in een blad moeten zitten (default = 1). Een blad is het uiteinde van een boom, die de voorspelde klasse weergeeft Garg ([2020](#)).
- *max_depth*: de maximale diepte van een boom (default = None).
- *bootstrap*: het gebruik van steekproeven bij het bouwen van de bomen (default = True)

	Geslacht	Leeftijd	Beide
Parameters	<i>n_estimators</i> = 230	<i>max_depth</i> =5, <i>n_estimators</i> =200	<i>n_estimators</i> =100
Accuracy	99,99%	29,71%	99,95%
Precision	99,99%	0,00%	99,93%
Recall	100%	11,21%	99,92%
Specificity	99,99%	0,00%	99,99%
F1	99,99%	0.00%	99,93%

Figuur [4.9](#) geeft de confusion matrices voor de 3 modellen weer. Uit de confusion matrices en de scores kunnen we afleiden dat de Random Forest Classifier uitstekend presteert voor de voorspelling van geslacht en beide. Voor de voorspelling van de leeftijdscategorie behaalt dit model slechts een score van circa 30%. Uit de confusion matrix van de leeftijdsvoorspellingen is ook af te leiden dat vrijwel enkel de leeftijdscategorie 20-29 wordt voorspelt. Dit kan te verklaren zijn door overfitting, omdat deze klasse het meest aanwezig is in de dataset is de kans het grootst dat de voorspelling juist is (Kreiger, [2020](#)). Het model presteert wel goed wanneer

het geslacht ook wordt toegevoegd. Dit kan te verklaren zijn door de levensfactoren en opvoeding, aangehaald in 2.7. Het model krijgt de context van het geslacht mee en kan hierdoor wel patronen herkennen.



Figuur (4.9)

Confusion Matrix voor Random Forest Classifier

4.4.2. XGBoost Classifier

Het XGBoost, Extreme Gradient Boosting, Classifier model wordt beschikbaar gesteld via *xgboost*. Ook aan dit model kan geen tekstuele data worden meegeven en worden de labels omgezet in numerische waarden, zoals in 4.4.1.

Voor de classificatie van beide labels met de MultiOutputClassifier vindt nog een preprocessing stap plaats. De tekstuele labels moeten omgezet worden naar numerische categorieën. Hiervoor wordt de LabelEncoder van *sklearn.preprocessing* gebruikt. De labels worden dan samengevoegd met *np.column_stack*. Codevoorbeeld 4.4 geeft weer hoe de labels worden aangepast voor de MultiOuputClassifier. Het label voor een vrouwelijke 0-2 jarige wordt zo [0,1].

Aan de Grid Search voor de XGBoost Classifier worden volgende parameters meegegeven:

n_estimators: het aantal bomen dat gebruikt wordt (default = 100).

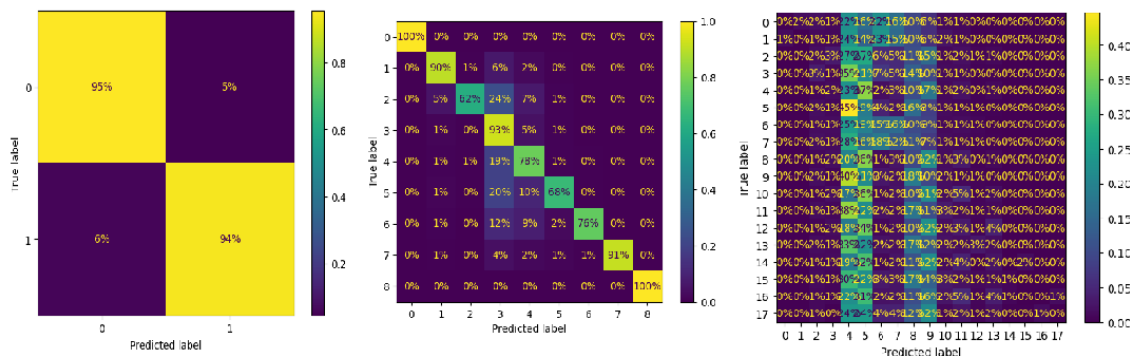
max_depth: de maximale diepte van een boom (default = 6).

```
1 age_class_encoded = label_encoder.fit_transform(y_train_age)
2 gender_encoded = label_encoder.fit_transform(y_train_gender)
3 y_combined = np.column_stack((age_class_encoded, gender_encoded))
4
```

Listing 4.4: Functie om labels aan te passen voor de MultiOutputClassifier (Roepke, 2024) en (Numpy, 2024)

	Geslacht	Leeftijd	Beide
Parameters	$n_estimators = 267, depth=5$	$n_estimators=300, depth=4$	$n_estimators=100$
Accuracy	94.76%	81,82%	10,23%
Precision	95.22%	90,98%	6,50%
Recall	94.09%	84,16%	5,35%
Specificity	95.41%	97,30%	94,32%
F1	94.65%	87,44%	5,87%

Figuur 4.10 geeft de confusion matrices voor de XGBoost classifier modellen weer. Hieruit kunnen we afleiden dat de XGBoost Classifier zeer goed presteert voor de aparte voorspelling van leeftijd en geslacht. Wanneer beide labels gelijktijdig worden voorspeld, presteert het model zeer slecht. Ook dit kan te verklaren zijn door overfitting, de aard van de data of het model kan geen patronen vinden.



Figuur (4.10)

Confusion Matrix voor XGBoost Classifier

4.4.3. Support Vector Machines

Het Support Vector Classifier (SVC) model wordt beschikbaar gesteld via *sklearn.svm*. Een SVM kan wel classificeren op basis van tekstuele data. Aan dit model wordt wel de originele dataset meegegeven en worden de labels niet gewijzigd.

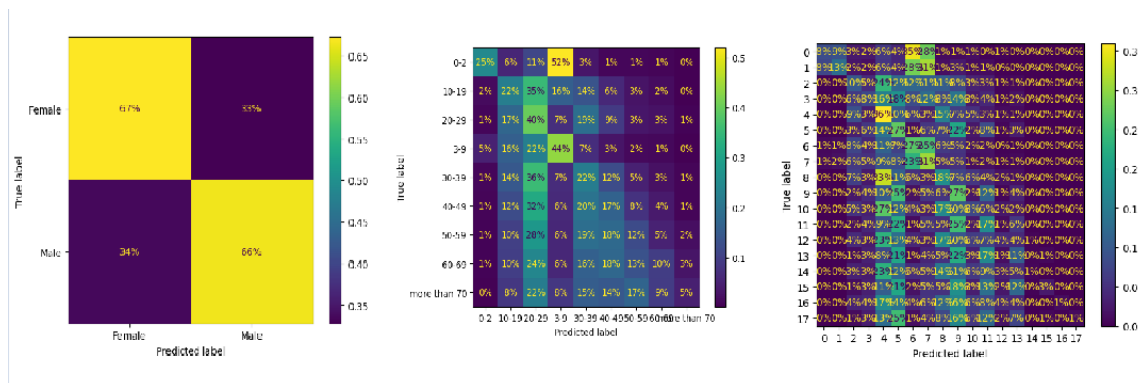
Aan de Grid Search voor het SVC model worden volgende parameters meegegeven:

- C: regularisatie parameter, afhankelijk van de grootte van de dataset (default = 1.0).
- gamma: coëfficiënt van de kernel (default = 'scale')
- kernel: kernel type die gebruikt wordt in het algoritme (default = 'rbf').

- degree: graad van de kernel (default = 3).

	Geslacht	Leeftijd	Beide
Parameters	C=125	C=125, kernel='linear', degree=1	C=25
Score model (train)	68.41%	21.52%	20.50%
Accuracy	66.79%	28.23%	21.55%
Precision	67.57%	22.12%	19.44%
Recall	66.40%	21.70%	13.87%
Specificity	67.19%	90.17%	95.15%
F1	66.98%	21.91%	16.18%

Figuur 4.11 geeft de confusion matrices voor de SVC modellen weer. Uit de scores kan afgeleid worden dat het SVC model slechter presteert dan voorgaande modellen. Voor de voorspelling van geslacht wordt hier slechts een score van circa 70% behaald. Ook de modellen voor de voorspelling van leeftijd en beide presteren slecht, met een score van slechts 20%. Ook uit de confusion matrices is af te leiden dat er veel foute voorspellingen zijn. Hieruit kan geconcludeerd worden dat een SVC niet het geschikte model is voor de classificatie van gezichtsafbeeldingen.



Figuur (4.11)

Confusion Matrix voor Support Vector Classifier

4.4.4. Voting Classifier

De Voting Classifier is een ensemble techniek die meerdere modellen combineert. De SVC, random forest en XGBoost Classifier worden tegelijk aangesproken om de accuracy van het model te verhogen. De parameters die het beste scoorden voor de afzonderlijke modellen worden hier hergebruikt. Codevoorbeeld ?? geeft weer hoe de Voting Classifier de verschillende modellen combineert.

	Geslacht	Leeftijd	Beide
Parameters	?	?	?
Accuracy	?%	?%	?%
Precision	?%	?%	?%
Recall	?%	?%	?%
Specificity	?%	?%	?%
F1	?%	?%	?%

5

Conclusie

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

Vivamus adipiscing. Curabitur imperdiet tempus turpis. Vivamus sapien dolor, congue venenatis, euismod eget, porta rhoncus, magna. Proin condimentum pretium enim. Fusce fringilla, libero et venenatis facilisis, eros enim cursus arcu, vitae facilisis odio augue vitae orci. Aliquam varius nibh ut odio. Sed condimentum condimentum nunc. Pellentesque eget massa. Pellentesque quis mauris. Donec ut ligula ac pede pulvinar lobortis. Pellentesque euismod. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent elit. Ut laoreet ornare est. Phasellus gravida vulputate nulla. Donec sit amet arcu ut sem tempor malesuada. Praesent hendrerit augue in urna. Proin enim ante, ornare vel, consequat ut, blandit in, justo. Donec felis elit, dignissim sed, sagittis ut, ullamcorper a, nulla. Aenean pharetra vulputate odio.

Quisque enim. Proin velit neque, tristique eu, eleifend eget, vestibulum nec, lacus. Vivamus odio. Duis odio urna, vehicula in, elementum aliquam, aliquet laoreet, tellus. Sed velit. Sed vel mi ac elit aliquet interdum. Etiam sapien neque, convallis et, aliquet vel, auctor non, arcu. Aliquam suscipit aliquam lectus. Proin tincidunt magna sed wisi. Integer blandit lacus ut lorem. Sed luctus justo sed enim.

Morbi malesuada hendrerit dui. Nunc mauris leo, dapibus sit amet, vestibulum et, commodo id, est. Pellentesque purus. Pellentesque tristique, nunc ac pulvinar adipiscing, justo eros consequat lectus, sit amet posuere lectus neque vel augue. Cras consectetur libero ac eros. Ut eget massa. Fusce sit amet enim eleifend sem dictum auctor. In eget risus luctus wisi convallis pulvinar. Vivamus sapien risus, tempor in, viverra in, aliquet pellentesque, eros. Aliquam euismod libero a sem. Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.



Onderzoeksvoorstel

Het onderwerp van deze bachelorproef is gebaseerd op een onderzoeksvoorstel dat vooraf werd beoordeeld door de promotor. Dat voorstel is opgenomen in deze bijlage.

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 pre-processing, 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.

A.1. Introductie

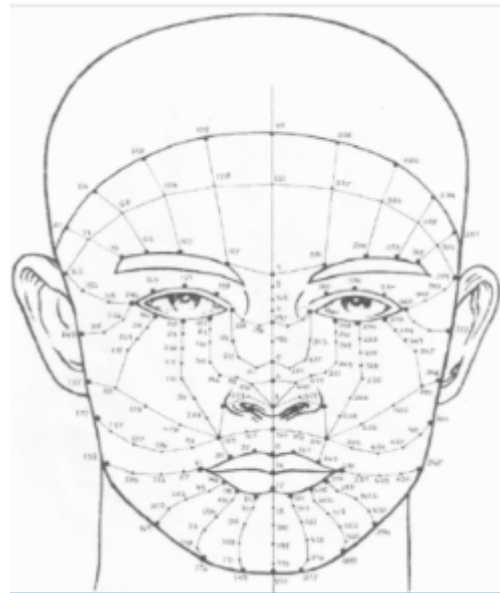
De onderzoeksvraag werd aangeboden door het bedrijf IntelliProve. IntelliProve biedt online gezondheidsoplossingen, een software die in staat is om binnen enkele seconden nauwkeurig gezondheidsparameters te bepalen, gebaseerd op een optische meting van het gezicht. Het doel van de bachelorproef is het 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 modeltraining. 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.

A.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 [A.2.1](#)



Figuur (A.1)

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

3. Classificatie: het analyseren en identificeren van de data na de feature extractie.

A.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 A.1 staan de belangrijkste antropometrische landmarks aangeduid. Het is noodzakelijk om de belangrijkste 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).

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

**Figuur (A.2)**

Gezichtsafbeeldingen voor normalisatie (C. Chen, 2011).

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 A.2 geeft een voorbeeld van een dataset met gezichtsfoto's. In figuur A.3 worden deze afbeeldingen genormaliseerd, waardoor er minder variaties overblijven in de afbeeldingen en we bijvoorbeeld al geen achtergrond overhouden.

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

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.

Adaptive Boosting

De Adaptive Boost, of AdaBoost, is typisch een classificatie tussen twee klassen. Om het herkenningsprobleem met meerdere klassen op te lossen, kan een majo-

**Figuur (A.3)**

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

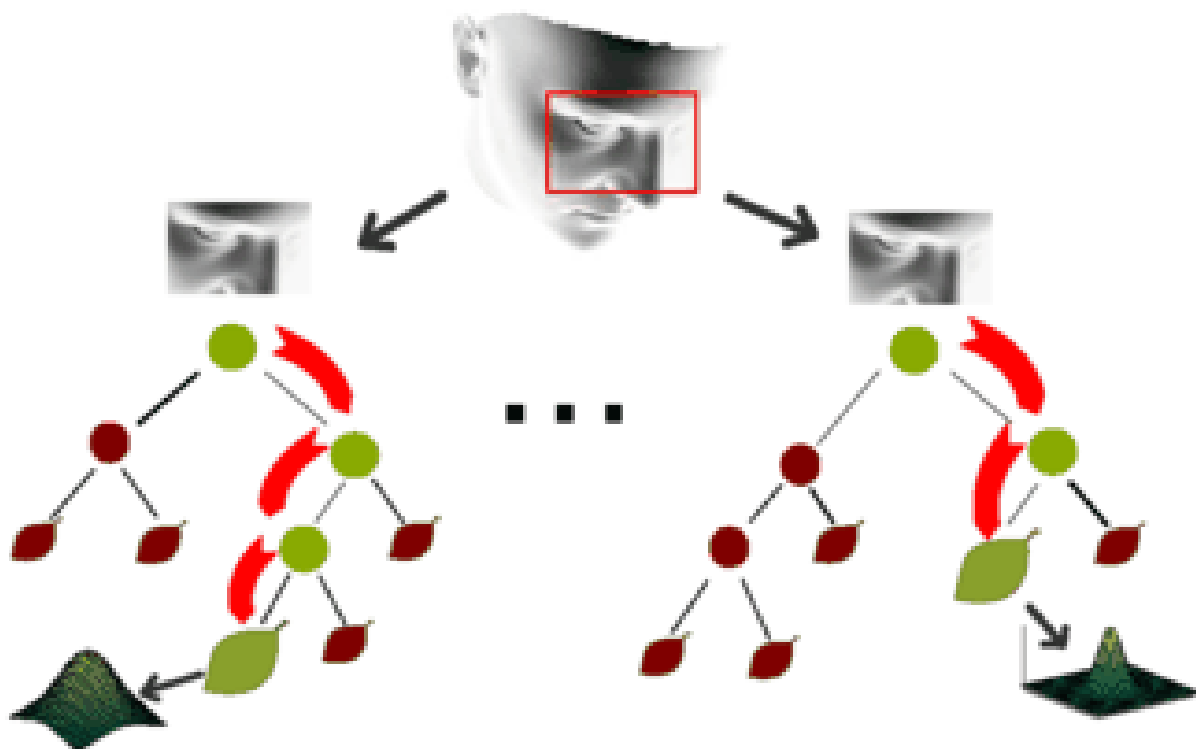
rity voting (MV) strategie worden gebruikt om alle paarsgewijze classificatieresultaten te combineren. Hierbij kiezen we de meest voorkomende klasse als voorspelling. AdaBoost is een adaptief algoritme om een reeks classificeerders te boosten, in die zin dat de gewichten dynamisch worden bijgewerkt op basis van de fouten in eerdere leerresultaten (G.-D. Guo & Zhang, 2001).

Random Forest Classifier

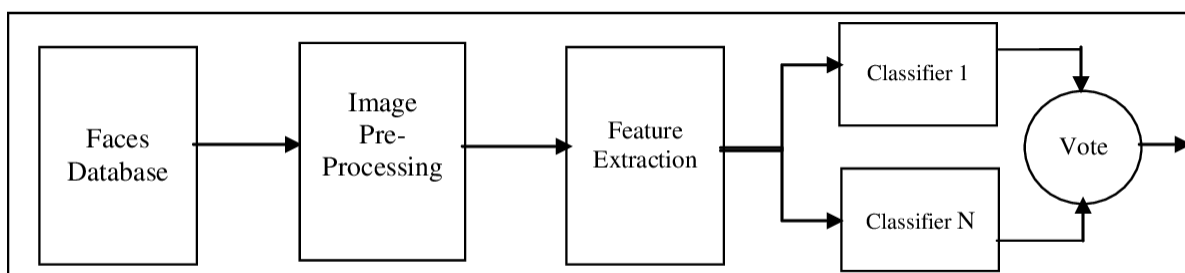
Random forests zijn verzamelingen decision trees die elk getraind zijn op een willekeurig gekozen deelverzameling van de beschikbare data. Deze decision trees worden willekeurig door de trainingsvoorbeelden die aan elke boom worden gegeven, maar ook door een willekeurige subset van tests die beschikbaar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

Ensemble

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

**Figuur (A.4)**

Voorbeeld van een random forest voor het schatten van de houding van het hoofd(Fanelli e.a., [2012](#)).

**Figuur (A.5)**

Robuust Gezichtsherkenning systeem(H. A. Khan, [2017](#)).

**Figuur (A.6)**

Methodologie voor het onderzoek.

A.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 trainingdata. Hierdoor zal het systeem falen wanneer we nieuwe data willen classificeren (Coppens, 2018).

A.3. Methodologie

Figuur A.6 geeft een visuele representatie van de opgestelde methodologie.

A.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 functionaliteiten 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.

A.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 extraheren 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.

A.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.

A.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.

A.3.5. Afwerken scriptie

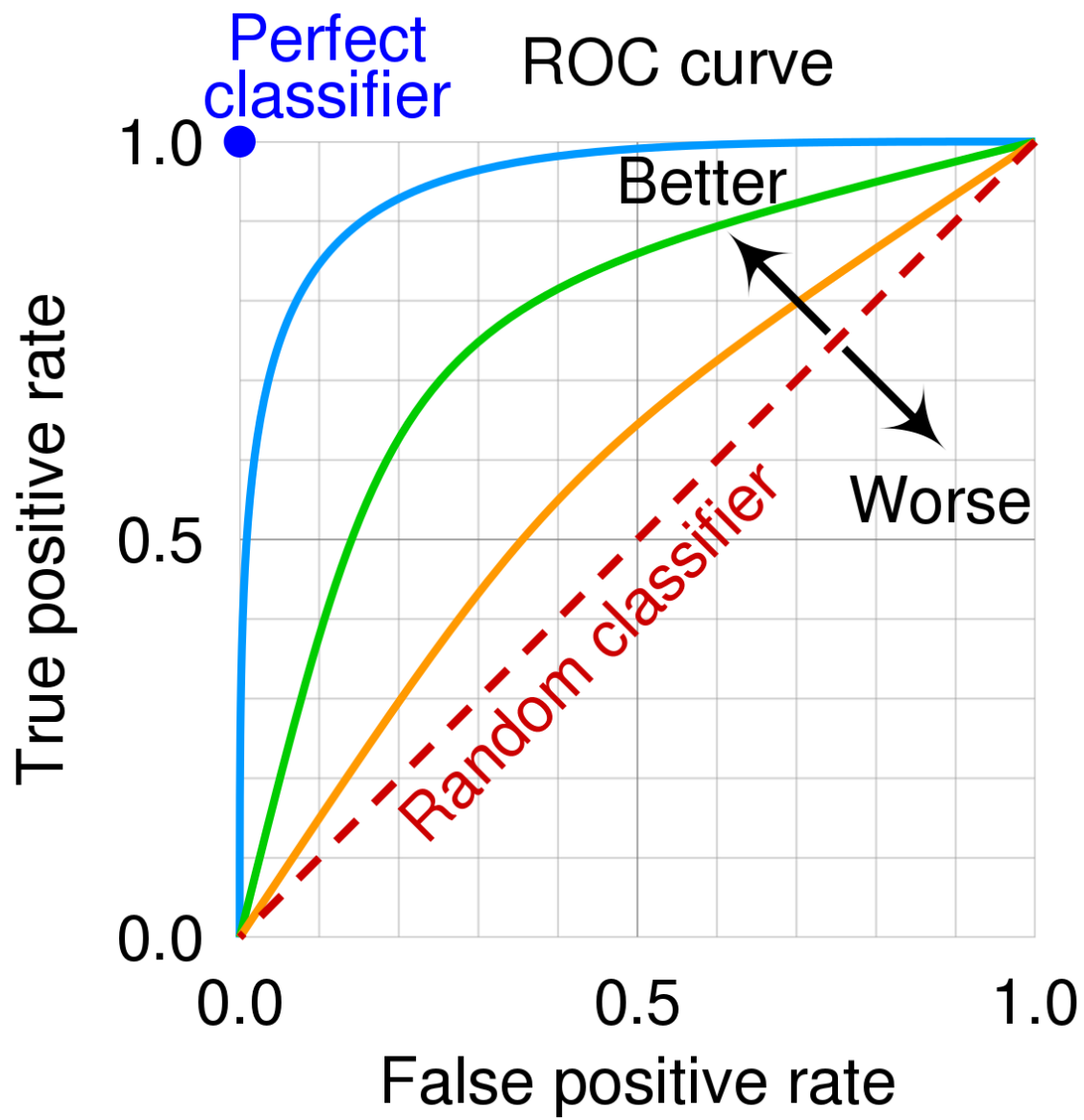
De laatste fase, die 2 weken duurt, omvat het afwerken van de bachelorproef. Dit is het eindresultaat van het geleverde onderzoek met een proof-of-concept die zal worden ingediend.

A.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 ba-

sis 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 A.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



Figuur (A.7)
ROC curve(Gomede, 2023).

Bibliografie

- Baruah, I. D. (2023). Dimensionality Reduction Techniques — PCA, LCA and SVD. *Nerd For Tech*. <https://medium.com/nerd-for-tech/dimensionality-reduction-techniques-pca-lca-and-svd-f2a56b097f7c>
- Basystiuk, O., Melnykova, N., & Rybchak, Z. (2023). Machine Learning Methods and Tools for Facial Recognition Based on Multimodal Approach. In M. Emmereich, V. Vysotska & V. Lytvynenko (Red.), *Proceedings of the Modern Machine Learning Technologies and Data Science Workshop (MoMLet 2023), Lviv, Ukraine, June 3, 2023* (pp. 161–170, Deel 3426). CEUR-WS.org. <https://ceur-ws.org/Vol-3426/paper13.pdf>
- Beheshti, N. (2022). Cross Validation and Grid Search: Using sklearn's GridSearchCV on random forest model. *Towards Data Science*. <https://towardsdatascience.com/cross-validation-and-grid-search-efa64b127c1b>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In S. A. Friedler & C. Wilson (Red.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91, Deel 81). PMLR. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Chen, C. (2011). *Gender Classification from Facial Images* [proefschrift, West Virginia University Libraries]. <https://doi.org/10.33915/etd.3449>
- 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>
- Déniz, O., Bueno, G., Salido, J., & De la Torre, F. (2011). Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, 32(12), 1598–1603. <https://doi.org/10.1016/j.patrec.2011.01.004>
- Educative.io. (2024, maart 18). *Hyperparameter tuning using GridSearchCV* (U. Irsahad, Red.). <https://www.educative.io/answers/hyperparameter-tuning-using-gridsearchcv>
- 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>

- García, V., Mollineda, R. A., & Sánchez, J. S. (2009). Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. In *Lecture Notes in Computer Science* (pp. 441–448). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02172-5_57
- Garg, A. (2020). Decision Trees in Machine Learning. *DataX Journal*. <https://medium.com/data-science-community-srm/decision-trees-in-machine-learning-7fdaf3d09a0>
- Géron, A. (2019, september 1). *Hands-On Machine Learning with Scikit-Learn, Keras Tensorflow*.
- 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., & Mu, G. (2014). A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, 32(10), 761–770. <https://doi.org/10.1016/j.imavis.2014.04.011>
- 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>
- Gupta, S. K., & Nain, N. (2022). Review: Single attribute and multi attribute facial gender and age estimation. *Multimedia Tools and Applications*, 82(1), 1289–1311. <https://doi.org/10.1007/s11042-022-12678-6>
- Ibtisamah, N. (2023). Tuning Multiple ML Models using GridSearchCV: A Project of Predicting Steel's Optimal Temperature to Reduce Electricity Consumption. *Medium*.
- 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
- Karkkainen, K., & Joo, J. (2021). FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. <https://doi.org/10.1109/wacv48630.2021.00159>
- Khan, H. A. (2017). Feature Fusion and Classifier Ensemble Technique for Robust Face Recognition, 3–8. <https://api.semanticscholar.org/CorpusID:203632225>
- Khan, S. A., Nazir, M., Riaz, N., Naveed, N., Zulfikar, S., & Bhutto, A. (2011). Computationally Intelligent Gender Classification Techniques: An Analytical Study. <https://api.semanticscholar.org/CorpusID:10305340>
- King, D. E. (Red.). (2024, april 8). <https://github.com/davisking/dlib-models>
- Kreiger, J. R. (2020). Evaluating a Random Forest model. *Analytics Vidhya*. <https://medium.com/analytics-vidhya/evaluating-a-random-forest-model-9d165595ad56>

- 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>
- Lakshmiprabha, N. S. (2016). Face Image Analysis using AAM, Gabor, LBP and WD features for Gender, Age, Expression and Ethnicity Classification. <https://doi.org/10.48550/ARXIV.1604.01684>
- Lin, D.-T. (2006). Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network. *Journal of Information Science and Engineering*.
- Mady, H., & Hilles, S. M. S. (2018). Face recognition and detection using Random forest and combination of LBP and HOG features. *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*. <https://doi.org/10.1109/icscee.2018.8538377>
- Narkhede, S. (2018, mei 9). *Understanding Confusion Matrix*. Towards Data Science. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
- Numpy. (2024, mei 5). *Numpy: The fundamental package for scientific computing with Python*.
- OpenCV. (2024, april 8). <https://docs.opencv.org>
- Othman, Z. A., & Adnan, D. A. (2014). Age classification from facial images system. *Internation Journal of Computer Science and Mobile Computing*, 3(10), 291–303.
- Ramon, A. (2024). Machine Learning Algorithms for Gender Prediction.
- Ramos, L. P. (2024, maart 31). *Python's zipfile: Manipulate Your ZIP Files Efficiently*. Real Python. <https://realpython.com/python-zipfile/>
- Roepke, B. (2024). Go Beyond Binary Classification with Multi-Class Plus Multi-Label. *Data Knows All*.
- Rustam, Z., & Ruvita, A. (2018). Application Support Vector Machine on Face Recognition for Gender Classification. *Journal of Physics: Conference Series*, 1108, 012067. <https://doi.org/10.1088/1742-6596/1108/1/012067>
- Salih Hasan, B. M., & Abdulazeez, A. M. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, 2(1). <https://doi.org/10.30880/jscdm.2021.02.01.003>
- Sánchez López, L. (2010). Local Binary Patterns applied to Face Detection and Recognition, 7–8.
- 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>
- Scikit-Learn. (2024, mei 6). *Scikit-learn: Machine Learning in Python*.

- Serengil, S. I. (2020, november 20). *Facial Landmarks for Face Recognition with Dlib*. <https://sefiks.com/2020/11/20/facial-landmarks-for-face-recognition-with-dlib/>
- Shah, P. M. (2012, juni 1). *Face Detection from Images Using Support Vector Machine* [proefschrift, San Jose State University Library]. <https://doi.org/10.31979/etd.wg5s-gyqn>
- 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>
- Tomasi, C. (2015). Histograms of Oriented Gradients. <https://api.semanticscholar.org/CorpusID:15782505>
- Wang, Y., Ricanek, K., Chen, C., & Chang, Y. (2010). Gender classification from infants to seniors. *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. <https://doi.org/10.1109/btas.2010.5634518>
- XGBoost. (2023). *Introduction to Boosted Trees*. <https://xgboost.readthedocs.io/en/stable/tutorials/model.html>
- Yavuz, S. (2014). Gender Recognition with Logistic Regression. https://www.academia.edu/6634049/Gender_Recognition_with_Logistic_Regression
- Yeh, M. (2022). An Efficient, Tolerance-Based Algorithm for the Truncated SVD. *UC Berkeley*.
- 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.