

Universaliteit van emoties

Automatische herkenning van gezichtsuitdrukkingen
bij verschillende geslachten en rassen

Koen Rodenburg – 3507696

Christien Visser – 3476952

Begeleider: Dr. Gerard Vreeswijk

13-07-2012

ECTS: 7,5

ABSTRACT

In dit onderzoek worden door middel van een artificieel neurale netwerk emotieherkenning bij blanke mannen vergeleken met emotieherkenning bij blanke vrouwen. Ditzelfde gebeurt met Aziatische vrouwen en blanke vrouwen. De prestaties bij de mannen en vrouwen verschillen niet van elkaar, terwijl er bij de Aziatische en blanke vrouwen wel een significant verschil te zien is.

De door ons geschreven broncode van het gebruikte artificieel neurale netwerk is te vinden op:
<http://codeviewer.org/view/code:28a3>

Inhoudsopgave

1 Inleiding	3
2 Eerder onderzoek	3
2.1 Kennis over emoties	3
2.2 Geautomatiseerde emotieherkenning	5
3 Hypothese.....	8
3.1 Conditie 1: Mannen en vrouwen.....	8
3.2 Conditie 2: Aziaten en blanken.....	9
4 Methode	9
4.1 Data	9
4.2 Netwerk	12
4.3 Uitvoering.....	14
5 Resultaten.....	15
5.1 Conditie 1: Mannen en vrouwen.....	16
5.2 Conditie 2: Aziaten en blanken.....	17
6 Discussie	18
6.1 Conditie 1: Mannen en vrouwen.....	18
6.2 Conditie 2: Aziaten en blanken.....	19
7 Conclusie.....	21
7.1 Conditie 1: Mannen en vrouwen.....	21
7.2 Conditie 2: Aziaten en blanken.....	21
8 Dankwoord	21
9 Referenties	22

1 Inleiding

In de huidige maatschappij wordt de communicatie tussen mensen en computers steeds belangrijker en daardoor steeds verder verfijnd. Niet langer zitten we alleen achter een terminal om de computer met commando's te vertellen wat hij moet doen, maar wordt er ook steeds meer gebruik gemaakt van onder andere spraakherkenning. Daarnaast zijn computers meer en meer in staat om te antwoorden in natuurlijke taal, geschreven dan wel gesproken.

Bij een door een mens gesproken boodschap wordt echter slechts 7% van de informatie overgebracht via de woorden en 38% via de intonatie. Er blijft 55% van de boodschap over, die via de gezichtsuitdrukking van de spreker wordt overgebracht (Mehrabian, 1968). Om deze reden zou een goede automatische herkenning van gezichtsexpressies de communicatie tussen mens en computer sterk kunnen verbeteren.

De eerste stap naar verbetering trachten wij in deze bachelorscriptie voor de studie Kunstmatige Intelligentie te zetten. We doen onderzoek naar emotieherkenning door middel van een Artificieel Neuraal Netwerk. Specifiek kijken we daarbij naar de universaliteit van gezichtsexpressies. Daarbij vragen we ons af of een Artificieel Neuraal Netwerk dat getraind wordt om emoties bij het ene geslacht te herkennen, even goed de emoties van het andere geslacht kan herkennen. Ditzelfde onderzoek doen we ook voor verschillende rassen. Door de beperkte beschikbaarheid van databases met gezichtsexpressies van personen van verschillende rassen beperken we ons hier tot het vergelijken van de prestaties bij Japanse vrouwen en blanke vrouwen.

We beginnen in hoofdstuk 2 met het bespreken van eerder onderzoek op het gebied van emotieherkenning in het algemeen en emotieherkenning met een Artificieel Neuraal Netwerk. In hoofdstuk 3 komen onze hypothesen ten aanzien van de onderzoeksvragen aan de orde. Vervolgens wordt in hoofdstuk 4 besproken hoe we het onderzoek aanpakken en worden in hoofdstuk 5 de

resultaten van het onderzoek gepresenteerd. Hoofdstuk 6 en 7 tenslotte bevatten respectievelijk de discussie en conclusie.

Allereerst is het nodig om heldere onderzoeksvragen te formuleren. We zullen in de rest van deze scriptie aan onze twee onderzoeken refereren als conditie 1 en conditie 2. Conditie 1 is het onderzoek waarbij we de prestaties van een Artificieel Neuraal Netwerk op emotieherkenning bij mannen en vrouwen bekijken. Bij conditie 2 doen we ditzelfde onderzoek met Japanse vrouwen en blanke vrouwen. Vanzelfsprekend heeft elke conditie een eigen onderzoeksvraag.

Conditie 1: Mannen en vrouwen

Wanneer een Artificieel Neuraal Netwerk getraind wordt om emoties bij het ene geslacht te herkennen, is het dan net zo goed in emotieherkenning bij het andere geslacht?

Conditie 2: Aziaten en blanken

Wanneer een Artificieel Neuraal Netwerk getraind wordt om emoties bij blanken te herkennen, is het dan net zo goed in emotieherkenning bij Aziaten vergeleken met blanken, en vice versa?

2 Eerder onderzoek

In dit onderdeel zal eerder onderzoek op het gebied van emotieherkenning besproken worden. Allereerst komen er recente onderzoeken naar de algemene uiting van emoties en het herkennen van emoties door mensen aan de orde. In het tweede onderdeel worden enkele recente onderzoeken uitgelicht waarbij emotieherkenning wordt geautomatiseerd, met name door middel van een Artificieel Neuraal Netwerk.

2.1 Kennis over emoties

Er wordt veel onderzoek gedaan naar menselijke emoties. Bij de meeste van deze onderzoeken ligt de focus op de verschillende

menselijke emoties. Andere onderzoeken richten zich meer op de manier waarop mensen die emoties vertonen en herkennen. In relatie tot het tonen van emoties door middel van het gezicht is veel gekeken naar variatie in uiting tussen mannen en vrouwen, daarnaast is er ook onderzoek gedaan naar verschillen tussen personen die niet in dezelfde cultuur zijn opgegroeid.

In 1972 onderzocht Ekman of emoties bij mensen universeel zijn. Ekman bezocht een geïsoleerd levende stam in Papoea-Nieuw-Guinea en legde de leden van deze stam afbeeldingen voor van mensen die een emotie vertoonden. Deze mensen leefden in culturen waar de stam niet bekend mee was. Toch waren de stamleden in staat bij de foto's de juiste emotie te benoemen. Uit de resultaten van het onderzoek concludeert Ekman dat er zes universele basisemoties bestaan, die elk mens kan vertonen en herkennen. Deze basisemoties zijn woede, afschuw, angst, vreugde, verdriet en verbazing.

2.1.1 Verschillen tussen mannen en vrouwen

Er wordt in veel vakgebieden onderzoek gedaan naar de verschillen tussen mannen en vrouwen. Zo is er ook al vaak onderzoek gedaan naar de verschillen in de manier van tonen van emotie tussen mannen en vrouwen. Daarnaast wordt er ook onderzoek gedaan naar de invloed van het geslacht van de persoon die de emoties vertoont op de herkenning van emoties.

In een onderzoek van Kring et al. (1998) werd aangetoond dat mannen en vrouwen emoties anders uiten maar niet per se anders ervaren. In hun onderzoek lieten ze studenten een emotionele film zien. Tijdens het bekijken van de film werd er door de onderzoekers vastgesteld welke emoties de proefpersonen met hun gezicht vertoonden. Achteraf moesten de proefpersonen aangeven welke emoties ze in welke mate ervaren hadden. Uit de resultaten bleek dat mannen en vrouwen emoties in dezelfde mate ervaren. De

onderzoekers constateerden echter dat de vrouwen tijdens het kijken van de film veel expressiever waren dan de mannen.

Hess et al. (2009) deden een onderzoek met getekende gezichten waaruit alle geslachtskenmerken weggelaten waren. Bij het onderzoek kreeg een proefpersoon tweemaal hetzelfde gezicht voorgelegd, de ene keer met een mannelijk kapsel en de andere keer met een vrouwelijk kapsel. Hess et al. rapporteren dat de proefpersonen het gezicht met het vrouwelijke kapsel eerder als boos beoordelen dan het gezicht met het mannelijke kapsel.

De resultaten van Hess et al. zijn tegenstrijdig met het stereotype dat mannen vaker boos zijn dan vrouwen. Toch zijn ze te verklaren doordat de neutrale gezichtsuitdrukking van een man van nature bozer is. Een man hoeft zijn gezichtsuitdrukking minder aan te passen om dezelfde boze gezichtsuitdrukking te bereiken dan een vrouw dat hoeft te doen. Het contrast met de neutrale uitdrukking is bij deze boze gezichtsuitdrukking bij de man dus kleiner dan bij de vrouw. Hierdoor veronderstelt de proefpersoon dat de vrouw bozer is dan de man; de vrouw doet meer moeite om haar gezichtsuitdrukking aan te passen en ervaart dus waarschijnlijk een intensere emotie.

Het onderzoek van Burton et al. (1993) toont aan dat er duidelijke verschillen in de gezichtsopbouw van mannen en vrouwen zijn. In het onderzoek werden afbeeldingen gebruikt van personen waarvan het haar verborgen was door een badmuts. In deze afbeeldingen werd de afstand tussen bepaalde punten in het gezicht, bijvoorbeeld op het voorhoofd en de neus, gemeten. Uit het onderzoek bleek dat met deze simpele afstanden het geslacht van de proefpersoon berekend kon worden. De opbouw van het gezicht is dus verschillend bij mannen en vrouwen.

Uit onderzoek blijkt dat er bij het vertonen en herkennen van emoties wel degelijk een verschil is tussen mannen en vrouwen. Vrouwen vertonen hun emotie uitbundiger dan mannen, hierdoor moet er bij het herkennen van emoties ook rekening gehouden worden met het

geslacht van de persoon die de emotie uit. Het blijkt uit het onderzoek van Hess et al. dat mensen automatisch rekening houden met het geslacht van de persoon die de emotie vertoont. Daarnaast is de gezichtsopbouw van mannen en vrouwen anders zoals het onderzoek van Burton et al. uitwijst.

2.1.2 Verschillen tussen culturen

Niet alleen het geslacht maar ook het ras en de cultuur waarin iemand opgroeit hebben invloed op de manier waarop iemand emoties vertoont. Uit de volgende onderzoeken blijkt dat er een duidelijk verschil is in de manier van het tonen van emoties tussen mensen die in verschillende culturen zijn opgegroeid. Hieronder wordt een aantal van deze onderzoeken besproken.

Camras et al. hebben in 2006 aangetoond dat de manier waarop iemand emoties toont afhankelijk is van de omgeving waarin die persoon opgroeit. De onderzoekers keken naar vier groepen meisjes met verschillende afkomst en cultuur: Chinese meisjes die opgroeiden (1) in China bij Chinese ouders, (2) in de Verenigde Staten bij Chinese ouders, (3) in de Verenigde Staten bij Amerikaanse ouders (geadopteerd). De vierde groep is een controlegroep met Amerikaanse meisjes die in de Verenigde Staten door Amerikaanse ouders werden opgevoed.

Uit het onderzoek van Camras et al. komt duidelijk naar voren dat er een verschil is tussen de meisjes die in een Chinees gezin zijn opgevoed en meisjes die in een Amerikaans gezin zijn opgevoed. Zowel Chinese als Amerikaanse meisjes die door Amerikaanse ouders in de Verenigde Staten opgevoed werden, tonen hun emoties uitbundig. Chinese kinderen die door Chinese ouders worden opgevoed, ongeacht of zij in China of in de Verenigde Staten wonen, doen dit veel minder. De emoties die hierbij het meest significant verschillen zijn vreugde en afschuw.

Yuki et al. (2005) toonden aan dat cultuur deels bepaalt welke delen van het gezicht er gebruikt worden in het tonen van emotie. Omdat de mond het meest expressieve deel van

het gezicht is, wordt in culturen waar het de norm is om openlijk emotie te tonen vooral gebruik gemaakt van de mond. Dit is het geval in bijvoorbeeld de Amerikaanse cultuur en andere Westerse culturen. Mensen die in deze culturen opgroeien zullen daarom naar de mond kijken om een emotie te herkennen. In andere culturen, zoals de Japanse, is het echter de norm om emoties niet openlijk te tonen. In deze culturen speelt de mond een kleinere rol bij het tonen van emoties en geeft het gebied rond de ogen de meeste informatie. Personen die opgroeien in deze culturen zullen daardoor vooral naar de ogen kijken om een emotie te herkennen.

Naast de manier waarop met emoties omgegaan wordt, is er nog een duidelijk verschil tussen Aziaten en blanken. Beide groepen hebben uiterlijke kenmerken die hen van elkaar en ander rassen onderscheiden. In een recent onderzoek (Zhuang et al., 2010) is onder andere vastgesteld dat blanken bij gelijk gewicht en gelijke leeftijd een significant smaller en langer gezicht hebben dan Aziaten. Ook is vastgesteld dat Aziaten een significant bredere en plattere neus hebben dan blanken.

Uit de eerste twee genoemde onderzoeken blijkt dat er in bepaalde Aziatische landen anders wordt omgegaan met emoties dan in Amerika en de Westerse wereld in het algemeen. Er is vooral een verschil te zien bij het tonen van emoties, wat wordt veroorzaakt door verschillende gebruiken die in deze culturen gelden in relatie tot het tonen van emoties. Daarnaast is ook vastgesteld dat er onderling significante verschillen in uiterlijk zijn.

2.2 Geautomatiseerde emotieherkenning

De combinatie tussen emotieherkenning en informatica is vaker gemaakt. Uit onderzoeken op dit gebied is duidelijk geworden dat het voorbereiden van de data vaak noodzakelijk is. Over deze voorbereiding zullen we het hier hebben. Daarnaast worden de twee meest-gebruikte methodes voor geautomatiseerde

emotieherkenning besproken; de template-gebaseerde aanpak en Artificiële Neurale Netwerken. Tot slot noemen we enkele veelbelovende onderzoeken op het gebied van automatische emotieherkenning.

2.2.1 Voorbewerking

Pantic en Rothkrantz (2000) geven een overzicht van de voortgang die gemaakt is gedurende de jaren 90 van de vorige eeuw. Veel van de onderzoeken die door hen genoemd worden beschrijven een systeem waarbij aan voorbewerking van de beelden gedaan wordt. Dit is noodzakelijk omdat veel van de huidige beschikbare databases niet volledig consistent zijn. Ook is het soms nodig om verschillende databases te gebruiken omdat één database niet gevarieerd genoeg is op het gebied van geslacht, leeftijd en uiterlijke kenmerken.

Een systeem dat aan gezichtsdetectie doet, herkent of er een gezicht in het beeld aanwezig is en waar het zich precies bevindt. Dit vergemakkelijkt het verkrijgen van geschikte beelden in grote mate, omdat deze niet gestandaardiseerd hoeven te zijn. Bij beeldsequenties (film) is uitgebreidere gezichtsdetectie nodig dan bij statische beelden (foto), omdat het gezicht in een video zelden continu op exact dezelfde positie blijft.

Naast gezichtsdetectie kan het ook nodig zijn om andere irrelevante eigenschappen van de afbeelding te neutraliseren. Voorbeelden hiervan zijn het omzetten van kleurenafbeeldingen naar zwart-witafbeeldingen, het corrigeren van over- of onderbelichting en het recht maken van de afbeeldingen.

2.2.2 Extractie van gezichtsdata

Bij de template-gebaseerde aanpak wordt een template of graaf over het gezicht gelegd. De punten van een dergelijke graaf liggen op vaste posities in het gezicht. Hierbij kan gedacht worden aan onder andere de mondhoeken, wenkbrauwen en kin. De expressie wordt herkend via het Facial Action Coding System (FACS) of door berekeningen te doen aan de

afstanden tussen deze punten. Het FACS is een systeem dat is ontwikkeld om subtiele veranderingen in de gezichtsexpressie in termen van 44 Action Units (AU's) vast te kunnen leggen. Uit de combinatie van waargenomen AU's kan de emotie worden afgelezen.

Een goede geautomatiseerde versie van FACS zou mogelijk een grote vooruitgang zijn op het gebied van automatische emotieherkenning. Uit het overzicht van Pantic en Rothkrantz (2000) blijkt echter dat er nog maar weinig initiatieven op dit gebied zijn. Daarnaast blijkt dat, omdat de AU's in veel gevallen over subtiele veranderingen in het gezicht gaan, het lastig is om een goed werkende automatisering van FACS te realiseren.

2.2.3 Artificieel Neuraal Netwerk

Artificiële Neurale Netwerken (ANN's) kunnen getraind worden om een bepaalde uitkomst te koppelen aan een invoer. Een ANN bestaat over het algemeen uit drie lagen, een invoerlaag, een tussenlaag en een uitvoerlaag, die zelf weer bestaan uit een bepaald aantal knopen of neuronen. Deze neuronen zijn verbonden met alle andere neuronen en elk van deze verbindingen heeft een gewicht. De uitvoerwaarde van een neuron hangt af van de uitvoerwaarden van de neuronen in de vorige laag en de gewichten van de verbindingen met die neuronen.

In de trainingsfase wordt een groot aantal keren invoer aan het Artificieel Neuraal Netwerk gegeven. Vervolgens wordt de uitvoer vergeleken met de gewenste uitvoer. Op basis van het verschil hiertussen wordt een aanpassing aan de gewichten gedaan. Doordat de trainingsfase zoveel voorbeelden gebruikt zijn ANN's robuust genoeg om om te kunnen gaan met ruis, zoals enkele onduidelijke afbeeldingen of foutieve labels, waardoor het niet altijd nodig is om (uitgebreide) voorbewerking toe te passen.

Gezichtsdetectie of gestandaardiseerde afbeeldingen zijn min of meer een vereiste voor het succes van een ANN bij emotieherkenning. De meest voor de hand liggende invoer voor het

Artificieel Neuraal Netwerk zijn de afbeeldingen zelf, omgezet in een getallenmatrix. Ook worden afstanden tussen referentiepunten in het gezicht als invoer voor een ANN gebruikt, maar dit lijkt de nauwkeurigheid niet noemenswaardig te verhogen (Pantic & Rothkrantz, 2000).

2.2.4 Recente ontwikkelingen

Het overzicht van Pantic en Rothkrantz is al ruim tien jaar oud en sindsdien zijn er enkele onderzoeken gedaan die het vermelden waard zijn. Hieronder worden deze drie onderzoeken besproken. Het eerste onderzoek (Franco & Treves, 2001) probeert met een extra tussenlaag de complexiteit van de invoer te verkleinen en zo betere en snellere resultaten te behalen. In het tweede onderzoek (Cantelmo, 2007) wordt geprobeerd een zo snel mogelijke correcte classificatie te bewerkstelligen. In het derde onderzoek (Youssif & Asker, 2011) tenslotte ligt de focus op een zo hoog mogelijk percentage correcte classificaties, door middel van uitgebreide voorbewerking en het gebruik van afstanden tussen punten in het gezicht.

Franco & Treves

Hoewel het gebruikelijk is om ANN's uit drie lagen te laten bestaan, heeft het netwerk dat Franco en Treves (2001) gebruikten vier lagen. De tweede tussenlaag, de derde laag van het netwerk, bestaat uit modules (groepen neuronen) die elk een bepaalde emotie herkennen. Door het gebruik van modules wordt de complexiteit van de invoer verkleind. De modules kunnen gezien worden als kleine ANN's die elk voor een andere emotie aangeven of deze in de afbeelding aanwezig is. De combinatie van de uitvoer van de modules geeft antwoord op de complexere vraag welke emotie de proefpersoon vertoont.

Het systeem van Franco en Treves bevat geen automatische voorbewerking van de afbeeldingen. Wel werden de afbeeldingen handmatig gecorrigeerd op helderheid en werd er een uitsnede gemaakt van 8 bij 24 pixels waarbij slechts de helft van het gezicht bewaard

bleef, ruwweg afgebakend door het midden van de neus en een mondhoek en ooghoek. Hierdoor wordt veel irrelevante informatie, zoals de oren, het haar en de achtergrond, weggelaten, waardoor de trainingsfase en de classificatie sneller verlopen. Ook kan deze irrelevante informatie de classificatie op deze manier niet beïnvloeden. Deze modulaire aanpak en het filteren van irrelevante informatie geeft een percentage correct geclassificeerde emoties van 83%.

Cantelmo

Cantelmo (2007) heeft een ANN gemaakt dat gebaseerd is op de beschrijving van Mitchell (1997). Hij gebruikt neuronen met een sigmoïde, of s-vormige, activatiefunctie. Zijn netwerk bestaat uit drie lagen, een invoerlaag, een tussenlaag en een uitvoerlaag. Cantelmo doet zijn onderzoek twee keer, de eerste keer met een tussenlaag van 5 neuronen en de tweede keer met een tussenlaag van 10 neuronen. Zijn doel is om te onderzoeken hoe een ANN met zo min mogelijk rekentijd een correcte classificatie kan realiseren.

Om het netwerk te testen werd de Japanese Female Facial Expression (JAFFE) database (Lyons et al. 1998) gebruikt. Het systeem gebruikt de onbewerkte afbeeldingen als invoer voor het ANN, er wordt geen voorbewerking toegepast. Cantelmo vergelijkt de resultaten van zijn ANN met de classificaties van een groep menselijke beoordelaars, in tegenstelling tot de meeste andere onderzoeken, die de resultaten vergelijken met de daadwerkelijke emotie bij het subject.

De classificaties door het ANN met 10 tussenneuronen komen gemiddeld in 73% van de gevallen overeen met de classificaties die door mensen werden gemaakt, terwijl het ANN met 5 tussenneuronen significant slechter presteerde. Hierdoor komt Cantelmo tot de conclusie dat, hoewel 5 tussenneuronen een snellere classificatie kunnen maken, er toch 10 nodig zijn voor de nauwkeurigheid van het systeem.

Youssif & Asker

Youssif en Asker (2011) hebben een diepgravendere aanpak, waarbij ze aan uitgebreide voorbewerking van de afbeeldingen doen. Hun Automatic Facial Expression Recognition (AFER) systeem bestaat uit drie stappen. De eerste stap is gezichtsdetectie, waarbij het gezicht gedetecteerd wordt en de afbeelding gestandaardiseerd wordt voor verdere verwerking. De tweede stap is extractie van geometrische gezichtskenmerken, hierbij worden de positie en oriëntatie van onder andere de mond en de wenkbrauwen bepaald. De coördinaten en hoek in graden van deze kenmerken vormen de invoer voor de laatste stap, een ANN. In het ANN gebeurt de daadwerkelijke classificatie, de getallen worden aan de juiste emotie gekoppeld.

Bij dit onderzoek werden afbeeldingen uit de Cohn-Kanade database (Cohn et al., 2000) gebruikt. Wanneer het netwerk werd getest op subjecten die ook in de trainingsset voorkwamen (person-dependent), herkende het de emotie gemiddeld in 96% van de gevallen correct. Wanneer de testsubjecten niet in de trainingsset voorkwamen (person-independent), werd in gemiddeld 93.5% van de gevallen de juiste emotie aangegeven.

3 Hypothese

In dit hoofdstuk bespreken we onze verwachtingen voor de uitkomsten van ons onderzoek. We beginnen met de hypothese voor de eerste conditie en daarna komt onze hypothese voor de tweede conditie aan de orde. De hypothesen worden opgesteld aan de hand van eerder genoemd onderzoek.

3.1 Conditie 1: Mannen en vrouwen

Bij deze conditie verwachten wij dat het Artificieel Neuraal Netwerk dat getraind is om emotie te herkennen bij het ene geslacht moeite zal hebben met de emotieherkenning bij het geslacht waar niet op getraind is. Deze

verwachting baseren we op eerder besproken onderzoek naar emotievertoning met het gezicht bij mannen en vrouwen. Onderzoeken die aan bod komen zijn onder andere van Hess et al. (2009) en Kring (1998).

Uit het onderzoek van Hess et al. (2009) bleek dat mensen vrouwen eerder als boos classificeren dan mannen. Dit is te verklaren aan de hand van de neutrale gezichtsuitdrukkingen van zowel mannen als vrouwen. Mannen kijken van nature al iets norser waardoor ze minder hun gezicht hoeven aan te passen om boos te kijken dan vrouwen. Wanneer het ANN getraind wordt op de mannelijke gezichten kan het zijn dat hierdoor hij de emoties bij de vrouwen minder goed kan herkennen en vice versa.

Een reden om aan te nemen dat het Artificieel Neuraal Netwerk moeite zal hebben om emoties te herkennen bij het ene geslacht wanneer het traint op het andere geslacht, is dat mannen een andere opbouw van het gezicht hebben dan vrouwen. Burton (1993) heeft bewezen dat de afstanden tussen bepaalde punten in het gezicht verschillen bij mannen en vrouwen. Door het meten van deze afstanden kan bepaald worden of het om het gezicht van een man of van een vrouw gaat. Dit kan ertoe leiden dat het netwerk leert de opbouw van een mannelijk gezicht te herkennen met de bijbehorende emotie, maar deze vervolgens niet bij een vrouw kan herkennen omdat de opbouw anders is. Het is mogelijk dat het ANN hierdoor de emoties bij mannen en vrouwen niet even goed herkent wanneer het traint op maar een van de geslachten.

Naast het verschil in uiterlijk kan er ook verschil ontstaan doordat mannen hun emoties anders tonen dan vrouwen. Volgens Kring (1998) voelen mannen en vrouwen wel emoties in dezelfde mate bij dezelfde gebeurtenissen, maar vertonen vrouwen deze veel uitbundiger dan mannen. Hieruit volgt dat het bijvoorbeeld mogelijk is dat het netwerk de gezichtsuitdrukking van een man foutief als neutraal classificeert als het op emotieherkenning bij vrouwen getraind is.

Het ANN zal zich aanpassen aan de kenmerken van de gezichten waarop het traint. Het netwerk kan emoties van gezichten met die kenmerken goed herkennen, maar vervolgens kan het moeite hebben met het herkennen van emoties in gezichten waar deze kenmerken niet aanwezig zijn. Daarnaast tonen vrouwen hun emoties uitbundiger dan mannen. Het ANN wordt getraind om emoties die uitbundig vertoond worden te herkennen en kan de emoties dan niet goed herkennen wanneer die subtiel vertoond worden. Het omgekeerde geldt wanneer het ANN getraind wordt om de subtiele emoties te herkennen.

3.2 Conditie 2: Aziaten en blanken

Bij de tweede conditie trainen we het netwerk op het ene ras en testen we het vervolgens op het andere ras. Wij verwachten bij deze conditie een verschil in de prestaties van het Artificieel Neuraal Netwerk te vinden. Voor deze verwachting hebben we twee belangrijke redenen.

Ten eerste verwachten we dat de verschillen in uiterlijke kenmerken invloed hebben op de prestaties. Aziaten en blanken hebben enkele karakteristieke uiterlijke kenmerken die ze van elkaar en van mensen van een ander ras onderscheiden. Zo hebben Aziaten een breder, maar minder hoog gezicht dan blanken. Ook hebben Aziaten een significant bredere en plattere neus dan blanken (Zhuang et al., 2010). Omdat het ANN getraind wordt op slechts één ras, zal het zich aanpassen aan bepaalde uiterlijke kenmerken. Wanneer het vervolgens getest wordt op mensen die andere uiterlijke kenmerken bezitten, heeft het mogelijk meer moeite om de emotie te herkennen.

Ten tweede blijkt uit meerdere onderzoeken dat er verschillen zijn in de manier waarop Aziaten en Westerlingen hun emoties uiten. De ogen zijn bij Japanners de belangrijkste informatiebronnen wat betreft emotie, terwijl bij Westerlingen de mond het meest veelzeggend is (Yuki et al., 2005). Daarnaast uiten Chinezen hun emoties veel ingetogener dan Amerikanen, die heel uitbundig zijn

(Camras et al., 2006). Hierbij moet wel gezegd worden dat in ons onderzoek alleen Japanse vrouwen worden onderzocht en hoewel ook zij bekend staan om hun ingetogenheid wat betreft emoties, wordt er in ons onderzoek geen wetenschappelijk bewijs daarvoor gebruikt.

Omdat het netwerk went aan de uiterlijke kenmerken en de manier van het uiten van emoties door het ene ras, verwachten wij dat het moeite zal hebben met het herkennen van emoties bij het andere ras. Omdat de verschillen tussen deze specifieke rassen vrij groot zijn, is het mogelijk dat het netwerk behoorlijk veel problemen heeft met de classificatie wanneer het getraind is op het ene ras en getest wordt op het andere ras.

4 Methode

Ons onderzoek bestaat uit het maken van een Artificieel Neuraal Netwerk, dat we trainen om emoties te herkennen van afbeeldingen. Vervolgens testen we hoe goed het ANN geworden is in het herkennen van emoties. Dit doen we op twee groepen: een groep individuen waar niet op getraind is, maar die hetzelfde geslacht en ras hebben als de trainingsgroep en een groep individuen die in geslacht of in ras verschillen van de trainingsgroep.

In dit hoofdstuk beschrijven we hoe we dit onderzoek uitvoeren. We beginnen met de data, waarbij we bespreken op welke manier we de afbeeldingen hebben verkregen en hoe we deze zullen gebruiken in het onderzoek. Vervolgens komt de beschrijving van het Artificieel Neuraal Netwerk aan de orde en tot slot beschrijven we hoe we het daadwerkelijke onderzoek exact uitvoeren.

4.1 Data

Dit onderdeel behandelt de data die we gebruiken: de afbeeldingen met personen die een emotie vertonen. We hebben twee databases gebruikt, deze worden in het eerste

gedeelte besproken. Daarna laten we zien welke voorbewerking er nodig is, zodat de databases voldoende met elkaar overeenkomen om samengevoegd te kunnen worden en zo één database te vormen, die we bij dit onderzoek gebruiken.

4.1.1 Databases

Er worden bij dit onderzoek twee databases gebruikt, de Japanese Female Facial Expression (JAFFE) database en de Extended Cohn-Kanade Facial Expression (CK+) database. Uit deze databases worden eerst de afbeeldingen geselecteerd die gebruikt kunnen worden. Vervolgens worden de geselecteerde afbeeldingen voorbewerkt zodat er geen irrelevante verschillen tussen de afbeeldingen bestaan.

Extended Cohn-Kanade Facial Expression Database

De CK+ database (Cohn et al., 2010) bevat afbeeldingen van 123 proefpersonen die emoties vertonen. Voor elke emotie is er een reeks foto's, deze reeks begint met de neutrale uitdrukking en eindigt met het hoogtepunt van de emotie. In totaal bevat de database 593 reeksen van een emotie. De database bevat afbeeldingen van zowel mannen als vrouwen. Hoewel er proefpersonen van verschillende rassen aanwezig zijn in deze database, is 81% van de afbeeldingen van een blanke proefpersoon.

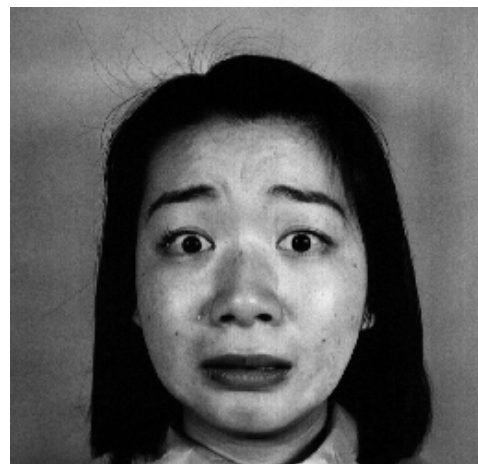


Figuur 1: Voorbeeldafbeelding uit de CK+ database

Deze database heeft de gezichtsuitdrukkingen verdeeld in zeven emoties en de neutrale gezichtsuitdrukking. De zeven emoties zijn boosheid, verachting, walging, angst, blijdschap, verdriet en verbazing. Voor 327 van de 593 reeksen is er aangegeven welke emotie wordt vertoond. In een los tekstbestand staat een getal tussen 0 en 7, dit geeft aan welke emotie in de bijbehorende afbeelding te zien is.

Japanese Female Facial Expression Database

De JAFFE database (Lyons et al., 1998) bestaat uit 213 afbeeldingen van de zes basisemoties (Ekman 1972) en de neutrale emotie. In de bestandsnaam van de afbeelding staat aangegeven om welke emotie het gaat. De emoties in deze database worden door tien verschillende Japanse vrouwen vertoond.



Figuur 2: Voorbeeldafbeelding uit de JAFFE database

Selectie

Als eerste vind er een selectie van de afbeeldingen plaats. Uit de JAFFE database zijn alle proefpersonen bruikbaar, bij de CK+ database wordt alleen gebruik gemaakt van de groepen blanke Amerikaanse mannen en blanke Amerikaanse vrouwen. De CK+ database bevat ook afbeeldingen met de emotie 'verachting'. Omdat deze emotie volgens Ekman (1972) niet tot de basisemoties behoort, vallen deze afbeeldingen voor dit onderzoek af.

De CK+ database bevat voor elke proefpersoon per emotie een reeks afbeeldingen. Deze reeks begint met de neutrale uitdrukking

en eindigt met het hoogtepunt van de emotie. Voor dit onderzoek zijn we alleen geïnteresseerd in de afbeeldingen van dit hoogtepunt. De JAFFE database bevat slechts één afbeelding per emotie per proefpersoon, dus hier is geen verdere selectie nodig.

4.1.2 Verwerking

Alle geselecteerde afbeeldingen moeten aan bepaalde eisen voldoen voordat ze ingevoerd kunnen worden in het netwerk. Om te zorgen dat ze aan deze eisen voldoen, wordt er een aantal bewerkingen uitgevoerd, waaronder het toevoegen van emotielabels en het voorbereiden van de afbeeldingen.

Eisen

Aan de afbeeldingen worden eisen gesteld zodat het verschil tussen de afbeeldingen beperkt blijft tot de verschillende emoties. Op deze manier wordt voorkomen dat het netwerk zich gaat richten op irrelevante verschillen, zoals de kleding of het kapsel van de proefpersonen.

Ten eerste moet de achtergrond van elke afbeelding egaal zijn. Ten tweede moet de proefpersoon zijn hoofd recht houden en in de richting van de camera kijken. Ten derde is het belangrijk dat het gezicht van een proefpersoon zich in elke foto op dezelfde positie bevindt.

Er kan voldaan worden aan de eisen door een consistente database van afbeeldingen te gebruiken of door een voorbereidingsstap aan het programma toe te voegen. Hier wordt ervoor gekozen om een consistente database te gebruiken om de hoeveelheid programmeerwerk binnen de perken te houden. De twee databases die gebruikt worden zijn echter nog niet consistent genoeg om direct te kunnen gebruiken, daarom moeten ze eerst aangepast worden.

Emotielabels

Om het ANN te trainen is het nodig om emotielabels bij de afbeeldingen te hebben, zodat het netwerk kan controleren of het de juiste classificatie maakt. Voor de meeste

afbeeldingen in de CK+ database zijn deze labels aanwezig. Ze staan in een apart bestand en bestaan uit een getal dat codeert voor de emoties. Bij de JAFFE database bestaan deze bestanden nog niet. Daarom wordt voor elke afbeelding in beide databases een bestand aangemaakt waarin staat welke emotie de foto bevat, als dit bestand nog niet aanwezig is.

In de bestanden die horen bij de afbeelding staat een reeks van zes cijfers die aangeven welke emotie afgebeeld wordt. Elk cijfer codeert voor een andere emotie. Een 1 geeft aan dat de corresponderende emotie aanwezig is in de afbeelding. Als alle getallen gelijk zijn aan 0 bevat de afbeelding geen enkele emotie en is de gezichtsuitdrukking van de proefpersoon neutraal.

Vorbewerking

De foto's dienen allemaal op dezelfde manier bijgesneden te worden zodat het gezicht van de proefpersoon goed in het midden van de afbeelding staat. Hier is ervoor gekozen om de afbeeldingen bij te snijden tot een verhouding van 3 bij 4. De randen komen zo te liggen dat de oren net van de foto verdwijnen. De onderkant van de afbeelding komt net onder de kin te liggen. Het formaat van de afbeeldingen dient ook aangepast te worden, zodat alle afbeeldingen dezelfde afmetingen hebben. Er is gekozen om een formaat van 66 bij 88 pixels te gebruiken.



Figuur 3: Voorbeeldafbeeldingen na bewerking

De laatste stap is het omzetten van de afbeeldingen naar een .pgm-bestand. De extensie .pgm staat voor Portable Gray Map. Wanneer een afbeelding wordt omgezet naar een .pgm-bestand wordt de afbeelding automatisch zwart-wit gemaakt als dit nog niet het geval was. Elke pixel in de afbeelding heeft dan een bepaalde grijstint met een waarde tussen de 0 en 255. Een .pgm-bestand bevat voor elke pixel de waarde van de grijstint.

4.2 Network

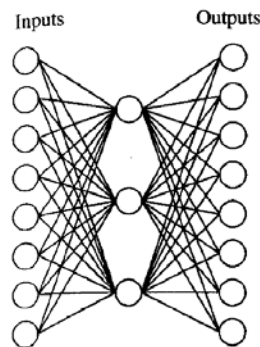
In dit onderdeel wordt het Artificieel Neuraal Netwerk beschreven dat we voor dit onderzoek gebruiken. Eerst komt de algemene structuur van een ANN aan de orde, daarna bespreken we de voor ons netwerk specifieke parameters. De broncode van het ANN is te vinden op: <http://codeviewer.org/view/code:28a3>

4.2.1 Structuur

Wij baseren ons ANN op hoofdstuk 4 van Mitchell (1997). Het bestaat uit drie lagen, een invoerlaag, een tussenlaag en een uitvoerlaag. De uitvoer van elk neuron wordt bepaald door een sigmoïdefunctie. Ons ANN is een feed-forward netwerk, dat leert door middel van backpropagation. Voor zover nog niet geïntroduceerd zullen deze noties in wat volgt aan de orde komen.

Lagen

De invoerlaag dient de waarden van de pixels uit de afbeelding te bevatten. Er zijn dus evenveel invoerneuronen als pixels in de afbeelding. De tussenlaag bestaat uit 10 neuronen. Dit is een aantal dat groot genoeg is om de gewenste resultaten te behalen (Cantelmo, 2007), maar niet zo groot dat het onnodig veel rekentijd vergt. Hoewel de



Figuur 4: Algemene ANN-structuur (Mitchell, 1997)

uitvoerlaag over het algemeen uit slechts één neuron bestaat, kiezen we er hier voor om zes uitvoerneuronen te gebruiken. Elk van deze uitvoerneuronen codeert voor één van de zes basisemoties die door Ekman (1972) beschreven zijn.

Uitvoer

Het ANN is een feed-forward netwerk. Dat houdt in dat de informatie in één richting, voorwaarts, doorgegeven wordt. Elk neuron in het netwerk is daarvoor verbonden met alle neuronen uit de volgende laag. Elke verbinding heeft een eigen gewicht, geïnitieerd op een willekeurige waarde uit het interval $[-0,05; 0,05]$. Een neuron berekent de lineaire combinatie van de neuronen uit de vorige laag en de bijbehorende gewichten. De sigmoïdefunctie, een S-vormige curve die zijn invoer projecteert op het interval $[0,1]$, fungeert als drempelfunctie. De uitkomst hiervan is de uitvoer van het neuron i :

$$o_i = \sigma(w_0 * x_0 + w_1 * x_1 + w_2 * x_2 + \dots)$$

$$\sigma(y) = \frac{1}{1 + e^{-y}}$$

In de functie voor de uitvoer staat o_i voor de uitvoerwaarde van neuron i , x_j voor de uitvoerwaarde van neuron j uit de voorgaande laag en w_j voor het gewicht van de verbinding tussen neuron i en j . De neuronen in de tussenlaag berekenen hun uitvoer aan de hand van de waarden uit de invoerlaag met de bovengenoemde formule. Vervolgens berekenen de neuronen in de uitvoerlaag hun uitvoer aan de hand van de uitvoerwaarden van de tussenlaag.

Afwijking

Een Artificieel Neuraal Netwerk leert van zijn eigen fouten, in dit geval gebruiken we voor het trainen van het ANN het backpropagation-algoritme. Daarbij wordt de uitvoer van het netwerk vergeleken met de gewenste uitvoer bij de huidige invoer. Dit wordt gedaan door de afwijking (error) van de uitvoerneuronen (k) en de tussenneuronen (h) te berekenen:

$$\delta_k = o_k(1 - o_k)(t_k - o_k)$$

$$\delta_h = o_h(1 - o_h) \sum_{k \in \text{uitvoeren}} w_{kh} \delta_k$$

In deze formules staat δ voor de afwijking het neuron, o voor de uitvoer (output) het neuron, t voor doelwaarde (target) het neuron en w_{kh} voor het gewicht (weight) tussen neuron k en neuron h .

De afwijking van de uitvoerneuronen hangt alleen af van het verschil tussen hun werkelijke uitvoer en de gewenste uitvoer. Omdat er voor de tussenneuronen geen vastgestelde gewenste uitvoer is, hangt de afwijking van de tussenneuronen af van de eigen uitvoer en de afwijking van de uitvoerneuronen.

Leren

De volgende stap van het backpropagation-algoritme is het aanpassen van de gewichten. Dit gebeurt aan de hand van de berekende afwijking van elk neuron in het netwerk. Het doel hiervan is dat de toekomstige uitvoer bij het huidige voorbeeld dichter bij de gewenste uitvoer ligt. Deze aanpassing verloopt volgens de volgende formule:

$$w_{ji} = w_{ji} + \Delta w_{ji}$$

$$\Delta w_{ji} = \eta \delta_j x_{ji}$$

Hierin is w_{ji} het gewicht van de verbinding tussen de neuronen i en j , δ_j is de bovengenoemde afwijking van neuron j . x_{ji} is de uitvoer van neuron i die de invoer van neuron j vormt.

4.2.2 Parameters

Er dienen nog enkele parameters voor het Artificieel Neuraal Netwerk gekozen te worden voordat het echte onderzoek kan beginnen. Deze parameters zijn de leersnelheid en het momentum.

Leersnelheid

De leersnelheid (learning rate), η , van het netwerk bepaalt hoe snel het leert. Bij een lage leersnelheid wordt er na elke evaluatie van een trainingsafbeelding een hele kleine aanpassing gedaan aan de gewichten van het ANN, waardoor het leren erg lang kan duren. Wanneer de leersnelheid hoog is, worden de gewichten elke keer sterk aangepast. Het gevaar van een hoge leersnelheid is dat de gewichten te veel toegespitst worden op het huidige voorbeeld, waardoor de eerder geleerde informatie over andere voorbeelden teniet gedaan wordt.

Na enkele proefrondes blijkt dat bij de gekozen omstandigheden een leersnelheid van 0,3, de gangbare waarde, niet optimaal werkt. Gezien de resultaten van de proefrondes kiezen we hier voor een leersnelheid η van 0,025. Dat deze waarde zoveel lager ligt dan de gangbare waarde komt onder meer door het gebruik van momentum, een notie die in de komende sectie geïntroduceerd wordt.

Momentum

Om te voorkomen dat het netwerk in een lokaal minimum terechtkomt wordt een mechanisme gebruikt dat momentum wordt genoemd. Het momentum is een deel van de vorige gewichtsaanpassing, dat bij de huidige gewichtsaanpassing wordt opgeteld. Hierdoor worden de gewichten ook aangepast als de huidige gewichtsaanpassing gelijk is aan 0, maar er bij de vorige aanpassing nog wel een verbetering nodig was. Op deze manier kan het netwerk uit een lokaal minimum komen. De totale gewichtsaanpassing op tijdstip t wordt nu:

$$\Delta w_{ji}(t) = \eta \delta_j x_{ji} + \alpha \Delta w_{ji}(t - 1)$$

De formule blijft grotendeels hetzelfde, de enige nieuwe variabele is α , de momentum-coëfficiënt. Bij deze variabele is het, net als bij de leersnelheid, belangrijk dat deze niet te hoog of te laag is. Wanneer de momentumcoëfficiënt te hoog is, loop je het risico dat het netwerk

blijft schommelen tussen het globale minimum en een lokaal minimum. Dit gebeurt omdat het netwerk elke keer als het echte minimum bereikt is, toch nog een (grote) gewichtsaanpassing gaat doen die ervoor kan zorgen dat het netwerk het globale minimum verlaat. Wanneer α te laag is of in het geheel niet gebruikt wordt, kan het netwerk niet uit een lokaal minimum komen en is het dus mogelijk dat er een suboptimale oplossing bereikt wordt.

In bovengenoemde proefrondes variëren we ook de momentumcoëfficiënt om uit te vinden welke waarde de beste resultaten geeft. Voor α is de gangbare waarde 0,05, maar voor ons systeem blijkt een waarde van 0,2 het beste te werken. Omdat door het gebruik van momentum in elke stap een grotere gewichtsaanpassing gedaan wordt, kan het nodig zijn om de leersnelheid lager te kiezen dan normaal.

4.3 Uitvoering

Nadat de data geschikt gemaakt is en het ANN klaar is kan het daadwerkelijke onderzoek uitgevoerd worden. Eerst dient het netwerk getraind te worden, waarna de prestaties geëvalueerd worden in de testfase. Bij de uitvoering wordt gewerkt met twee groepen afbeeldingen, de trainingsset en de testset.

4.3.1 Training

Aangezien alle verbindingen in het netwerk initieel willekeurige gewichten hebben, dient het netwerk getraind te worden om bepaalde invoer, in dit geval foto's van gezichten, te koppelen aan bepaalde uitvoer, in dit geval emotielabels. Hiervoor worden de afbeeldingen in de trainingsset één voor één voorgelegd aan het netwerk. Vervolgens wordt de uitkomst van het netwerk vergeleken met het label van de afbeelding en de gewichten worden waar nodig aangepast.

Een 'ronde', waarin alle trainingsafbeeldingen eenmaal door het netwerk gezien worden, wordt een *epoch* genoemd. Omdat ANN's hun gewichten telkens maar lichtelijk

aanpassen in de richting van de optimale gewichten, zijn er veel epochs nodig om een goed getraind netwerk te krijgen. Het exacte aantal epochs is onder andere afhankelijk van het aantal afbeeldingen in de trainingsset. Hoe meer afbeeldingen het netwerk per epoch ziet, hoe minder epochs er over het algemeen nodig zijn om een goed resultaat te bereiken.

Wanneer het aantal epochs te klein is, kan het zo zijn dat het netwerk een betere prestatie kan verrichten dan de resultaten doen denken. Daarom is het belangrijk om het programma meer epochs door te laten lopen dan het netwerk naar schatting nodig heeft om het optimale resultaat te behalen. Een te groot aantal epochs levert het gevaar van *overfitting* op. Hierbij leert het netwerk zo goed om te gaan met de trainingsafbeeldingen, dat het percentage correcte classificaties op testafbeeldingen, die niet in de trainingsset zitten, weer omlaag gaat.

Hier is er voor gekozen om het programma gedurende 5.000 epochs te laten trainen. In combinatie met de grootte van de trainingssets, de gekozen leersnelheid en het momentum heeft het netwerk zo genoeg mogelijkheden om de gewichten aan te passen en daarmee zo veel mogelijk emoties correct te herkennen. Bij een groter aantal epochs vindt er geen verbetering van de resultaten meer plaats en beginnen de percentages zelfs terug te lopen.

4.3.2 Testen

Om de prestaties van het netwerk te evalueren, worden afbeeldingen uit de testset ingevoerd in het netwerk. Deze afbeeldingen komen niet voor in de trainingsset en zijn dus nieuw voor het netwerk. Nadat de afbeeldingen zijn voorgelegd aan het netwerk wordt de uitvoer van het netwerk gecontroleerd aan de hand van het label dat bij de afbeelding hoort. In deze fase wordt alleen gekeken of de classificatie klopt, er worden geen gewichten aangepast.

Classificatie

Zoals eerder besproken berekent het netwerk zes uitvoerwaarden. Elk van die waarden codeert voor één emotie. Omdat we uitgaan van de zes basisemoties en er maar één emotie in een afbeelding aanwezig kan zijn, kijken we op dit moment alleen naar de hoogste uitvoerwaarde. Dit is de methode die Cantelmo (2007) ook hanteert. Het toevoegen van ingewikkeldere, gemengde emoties kan gemakkelijk worden bereikt door de evaluatiefunctie daar op aan te passen.

Als de hoogste uitvoerwaarde boven de 0,3 ligt, betekent dit dat de afbeelding de emotie bevat die correspondeert met deze uitvoerwaarde. Als geen enkele van de uitvoerwaarden hoger dan 0,3 is, betekent dit dat het netwerk de emotie classificeert als neutraal.

Het netwerk heeft de emotie goed herkend als de classificatie overeenkomt met het label. Tijdens de testfase wordt bijgehouden welk percentage van de afbeeldingen correct is beoordeeld. Aan het eind van de testfase is dit percentage een graadmeter voor de prestaties van het netwerk.

Conditie 1: Mannen en vrouwen

Bij deze conditie onderzoeken we of het ANN na training op het ene geslacht slechter presteert op het herkennen van emoties bij het andere geslacht. Er zijn twee subcondities te onderscheiden, een waarbij het netwerk getraind wordt op mannen en een waarbij het netwerk getraind wordt op vrouwen. Bij elke subconditie wordt het netwerk tienmaal getraind en getest om zo te kunnen berekenen of er een significante afwijking is tussen de prestaties van het netwerk bij mannen en vrouwen. Elke keer als het programma 50 epochs heeft doorgemaakt, wordt het netwerk getest waardoor de leercurve van het netwerk te zien is.

Bij de eerste subconditie wordt het netwerk getraind op blanke mannen. Vervolgens wordt het tweemaal getest, zowel op (andere) blanke mannen als op blanke vrouwen. Bij de tweede subconditie wordt het netwerk getraind op blanke vrouwen, waarna het getest wordt op

zowel (andere) blanke vrouwen als op blanke mannen. Uit elke subconditie komen twee percentages correcte classificaties. Aan de hand van deze percentages kan bepaald worden of er een significant verschil tussen de twee percentages is.

Conditie 2: Aziaten en blanken

Bij deze tweede conditie onderzoeken we of er een verschil in de resultaten is wanneer het ANN getraind wordt op het ene ras en tweemaal getest, zowel op dat ras als op een ander ras. Specifiek kijken we bij dit onderzoek naar Japanse vrouwen en blanke Amerikaanse vrouwen. Er zijn dus opnieuw twee subcondities te onderscheiden. Bij beide subcondities wordt het netwerk tien keer getraind en getest, op deze manier kan er berekend worden of het netwerk significant anders presteert bij Aziaten dan bij blanken.

Bij de eerste subconditie wordt het netwerk getraind op de Japanse vrouwen. Vervolgens wordt het getest op (andere) Japanse vrouwen en op blanke vrouwen. De tweede subconditie bestaat uit het trainen op blanke vrouwen, waarna het netwerk getest wordt op (andere) blanke vrouwen en op Japanse vrouwen.

Uit elke subconditie komen opnieuw twee percentages. Om robuuste resultaten te verkrijgen trainen we het netwerk 10 keer opnieuw voor elke subconditie. Na elke 50 epochs die het programma doormaakt wordt het netwerk getest waardoor de leercurve van het netwerk inzichtelijk gemaakt kan worden.

5 Resultaten

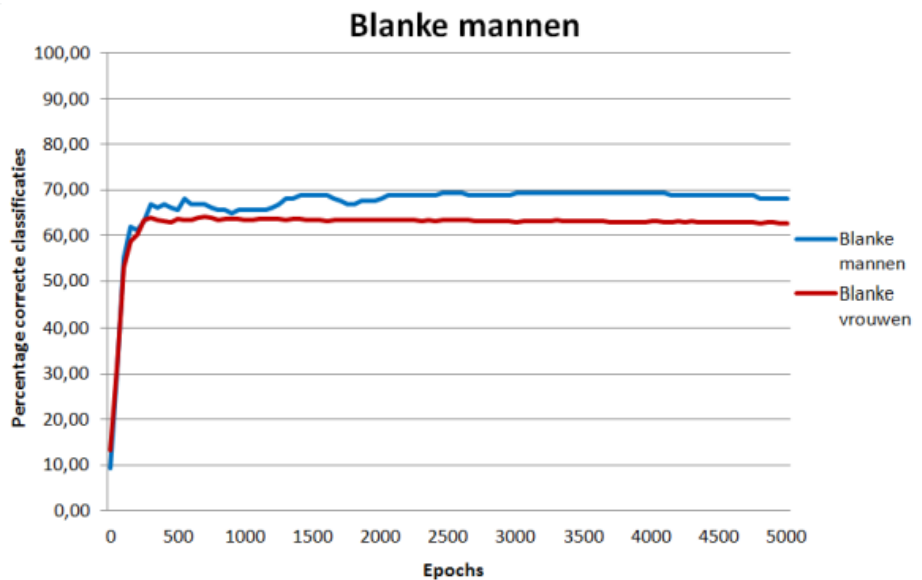
In dit onderdeel worden de resultaten van het onderzoek besproken. Als eerste komen de resultaten van de conditie met blanke mannen en vrouwen aan de orde en daarna komen we op de resultaten van de conditie met Aziatische vrouwen en blanke vrouwen.

5.1 Conditie 1: Mannen en vrouwen

Voor dit onderdeel zijn er twee subcondities die zijn onderzocht. Bij de eerste subconditie werd het netwerk getraind op afbeeldingen van mannen en vervolgens werd er gekeken of het netwerk ook de juiste emotie kon koppelen aan afbeeldingen van vrouwen en van mannen die niet in de trainingsgroep zaten. Bij de tweede subconditie werd het netwerk getraind op vrouwen en vervolgens getest op mannen en (andere) vrouwen.

Blanke mannen

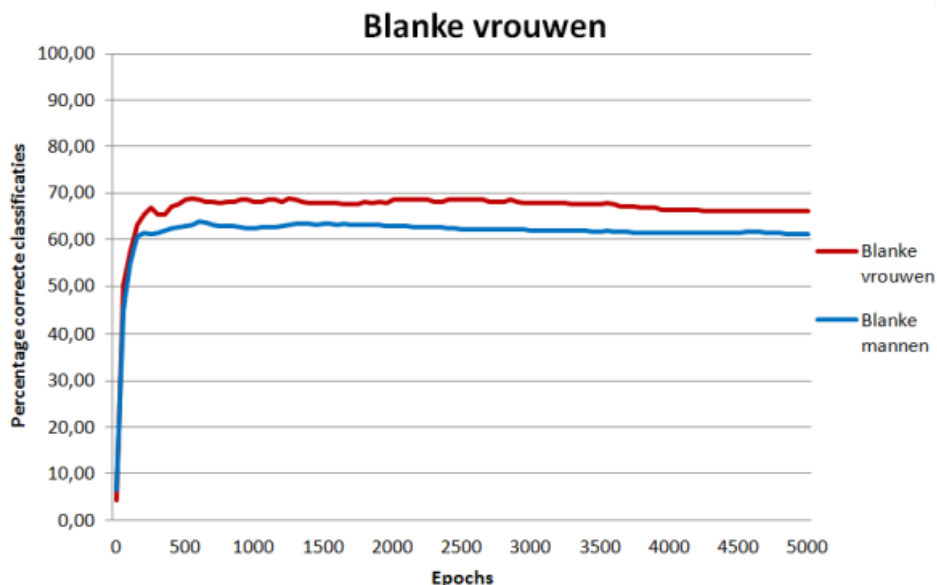
Bij deze subconditie werd het netwerk getraind op blanke mannen en vervolgens werd getest hoe goed het emoties herkende bij (andere) blanke mannen en blanke vrouwen. In de grafiek is te zien dat het netwerk al vrij snel de emoties bij de mannen iets beter kan herkennen dan bij de vrouwen. Gemiddeld classificeert het netwerk 67% van de emoties bij de mannen correct en 62% van de emoties bij de vrouwen. Er is dus een gemiddeld verschil van 5% tussen de prestaties van het netwerk op mannen en vrouwen.



Figuur 5: Leercurve van training op blanke mannen

Blanke vrouwen

Bij de tweede subconditie is het netwerk getraind op het herkennen van emoties bij blanke vrouwen vervolgens werd het getest op (andere) blanke vrouwen en blanke mannen. Het percentage correct geclassificeerde emoties bij de vrouwen ligt hoger dan dat bij de mannen. Dit is conform de resultaten van de eerste subconditie. Het netwerk classificeert gemiddeld 66,5% van de vrouwen goed terwijl het gemiddeld maar 61,5% van de mannen goed doet. Ook hier is dus een verschil van gemiddeld 5% in het aantal correcte classificaties.



Figuur 6: Leercurve van training op blanke vrouwen

5.2 Conditie 2: Aziaten en blanken

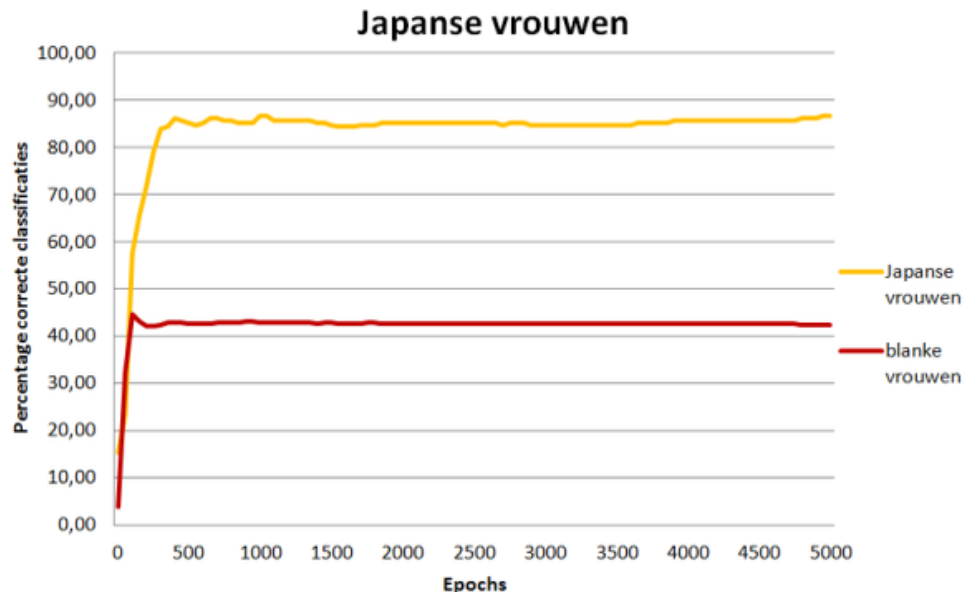
Hier worden de resultaten van het onderzoek met Japanse vrouwen en blanke vrouwen besproken. Daarbij worden de resultaten opnieuw onderverdeeld in twee subcondities, een waarbij het netwerk getraind is op het herkennen van emoties bij de Japanse vrouwen en een waarbij het getraind is op het herkennen van emoties bij de blanke vrouwen.

Japanse vrouwen

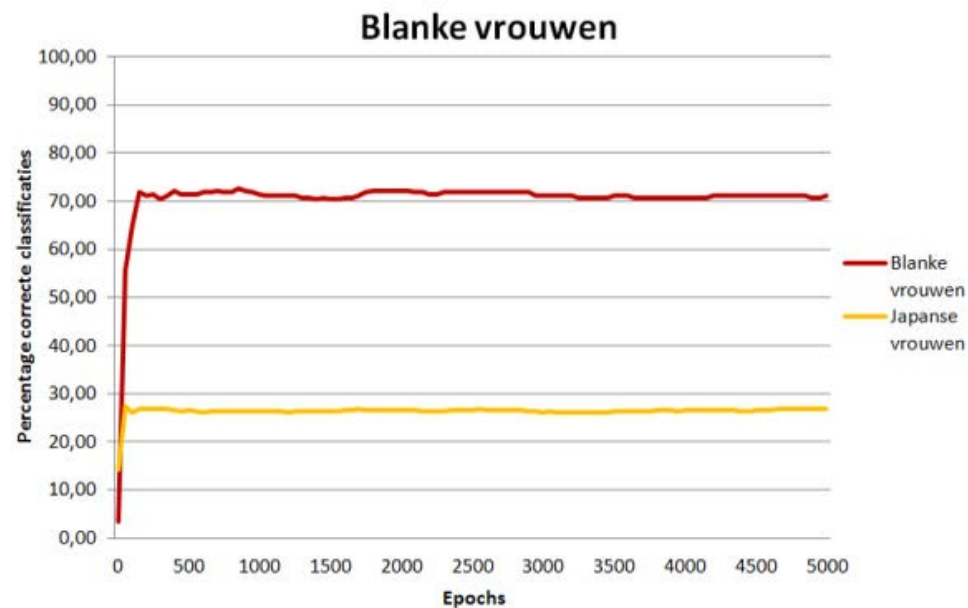
De eerste subconditie is die waarbij het netwerk getraind werd op het herkennen van emoties bij Japanse vrouwen. Het netwerk werd vervolgens getest op (andere) Japanse vrouwen en op blanke vrouwen. Bij deze test presteerde het netwerk gemiddeld 41% slechter op het herkennen van emoties bij de blanke vrouwen dan bij de Japanse vrouwen. Bij de Japanse vrouwen classificeerde het netwerk gemiddeld 83% van de emoties correct, terwijl dit bij de blanke vrouwen slechts in gemiddeld 42% van de gevallen zo was.

Blanke vrouwen

Bij de tweede subconditie werd het netwerk getraind op het herkennen van emoties bij blanke vrouwen. Het werd getest op zowel (andere) blanke vrouwen als op Japanse vrouwen. Het gemiddelde verschil tussen de prestaties van het netwerk is hier ongeveer even groot als bij de andere conditie, namelijk 44%. Het percentage correct geclassificeerde emoties bij de blanke vrouwen die niet in de trainingsgroep voorkwamen lag gemiddeld op 70%, terwijl dit percentage bij de Japanse vrouwen slechts op gemiddeld 26% lag.



Figuur 7: Leercurve van training op Japanse vrouwen



Figuur 8: Leercurve van training op blanke vrouwen

6 Discussie

In dit hoofdstuk bespreken we de resultaten uit hoofdstuk 5. Het is opgesplitst in twee onderdelen, een voor elk van de condities. In beide onderdelen beginnen we met de significantie van de resultaten en vervolgens verbinden we conclusies aan de al dan niet significante resultaten. Tot slot bespreken we voor elke conditie de beperkingen van dit onderzoek en bekijken we enkele mogelijke toekomstige onderzoeklijnen.

6.1 Conditie 1: Mannen en vrouwen

Bij deze conditie is onderzocht of er een verschil is tussen emotieherkenning bij mannen en vrouwen. Hier wordt besproken of er significante verschillen in de resultaten zijn en wat er uit de resultaten geconcludeerd kan worden. Ook wordt er besproken wat deze resultaten betekenen voor de toekomst.

6.1.1 Significantie

Bij beide subcondities is er door middel van een t-test berekend of de gevonden verschillen tussen de percentages correct beoordeelde emoties significant verschillen. We hebben een tweezijdige t-test uitgevoerd op de eindwaarden van de tien training rondes die bij elke subconditie gedaan zijn. Voor α , het significantieniveau, hebben we gekozen voor een waarde van 0,001.

Blanke mannen

Bij deze subconditie, waar er getraind werd op blanke mannen, vinden we een t-waarde van 1,941. In combinatie met het aantal vrijheidsgraden van 9 komen we uit op een p-waarde van 0,084. Dit betekent dat er 8,4% kans is dat het gevonden verschil niet een werkelijk verschil is, maar veroorzaakt is door toeval. De p-waarde is bij deze subconditie een stuk groter dan α waardoor we kunnen concluderen dat het verschil niet significant is. Dit betekent dat, hoewel de gemiddelde waarden iets lager liggen, er bij deze subconditie geen significant

verschil is gevonden in de prestaties van het netwerk wanneer het werd getraind op blanke mannen en daarna getest op (andere) blanke mannen en blanke vrouwen.

Blanke vrouwen

Bij de tweede subconditie wordt het netwerk getraind op blanke vrouwen. Hier vonden we een t-waarde van 1,470, samen met de vrijheidsgraden komt hier een p-waarde van 0,176 uit. Dit betekent dat er een kans van 17,6% is dat het gevonden verschil niet is ontstaan door verschillen in emotieherkenning maar een andere oorzaak heeft. De p-waarde ligt ook bij deze subconditie ruim boven het gekozen significantieniveau van 0,001. Hier kunnen we dus ook concluderen dat er geen significant verschil in emotieherkenning is gevonden tussen vrouwen en mannen. Dit komt overeen met de resultaten van de eerste subconditie.

6.1.2 Implicaties

Uit het onderzoek blijkt dat er geen significant verschil is tussen de resultaten van beide subcondities. We kunnen dus concluderen dat het netwerk, dat getraind is op het ene geslacht en vervolgens zowel getest op het andere geslacht als op andere afbeeldingen van hetzelfde geslacht, de emoties bij de getoonde mannen en vrouwen even goed kan herkennen. De mannelijke en vrouwelijke proefpersonen vertonen zoveel overeenkomsten in het tonen van hun emoties dat het netwerk de emoties in de meeste gevallen goed kan classificeren. In tegenstelling tot de verwachting beïnvloeden ook de uiterlijke verschillen tussen mannen en vrouwen de prestatie van het netwerk niet significant.

Dit onderzoek is uitgevoerd met blanke mannen en blanke vrouwen die allemaal in de leeftijdsklasse van 18 tot 30 jaar oud vallen. De verkregen resultaten gelden dan ook alleen voor deze groep. Zonder verder onderzoek kan niet worden aangenomen dat deze resultaten ook gelden voor personen buiten deze leeftijdsklasse of personen van een ander ras.

Naast het feit dat de resultaten alleen voor deze specifieke groep gelden, is het mogelijk dat met een andere onderzoekstechniek andere resultaten worden verkregen. Het ANN dat wij bij dit onderzoek gebruikt hebben kan goed omgaan met de verschillen tussen mannen en vrouwen. Een ANN waarbij technieken voor het aanpassen van de gewichten of andere belangrijke elementen verschillen van de door ons gebruikte techniek zou mogelijk andere resultaten op kunnen leveren.

6.1.3 Beperkingen

De CK+ database bevat vooral blanke proefpersonen, het aantal proefpersonen van een ander ras is erg beperkt. Hierdoor hebben we alleen het verschil tussen blanke mannen en vrouwen kunnen onderzoeken en niet het verschil tussen mannen en vrouwen van andere rassen. Om deze reden kunnen we geen uitspraak doen over mannen en vrouwen in het algemeen, maar alleen over het door ons onderzochte ras.

Zoals hierboven ook genoemd, zijn alle proefpersonen uit de database studenten aan de universiteit. De variatie in leeftijd is dus erg beperkt, hierdoor kon er niet vastgesteld worden of de resultaten voor alle leeftijden gelden. We kunnen daarom geen uitspraak doen over mannen en vrouwen met een leeftijd die buiten de leeftijdsklasse van de CK+ database valt.

Een andere beperking van dit onderzoek is het aantal afbeeldingen in de CK+ database. Het samenstellen van een database met gezichts-expressies kost erg veel tijd, hierdoor is het formaat van de meeste databases vrij beperkt. Wanneer er wordt gewerkt met een ANN is het echter wenselijk om zo veel mogelijk trainingsvoorbeelden ter beschikking te hebben.

Een oplossing voor deze problemen is om een grotere database te gebruiken. Deze database zou niet alleen meer afbeeldingen moeten bevatten, maar ook meer proefpersonen van verschillende rassen. In een optimale database zouden alle rassen evenveel

voorkomen. Bovendien zouden de personen in deze database van een bredere leeftijdsklasse zijn.

6.1.4 Vervolg

Bij het doen van onderzoek naar gezichtsuitdrukkingen wordt er vaak tegen een of meerdere van bovengenoemde beperkingen aangelopen. Voor onderzoeken op dit gebied zou het dus een grote vooruitgang zijn om over betere databases te beschikken. Het samenstellen van een database vergt echter zeer veel tijd. Proefpersonen moeten worden gezocht en vervolgens onder exact dezelfde omstandigheden gefotografeerd worden.

Wanneer een dergelijke database bestaat is het interessant om te onderzoeken of dezelfde resultaten behaald kunnen worden met andere rassen. Hierdoor kan vastgesteld worden of het ANN emoties bij mannen en vrouwen binnen hetzelfde ras even goed kan herkennen. Daarnaast zou er ook onderzoek gedaan kunnen worden met een bredere leeftijdsklasse. Op deze manier kan vastgesteld worden of het gevonden resultaat geldt voor alle leeftijden.

6.2 Conditie 2: Aziaten en blanken

Bij de tweede conditie werd het netwerk getraind op Japanse vrouwen en vervolgens getest op zowel Japanse vrouwen als op blanke vrouwen. In dit onderdeel worden de resultaten uit het vorige hoofdstuk besproken en wordt er gekeken naar de beperkingen van dit onderzoek. Tot slot worden enkele mogelijkheden voor toekomstig onderzoek genoemd.

6.2.1 Significantie

Ook bij deze conditie bepalen we door middel van een tweezijdige t-test met een significantieniveau $\alpha = 0,001$ of de percentages per subconditie van elkaar verschillen. Hiervoor hebben we de tweemaal tien percentages gebruikt die we voor elke subconditie hebben verkregen uit het onderzoek. Omdat de

grootte van de steekproef tien is, is het aantal vrijheidsgraden 9.

Japanse vrouwen

Bij deze subconditie werd het netwerk getraind op Japanse vrouwen en getest op zowel (andere) Japanse vrouwen als op blanke vrouwen. De t-waarde die we bij dit onderzoek vonden is 13,681. De p-waarde die uit de combinatie van de t-waarde en de vrijheidsgraden komt is $2,5 \times 10^{-7}$. De p-waarde ligt ruim onder α en het verschil in de prestaties van het netwerk is dus significant.

Blanke vrouwen

Ook bij de subconditie waar het netwerk getraind werd op blanke vrouwen is de t-test uitgevoerd. Hierbij zijn de resultaten gebruikt van tienmaal testen op zowel (andere) blanke vrouwen als op Japanse vrouwen. De bijbehorende t-waarde is 16,886. In combinatie met de vrijheidsgraden levert dit een p-waarde van 4×10^{-8} op. Ook deze p-waarde ligt lager dan het significantieniveau waardoor gesteld kan worden dat het verschil in de prestaties bij de twee groepen significant is.

6.2.2 Implicaties

Uit het feit dat er bij beide subcondities een significant verschil in de prestaties van het ANN gevonden is, kan worden geconcludeerd dat het netwerk veel moeite heeft met het herkennen van emoties bij Japanse vrouwen als het getraind is op blanke vrouwen en vice versa.

Het lijkt voor de hand te liggen om te concluderen dat de grote verschillen van respectievelijk 41% en 44% veroorzaakt zijn door de verschillen in uiterlijke kenmerken tussen blanken en Aziaten en de manier waarop Aziaten en Westerlingen hun emoties uiten. Hiermee moeten we echter voorzichtig zijn. Dit onderzoek laat alleen zien dat het ANN met de specifieke kenmerken die wij gekozen hebben significant slechter presteert op een van de twee rassen wanneer het getraind is op het andere ras. We kunnen slechts speculeren over de oorzaken hiervan.

Ook kunnen er geen conclusies worden getrokken over verschillen in emotieherkenning tussen rassen in het algemeen, aangezien in dit onderzoek alleen blanken en Aziaten gebruikt zijn. Voordat dergelijke conclusies getrokken kunnen worden, zal er eerst uitgebreider onderzoek gedaan moeten worden met meerdere rassen.

6.2.3 Beperkingen

Vanwege de beperkte beschikbaarheid van bruikbare databases met gezichtsexpressies zijn er voor deze conditie alleen vrouwen gebruikt. Voordat er sterke conclusies getrokken kunnen worden met betrekking tot verschillen tussen Aziaten en blanken in het algemeen dient er eerst een soortgelijk onderzoek gedaan te worden waar ook Aziatische en blanke mannen in voorkomen.

Ook zijn de databases niet erg gevarieerd wat betreft leeftijd. Alle personen in zowel de CK+ database als de JAFFE database zijn universitair studenten en dus jongvolwassenen. De betrouwbaarheid en bruikbaarheid van de resultaten zouden een stuk groter zijn wanneer het onderzoek gedaan zou worden met proefpersonen van alle leeftijden.

Een andere beperking van dit onderzoek is dat de Aziatische vrouwen en de blanke vrouwen uit twee verschillende databases komen. Hoewel wij ons best hebben gedaan om de afbeeldingen zoveel mogelijk overeen te laten komen, kan dit toch gezien worden als een slecht te controleren variabele.

6.2.4 Vervolg

Uit het bovenstaande komen enkele mogelijke vervolgonderzoeken bovendien. De eerste daarvan is het doen van een zelfde onderzoek waarbij de groepen Aziaten en blanken elk bestaan uit mannen en vrouwen, in plaats van alleen vrouwen. Ook zou er meer variatie in de leeftijd van de proefpersonen gebracht kunnen worden. Een tweede is het gebruiken van slechts één goede, consistente database die alle proefpersonen bevat, om zo

veel mogelijk twijfel over de resultaten weg te nemen. Op dit moment lijkt een dergelijke database echter niet beschikbaar te zijn. Wanneer dit wel het geval is kunnen deze twee genoemde onderzoekslijnen wellicht samen genomen worden.

We willen hier ook refereren aan het eerder genoemde onderzoek van Camras et al. (2006), waarbij gekeken werd naar Chinese kinderen die in China of in de Verenigde Staten opgroeiden bij Chinese dan wel Amerikaanse ouders. Het zou een interessante lijn van onderzoek zijn om de geautomatiseerde emotieherkenning op deze groepen te trainen en testen. In een dergelijk onderzoek zou zowel emigratie als adoptie onderzocht kunnen worden. Wellicht kan op die manier inzichtelijk worden in welke mate de uiting van emoties aangeboren is en welk deel wordt bepaald door de cultuur waarin iemand opgroeit.

7 Conclusie

In dit hoofdstuk wordt voor beide condities een conclusie getrokken. Daarbij geven we een antwoord op de onderzoeksvraag en kijken we of de resultaten in lijn zijn met de in hoofdstuk 3 gestelde hypothese.

7.1 Conditie 1: Mannen en vrouwen

Uit de resultaten is naar voren gekomen dat het netwerk dat getraind is op maar een van de geslachten, emoties bij beide geslachten even goed kan herkennen. Uit deze resultaten kan niet de conclusie getrokken worden dat er geen verschillen in het tonen van emoties tussen mannen en vrouwen zijn. Wel kan hier worden geconcludeerd dat er genoeg overeenkomsten tussen mannen en vrouwen zijn, waardoor het netwerk bij beide geslachten emoties kan herkennen, zelfs als het traint op maar een van de geslachten.

Uit deze resultaten wordt duidelijk dat het antwoord op de onderzoeksvraag niet hetzelfde is als de verwachting was bij de hypothese. Het gebouwde netwerk is wel degelijk in staat om

emoties even goed te herkennen bij beide geslachten wanneer het getraind is op maar één geslacht. Het netwerk kan dus goed omgaan met eventuele verschillen die bestaan tussen mannen en vrouwen wanneer deze emotie vertonen.

7.2 Conditie 2: Aziaten en blanken

De onderzoeksvraag die we bij deze conditie hebben geprobeerd te beantwoorden is: Wanneer een Artificieel Neuraal Netwerk getraind wordt om emoties bij blanken te herkennen, is het dan net zo goed in emotieherkenning bij Aziaten vergeleken met blanken, en vice versa? Het antwoord op deze vraag is: 'Nee'.

Er is een significant verschil gevonden in de prestaties van het Artificieel Neuraal Netwerk bij beide subcondities. Deze verschillen waren gemiddeld ruim 40%. Daaruit kan geconcludeerd worden dat er te weinig relevante overeenkomsten zijn tussen de geselecteerde Japanse vrouwen en blanke vrouwen, waardoor het netwerk, wanneer het getraind is op de ene groep, moeite heeft met het herkennen van emoties bij de andere groep.

In de hypothese werd voorspeld dat er een verschil in de prestaties van het netwerk bij de twee groepen zou zijn, wanneer het netwerk getraind werd op slechts een van de groepen. De resultaten zijn in lijn met deze hypothese. In de hypothese werd aangedragen dat deze verschillen veroorzaakt worden door het verschil in uiterlijk en de manier waarop door Aziaten en blanken met emoties wordt omgegaan. Dit zullen voor nu echter speculaties blijven, omdat we geen onderzoek hebben gedaan naar de oorzaken van dit verschil.

8 Dankwoord

We willen alle onderzoekers die aan de Extended Cohn-Kanade database meegewerkt hebben bedanken dat we gebruik mochten maken van hun database. Daarnaast willen we ook alle onderzoekers die betrokken waren bij

het maken van de Japanese Female Facial Expression database bedankt voor het beschikbaar stellen van hun database voor dit onderzoek. Tenslotte willen we dr. Gerard Vreeswijk bedanken voor zijn begeleiding tijdens het onderzoek. Zonder deze mensen hadden we dit onderzoek niet met succes kunnen afronden.

9 Referenties

Burton, A. M., Bruce, V.m Dench, N. (1993). What's the difference between men and women - Evidence from facial measurement. *Perception* 22 (2), pp. 153-176.

Camras, L., Bakeman, R., Chen, Y., Norris, K., Cain, T. (2006). Culture, Ethnicity, and Children's Facial Expressions: A Study of European American, Mainland Chinese, Chinese American, and Adopted Chinese Girls. *Emotion*, Vol. 6, No. 1, pp. 103-144.

Cantelmo, N. (2007). *Rapid Facial Expression classification Using Artificial Neural Networks*. Northwestern University [online], http://steadystone.com/research/ml07/ncantelmo_f inal.doc

Cohn, J. F., Kanade, T., & Tian, Y. (2000). *Comprehensive database for facial expression analysis*. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

Cohn, J. F., Kanade, T., Lucay, P., Saragih, J., Ambadar, Z., & Matthews, I. (2010). *The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression*. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.

Ekman, P. (1972). *Universals and Cultural Differences in Facial Expressions of Emotion*. In: Cole, J. (ed). *Nebraska Symposium on Motivation*, vol. 19. Lincoln University of Nebraska Press, pp. 207-287.

Franco, L., Treves, A. (2001). *A Neural Network Facial Expression Recognition System using Unsupervised Local Processing*. *Image and Signal Processing and Analysis*, 2001, pp. 628-632.

Hess, U., Adams, R., Kleck, R. (2004). *Facial Appearance, Gender, and Emotion Expression*. *Emotion*, Vol. 4, No. 4, pp. 378-388

Kring, A. M., Gordon, A. H. (1998). *Sex differenced in emotion: Expression, experience, and physiology*. *Journal of Personality and Social Psychology*, Vol 74(3), Mar 1998, 686-703.

Lyons, M. J., Akamatsu, S., Kamachi, M., Gyoba, J. (1998). *Coding Facial Expressions with Gabor Wavelets*. Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.

Mehrabian, A. (1968). *Communication without Words*, *Psychology Today*, vol. 2, no. 4, pp. 53-56.

Mitchell, T. (1997). Chapter 4: Artificial Neural Networks. *Machine Learning*, pp. 81-127.

Pantic, M. & Rothkrantz, J.M. (2000). *Automatic Analysis of Facial Expressions: The State of the Art*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22., no. 12.

Youssif, A., Asker, W. (2011). *Automatic Facial Expression Recognition System Based on Geometric and Appearance Features*. *Computer and Information Science*, Vol. 4, No. 2.

Yuki, M., Maddux, W., Masuda, T. (2005). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology* 43 (2007), pp. 303-311.