

Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

Opplæring i psykisk helsevern: Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

En evaluering av Microsoft Cognitive API

Agathon Maximillian Skei-Hart



Masteroppgave i Arbeids- og Organisasjonspsykologi
ved Psykologisk Institutt

UNIVERSITETET I OSLO

15.05.2018

© Agathon Maximillian Skei-Hart

2018

Opplæring i psykisk helsevern: Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

Agathon Maximillian Skei-Hart

<http://www.duo.uio.no>

Trykk: Reprosentralen, Universitetet i Oslo

Sammendrag

Trening og opplæring er et felt de aller fleste organisasjoner bruker mye ressurser på. Derfor ønsker mange organisasjoner å finne kostnadsbesparende opplæringsintervensjoner som har høy kvalitet og er tilgjengelig for alle. En form for opplæring som har fått mye oppmerksomhet i det siste er opplæring på tekniske og virtuelle plattformer. Å digitalisere opplæring er også noe Simuleringssenteret på Hamar ønsker å gjøre. For å få til dette har Simulering senteret på Hamar startet å undersøke om man kan bruke teknologi som identifiserer ansiktsemosjoner hos andre. Forskningsspørsmålet i denne oppgaven er å undersøke om ansiktslesnings programmet Microsoft Cognitive API (MC-API) kan identifisere emosjoner like riktig og nøyaktig som individer kan. Dette forskningsspørsmålet introduseres for å besvare problemstillingen ”kan ansiktslesnings teknologi brukes som et verktøy for å øke kvaliteten av opplæring av relasjonelle ferdigheter hos Simuleringssenteret Innlandet”. For å undersøke dette har kandidaten designet en studie der deltagerne skal identifisere emosjoner i et utvalg bilder. Bildene inneholder 1 av 8 forskjellige emosjoner. For å svare på forskningsspørsmålet ble deltagernes svar sammenlignet med MC-API sine svar i en statistisk analyse.

Studien fant at MC-API klarer å identifisere emosjonene glad og nøytral nøyaktig og riktig. Til gjengjeld klarer ikke MC-API å kjenne igjen noen andre emosjonene. Dette skiller seg deltagerne i studien som i tillegg til å nøyaktig kunne identifiserer nøytral og glad, også identifiserer de andre 6 emosjonene som blir presentert i studien.

Nøkkelbegreper: Universale emosjoner, Trening og opplæring, Maskinlæring, Non-verbal kommunikasjon

Forord

Denne oppgaven skrives som en del av MUN-Light prosjektet på Simuleringsenteret på Hamar. Oppgaven er et samarbeid mellom UiO og Simuleringsenteret på Hamar. Jeg vil gjerne få lov til å takke Førsteamanuensis Cato Bjørkli ved psykologisk institutt som har vært helt sentral for gjennomføringen av denne oppgaven, veiledet og kommet med nyttige tilbakemeldinger og perspektiver.

Jeg vil også takke alle ved Simuleringsenteret på Hamar som har hjulpet til med den praktiske gjennomføringen av studiet som ble laget. Jeg vil gjerne rette en spesiell takk til Johnny Sandaker, Marianne Lundgård og Egil Utheim som har vært spesielt delaktige for utvikling, gjennomføring og innsamling i studien.

Tilslutt vil jeg gjerne takke alle studenter ved NTNU Gjøvik, ved Høyskolen Innlandet og alle ansatte i stange kommune som deltok i studien. Jeg vil også gi en spesiell takk til Camilla Temmen Evensen og Stig Ekeberg fra Stange kommune som var spesielt delaktige ved innsamling av data hos de ansatte i stange kommune.

Oslo 15.Mai.2018

Agathon Maximillian Skei-Hart

Innholdsfortegnelse

Introduksjon	1
Teori	3
Emosjoner	3
Maskinlæring	8
Formål med oppgaven	10
Metode	11
Prosjektet	11
Utvalg	11
Materiale	12
Målevariabler	14
Prosedyre	14
Analyse	16
Etikk	17
Resultat	17
Deskriptiv statistikk	17
Sluttiningsstatistikk	19
Diskusjon	23
Oppsummering av data	23
Hvordan svarer gruppene	24
Praktisk implikasjon	27
Teoretisk implikasjon	29
Begrensninger ved studien	30
Videre forskning	31
Konklusjon	33
Litteraturliste	34

Vedlegg 1: Informasjon- og samtykkeskjema	38
Vedlegg 2: Spørreskjema for undersøkelsen.....	40
Vedlegg 3: Skjema for evaluering av bildene i studien	45

Introduksjon

Formålsparagrafen i helsepersonell loven sier at loven skal gi sikkerhet for pasientene, sikre kvalitet i helse- og omsorgstjenesten samt fostre tillit til helsepersonell og helse- og omsorgstjenesten, (Helsepersonell loven, 1999, §1) blant annet ved at helsepersonell skal utføre sitt arbeid i samsvar med faglig forsvarlighet og omsorgsfull hjelp (Helsepersonell loven 1999, §4). Det inngående fokuset på faglig kvalitet kommer også frem i helse sør-øst (HSØ) sine etiske retningslinjer fra 2012 (Styret i HSØ, 2012). I retningslinjene står det at ansatte i HSØ skal bestrebe høy faglig kvalitet og har rett på veiledning og opplæring for å oppnå dette. I tillegg skal HSØ drive aktiv undervisning, forskning og innovasjon for å bedre helsetjenesten i sin helhet (Styret i HSØ, 2012). Det kommer tydelig frem at høy kvalitet er helt essensielt i helse- og omsorgstjenesten generelt.

Høy kvalitet er derimot et komplisert begrep, som er sammensatt av flere forskjellige faktorer og prosesser. For en stor organisasjon som HSØ vil begrepet faglig kvalitet bli desto mer komplisert, og vil kunne variere i betydning. For de administrative avdelingene i HSØ kan høy faglig kvalitet sikte til ryddige administrative prosesser, mens høy faglig kvalitet i en omsorgsbolig kan være å ha ansatte med gode evner til å skape relasjon til og vise omsorg for pasientene. Dette illustrerer kompleksiteten rundt begrepet faglig kvalitet i en organisasjon som HSØ.

Siden formålet til HSØ i tråd med helsepersonell loven er å sikre kvalitet i helse- og omsorgstjenesten, blir spørsmålet på hvilken måte man skal oppnå høy kvalitet. Dette er et spesielt viktig spørsmål siden mange faktorer har vist seg å kunne øke den faglige kvaliteten i en organisasjon. For eksempel har det blitt funnet sammenheng mellom opplæring, ledelse, sosialt miljø og stress på arbeidsprestasjon, kunnskapsdeling og faglig kvalitet (Amin, Basri, Hassan & Rehman, 2011; Borges, 2012; Wang, Oh, Courtright & Colbert, 2011). Dette illustrerer at flere faktorer kan påvirke faglig kvalitet i en organisasjon. Likevel er trening og opplæring en faktor som er helt nødvendig for å øke kompetanse, kvalitet og kunnskapsnivå i en organisasjon (Blume, Ford, Baldwin & Huang, 2010). Derfor er trening og opplæring sentralt for å oppnå målet om økt faglig kvalitet i helsepersonell loven og i HSØ sine retningslinjer. Oppgaven kommer derfor til å ta utgangspunkt i trening og opplæring for å oppnå formålet om økt kvalitet slik helsepersonell lovens § 1 sier.

Det Amerikanske forbundet for trening og utvikling (ASTD), fant i sin årlige rapport fra 2007 at amerikanske selskaper bruker mer enn 126 milliarder dollar på opplæring og utvikling av ansatte i året (Aguinis & Kraiger, 2009). Dette illustrer hvor kostnadskrevene opplæringsprosesser er for en organisasjon. Siden Georgenson fant at kun 10 % av ressursene man bruker på opplæring fører til at man aktivt bruker kunnskapen man har trent på i jobb (Georgenson, 1982), er det viktig at man ikke bare starter å trene uten god planlegging og en klar oversikt over forskningen som er blitt gjort på området (Burke & Huchins, 2007). Spesielt siden overføringen av kunnskap fra læring til jobb er sterkt korrelert med planlegging, design av opplæringsprogram og organisatoriske faktorer ellers (Aguinis & Kraiger, 2009; Blume et al., 2010; Burke & Huchins, 2007; Colquitt, LePine & Noe, 2000). Høye kostnader ved opplæring er ikke bare tilfellet for Amerikanske bedrifter. NIUF beregnet at lønnskostnadene alene for opplæring i norske bedrifter, var på rundt 46 milliarder kroner for kursing og videreutdanning av arbeidstagere i 2017 (Mark, Børing og Fevolden, 2018).

De høye kostnadene ved trening og opplæring gjør at det er spesielt interessant for organisasjoner å effektivisere opplæringen, slik at kostnaden går ned og tilgjengeligheten økes uten at kvaliteten av opplæringen faller. For å oppnå dette har man i nyere tid sett på bruken av opplæring på virtuelle flater som pc-spill, nettopplæring og andre typer VR-Løsninger (Cohen et al., 2003; Deterding, Dixon, Khaled & Nacke, 2011). Siden majoriteten har tilgang på det teknologiske utstyret som gjør det mulig for alle å drive denne typen opplæring, er dette en kostnadsbesparende og tilgjengelig måte å drive opplæring på (Valmaggia, Latif, Kempton, & Rus-Calafell, 2016). Selv om dette kan være en kostnadsbesparende måte å drive opplæring på, står faglig kvalitet i opplæring helt sentralt i helsetjenesten. I følge Helsepersonell loven er Kvaliteten i opplæringen er dermed et premiss som må være tilstede i opplæringen som blir gjennomført.

Å opprettholde denne kvaliteten kan være spesielt utfordrende innen helsetjenesten, siden mye av opplæringen i denne bransjen også er opplæring av relasjonelle ferdigheter. I Helsepersonell lovens § 4, blir forsvarlig og omsorgsfull hjelp understreket. I motsetning til å trene fagkunnskap om hvordan et snitt for å fjerne en føflekk gjøres, handler relasjonell ferdigheter om å trene på kommunikasjon, omsorg og hensiktsmessige responser i sosial interaksjon med andre. Dette vil si at komplekse

menneskelige funksjoner som kroppsspråk, ansiktsemosjoner, tanker og atferd må kunne representeres virtuelt for at virtuell opplæring skal kunne bringe høyere faglig kvalitet.

En type teknologi som er blitt introdusert for å møte denne utfordringen er algoritmer som bruker maskinlæring til å identifisere emosjoner mennesker viser i bilder eller videoer. Denne teknologien gir for eksempel utviklere muligheten til å lage et opplæringsprogram som responderer på den som trener sin emosjonelle tilstand. Teknologien kan derfor potensielt være et verktøy for å sikre høyere faglig kvalitet ved trening på relasjonelle ferdigheter i en organisasjon. En av aktørene som har presentert teknologi for å identifisere emosjoner er Microsoft med programmet Microsoft Cognitive API (MC-API). Selv om flere aktører tilbyr lignende programmer, kommer denne oppgaven til å fokusere på Microsoft sitt program MC-API.

For å undersøke dette kommer oppgaven til å undersøke om MC-API kan identifisere emosjoner hos individer like riktig og nøyaktig som mennesker ellers. Dette gjøres for å undersøke om MC-API er et verktøy man kan bruke til å øke kvalitet av opplæring i en organisasjon med fokus på trening av både faglige og relasjonelle ferdigheter.

Teori

Flere store fagfelt er sentrale i forbindelse med denne oppgaven, men det er ikke praktisk mulig å dekke alle fagfeltene i denne oppgaven. I teoridelen av oppgaven kommer forskningen som er gjort på emosjoner å bli forsøkt sammenfattet, innforstått med at et så omfattende fagfelt ikke kan dekkes i helhet i denne oppgaven. Deretter vil de overordnede prinsippene bak algoritmer og maskinlæring bli gjennomgått, før oppgaven kort nevner forholdet mellom verbal og non-verbal kommunikasjon for kommunikasjon mellom mennesker.

Emosjoner

I følge James A Russell (2003), vil majoriteten av alle spørsmål i verden være påvirket av emosjoner. Emosjoner har lenge vært godtatt som en form for "common-sense" folketeori som har blitt brukt til å predikere og forstå ulike spørsmål, eller blitt forklart som en konsekvens av en situasjon. En rekke hverdagslige ord som sinne, glede, overraskelse og tristhet har blitt brukt til å kategorisere emosjonene (Russell, 2003), men forskere på området

er fortsatt uenige om hva emosjoner egentlig er (Lazarus, 1991; Russell, 2003). Det er gjort mye forskning på emosjoner, og det er blitt gjort mange forsøk på å presentere modeller for å forstå emosjoner. Modellene varierer ofte på et spekter fra en rent behavioristisk modell slik Ekman foreslo (Ekman, 1977) til en mer sosial konstruktivistisk modell slik Carl Ratner foreslo i 1989.

Selv om det finnes flere modeller, kommer oppgaven til å ta for seg to av de mest fremtredende modellene for emosjoner. Den første modellen omtales som vurderingsmodellen for emosjoner, og er en modell som beskriver emosjoner som et resultat av et individ sin vurdering av ytre faktorer (Lazarus, 1991). Den andre modellen ble presentert av James A Russell (2003), og omtales gjerne som psykologisk konstruktivistisk modell for emosjoner. Denne modellen argumenterer for at samlebetegnelsen for emosjoner slik de er i dag ikke er gode nok, og at man må ha en mer dimensjonal forståelse av emosjoner (Russell, 2003).

Vurderingsmodellen. Vurderingsmodellen til Lazarus (1991) kan ses på som en forlengelse av Ekman sin behavioristiske modell for emosjoner (Russell, 2003). Den behavioristiske modellen til Ekman (1977) argumenterer for at emosjoner er et resultat av universale biologiske prosesser. Han anerkjenner likevel at ulike kulturelt betingede faktorer påvirker forholdet mellom en situasjon og emosjonen (Lazarus, 1991). Ekman (1977) argumenterer for at hvis et individ opplever en situasjon på en bestemt måte, vil dette individet nesten uten unntak oppleve en bestemt universell emosjonell respons uavhengig av situasjonen ellers (Lazarus, 1991). Lazarus derimot argumenterer for at dette ikke er en god nok modell for å beskrive hvordan emosjoner oppstår i et menneske, og introduserer derfor vurderingsmodellen.

Lazarus beskriver forholdet mellom en situasjon og den emosjonelle responsen som påvirket av en person-miljø interaksjon. Med dette mener Lazarus, at emosjoner i seg selv ikke skapes av miljøet eller en fysiologisk prosess, men samspillet mellom dem. Han kaller selv modellen en relasjonell, motivasjons og kognitiv modell for emosjoner. Med dette mener han at relasjonelle, motiverende eller kognitive faktorer er et premiss for den emosjonelle responsen til den gitte situasjonene (Lazarus, 1991).

Med relasjonelle faktorer refererer Lazarus til at emosjoner alltid er om person-miljø forhold der det er sjanse for skade eller fordeler, og dette i tur påvirker om man vil oppleve en positiv eller negativ emosjon (Lazarus & Folkman, 1987). Med motivasjons faktorer snakker Lazarus om emosjoner som er en reaksjon som resultatet av det vi ønsker å oppnå. Han argumenterer videre for at motivasjonen gir en forståelse av hva som er en tilpasningsdyktig

respons til situasjonen, og derfor også en kilde til forståelse om skade eller fordeler på samme måte som relasjonelle forhold (Lazarus, 1991). De kognitive faktorene består av et individ sin vurdering og kunnskap om hvordan man skal møte en situasjon, for å respondere på en tilpasningsdyktig måte. Kunnskap består av en generell tanke om hvordan ting fungerer og vurderinger er en forståelse om hva som kommer til å skje når et individ interagerer med situasjonen på en bestemt måte.

Selv om Lazarus nevner faktorene over som premissene for en emosjon, sier han at dette alene ikke er et verktøy som er godt nok til å gjøre en detaljert analyse av hver emosjon. Han introduserer derfor individer sitt individuelle vurderingsmønster (appraisal pattern) som et verktøy for å kunne forklare fremkomsten av spesifikke emosjoner. Lazarus presenterer primær og sekundær vurdering av en situasjon, som verktøy for å forklare forekomsten av de gitte emosjonelle responsene.

Primær vurdering beskriver Lazarus som vurderinger angående risikoen som kommer med utfallet av en gitt situasjon. Han argumenterer for at denne er primær fordi det ikke er grunnlag for en emosjonell respons hvis ikke det er risiko tilstede i situasjonen. Lazarus never at det finnes tre former for primær vurderinger kalt "goal relevance", "goal congruence" og "goal content". For en dypere beskrivelse av de tre vurderingsformene se Lazarus, 1991.

Den andre formen for vurderinger kaller Lazarus for sekundær vurderinger. Dette er vurderinger som omhandler et individ sine evner og muligheter for å kunne takle situasjonen de befinner seg i. Lazarus kaller de tre formene for sekundær vurderingene "Blame and credit", "Coping potential" og "future expectations".

En blanding av disse primær og sekundær vurderingene mener Lazarus er grunnlaget for den spesifikke emosjonelle responsen hos et individ. Dette forklarer hvordan flere individer kan være tilstede i samme situasjon uten å få den samme emosjonelle responsen, siden den kognitive informasjonen og vurderingen av situasjonen er ulik fra person til person (Lazarus, 1991).

Lazarus mener likevel at alle mennesker opplever de samme emosjonene og at dette er noe som går utover sosiale begrensninger som språk og kultur. I en videreføring av dette mener han at alle emosjoner er bygget opp av en rekke spesifikke primær og sekundær vurderinger, og at vurderingene mennesket gjør av en situasjon avgjør om man føler glede eller sinne i den spesifikke situasjonen.

James A. Russell kritiserer Vurderingsmodellen og mener at modellen baserer seg på et teoretisk rammeverk om at det "finnes" et begrenset antall følelser. Han sammenligner

dette ved å påstå at Lazarus fremmer et syn om at følelsene er begrenset til en periodisk tabell av emosjoner (Russell, 2003). Russell sin dimensjonale modell argumenterer for at man ikke kan forstå emosjoner som kategorier som glad eller sint. Følelser må ses som er en sammensetning av flere faktorer som individuelle vurderinger, sosiale- og kulturelle faktorer. Dermed presenterte Russell Den psykologisk konstruktivistiske modellen som en alternativ forståelse av emosjoner (Russell, 2003)

Den psykologisk konstruktivistiske modellen. Russell argumenterer for at man i psykologien har tatt forgitt begreper som sinne, redsel, angst osv. men at det i realiteten egentlig er flere følelser som befinner seg på en dimensjonal flate. I stedet for å se en person som trist, må man se emosjonene som en som en blanding av alle kulturelle, sosiale, vurderinger og andre faktorer som fremprovoserer den spesifikke prototype emosjonen de føler (Pinker, 1997; Russell, 2003). Russell forklarer det slik: Hvis Erik er ute å går i skogen og støter på en brunbjørn, er frykt en naturlig emosjonell reaksjon. Erik sin naturlige reaksjon vil være å skjelve, løpe eller skrike. Derimot når han en uke senere går han på kino og betaler for å se en skummel film opplever han en annen form for frykt. I motsetning til møtet med brunbjørnen vet Erik at denne situasjonen ikke er farlig, og løper derfor ikke fra kinoen. I stedet blir han underholdt av situasjonen og betaler gjerne for å se filmen igjen (Russell, 2003). Han argumenterer derfor for at de to emosjonene Erik føler i de to situasjonene kun deler en merkelappen redd, men de er utover dette ikke like. Man kan derfor heller ikke tenke på emosjoner slik man gjør per dags dato, men må se de langs en dimensjon. Selv om begge de emosjonelle responsene omtales som frykt, mener Russell at denne frykten ikke er det samme, og kanskje til og med er forskjellige emosjoner. Responsen kan derfor ikke klassifiseres som en "common-sense" kategori av frykt, slik Lazarus prøver å gjøre i sin modell (Russell, 2003). Uansett hvilken teoretisk posisjon leseren velger å ta, kan det ikke stikkes under en stol at mennesker generelt er gode til å kjenne igjen emosjoner hos andre (Ekman & Freisen, 1986; Prinz, 2004; Russell, 1994; Van Hemert, Poortinga & van de Vijver, 2007).

Universale emosjoner. I forskning på emosjoner finner man bred støtte for at noen universale emosjoner blir kjent igjen på tvers av både sosiale og kulturelle forskjeller slik Ekman, Sorensen og Friesen fant i 1969 (Ekman, Sorenson & Friesen, 1969). Matsumoto (1990) sier at de ikke lenger er en uenighet om at det finnes universale emosjoner i psykologien. Debatten går i større grad ut på hva implikasjonen av dette er i praksis (Turner & Ortony, 1992). Denne diskusjonen kommer ikke oppgaven til å gå nærmere inn på, da

premisset for oppgaven er at universale emosjoner eksisterer og at mennesker har en evne til å identifisere disse emosjonene. Utgangspunktet for oppgaven blir dermed at Ekman, Freisen og Izard sine teorier om grunnleggende identifiserbare universale emosjoner stemmer.

Leseren kan selv bestemme hvilken teoretisk modell han eller hun ønsker å legge til grunn, men hypotesen om at universale identifiserbare emosjoner eksisterer slik forskningen tyder på er satt som premiss for oppgaven (Ekman & Freisen, 1986; Prinz, 2004; Russell, 1994; Van Hemert, Poortinga & van de Vijver, 2007).

Det har blitt presentert flere teorier om hvilke emosjoner som er universale. Ekman og Freisen (1986) argumenterer for de seks universale emosjonene de presenterte i 1971. Dette er emosjoner man finner en høy pan-kulturell enighet i (Ekman, Sorensen & Freisen, 1969). De seks emosjonene Ekman presenterer er: glede, sinne, tristhet, overraskelse, avsky og redsel (Ekman & Freisen, 1986). Derimot fant de ikke like mye hold for andre emosjoner, blant annet for Izard sine to supplerende emosjoner interesse og skam fra 1969 og 1971 (Ekman & Freisen, 1986).

I ettertid er det disse seks emosjonene det har blitt funnet bredest støtte for i forskningen (Prinz, 2004; Russell, 2003; Sprung, Münch, Harris, Ebesutani & Hofman, 2015). De seks emosjonene populært kalt "the big six" er derfor blitt bredt anerkjent som de grunnleggende universale emosjonene (Prinz, 2004). I ettertid har det blitt gjort forsøk på å legge til emosjoner på denne listen, blant annet av Ekman & Freisen i 1986 da de fant resultater som tydet på at forakt også er en universal emosjon (Ekman & Freisen, 1986). Likevel er det i forskningen som er gjort til nå, "the big six" som får bredest støtte. Det er i hovedsak disse 6 emosjonene, pluss emosjonene forakt og nøytral som kommer til å brukes i studien som gjøres i oppgaven. Dette skyldes det er MC-API som skal evalueres i denne oppgaven, har emosjonene i "the big six" samt nøytral og forakt som mulige responsalternativer for evaluering av emosjoner hos andre.

Siden forskningen tyder på at universale emosjoner eksisterer, har det vært av interesse å lage modeller som beskriver hvilke muskelbevegelser som kjennetegner de ulike ansiktsemosjonene. Det mest kjente rammeverket er FACS (Facial Action Coding System) som ble presentert av Ekman & Friesen i 1978.

FACS. FACS er et system for å kode muskulære bevegelser i ansiktet. FACS forsøker å lage et system som kan brukes til å identifisere forskjellige ansiktsbevegelser ved å dele alle mulige muskelbevegelser i ansiktet inn i bevegelses områder (action units) (Ekman & Friesen, 1978). Bevegelses områdene (BO) er systematisert slik at alle muskler går inn under et

spesifikt område. For eksempel er BO 12 området på ytterkanten av leppene, og en aktivering i dette bevegelsesområdet kan resultere i at man løfter på hjørnene av leppene. FACS består av 44 bevegelses områder (Ekman & Friesen, 1978; Kanade, Cohn & Tian, 2000), og hver BO forteller om hver enkelt muskel sin mulighet til bevegelse (Ekman & Friesen, 1978).

Ved å kode hver BO på denne måten, har Ekman og Friesen gjort det mulig å lage et system som kan fortelle noe om hvilke muskulære bevegelser som kjennetegner de ulike emosjonene. Man må likevel være klar over at FACS bevegelses områder kun beskriver muskulære bevegelser, og at emosjonelle uttrykk i sin helhet er kodet inn i andre systemer som EMFACS (Kanade, Cohn & Tian, 2000). Selv om FACS kun beskriver enkelte muskulære bevegelser kan en sammensetning av aktivering i forskjellige BO være kjennetegn på ulike emosjoner. For eksempel kan glede kjennetegnes ved muskulær aktivering ved å heve kinnene (BO 6), og samt aktivering ved å trekke opp hjørnene på leppene (BO 12).

Å kategorisere enkelte ansiktsbevegelser, eller en sammensetning av disse slik at de utgjør en emosjon gjør at man ved ren mønstergjenkjenning kan identifisere de universale emosjonene som finnes. Siden maskinlæring kan brukes til all form for mønstergjenkjenning, skal det ved bruk av maskinlæring være mulig å identifisere de universale emosjonene som eksisterer (Lien, Kanade, Cohn & Chung Li, 1998). Siden det er mulig å bruke maskinlæring til ansiktsgjenkjenning, vil det være interessant å se nærmere på bruken av maskinlæring som en metode å gjenkjenne emosjoner hos andre.

Maskinlæring

I boken "Machine Learning, a probabilistic perspective" (Robert, 2014) defineres maskinlæring som et sett med metoder som automatisk kan oppdage mønstre, for deretter å kunne bruke mønstret til å forutse fremtidige data eller utføre alternative beslutninger hvis det ikke er klart for programmet hva som skal gjøres. Dette er en av flere definisjonen for maskinlæring, men er den som vil brukes i denne oppgaven. Kort fortalt vil dette si at en algoritme i et system er designet slik at programmet kan endre respons, når den over tid oppdager at en annen respons enn den opprinnelige er mer hensiktsmessig.

Man har forsøkt å implementere maskinlæring i flere studier som har brukt programmer til å identifisere ansiktsemosjoner slik Microsoft Cognitive API gjør. Alle studiene har brukt ulike metoder for å oppnå maskinlæring i disse programmene (Cootes, Edwards & Taylor, 2001; Viola & Jones, 2001; Yu & Zhang, 2015). I

denne oppgaven kommer det ikke til å bli forklart nærmere hvordan ulike algoritmer for maskinlæring i MC-API eller andre lignende programmer fungerer, siden dette blir et mer teknisk spørsmål enn det oppgaven skal ta for seg. Derimot søker oppgaven å undersøke hvor godt MC-API presterer sammenlignet med menneskelige deltagere for identifisering av emosjoner hos andre.

Så hvorfor er det av interesse å undersøke om MC-API funker? Hvis man ser nærmere på læringsmålene i faget "Relasjonelle ferdigheter og psykisk helsevern" ved HiOA skal studenten lære å "skape gode relasjoner til aktuelle samarbeidspartnere" (HiOA, 2017). Caris-Verhallen, Kerkstra og Bensing (1999) fant at relasjonen og kommunikasjonen mellom pasienten og sykepleiere er essensielt for å kunne oppdage behovene en pasient har, samtidig som det er viktig for å kunne vise empati og omsorg for pasienten. Dermed blir de relasjonelle ferdighetene helt avgjørende, for at helsepersonell skal kunne utføre arbeidet sitt i samsvar med den faglig forsvarligheten og den omsorgsfulle hjelpen § 4 i helsepersonell loven krever (Helsepersonell loven 1999, §4).

Chambers argumenterer for at all kommunikasjon involverer bruk av både non-verbal og verbal kommunikasjon (Chambers, 2003). Non-verbal kommunikasjon defineres som all form for kommunikasjon som ikke kontrolleres av talen (Kacperck, 1997). Chambers argumenterer videre for at den non-verbale kommunikasjonen har opptil 5 ganger mer påvirkning på en situasjon enn den verbale kommunikasjonen har. For arbeidstagere som er avhengige av å ha gode relasjonelle ferdigheter kommer det tydelig frem at kommunikasjon, deriblant non-verbal kommunikasjon er en sentral ferdighet. Dermed vil trening og opplæring av kommunikasjon, både verbal og non-verbal stå sentralt for opplæring i helsetjenesten. Hvis man ønsker å drive trening og opplæring av relasjonell kunnskap på et virtuelt plan, blir man dermed også nødt til å gjenskape non-verbal kommunikasjon på en virtuell flate.

Denne oppgaven er en del av et prosjektet MUN-Light som ønsker å digitalisere opplæringen av blant annet relasjonelle ferdigheter for Simuleringscenteret Innlandet (SIM-innlandet). Oppgaven vil i forlengelse av dette prosjektet undersøke om MC-API kan brukes som et verktøy for å oppnå dette målet.

MUN-Light. MUN-Light prosjektet gjennomføres i regi av Simulering senteret på Hamar. Senteret jobber med simuleringstrening av ulike situasjoner ansatte i psykiatri,

ambulansse eller annen form for psykisk helsevern møter i hverdagen. Opplæringen fokuserer på utvikling av både faglig og relasjonelle ferdigheter. Derfor er teknologi for å øke kvalitet i relasjonell opplæring av de ansatte av spesiell interesse for SIM-innlandet. Per dags dato bruker SIM-innlandet scenarier der ansatte spiller ut for eksempel utageringssituasjoner. Scenariene som spilt ut blir filmet slik at markører (de som spiller ut scenariet) og andre kolleger kan diskutere situasjonen, og bli enige om hva som ble håndtert godt og hvilke kommunikative eller faglige vurderinger som kan forbedres.

I 2015 startet SIM-Innlandet opp prosjektet MUN-LIGHT for å utvikle og digitalisere opplæringen som blir tilbudt i dag. Tanken bak prosjektet er å øke kvaliteten på opplæringen i psykiatri og andre helsesektorer ved å øke tilgjengeligheten av god opplæring for de ansatte, samt redusere kostnaden ved å tilby denne opplæringen. Selv om SIM-innlandet er kommet et stykke på vei, ønsker de med MUN-LIGHT prosjektet å fortsette å utvikle opplæring på inn på en digital plattform. Et av verktøyene prosjektgruppen har vurdert å ta i bruk for å oppnå dette målet er ansiktslesningsprogrammet Microsoft Cognitive API (MC-API).

Formål med oppgaven

Formålet med oppgaven er å undersøke om programmet MC-API kan identifisere emosjoner like nøyaktig som yrkesutøvere i psykiatrien gjør. Det er enighet i forskningen om at mennesker på et krysskulturelt plan, er gode til å kjenne igjen noen universale emosjoner hos andre mennesker (Ekman & Freisen, 1986; Prinz, 2004; Russell, 1994; Van Hemert, Poortinga & van de Vijver, 2007). For at MC-API skal oppfylle kravet for å kunne fungere som et verktøy for å øke kvaliteten av opplæring, må programmet i likhet med mennesker effektivt kunne identifisere emosjoner hos andre mennesker. Den første hypotesen i oppgaven (H1) blir dermed:

H1: Det er ingen forskjell i MC-API og deltagerne sin evne til nøyaktig og riktig identifisere andre sine ansiktsemosjoner.

Slik studien er laget er det en fasit på bildene som er blitt vist. Siden det praktisk ikke lot seg gjøre å få tilgang på et allerede eksisterende teoretisk rammeverk som FACS (Ekman & Friesen, 1976) eller annet, ble 4 sykepleiere/miljøterapeuter brukt til å evaluere bildene. De bildene det var bredest enighet om ble tatt med og vurderingen til de 4 som evaluerte bildene ble satt som fasit (mer inngående beskrivelse i metodedel). Den andre og tredje hypotesen (H2 og H3) i oppgaven blir dermed:

H2: Det er ingen forskjell i deltagerne og fasiten sin evne til nøyaktig og riktig

identifisere andre sine ansiktsemosjoner.

H3: Det er ingen forskjell i MC-API og fasiten sin evne til nøyaktig og riktig identifisere andre sine ansiktsemosjoner.

Metode

Prosjektet

I forlengelse av MUN-Light prosjektet ved SIM-innlandet ble det forespeilet om oppgaven kunne evaluere MC-API. Målet har vært å undersøke om programmet potensielt kan være et verktøy for å digitalisere opplæringen av relasjonelle ferdigheter i psykiatrien. For å adressere dette spørsmålet har kandidaten designet, organisert og samlet inn data for å undersøke om MC-API kan identifisere emosjoner like riktig og like nøyaktig som deltagerne i studiet gjør. Oppgaven er en av flere artikler og oppgaver som skal gi SIM-Innlandet et godt faglig fundament å utvikle opplæringsintervensjonene sine videre på. Den samlingen av oppgaver og artikler SIM-innlandet ønsker å skape, er en blanding av forskning på forskjellige temaer og fagområder i samarbeid med flere utdanningsinstitusjoner, blant annet Universitetet i Oslo, Høgskolen Innlandet og NTNU Gjøvik.

Utvalg

Utvalget av deltagere i studien består av 115 menneskelige deltagere (heretter omtalt som deltagerne) pluss MC-API sine svar. Deltagerne i studien er en blanding av studenter og ansatte i en omsorgsbolig i Stange Kommune. Alle deltagerne er i en viss grad tilknyttet MUN-LIGHT prosjektet. Eksempelvis har stange kommune flere representanter som er med i prosjektet for å kartlegge hvilke opplæringsbehov ansatte som trener hos SIM innlandet har. Studentene som har gjennomført studien studerer ved to av utdanningsinstitusjonene som er tilknyttet prosjektet, henholdsvis Høgskolen Innlandet og NTNU Gjøvik.

Utvalget for MC-API i studien er kun 1. Dette skyldes at MC-API sine svar i studien var identiske ved 4 gjennomganger, og det er dermed rimelig å anta at svarene til MC-API vil være identiske uansett hvor mange gjennomganger som gjennomføres. Dette vil si at man ikke finner varians eller std.avvik i MC-API sine svar uansett hvor mange ganger programmet evaluerer de 17 bildene. Dermed er det det ikke noen argumenter for å sette en høyere N, og derfor er N=1 for MC-API.

Materiale

For å undersøke hypotesene i oppgaven laget kandidaten et eksperiment med 17 bilder deltagerne og MC-API skulle identifisere. Bildene er basert på Ekman, Sorenson & Frieisen (1969) og Ekman (1971) sine big six emosjoner: Sint, Glad, Overrasket, Trist, redd og avsky, pluss emosjonene forakt og nøytral. De to siste emosjonene ble i hovedsak inkludert i studien siden MC-API bruker disse to emosjonene i sin algoritme. Av de 17 bildene som ble tatt med er det 2 bilder av emosjonene glad, overrasket, avsky, forakt, sint, nøytral og redd, mens det ble tatt med tre bilder av emosjonen trist.

For å produsere bildematerialet ble det brukt 6 frivillige fotografert mens de simulerte de 8 emosjonene i eksperimentet. Bildematerialet ble samlet inn fra 3 menn og 3 kvinner, for å minimere sjansen for skjevhet i bildematerialet. Optimalt burde de frivillige som ble fotografert variert mer i alder også, men dette lot seg ikke gjøre og kan potensielt være en svakhet ved bildemateriale i denne studien.

Kohler et al., (2004) argumenterer for at det finnes to måter å samle inn bilder av emosjoner på. Den ene er omtales som Stanislawski's teknikk for å fange fremprovoserte ansiktsemosjoner. Ved å bruke Stanislawski's teknikk ønsker du å fremprovosere en emosjon som for eksempel frykt ved å få individer til å se en film, for så å ta bilder når den emosjonen man ønsker å fange oppstår. Den andre er ved å få modellene til å "posere" emosjonen du ønsker å ta bilde av. Kohler et al., (2004) fant at det er betydelig overlapp mellom de to metodene (fremprovoserte og poserte emosjoner), spesielt når det kommer til emosjonen glad og overrasket. En svakhet ved å bruke poserte ansikter, er at den man tar bilde av velger seg ut noen bestemte ansiktsbevegelser for å simulere emosjonen. Derfor får man ikke en like realistisk emosjon som man ville gjort ved å ta bilde av en fremprovosert ansiktsemosjoner (Kohler et al., 2004). I denne oppgaven ble Stanislawski's teknikk for fremprovosere emosjoner brukt. For å gjøre dette ble det presentert utsagn som skulle hjelpe modellene med å sette seg inn i en bestemt situasjon, der det hadde vært naturlig å vise de emosjonene man ønsket å fange i bildene.

En uttalelse for emosjonen glad kan for eksempel se slik ut: *"Du har søkt på drømmejobben, den du har hatt lyst på i mange år, og har nylig fått beskjed om at du har fått den. Du føler deg glad, hvordan viser du dette?"*.

Når bildene ble tatt var det viktig å passe på at lysforhold var mest mulig standardiserte. I tillegg fikk alle modellene fikk beskjed om å holde skuldre eller annen for non-verbal kommunikasjon som ikke var ansiktsemosjoner på et minimum. Dette ble gjort for

i høyest mulig grad å sikre at deltagerne i studien gjorde bedømmelse basert på ansiktsemosjoner alene.

Det ble tatt 2 bilder av hver modell per emosjon og dermed samlet inn totalt 96 bilder. Før de 96 bildene ble bedømt av uavhengige ratere, ble alle 96 bildene randomisert for å minimere sjansen for å oppdage noe mønster i bildene. De fire uavhengige raterne som ble brukt er ansatt i psykiatrisk helsetjeneste i Akershus, for å redusere sjansen for at de kjente noen av modellene fra Hamar. For å bestemme hvilke bilder skulle tas med ble det undersøkt hvilke bilder det var høyest overenstemmelse om (for bilde 10 for eksempel svarte alle 4 glad).

I tidligere forskning på området brukes eksperter på rammeverk til å identifisere emosjoner for å velge hvilke bilder som presenteres i den gitte studien. Det ledende rammeverket er FACS (Facial Action Coding System) som er en objektiv metode for å kvantifisere ansiktsbevegelser for ulike komponenter i ansiktet (Donato, Bartlett, Hager, Ekman & Sejnowski, 1999). Oppgaven hadde ikke tilgang på eksperter på dette rammeverket, og det ble istedenfor brukt uavhengige ratere som ikke har opplæring i bruk av slike rammeverk. Dette er en potensiell svakhet ved oppgaven (dette kommer til å bli diskutert nærmere i begrensninger med oppgaven), men siden det var det ikke praktisk gjennomførbart å finne uavhengige ratere med denne kompetansen ble ratere uten denne kompetansen brukt.

I forkant av denne utvelgelsesprosessen ble hvert bilde redigert for å standardisere bildene. Bilder ble redigert slik at det var like mye luft over hodet til modellen på hvert bilde. I tillegg skulle like mye av skulder og overkropp være synlig på hvert bilde, og hodet skulle oppta tilnærmet like stor andel av hvert bilde. Dette ble gjort for å sikre at identifiseringen av emosjonene i størst mulig grad blir gjort basert på ansiktsuttrykk og ikke annen tilgjengelig informasjon i bildene. For at de uavhengige ratere og deltagerne skal kunne identifisere bildene har hvert bilde blitt tildelt en randomisert nummer mellom 1 og 96.

De 17 bildene som samsvarte best av de 96 bildene ble tatt ut og brukt i studien. For emosjonene forakt og avsky, er det 1 bilde der bare 2 av raterne var enige. Utenom dette hadde alle bildene som ble tatt med i studiet minimum 3 av 4 uavhengige ratere med samme vurdering av emosjonen i bildene. I utvelgelsen av de endelige bildene ble det forsøkt så godt det lot seg gjøre å holde fordelingen mellom kjønn så lik som mulig. Sluttresultatet ble de 17 bildene deltagerne og MC-API bedømte i studien.

Målevariabler

Dataene for hver variabel som ble overført til SPSS ble kodet om til tallverdier. Dette ble gjort ved å gi variabler hver emosjon, kjønn, erfaring og gruppe sin egen tallskåre. For emosjoner fikk for eksempel emosjonen glad fikk tallskåre = 1, overasket = 2 osv. For å kunne sammenligne MC-API med deltagerne i studien, ble raterene sin vurdering av bildene satt som fasit. Evalueringen deltagerne gjorde av bildene ble omgjort til verdien 1 hvis emosjonen ble identifisert riktig av deltager eller MC-API i forhold til fasiten, og ble skåret med tallet 0 hvis emosjonen ble feilidentifisert. Dermed vil responsene på bildene enten gi verdien 1 eller 0 (1 for treff og 0 for feilidentifisering) i SPSS. Denne måten å operasjonalisere svarene på gjør det mulig å sammenligne hvor nøyaktig og riktig de to gruppene klarer å identifisere emosjoner hos andre i bilder i forhold til fasiten. Siden målet med oppgaven er å undersøke hvor nøyaktig og riktig MC-API svarer i forhold til deltagerne, er det ikke tatt høyde for noen feil svar kan være mer ”riktige” enn andre. For eksempel kan avsky og forakt kan flyte litt over i hverandre, og forakt kan potensielt være en mer riktig vurdering av et bilde som viser avsky enn det glad ville vært. Det kunne potensielt gitt spennende informasjon, men denne informasjonene har oppgaven ikke tatt høyde for i hypotesene som er satt. Dermed kommer oppgaven heller ikke til å undersøke dette mer inngående.

Prosedyre

Denne masteroppgaven presenterer et kvasi eksperimentelt design, med et gruppe post test studie (One group post test only study). Dette er en type design der gruppene blir utsatt for en intervensjon (bilder av emosjonene som ble presenter), og målingen som gjøres måler resultatet av denne intervensjonen. I dette designet får man ikke i like høy grad som ved for eksempel et pre-post test design tatt høyde feilvarians grunnet tredjevariabler eller noe annet (Cook, Campbell & Shadish, 2002). Innsamling av data i denne studien er designet slik at det er mulig å samle inn data fra flere deltagere samtidig. Argumentet for å gjøre det slik er at det blir lettere for deltagerne å delta i studien og at antallet deltagere derfor blir høyere. Eksperimentet tok ca. 30 minutter når det ble designet slik at kun én person kunne gjennomføre studien om gangen. For å få et tilstrekkelig utvalg med et slikt design, ville det tatt mye tid og ressurser fra både undertegnede og deltagerne. Resultatet hadde dermed blitt at man i oppgaven hadde stått i fare for å få et veldig begrenset utvalg. Ved å samle inn fra flere respondenter på samme tid, vil flere respondenter være villige til å delta i studien og resultatet

blir et større utvalg i studien. Det ferdige studiet tok ca. 35 til 40 minutter per gruppe, i stedet for ca. 30 minutter per person.

De 115 deltagerresponsene ble samlet inn ved 4 forskjellige anledninger. For å holde forholdene konstant fikk alle respondentene identiske skjemaer for bedømming av bilder, spørreskjemaer for utdypende informasjon og informasjons- og samtykkeskjema. Ved alle tilfellene ble deltagerne informert om at det var frivillig å delta i studien, før de fikk informasjon om MUN-LIGHT prosjektet. Etter dette fikk alle deltagerne utdelt et informasjons- og samtykkeskjema (Se vedlegg nr.1), der deltagerne fikk konkret skriftlig informasjon om hva studien gikk ut på.

Før selve studien fikk alle respondentene et skjema for og vurdere bildene (se vedlegg nr.3) og et skjema for å samle inn mer utdypende demografisk informasjon som kjønn, alder osv. (Se vedlegg 2 for mer informasjon). All informasjon som ble samlet inn, ble samlet inn anonymt. Dette ble gjort ved å gi alle skjemaene som ble brukt til å evaluere bilder/videoer et tilfeldig tall mellom 1 og 200(ingen av skjemaene hadde samme ID-nøkkel). Denne ID-nøkkelen ble alle respondentene bedt om å skrive på skjemaet som samlet inn demografisk informasjon, slik at tallene på de to skjemaene korresponderte. Alle respondenter som ikke hadde skjemaer med to korresponderende ID-Nøkler (identifikasjon av bilder og spørreskjema), ble fjernet fra studien.

For å unngå misforståelser og uklarheter om evaluering av bilder eller besvarelser av spørreskjemaet, ble det forklart hvordan skjemaene skulle fylles ut. Dette ble ansett som nødvendig siden piloten som ble gjennomført før studien, fant at noen av spørsmålene og skalaene ikke var intuitive nok.

For å standardisere eksperimentet ble hvert bilde vist i 20 sekunder i Google Drive. Siden data-systemene var forskjellige ved de ulike lokasjonene dataene ble samlet inn fra, var dette den enkleste måten å standardisere fremvisningen av bildene i studien. Spesielt siden filene lar seg vises på samme måte uansett om man må koble til en egen data, eller bruker de tekniske systemene som allerede ligger på lokasjonen der man samler inn data (for eksempel skolemaskiner eller egne datasystemer for forelesningssaler).

All informasjon som ble samlet inn, ble samlet inn på papir for så å bli kodet inn i SPSS manuelt. Siden det kreves at alle deltagerne har tilgang på en data for å samle inn data direkte på en teknisk plattform, ble det vurdert mer effektivt å samle inn på papir, for så å legge tallene inn i SPSS i ettertid.

Analyse

Analysene som ble gjennomført i oppgaven ble gjort ved å bruke en "Independent Sample t-test". Bordens og Abbott (2011) beskriver t-testen som en test man kan bruke når den uavhengige variabelen (variabelen man er interessert i virkningen av) kun består av to nivåer (deltager og MC-API i denne oppgaven). Independent sample t-test kan brukes for å fastslå om det er en statistisk signifikant forskjell mellom to grupper.

Denne formen for analyse faller innenfor slutningsstatistikk, og omtales som en variansanalyse eller ANOVA (Analysis of Variance). En variansanalyse er en statistisk test man bruker når man ønsker å sammenligne gjennomsnittet to eller flere grupper, med kun en uavhengig variabel (Field, 2013). I denne oppgaven vil den uavhengige variabelen være bildene som presenteres og den uavhengige variabelen være treffsikkerheten til gruppene MC-API og deltagerne.

Når det undersøkes hvorvidt MC-API og deltagerne oppfatter det samme når de ser et bilde av en emosjon, vil en signifikant forskjell mellom gruppene tyde på at det er en forskjell i gruppene sin evne til å riktig identifisere en emosjon. Derfor må man vurdere det slik at de MC-API og deltager er så forskjellige i sin evne til å identifisere emosjoner, at man ikke kan beholde nullhypotesen.

I denne oppgaven lar det seg ikke gjøre å bruke en form for t-test som forutsetter at begge gruppene har indre varians. Dette er nødvendig for å regne ut både standardavvik, varians og annen informasjon man trenger for å gjennomføre en t-test mellom to grupper. I studien oppfylder kun gruppen med deltager dette kravet, siden svarene til MC-API er identiske ved flere gjennomganger. Dette resultatet er å forvente av en algoritme som bruker mønstergjenkjenning til å identifisere emosjoner, men siden man alltid får det samme svaret er det ikke mulig å gjøre en analyse der man går ut fra at begge gruppene har indre varians.

Løsningen er å gjøre en t-test mot faste skårer. Dette er en test man vanligvis bruker når man skal undersøke om en gruppe er signifikant forskjellig fra gruppegjennomsnittet i en populasjon, uten å vite annet enn gjennomsnittet i populasjonen man sammenligner mot (Bordens og Abbott, 2011). Et eksempel på dette kan være om man ønsker å sammenligne vold og utagering på en bestemt psykiatrisk sykehus, når man vet at gjennomsnittet på de andre psykiatriske sykehusene i landet er 14 per måned. Ved å bruke denne formen for t-testing kan man gjennomføre en t-test ved å sammenligne én gruppe med varians mot en gruppe som bare har et satt gjennomsnitt.

Etikk

I oppgaven ble det samlet inn demografisk data fra alle deltagerne i studien. Dette er blant annet informasjon som alder, kjønn og erfaring i helsetjenesten (Se vedlegg 2 for mer informasjon). All data ble samlet inn anonymt og dataene kan dermed ikke identifisere noen av deltagerne i studien. Det ble ikke samlet inn noen form for sensitive persondata eller helsedata i denne studien. Derfor var det ikke nødvendig å sende en søknad til Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK) for denne studien. Prosjektet ble meldt inn til Norsk senter for forskningsdata (NSD) og godkjent under prosjektnummer 56414. Siden ingen av deltagerne kan knyttes til svarene i studien og det ikke ble samlet inn helsedata eller personsensitive opplysninger er det rimelig å anta at dataene ikke har en negativ påkjenning for noen av deltagerne. Før deltagerne tok studien måtte alle deltagerne avgi et informert og frivillig samtykke om å delta i studien. Dette vil si at deltagerne fikk informasjon om gjennomføringen av studiet samtidig som de ble informert om at de uten å oppgi grunn hadde mulighet til å trekke seg fra studien. For oppbevare dataene som ble samlet inn på en sikker måte ble dataene lagret på UiO sine servere. Det er dermed ingen grunn til å anta at det er noen etiske problemstillinger ved innsamling og gjennomføring av dette studiet.

Resultat

Av de 115 deltagerne i studien var 73 av deltagerne som tilsvarer 63% av total N studenter, mens 42 av deltagerne (37%) er arbeidstagere i Stange kommune. I utvalget er kjønnsfordelingen slik at 74 av deltagerne (64%) er kvinner og 41 av deltagerne (36%) deltagerne er menn.

Deskriptiv statistikk

Tabell 1 viser de deskriptive dataene for hvordan MC-API og deltagerne svarer for hver av de 8 emosjonene som ble presentert i studien.

I tabellen forteller gjennomsnitt hvor riktig og nøyaktig hver av gruppene identifiserer hver av emosjonene som presenteres. Emosjonene glad, overrasket, redd, avsky, forakt, sinne og nøytral består av 2 bilder, mens emosjonen trist består av 3. Ut fra tabell 1 kommer det frem at MC-API treffer veldig godt på emosjonene glad og nøytral, med 100% riktig identifikasjon av emosjonene. Derimot klarer ikke MC-API å identifisere noen av de andre emosjonene og treffer i 0 % av tilfellene for emosjonene overrasket, trist, redd, avsky, forakt og sint.

Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

Emosjon	Glad	Overrasket	Trist	Redd	Avsky	Forakt	Sinne	Nøytral
Fasit								
<i>Gjennomsnitt</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<i>N</i>	1	1	1	1	1	1	1	1
<i>Std. avvik</i>
Deltager								
<i>Gjennomsnitt</i>	.991	.570	.725	.352	.465	.361	.339	.830
<i>N</i>	115	115	115	115	115	115	115	115
<i>Std. avvik</i>	.066	.310	.280	.296	.342	.328	.285	.288
MC-API								
<i>Gjennomsnitt</i>	1.00	.000	.000	.000	.000	.000	.000	1.00
<i>N</i>	1	1	1	1	1	1	1	1
<i>Std. avvik</i>

Tabell 1: Deskriptiv statistikk for MC-API og Deltagerne for alle emosjonene. Utdrag av deskriptive data og gjennomsnittlig treff på bilde identifikasjon hentet fra SPSS.

For deltagerne ser man i tabell 1 at den gjennomsnittlige treffprosenten for deltagerne er forskjellig fra emosjon til emosjon. Deltagerne treffer nøyaktig og riktig for identifisering av glad med en gjennomsnittlig treffprosent på .991. Treffprosenten er noe lavere for deltagergruppen for nøytrale ansiktsemosjoner, der deltagerne har en gjennomsnittlig treff på .830. Selv om deltagerne treffer noe dårligere på emosjonene glad og nøytral, treffer deltagerne gjennomsnittlig mer nøyaktig på de andre emosjonene MC-API ikke klarer å identifisere.

Deltagerne gjenkjenner overrasket og trist i godt over .50 av tilfellene i studien. Avsky har en treffrate rett under .50 med en gjennomsnittlig treff på .465. For bildene som viser emosjonene redsel, forakt og sinne (henholdsvis .352, .361 og .339), viser disse en treff rate som er mer enn tre ganger så høy som ved ren tilfeldighet på .111. Når man ser nærmere std.avvik for emosjonene varierer tallene veldig mye. Glad har et lavt std.avvik på .066, noe som tyder på at deltagerne svarer veldig likt. For emosjonen nøytral, er det en høy treffrate, men standardavviket er også ganske høy. Dette kan tyde på at selv om man har en ganske høy treff på disse bildene, er det det stor forskjell på hvordan hver deltager svarer i forhold til hverandre. Selv om det er vanskelig å si noe konkret om dataene basert bare på denne informasjonen, gjør denne informasjonene at det kan være interessant å undersøke hvert bilde for seg selv. I tabell 2 presenteres derfor hvert bilde for seg selv for å undersøke om det er stor forskjell på treffprosenten mellom bildene innad i hver emosjon.

Tabell 2 viser at det er ganske stor variasjon mellom noen av bildene innad i hver emosjon. Dette er et faktum spesielt for bildene i emosjonen redd og sinne. Bilde 40 (redd)

Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

har en gjennomsnittlig treffprosent som tilsvarer ren gjetting på .11 mens bilde nummer 42 har en gjennomsnittlig treffprosent på .59. Denne tendensen finner man også for bildene for emosjonen sinne der bilde 83 har gjennomsnittlig treff på .08 mens bilde 52 har .60. Flere grunner kan være med å forklare hvorfor dette er tilfellet og noen av dem vil bli undersøkt nærmere i diskusjonsdelen av oppgaven. Ser man på tabell 2 er det likevel ingen tvil om at det er stor variasjon i deltageres nøyaktige identifikasjon av bildene innad i emosjonene som presenteres.

Fra tabell 1 kommer overordnet fram at gjennomsnittlig treff for MC-API og deltagerne er høy for emosjonene glede og nøytral. De andre emosjonene derimot klarer ikke MC-API å identifisere med en treffrate på 0%. For deltagerne derimot viser tabell 1 at alle emosjonene uavhengig av enkeltbilder med lav treffrate, gjenkjennes godt over raten for ren gjetning. Dette er ikke konkluderende data og en statistisk analyse blir likevel nødvendig for å undersøke om det er en signifikant forskjell mellom MC-API og deltageres evne til å identifisere emosjoner hos andre.

Bilde:	<i>Glad</i>		<i>Overrasket</i>		<i>Trist</i>			<i>Redd</i>		<i>Avsky</i>		<i>Forakt</i>		<i>Sinne</i>		<i>Nøytral</i>	
	10	12	28	33	13	19	21	40	42	57	76	73	78	52	83	88	90
Fasit																	
<i>Gjennomsnitt</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<i>N</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Std. avvik</i>
Deltager																	
<i>Gjennomsnitt</i>	1.00	.98	.31	.83	.62	.73	.83	.11	.59	.30	.63	.43	.29	.60	.08	.85	.81
<i>N</i>	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115	115
<i>Std. avvik</i>	.000	.131	.466	.381	.488	.446	.381	.318	.494	.462	.486	.498	.454	.492	.270	.356	.395
MC-API																	
<i>Gjennomsnitt</i>	1.00	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	1.00	1.00
<i>N</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>Std. avvik</i>

Tabell 2: Deskriptiv statistikk for MC-API og deltagerne, alle bildene som ble tatt med i studien. Utdrag av deskriptive data og gjennomsnittlig treff på bilde identifikasjon hentet fra SPSS.

Slutningsstatistikk

For å undersøke hypotesene blir det gjennomført det to t-tester i denne oppgaven. I den første blir deltagerne sammenlignet med MC-API. Den andre t-testen undersøker om det

er signifikante forskjeller mellom fasiten og deltagerne sin evne til å identifisere emosjoner. For å undersøke fasiten og MC-API er det ikke mulig å gjøre en t-test siden ingen av gruppene har en in-gruppe varians. For å fastslå om det er en god overenstemmelse mellom MC-API og fasiten, vil oppgaven gjøre en visuell inspeksjon av de deskriptive dataene for å sammenligne disse gruppene.

MC-API vs. deltagerne. I tabell 3 presenteres resultatene av t-testen for sammenligningen av deltagerne i studien og MC-API.

	<i>t</i>	<i>df</i>	<i>Sig. (2-tailed)</i>	<i>Gjennomsnitts varians</i>	<i>95% Konfidensintervall</i>	
					<i>Øvre</i>	<i>Nedre</i>
Glad	-1.420	114	.158	-.0087	-.021	.003
Overrasket	19.714	114	.000	.570	.512	.627
Trist	27.739	114	.000	.725	.673	.776
Redd	12.760	114	.000	.352	.298	.407
Avsky	14.573	114	.000	.465	.402	.529
Forakt	11.794	114	.000	.361	.300	.422
Sint	12.751	114	.000	.339	.286	.392
Nøytral	-6.318	114	.000	-1.696	-.223	-.116

Tabell 3: *t*-test MC-API og deltagerne

I tabell 3 ser man at det er klare signifikante forskjeller mellom MC-API og deltagerne for alle emosjonene utenom glad. Dermed kan man ikke gå ut fra at gruppeforskjellen for riktig og nøyaktig identifisering av emosjoner skyldes tilfeldigheter, men en faktisk forskjell i de to gruppene sin evnen til å identifisere ansiktsemosjoner for alle emosjoner utenom glad. Siden studien ble designet slik at 1 er treff og 0 er bom på hvilken emosjon som blir vist i de ulike bildene, kan man ved å se på *t* verdiene fastslå hvilken gruppe som har den høyeste treffraten.

En *t*-verdi på -.1420 for emosjonen glad i tabell 3 vil tilsi at deltagerne i studien gjorde færre riktige vurderinger av bildene enn MC-API. I tabell 3 viser *t*-verdiene at deltageren gjorde færre riktige vurderinger av bildene for glad og nøytral emosjonene, men at de flere riktige identifiseringer av de resterende emosjonene.

Siden tabell 3 viser at det ikke er en signifikant forskjell mellom gruppene for identifisering av emosjonen glad, kan ikke forskjellen mellom de to gruppene tilskrives gruppens faktiske evne til å identifisere ansiktsemosjonen. For de resterende emosjonene tilsier tabell 3 at man må anta MC-API og deltagerne sin evne til identifisere emosjoner

Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

nøyaktig er ulik. Med signifikante verdier på $p < 0.01$ for de signifikante emosjonene kan man med en høy grad av sikkerhet si at dette er et resultat som ikke er grunnet tilfeldigheter.

For emosjonene trist, overrasket, redd, avsky, forakt og sinne som har en positiv t-verdi (Eksempelvis forakt t. 11.794) viser dataene at deltagerne er signifikant bedre enn MC-API. Derimot for emosjonen nøytral som har en negativ t-verdi på -6.318, er deltagerne signifikant dårligere enn MC-API. Dette underbygges av konfidensintervallene for de 8 emosjonene som blir presentert i tabell 3.

Deltagere vs. Fasit. I tabell 4 presenteres resultatene for t-testen for sammenligningen av de uavhengige raterne (fasiten) og deltagerne i studien.

	<i>t</i>	<i>df</i>	<i>Sig. (2-tailed)</i>	<i>Gjennomsnittss varians</i>	<i>95% konfidens intervall</i>	
					<i>Øvre</i>	<i>Nedre</i>
Glad	-1.420	114	.158	-.0087	-.021	.003
Overrasket	-14.898	114	.000	-.4304	-.488	-.373
Trist	-10.541	114	.000	-.2754	-.327	-.224
Redd	-23.472	114	.000	-.6478	-.703	-.593
Avsky	-16.752	114	.000	-.5348	-.598	-.472
Forakt	-20.888	114	.000	-.6391	-.700	-.579
Sint	-24.848	114	.000	-.6609	-.714	-.608
Nøytral	-6.318	114	.000	-.1696	-.223	-.116

Tabell 4: t-test Deltagere og fasit

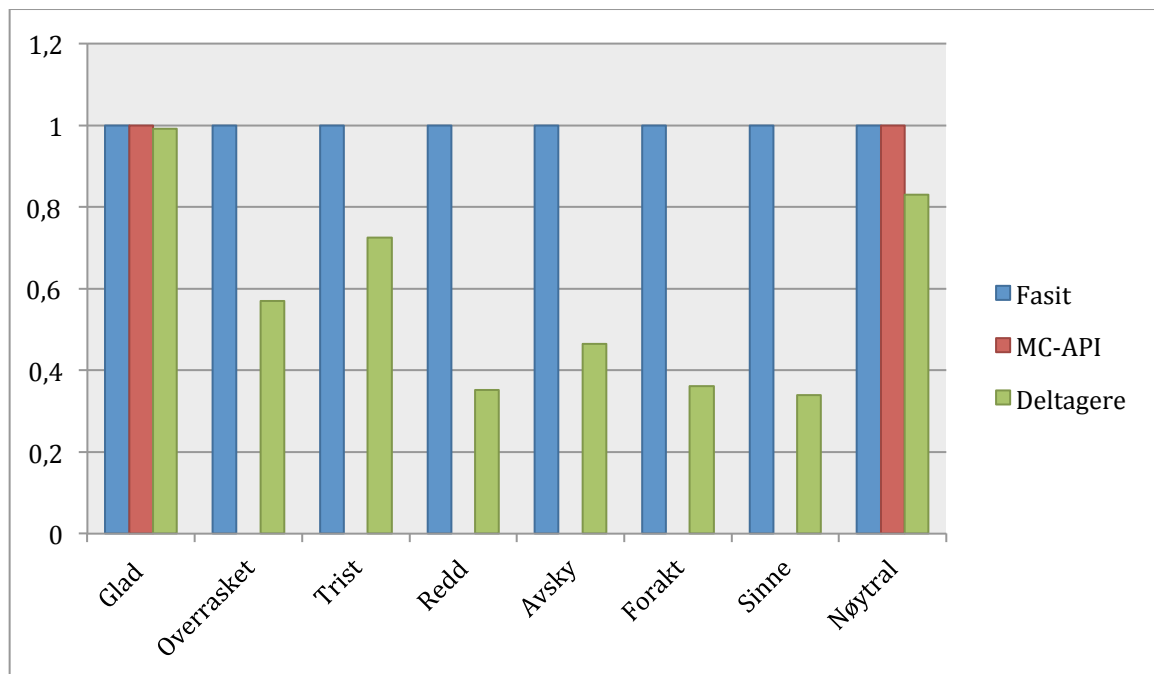
Resultatene som presenteres i tabell 4 viser på samme måte som tabell 3 at det ikke er noen signifikant forskjell mellom deltagerne og fasiten sin evne til å identifisere emosjonen glad. Siden de uavhengige raterne er brukt som fasit i denne oppgaven, vil det vanskelig la seg gjøre å svare mer korrekt enn dem. Derfor er alle t verdiene som presenteres i tabell 4 negative og viser at deltagerne gjør færre riktige vurderinger enn fasiten for alle emosjoner. For emosjonen glad er det ikke en signifikant forskjell mellom fasiten og deltagerne i studien. Forskjellen for identifisering av emosjonen glad kan dermed ikke forklares med en klar gruppeforskjell, derfor må man konkludere med at forskjellen mellom gruppene skyldes noe annet enn gruppenes evne til å identifisere emosjonen glad.

For de andre emosjonene derimot (overrasket, trist, redd, avsky, forakt, sinne og nøytral) er gruppene signifikant forskjellige. Alle emosjonene er signifikant forskjellige med en $p < 0.01$ og man må derfor ut fra dataene konkludere med at deltagerne i studien er signifikant dårligere til identifisere de andre emosjonene i studien. P-verdiene (Sig.) som presenteres i tabell 4 underbygges av resultatene man ser i konfidensintervallet som

presenteres for hver av emosjonene.

Siden studien er designet slik den er, kan dette resultatet vel så godt beskrives som en metodisk svakhet i designet som at det faktisk er en signifikant forskjell mellom de to gruppene. Dette er en problematikk oppgaven kommer til å diskutere nærmere i diskusjonsdelen av oppgaven.

MC-API vs. Fasit. Etter en visuell inspeksjon av de deskriptive dataene i tabell 1 kommer det frem at MC-API treffer veldig godt på emosjonene glad og nøytral. MC-API klarer derimot ikke å identifisere noen av de andre emosjonene som blir presentert på bildene i studien. For en visuell representasjon av tabell 1, se modell 1:



Dette resultatet ble fastslått ved å følge Microsoft Cognitive API sin instruksjon om hvordan man skal tolke resultatet av vurderingen til MC-API (Microsoft, 2017). I følge Microsoft skal man skal man vurdere outputen slik at den emosjonen som får den høyeste skåren mellom 0 og 1 av de 8 emosjonene blir svaret. Når man følger denne instruksjonen for identifisering av bildene er det 100% overenstemmelse for emosjonene glad og nøytral, som tyder på at MC-API klarer å identifisere disse emosjonene riktig og nøyaktig. Derimot klarer ikke MC-API å skille noen av de andre emosjonene fra hverandre og kategoriserer alle de andre bildene som enten nøytrale eller glade. I svarene til MC-API kommer det fram av 17

bilder at alle blir identifisert som nøytrale, med unntak av bilde 10, 12, 33 og 57 som ble identifiserte som glad.

Etter dette å dømme er MC-API kun nøyaktig for gjenkjenning av emosjonene glad og nøytral, hvis man leser MC-API sine responser slik Microsoft instruksjoner sier.

Diskusjon

Oppsummering av data

Formålet med denne studien er å undersøke om programmet MC-API kan identifisere emosjoner like nøyaktig som individer gjør. Studien har belyst dette ved å analysere spørredata fra deltagerne og MC-API. For å oppsummere resultatene kommer oppgaven undersøke om resultatene i studien finner støtte for de 3 hypotesene eller tyder på at hypotesene bør forkastes (H1,H2 og H3).

For hypotese 1 (H1) som undersøker om MC-API og deltagerne identifiserer emosjoner like riktig og nøyaktig i bilder, fant studien at de to gruppene er signifikante forskjellige for alle emosjoner utenom glad. Dette vil si at studien finner støtte for H1 i emosjonen glad, men ingen av de andre emosjonene. Overordnet finner oppgaven ikke støtte for H1 i denne studien, og hypotesen må forkastes.

For H2 som sammenligner deltagerne og fasiten sin evne til riktig og nøyaktig identifiserer andres emosjoner i bilder finner studien det samme resultatet. Det er signifikant forskjell for alle emosjonene som ble presentert i studien, med unntak av emosjonen glad. Overordnet finner oppgaven ikke støtte for H2, og hypotesen forkastes.

For H3 som sammenligner MC-API og fasiten sin evne til å identifiserer andres emosjoner riktig og nøyaktig i bilder, er resultatene litt annerledes. Det var ikke mulig å gjennomføre en t-test for å undersøke H3, siden ingen av gruppene som ble sammenlignet har in-gruppe variasjon. Etter en visuell inspeksjon av de deskriptive statistikken kommer det frem at fasiten og MC-API er like gode til å identifisere emosjonene nøytral og glede. Studien finner derimot klare forskjeller mellom gruppene for emosjonene overrasket, trist, redd, avsky, forakt og sinne. Overordnet finner studien ikke støtte for H3, og hypotesen forkastes.

Dermed forkastes alle hypotesene og man kan konkludere med at det er en forskjell i deltagerne, MC-API og fasiten sin evne til å identifisere emosjoner hos andre.

I den neste delen av oppgaven vil MC-API sin prestasjon i forhold til fasit og deltagere diskuteres nærmere. Oppgaven vil se nærmere på hva dette vil si for bruken

av teknologi som dette i MUN-Light prosjektet, og bruken av denne teknologien ellers. Siden det ikke har vært mulig å samle inn data som sier noe om hvordan andre programmer som leser emosjoner presterer, vil diskusjonen som presenteres i oppgaven kun bruke MC-API sine responser som utgangspunkt. Etter dette kommer oppgaven til å diskutere kort hvordan resultatene av dette studiet påvirker fagfeltet på teknologi som identifiserer emosjoner, før studiens begrensninger diskuteres nærmere. Tilslutt vil oppgaven komme med noen forslag til videre forskning på området, før oppgaven kommer med en konklusjon av studien som er blitt gjennomført.

Hvordan svarer gruppene

Selv om studien ikke fant støtte for noen av hypotesene som ble presentert, ligger det mye informasjon i svarene som ble avgitt i studien. For svare på om bruken av MC-API kan være et verktøy som kan øke kvalitet av opplæring, kommer oppgaven til å se nærmere på dataene som ble samlet inn. Svarene til deltagerne i studien og svarene til MC-API vil bli diskutert nærmere, ved å gå inn i den deskriptive statistikken av oppgaven. Først vil svarene til deltagerne undersøkes nærmere, før svarene MC-API blir undersøkt mer inngående.

Deltagerne. I resultatdelen av oppgaven kommer det frem at det er signifikante forskjeller mellom gruppene for alle emosjonene med unntak av glad for deltagerne og MC-API og deltagerne og fasiten. For fasiten og MC-API svarte de like nøyaktig for glad og nøytral, men MC-API klarte ikke å identifisere noen av de andre emosjonene. Dette førte til at alle hypotesene måtte forkastes, og man må regne med at verken MC-API eller deltagerne svarer på samme måte som fasiten gjør. Hvis man ser nærmere på den deskriptive statistikken derimot er det tydelig at deltagerne svarer godt over ren sjanse på alle emosjonene som er blitt presentert. I studien ble alle deltagerne og MC-API presentert med 9 mulige valg. Det vil si at man ved ren gjetning skal ha en treffprosent på 11%. I tabell 1 vil dette tilsi at gjennomsnittet for alle følelsene bør være .11, hvis resultatet er et resultat av ren gjetning. Ser man på hvordan deltagerne i studien svarer er ikke dette tilfellet. Alle emosjonene fikk over 3 ganger så høy gjennomsnittlig treff, med redd som har lavest treff på .352. Denne treffprosenten er såpass mye høyere enn ren gjetning at det er vanskelig å forklare disse resultatene ved ren sjanse.

Selv om dette er godt over ren sjanse, er en identifiseringsrate på mellom .30 og

.40 relativt lav. Det kan være mange grunner til at man finner signifikante forskjeller for en så stor del av emosjonene som blir presentert. I denne oppgaven blir spesielt to av disse grunnene diskutere nærmere. Den første grunnen er at noen av bildene kan være dårlige og gir vage representasjoner av den emosjonen man ønsker å vise. Det kan derfor argumenteres for at disse bildene ikke burde vært en del av studien i utgangspunktet. Denne problematikken vil diskuteres nærmere i kapittelet om begrensninger ved denne studien.

Den andre grunnen kan være knyttet til problematikken ved å ha et satt utvalg uten varians å sammenligne deltagerne mot. På et teoretisk plan er dette en form for t-test man bruker når man sammenligner en gruppe med in-gruppe varians mot et gjennomsnitt. Hadde dette vært et gjennomsnitt der skårene kunne samlet seg både over og under gjennomsnittet, hadde utgangspunktet vært et annet. Når man måler for eksempel IQ i gruppe med 115 deltagere, vil noen deltagere ha 120 i IQ og noen med 80. Hvis man bare vet at snittet i populasjonen er 100 vil en t-test mot et satt utvalg likevel kunne være en beskrivende analyse. For utvalget i denne oppgaven kan ikke utvalget befinne seg over fasiten (gjennomsnittet) siden fasiten alltid vil være 100% riktig. Dette gjør at alle svar som ikke er veldig nære 1.00 i gjennomsnittlig treff, gir resultater som ikke er signifikante. Et eksempel som illustrere dette godt i tabell 4 er nøytral. Det er en signifikant forskjell mellom fasiten og deltagerne sin evne til å identifisere emosjonen nøytral. Hvis man ser på hvor ofte de 115 deltagerne identifiserer nøytrale uttrykk skjer dette med en gjennomsnittlig treff på .830. Dette vil si at de 115 deltagerne identifiserer 2 nøytrale bilder i 83% av tilfellene. Dermed identifiserer deltagerne emosjonen nøytral korrekt i 191 ganger av 230 mulige men er likevel signifikant dårligere enn fasiten til å identifisere emosjonen nøytral. Av denne grunnen er det viktig å se på gjennomsnittlige treffandelen også. Riktig identifikasjon av nøytral i 83% av bildene, med et utvalg med 115 deltagere er så høyt at man må argumentere for at deltagere er presise i sin identifikasjon av emosjonen nøytral, selv om de er signifikant dårligere enn fasiten. Selv om dette er den emosjonen der man ser signifikante forskjeller mellom fasit og deltager som har høyest gjennomsnittlig treffprosent, ser man samme tendensene av de deskriptive dataene for de andre emosjonene også. Spesielt siden alle emosjonene identifiseres godt over ren sjanse. Derfor blir det også spesielt viktig å diskutere de deskriptive dataene i denne studien også mer inngående.

MC-API. Som nevnt i metodekapittelet svarer MC-API alltid likt, uansett hvor mange ganger du viser et bilde. Dette vil si at uansett hvor mange ganger man kjører programmet vil den svare uten å få indre varians i utvalget. For det første gjør det at man ikke kan bruke en t-test for å sammenligne MC-API med fasiten siden ingen av gruppene har indre varians. Det vil det også si at man ut fra bildene presentert i denne studien alltid vil få 100% treff på emosjonene glad og nøytral, og at MC-PAI alltid vil få feilidentifikasjon av de andre emosjonene som presenteres i studien. MC-API er dermed ut fra denne studien bedre til å identifisere glad og nøytral enn deltagerne. Derimot klarer MC-API ikke å identifisere emosjonene overrasket, trist, redd, avsky, forakt og sint.

Ser man nærmere på svarene finner man også at MC-API kun benytter seg av 2 av de 8 mulige svaralternativer. Ved 4 anledninger i studien identifiserer MC-API emosjonen i bildene som glad (Bilde 10,12, 33 og 57), mens resten av bildene ble identifisert som nøytrale. Dette til tross for at for eksempel bilde 21 ble riktig identifisert av deltagerne med en treffrate på .83, ble dette bildet skåret som nøytralt av MC-API. Dette tyder på at selv for bilder som i høy grad identifiseres riktig av deltagerne, vil ikke MC-API klare å gjøre en korrekt vurdering hvis bildet inneholder andre emosjoner enn glad eller nøytral.

Tar man dette i betraktning, blir det vanskelig å trekke kvaliteten av bildene alene som grunn til at MC-API ikke klarer å identifisere emosjonen i bildene riktig. Ut fra svarene MC-API har avgitt kan det virke som om programmet er spesielt god til å gjenkjenne spesielt to emosjoner, mens den sliter med å gjenkjenne de andre emosjonene.

Hva vil dette si? Hva tyder egentlig disse resultatene på? Dette tyder på at deltagerne som mennesker ellers, har en evne til å identifisere emosjoner hos andre. Dette støtter Ekman og Freisen (1986) sin tanke om at det finnes universale emosjoner som mennesker har en evne til å identifisere. Resultatene av studien tyder også på at det er mulig for ulike programmer som bruker algoritmer og maskinlæring slik MC-API gjør å identifisere emosjoner hos andre. Derimot når studien sammenlignet deltagerne mot MC-API kommer det frem at MC-API ikke er optimalisert enda. Spesielt siden programmet slik det er i dag ikke klarer å identifisere andre emosjoner enn glad og nøytral.

Selv om man kan trekke designet av denne studien i spørsmål, er det likevel slik at emosjoner som av deltagerne identifiseres i over 80% av tilfellene ikke identifiseres riktig av MC-API. Hvordan man ønsker å bruke programmet er selvsagt avgjørende for om programmet lar seg bruke slik det er i dag, slik oppgaven vil diskutere nærmere i diskusjonsdelen. Det er likevel slik at studien som har blitt gjennomført, tydelig konkluderer med at MC-API identifiserer emosjoner i bilder dårligere enn det deltagerne gjør. Unntaket for dette gjelder for emosjonen nøytral der MC-API identifiserer nøytral riktig i 100% av tilfellene mot 83% for deltagerne.

Selv om man ikke kan skylde på tilfeldighet når det kommer til en algoritme, er det ikke å stikke under en stol at 13 av 17 bilder (76%) identifiseres som nøytrale. Dette tyder på at MC-API er mer utsatt for å velge å indentifisere bilder som nøytrale. Det er vanskelig å tolke om responsen nøytral blir gjort fordi den identifiserer bildet som nøytralt, eller at algoritmen ikke finner en annen emosjon i bildet og derfor er programmert til å svare nøytral.

Praktisk implikasjon

Det neste spørsmålet som er naturlig å stille er hvordan disse resultatene påvirker bruken av MC-API for sykehuset innlandet i MUN-Light prosjektet, og om teknologien kan brukes?

Som oppgaven har kastet lys på i metodedelen er målet med MUN-Light prosjektet å digitalisere opplæring i psykiatri og annen form for helsetjeneste. Hensikten med å gjøre dette står i tråd med formålsparagrafen i helsepersonell loven (§1, helsepersonell loven, 1999) som snakker om blant annet og øke kvalitet i helse- og omsorgstjenester. I MUN-Light prosjektet gjør man dette ved å prøve å øke kvaliteten i opplæring.

Hvordan MC-API eller lignede teknologi kan brukes, eller om det kan brukes er spørsmål som er vanskelig å svare på. Men denne studien gir informasjon som bør tas med i denne diskusjonen. Som oppgaven har forsøkt å argumentere for gjennomgående har denne typen programmer potensiale til å kunne øke kvaliteten i opplæringen som gis i norsk helsetjeneste. Man kan se for seg flere mulige bruksområder for et slikt program, spesielt når det per dags dato ikke kommer nøyaktig frem hvordan Simuleringscenteret innlandet har tenkt til å bruke teknologien. Oppgaven kommer til å presentere to mulige bruksområder, selv om bare fantasien setter grenser for hva man faktisk kan bruke teknologien til.

Den første og kanskje mest logiske vil være å bruke programmet til å skape en

interaktiv opplæring, der den emosjonelle responsen til den som trener avgjør hvordan et opplæringsscenario utarter seg. Et annet mulig bruksområde er å bruke programmer som kan identifisere følelser hos andre som en database for å dobbeltsjekke om de opplæringsscenariene som eksisterer er realistiske nok. For dette eksempelet kan teknologien brukes til å evaluere film, spill eller annet opplæringsinnhold på alle andre tekniske flater som ønsker å kommunisere ansiktsemosjoner. Som sagt er det bare fantasien som setter grenser her, men i begge tilfellene kan denne typen teknologi fungere som et verktøy for å øke kvaliteten av opplæringen i en organisasjon som HSØ.

Derfor er det ingen tvil om at teknologien kan brukes. Det er likevel med forbehold at teknologien riktig og nøyaktig klarer å identifisere emosjoner hos andre i bilder eller videoer. Selv om potensialet er der, ønsker denne oppgaven å undersøke om MC-API kan brukes slik det fungerer i dag. Slik resultatene viser klarer ikke MC-API å identifisere emosjoner hos andre like godt som deltagerne klarer (med unntak av glad og nøytral). MC-API er derfor ikke tilstrekkelig nøyaktig til å integreres for å øke kvaliteten slik det fungerer i dag. Dette vil nødvendigvis ikke si at programmet ikke kan brukes en gang i fremtiden, da det bygger på en maskinlærings algoritme og at det ligger i programmet sin natur å forbedre seg. Dermed kan MC-API i fremtiden kan bli et program som er nøyaktig nok til å kunne integreres i MUN-Light prosjektet. Men i følge studien som er blitt gjort i denne oppgaven er ikke dette tilfellet i dag, og det er derfor heller ikke et verktøy som vil kunne løfte kvaliteten av opplæringen i MUN-Light prosjektet enda.

Så hvordan påvirker dette MUN-Light? Siden MUN-Light prosjektet er tidlig i en utviklingsprosess og MC-API er en teknologi som ikke enda har blitt tatt i bruk, vil det i liten grad påvirke prosjektet. Resultatene i studien foreslår likevel at MC-API ikke integreres som en sentral del av opplæringen som gjøres på sykehuset innlandet per dags dato. Dette vil som sagt ikke si at SI skal utelukke bruken av MC-API eller lignende verktøy i fremtiden, spesielt siden potensialet er så stort som det er. Siden denne oppgaven kun har tatt for seg og testet MC-API, kan man basert på denne oppgaven ikke gjøre vurderinger eller prediksjon om nøyaktigheten av andre former for ansikts emosjons programmer. Dermed vil alle konklusjonene som blir gjort på bakgrunnen av resultatene i denne studien, kun gi informasjon om bruken av MC-API.

Teoretisk implikasjon

Hva vil dette si for forskningen på området? Målet til Microsoft er å kunne tilby teknologien i MC-API som et verktøy for å lage apper eller programmer som basert på din emosjonelle tilstand skal kunne skreddersy en respons. Helt konkret skriver Microsoft at du med Microsoft Cognitive API kan:

”Gi appene, nettstedene og bot-ene dine intelligente algoritmer for å se, høre, snakke, forstå og oversette brukernes behov gjennom naturlige kommunikasjonsmetoder. Forvandle bedriften din med AI i dag.” (Microsoft Azure, 2018)

Under alle verktøyene Microsoft tilbyr i dette programmet finner man verktøyet MC-API. For Microsoft er det av interesse at dette verktøyet fungerer godt og gjør det den skal gjøre. I følge denne studien gjør MC-API derimot ikke det. Dette er ikke nødvendigvis i alle tilfeller et problem, siden det ikke alltid er like essensielt at programmet gjør en riktig vurdering av brukeren sin emosjonelle tilstand til en hver tid. Sett at man bruker verktøyet til å generere reklame til en person som bruker app-en snapchat, er resultatet av en unøyaktig vurdering av emosjonen til et individ at man genererer reklame for et produkt individet ikke vil ha der og da. Resultatet blir kanskje at personen ikke velger å se nærmere på dette tilbudet denne gangen, men utsetter det. Det er helt klart tapt inntekt for de som ikke får solgt produktet sitt. Men den gangen MC-API gjør en riktig identifisering av emosjon, stemme eller annen form for kunnskap som skaper en forståelse av brukerens behov, vil man få opp ”riktig” reklame. Dermed vil man etter et visst antall forsøk få opp riktig reklame for brukerens behov der og da og derfor også få nytte av kunnskapen og informasjonen denne typen verktøy per dags dato har muligheten til å tilby.

Hvis man ønsker å bruke dette i et opplæringsprogram slik Sykehuset Innlandet ønsker i MUN-Light prosjektet er et uvisst antall feilvurderinger ikke tilstrekkelig, og kanskje til og med mot sin hensikt. Definisjonen av læring å endre adferd eller respons basert på tidligere tilegnet kunnskap eller erfaring (Cascio & Aguinis, 2014). Hvis denne typen teknologi blir integrert i en form for opplæring om hvordan man skal møte pasienter, og at opplæringsprogrammet genererer en respons til at du som sykepleier ser nøytral ut (som tilsvarte 76% av responsene til MC-API), kan dette gjøre overføringen vanskelig. Hvis det er slik at pasienten opplever deg som fravende og ikke interessert siden du ”er” nøytral i opplærings situasjonen, og derfor velger å utagere som et tegn på oppgitthet eller en følelse av å bli oversett kan dette være et problem. Spesielt hvis man som deltager i denne opplæringen egentlig viste en form

for tristhet eller sinne, og dette hadde vært en mer hensiktsmessig respons i den gitte situasjonen. Selv om dette et tenkt scenario som vanskelig lar seg implementere i opplæringen per dags dato, er dette et av potensialene som ligger i bruken av denne teknologien i fremtiden. Etter definisjonen av læring som Cascio & Auginis (2014) gir, er poenget å kunne bruke tidligere tilegnet kunnskap og tidligere tilegnet erfaring til å endre en respons. Denne responsen skal sikre økt kvalitet i helse- og omsorgstjenesten så vel som i andre sektorer. Hvis verktøyet i 76 % av tilfellene identifiserer nøytral og resten identifiserer glad, vil man ikke kunne bruke dette til å oppnå den erfaringen og kunnskapen man trenger for å endre respons eller øke kvaliteten av trening og opplæring.

I Denne situasjonen er riktig identifikasjon i hver emosjon er essensielt. Man kan videre argumentere for at det i større grad er viktig i opplæring enn for generering av reklame eller forslag til søkeord på Google. Spesielt siden informasjonen man får under opplæring i helse- og omsorgstjenestene skal brukes til å "lage" et riktig responsmønster, og forskjellen på respons 1 og respons 2 kan bety at en pasient ikke dør, at en pasient føler seg godt ivaretatt eller ikke går til angrep på personal eller andre pasienter.

Selv om resultatet av én studie ikke er verken konkluderende eller generaliserbart gir det fortsatt et bilde over muligheter og begrensninger ved å ta i bruk en sånn teknologi i trening og opplæring. Den teoretiske implikasjonen av denne studien blir at det må gjøres mer forskning på feltet for å sikre at MC-API klarer å identifisere de 8 emosjonene den skal identifisere. Dette bør gjøres på et generelt grunnlag for å ha et program som er av god kvalitet. Men spesielt før man kan integrere denne typen teknologi i felt som trening/ opplæring og læring.

Begrensninger med studien

Oppgaven vil belyse de 3 begrensningene ved denne oppgaven som størst. Selv om kandidaten er innforstått med at dette ikke dekker alle begrensningene i denne studien.

Innsamlingsmetodikk. Slik det ble opplyst i metoddelen ble dataen samlet inn data fra de 115 deltagerne ved 4 forskjellige anledninger. Dette vil si at i gjennomsnitt ble i underkant av 30 besvarelser samlet inn ved hver anledning. Siden denne studien var tidskrevende (ca. 30 min per pers) var dette den mest hensiktsmessige måten å gjennomføre studien på. For det første hadde kostet deltagerne i studien så store tidsmessige ressurser å delta, at det hadde vært usannsynlig å få samlet inn et tilstrekkelig utvalg hvis ikke det hadde blitt gjort slik. For det andre er det begrensninger ved tiden tilgjengelig for oppgaven, så å

Samle inne data fra én og én deltager hadde spist så mye tid at det heller ikke hadde vært mulig å samle inn tilstrekkelig mengde data fra kandidaten sin side. Etter en avveining ble det derfor bestemt det var mest hensiktsmessige å samle inn data fra flere deltagere om gangen.

Med dette sagt er det et faktum at når så mange gjennomfører studien samtidig, kan dette kan føre til at folk snakker sammen og avtaler svar. Selv om det eksplisitt ble sagt at man skulle svare individuelt, kan man ikke ta høyde for dette statistisk i dataene. Selv om dette i seg selv ikke sier noe om man identifiserer bildene mer eller mindre riktig, er dette en skjevhet i dataene oppgaven ikke klarer å ta høyde for.

FACS. Den andre store begrensningen med studien som ble presentert i denne oppgaven er at det ikke ble brukt sertifiserte FACS spesialister til å evaluere bildene som ble presentert i studien. Dette ble løst ved å la 4 uavhengige respondenter som ikke var tilknyttet studien gjøre en vurdering av 96 bilder slik at 17 bilder av 8 emosjoner sto igjen. Å sette responsen til de 4 som fasit, er i seg selv en begrensning ved oppgaven spesielt siden de ikke har noen form for tidligere erfaring med den allerede godt etablerte forskningen på universale emosjoner. For å få et best mulig utvalg av bilder hadde det vært optimalt om disse bildene ble evaluert i tråd med forskningen og på bakgrunn av det man allerede vet.

Bildene. Som en mulig forlengelse av dette var det noen av bildene som var vanskelig å identifisere. Som man ser i tabell 2 er to av bildene så vanskelig å identifisere at et de gjenkjennes ved ren gjetning eller dårligere (Bilde 40 med .11 og Bilde 83 med .08). Selv om dette er de to desidert vanskeligste bildene å identifisere riktig, er også noen av de andre bildene vanskelige å identifisere med en korrekt identifiseringsrate på .30. Dette er en begrensning med oppgaven som kunne vært unngått ved bruk av en bedre seleksjonsprosess for bildene som ble brukt. Denne skjevheten opptrer som en klar begrensning ved studien.

Det er helt klart noen tydelige metodiske svakheter i studien i denne oppgaven. Det var ikke praktisk gjennomførbart i denne oppgaven å gjøre noe med disse begrensningene og derfor ble resultatet av studien slik den ble. Oppgaven oppfordrer likevel andre som gjør forskning på området, til å undersøke hypotesene som er satt i denne studien, men å ta høyde for og utbedre studien etter begrensningene ved denne oppgaven.

Videre forskning

Det var ikke praktisk mulig å bruke sertifiserte FACS spesialister til å evaluere og velge ut bildene som ble presentert i denne studien. Dette er en svakhet siden oppgaven ikke

hadde et optimalt design for å bygge på den kunnskapen man allerede har om universale emosjoner. I oppgaven hadde noen av enkeltbilder veldig lav treffrate, som kan være problematisk for resultatene av studien og et direkte resultat av evalueringsmetoden som ble brukt for å inkludere bildematerialet i studien. I videre forskning på området bør det brukes spesialister innen FACS rammeverket for å evaluere bildene, slik at utvelgelsen av bildemateriale er i tråd med tidligere forskning og kan sikre høy kvalitet på bildene som presenteres.

Selv om denne oppgaven kun var interessert i hvorvidt deltagerne og MC-API identifiserte emosjoner nøyaktig og riktig på det bildene som ble presentert, hadde det vært interessant å undersøke om ulike grupper har en tendens til å identifisere forskjellige emosjoner. Her kan man for eksempel undersøke om arbeidstakere innen psykiatri har en tendens til å identifisere en emosjon oftere enn en ansatt ved en somatisk avdeling ved et sykehus vil.

En annen form for spørsmål denne studien ikke kunne undersøke er hvorvidt forskjellen mellom MC-API og deltagerne i studien er annerledes for video enn for bilder. Ekman og Friesen (1976) presenterer et sterkt argument for å undersøke dette. Forfatterne argumenterer for at de 44 BO som identifiseres i FACS kun sier noe om hvordan ulike muskler i ansiktet kan bevege seg. Det er forøvrig slik at et nøytralt avslappet ansikt er utgangspunktet for å kunne gjenkjenne disse bevegelsene. Det vil si at man for best mulig gjenkjenning vil være optimalt å kjenne det ”nøytrale” ansiktet før man vurderer emosjonen som blir presentert.

Dette argumentet gjør det interessant å undersøke om mennesker er bedre til å gjenkjenne ansiktsemosjoner hvis emosjonene blir presentert i en video med et nøytralt ansikt som referanse.

I forlengelse av dette hadde det også vært interessant å undersøke om ulike emosjons identifiseringsprogrammer som MC-API, identifiserer emosjoner mer nøyaktig hvis emosjonene blir presentert i et videoklipp.

Den siste forslaget til videre forskning i denne oppgaven er å sammenligne menneskelige deltagere mot andre former for programmer som skal identifisere emosjoner hos andre. Dette skyldes at denne studien kun undersøkte MC-API. Siden flere aktører har lignende programmer er det viktig å undersøke hvor nøyaktige disse programmene i forhold til MC-API, ved å gjøre en sammenligning mot individer slik det er blitt gjort i denne oppgaven. Dette vil føre til en bedre forståelse av et nytt spennende felt i stadig utvikling.

Konklusjon

Målet med oppgaven har vært todelt. Først forsøker studien å undersøke om MC-API klarer å identifisere ansiktsemosjoner like riktig og nøyaktig som individer gjør. Det andre målet oppgaven undersøke er om MC-API slik det er i dag, kan brukes som et verktøy for å øke kvaliteten av den relasjonelle opplæringen som gjøres på Simuleringscenteret innlandet gjennom MUN-Light prosjektet.

For å undersøke dette ble tre hypoteser introdusert (H1, H2 og H3). Studien fant bare delvis støtte for alle hypotesene. For emosjonen glad og nøytral traff alle gruppene nøyaktig og riktig, selv om deltagerne var signifikant dårligere enn fasiten og MC-API for nøytrale emosjoner. De andre emosjonene klarte ikke MC-API å identifisere, og deltagerne var signifikant dårligere til å identifisere disse emosjonene en fasiten. Overordnet må dermed alle tre hypotese forkastes, og man må konkludere med at MC-API ikke identifiserer emosjoner hos andre like rett og nøyaktig som deltagerne gjør.

MC-API treffer godt på glad og nøytral, som tyder på at programmet har potensialet til å kunne gjøre nøyaktige og riktige vurderinger av emosjoner hos andre. Det er derfor et viktig område å forske videre på for å sikre at MC-API klarer å identifisere de andre emosjonene like nøyaktig som glad og nøytral. Slik MC-API er i dag er bruken av verktøyet noe begrenset, og MC-API klarer kun å identifisere emosjonene glad og nøytral.

Hvis målet med å implementere verktøyet er å bruke hele spekteret av følelser, vil ikke MC-API prestere tilstrekkelig per dags dato. For MUN-Light prosjektet som ønsker å benytte hele verktøyets bredde, må MC-API utbedres før det kan brukes til å oppnå målet om økt kvalitet i opplæring. Konklusjonen blir derfor at MC-API slik det fungerer i dag, ikke er et verktøy som kan brukes til å øke kvaliteten av den relasjonelle opplæringen i MUN-Light prosjektet.

Litteraturliste

- Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual review of psychology*, 60, 451-474. DOI: <https://doi.org/10.1146/annurev.psych.60.110707.163505>
- Amin, A., Basri, S., Hassan, M. F., & Rehman, M. (2011). Software engineering occupational stress and knowledge sharing in the context of global software development. In National Postgraduate Conference (NPC), 2011 (pp. 1-4). IEEE. DOI: 10.1109/NatPC.2011.6136269
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of management*, 36(4), 1065-1105. DOI: <https://doi.org/10.1177/0149206309352880>
- Abbott, B. B., & Bordens, K. S. (2011). Research design and methods: A process approach.
- Borges, R. (2012). Tacit knowledge sharing between IT workers: The role of organizational culture, personality, and social environment. *Management Research Review*, 36(1), 89-108. DOI: <https://doi.org/10.1108/01409171311284602>
- Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human resource development review*, 6(3), 263-296. DOI: <https://doi.org/10.1177/1534484307303035>
- Caris-Verhallen, W. M., Kerkstra, A., & Bensing, J. M. (1999). Non-verbal behaviour in nurse-elderly patient communication. *Journal of advanced nursing*, 29(4), 808-818. DOI: <https://doi.org/10.1046/j.1365-2648.1999.00965.x>
- Cascio, W. F., & Aguinis, H. (2014) (7. Utgave). Applied psychology in human resource management.
- Chambers, S. (2003). Use of non-verbal communication skills to improve nursing care. *British journal of nursing*, 12(14), 874-878. DOI: <https://doi.org/10.12968/bjon.2003.12.14.11412>
- Cohen, D., Sevdalis, N., Taylor, D., Kerr, K., Heys, M., Willett, K., ... & Darzi, A. (2013). Emergency preparedness in the 21st century: training and preparation modules in virtual environments. *Resuscitation*, 84(1), 78-84. DOI: <https://doi.org/10.1016/j.resuscitation.2012.05.014>
- Cook, T. D., Campbell, D. T., & Shadish, W. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.
- Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6), 681-685. DOI: 10.1109/34.927467
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of applied psychology*, 85(5), 678. DOI: 10.1037/0021-9010.85.5.678

Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*(pp. 9-15). ACM. DOI: 10.1145/2181037.2181040
- Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. *IEEE Transactions on pattern analysis and machine intelligence*, 21(10), 974-989. DOI: 10.1109/34.799905
- Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*. University of Nebraska Press.
- Ekman, P. (1977). Biological and cultural contributions to body and facial movement. *1977*, 34-84.
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1), 56-75. DOI: <https://doi.org/10.1007/BF01115465>
- Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.
- Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and emotion*, 10(2), 159-168. DOI: <https://doi.org/10.1007/BF00992253>
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875), 86-88. DOI: 10.1126/science.164.3875.86
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.
- Georgenson, D. L. (1982). The problem of transfer calls for partnership. *Training & Development Journal*.
- Helsepersonell loven. (1999). Lov om helsepersonell m.v. av 7 februar 1999, nr. 64. Hentet fra: <https://lovdata.no/dokument/NL/lov/1999-07-02-64> Den 18. April. 2018
- Helse Sør-Øst. (2012) *Etiske retningslinjer i Helse Sør-Øst*: Helse Sør-Øst
- HiOA (2017) *VEPRA20 Ferdighets- og atferdslaering*. Hentet fra: <http://www.hioa.no/Studier-og-kurs/HF/Bachelor/Vernepleie/Programplan-for-Bachelorstudium-i-vernepleie-2017/VEPRA20-Ferdighets-og-atferdslaering-2017> Den 10. Mai 2018
- Kacperck, L. (1997). Non-verbal communication: the importance of listening. *British Journal of Nursing*, 6(5), 275-279. DOI: <https://doi.org/10.12968/bjon.1997.6.5.275>
- Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (pp. 46-53). IEEE. DOI: 10.1109/AFGR.2000.840611
- Kohler, C. G., Turner, T., Stolar, N. M., Bilker, W. B., Brensinger, C. M., Gur, R. E., & Gur, R. C. (2004). Differences in facial expressions of four universal emotions. *Psychiatry research*, 128(3), 235-244. DOI: <https://doi.org/10.1016/j.psychres.2004.07.003>

- Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American psychologist*, 46(8), 819. DOI: <http://dx.doi.org/10.1037/0003-066X.46.8.819>
- Lazarus, R. S., & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of personality*, 1(3), 141-169. DOI: <https://doi.org/10.1002/per.2410010304>
- Lien, J. J., Kanade, T., Cohn, J. F., & Li, C. C. (1998). Automated facial expression recognition based on FACS action units. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* (pp. 390-395). IEEE. DOI: 10.1109/AFGR.1998.670980
- Mark, M. S., Børing, P., & Fevolden, A. M. (2018). Forstudie om avkastning av kompetanseinvesteringer i arbeidslivet.
- Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and emotion*, 14(3), 195-214. DOI: <https://doi.org/10.1007/BF00995569>
- Microsoft (2017). Emotion API. Hentet fra: <https://docs.microsoft.com/nb-no/azure/cognitive-services/emotion/home> Den 30. April 2018
- Microsoft Azure (2018). Kognitive tjenester. Hentet fra: <https://azure.microsoft.com/nb-no/services/cognitive-services/> Den 10. Mai 2018
- Pinker, S. (1997). *How the mind works*. New York: W. W.
- Prinz, J. (2004). Which emotions are basic. *Emotion, evolution, and rationality*, 69, 88.
- Robert, C (2014) Machine Learning, a Probabilistic Perspective, *CHANCE*, 27:2, 62-63, DOI: 10.1080/09332480.2014.914768
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, 115(1), 102. DOI: <http://dx.doi.org/10.1037/0033-2909.115.1.102>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, 115(1), 102. DOI: <http://dx.doi.org/10.1037/0033-2909.115.1.102>
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1), 145. DOI: <http://dx.doi.org/10.1037/0033-295X.110.1.145>
- Sprung, M., Münch, H. M., Harris, P. L., Ebesutani, C., & Hofmann, S. G. (2015). Children's emotion understanding: A meta-analysis of training studies. *Developmental Review*, 37, 41-65. DOI: <https://doi.org/10.1016/j.dr.2015.05.001>
- Turner, T. J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge?. DOI: 10.1037/0033-295X.99.3.566
- Valmaggia, L. R., Latif, L., Kempton, M. J., & Rus-Calafell, M. (2016). Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. *Psychiatry Research*, 236, 189-195. DOI:

<https://doi.org/10.1016/j.psychres.2016.01.015>

Van Hemert, D. A., Poortinga, Y. H., & van de Vijver, F. J. (2007). Emotion and culture: A meta-analysis. *Cognition and emotion*, 21(5), 913-943. DOI:

<https://doi.org/10.1080/02699930701339293>

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features.

In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE. DOI:

10.1109/CVPR.2001.990517

Wang, G., Oh, I. S., Courtright, S. H., & Colbert, A. E. (2011). Transformational leadership and performance across criteria and levels- A meta-analytic review of 25 years of research. *Group & Organization Management*, 36(2), 223-270. DOI:

<https://doi.org/10.1177/1059601111401017>

Yu, Z., & Zhang, C. (2015). Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 435-442). ACM. DOI: 10.1145/2818346.2830595

Vedlegg 1: Informasjon- og Samtykkeskjema

Forespørsel om deltakelse i forskningsprosjektet

”Bruk av ansiktslesningsalgoritmer innen opplæring i psykiatrien”

Bakgrunn og formål

Formålet studien er å undersøke om det er en forskjell mellom teknologi og individer, når det kommer til deres evne å oppfatte og forstå hvilken emosjonell tilstand et annet individ er i. Resultatene av studien vil blant annet være med og sette rammene for hvordan scenario trening og opplæring bør gjennomføres innen psykiatri, ambulanse og eldreomsorgen i fremtiden. Undersøkelsen vil også være en del av et masterprosjekt for Universitetet i Oslo, der Sykehuset Innland og SIM-innlandet fungerer som ekstern oppdragsgiver.

Hva innebærer deltakelse i studien?

Deltagelse i denne studien innebærer at all informasjon som blir registrert vil bli brukt i en vitenskapelig analyse. Dette vil brukes inn i en masteroppgave, men vil også være med å undersøke hvilke forskjeller det er mellom ansiktslesningsteknologi og arbeidstageres forståelse av ansiktsuttrykk for å lage en rapport som skal hjelpe SIM innlandet å utvikle og forbedre sin opplæring innen scenariotrening.

Undersøkelsen vil bestå av to deler. Den første delen vil bestå av et spørreskjema, der du fyller ut informasjon om seg selv. Innsamlingen av dataene vil skje anonymt, og ingen personsensitive opplysninger vil samles inn. Hvis du likevel ikke ønsker å bli med i studiet vil du så langt det lar seg gjøre når som helst kunne trekke deg uten å oppgi noen spesiell grunn. Skjemaet vil samle inn generell demografisk informasjon om deg og omhandle blant annet kjønn, alder og informasjon ellers om arbeidserfaring.

I den andre delen av eksperimentet vil du gjennomføre en praktisk oppgave der du sitter foran en pc-skjerm og evaluerer ansikter etter hvilken emosjon du oppfatter at hvert ansiktet viser. Her vil vi vise en rekke ansikter med forskjellige uttrykk du skal evaluere. Evalueringen av hvert ansikt skal besvares på papir på et skjema du får utdelt i andre delen av studiet.

Hele prosessen er forventet å ta ca.15-20 minutter per person.

Hva skjer med informasjonen om deg?

Alle opplysningene som blir samlet inn vil bli registret anonymt og vil derfor aldri kunne kobles til deg. Kun prosjektleder ved universitetet, studenten som gjennomfører studiet og prosjekteier hos sykehuset vil ha tilgang på dataene som blir samlet inn.

Du vil ikke som deltager i studien kunne godkjenne publikasjon av masteroppgaven, men du vil stå fritt til å trekke deg så langt dette lar seg gjøre.

Masteroppgaven skal etter planen avsluttes 15.mai 2018. Men dataene vil gjennom hele prosjektet lagres anonymt og vil derfor ikke kunne brukes til å personidentifikasjon av deg eller andre individer som deltar i denne studien.

Frivillig deltakelse

Det er frivillig å delta i studien og du kan når som helst så sant det lar seg gjøre, trekke deg fra studien uten å oppgi noen spesiell grunn. Dersom du trekker deg, vil alle opplysninger som lar seg slette bli trukket fra studien.

Dersom du har andre spørsmål til studien, ta kontakt med Agathon Maximillian Skei-Hart på telefon: 97485332 eller på e-mail: Agathons@student.sv.uio.no
<mailto:Agathon.hart@gmail.com> eller Cato Bjørkli på e-mail:
cato.bjorkli@psykologi.uio.no.

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Samtykke til deltakelse i studien

Jeg har mottatt informasjon om studien, og er villig til å delta

Sted

Signert av prosjektdeltaker, dato

Vedlegg 2: Spørreskjema for studien

Spørreskjema

Generell informasjon

-Kjønn

- Mann
- Kvinne

-Alder

- Under 20 20- 23 24-27 28-31 32-35
- 36-39 40-43 44-47 48-51 Over 51

-Arbeidssted per dags dato

- Ambulansetjenesten
- Akutt psykiatrisk avdeling
- Distrikt psykiatrisk/ og alders psykiatrisk senter
- Student med jobb innen psykiatri eller en annen form for helseforetak
- Student
- Annet

Arb. tittel

-Høyest fullførte utdanning

- Fullført grunnskole og/eller Videregående
- Bachelor
- Master
- Doktorgrad
- Annet

-Erfaring fra arbeid innen psykiatri, ambulansetjeneste eller annen form for helsetjeneste.

- 0-3 år
- 4-7 år
- 8-11 år
- 12-15 år
- 15 år eller mer

Informasjon om viktighet av pasienten sin sinnstilstand

Vi samlet inn litt demografisk informasjon om deg. I den neste delen av spørreskjemaet, ønsker jeg å samle inn informasjon om hvor viktig du og din arbeidstager opplever sinnstilstanden hos en pasient.

Noen av de som gjennomfører studien kan være studenter eller ikke enda ute i arbeid, og denne delen kan derfor for noen være vanskelig å svare på. Du svarer derfor etter beste evne på spørsmålene under.

-Hvor viktig oppfatter du at sinnstilstanden til pasienten du jobber med er.

- Uviktig
 - Litt viktig
 - Viktig
 - Veldig viktig
 - Vet ikke
 - Er student eller ikke enda ute i jobb
 - Annet
-

-Hvor viktig opplever du at det er at du opptrer ”riktig”, for å unngå utagering hos pasienten du jobber med.

- Uviktig
 - Litt viktig
 - Viktig
 - Veldigviktig
 - Vet ikke
 - Er student eller ikke enda ute i jobb
 - Annet
-

-Oppfatter du at sykehus eller andre arbeidsgivere legger vekt på en pasienten sin sinnstilstand, og hvordan dette kan påvirke en situasjon i opplæringen du får på arbeidsplassen.

- I ingen grad
 - I liten grad
 - I noen grad
 - I stor grad
 - Vet ikke
 - Er student eller enda ikke ute i jobb
 - Annet
-

Evaluerings av eksperimentet

NASA-TLX

<u>Skala definisjoner</u>		
Skala navn	Skala ytterpunkter	Beskrivelse
Mentale krav	Lavt/Høyt	Hvor mye mental og perseptuell aktivitet (tenking og avgjørelser) krevde eksperimentet av deg? Var oppgaven lett (lav) eller vanskelig (høy)?
Tidspress	Lavt/Høyt	Hvor høyt tidspress følte at eksperimentet forårsaket, som resultat av hastigheten eksperimentet ble gjennomført i. Følte du at eksperimentet ble gjennomført rolig å behagelig (lavt), eller rakt og stressende (høyt)?
Måloppnåelse	Lav/Høy	Hvor riktig føler du at du var i eksperimentet du gjennomførte. Føler du at du hadde høy måloppnåelse og gjorde mange riktige valg (høy) eller føler du at du hadde en lav måloppnåelse (lav) med mange feil?
Anstrengelse	Lav/Høy	Hvor hard måtte du anstrenge deg mentalt for å kunne oppnå måloppnåelsen du hadde i dette eksperimentet. Måtte du anstrenge deg mye (høy) eller krevde oppgaven liten anstrengelse (lav)
Frustrasjonsnivå	Lavt/Høyt	Var du sikker (lavt frustrasjonsnivå) eller usikker (høyt frustrasjonsnivå) på valgene du tok i eksperimentet

Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

Mentale krav

	1	2	3	4	5	6	7	
Lavt								Høyt

Tidspress

	1	2	3	4	5	6	7	
Lavt								Høyt

Måloppnåelse

	1	2	3	4	5	6	7	
Lav								Høy

Anstrengelse

	1	2	3	4	5	6	7	
Lav								Høy

Frustrasjonsnivå

	1	2	3	4	5	6	7	
Lavt								Høyt

Vedlegg 3: Skjema for evaluering av bildene i studien

Evaluering av bilder

Kryss av for den følelsen du oppfatter at bildet/videoen viser. Kun et kryss per bilde.

Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bruk av universale emosjoner og teknologi for å øke kvalitet i opplæring

Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke

Bilde/video nr.:



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke



Bilde/video nr.:

- Glad
- Overrasket
- Trist
- Redsel
- Avsky
- Forakt
- Sinne
- Nøytral
- Vet ikke

