



Universitat de les
Illes Balears



Treball Final de Grau

GRAU D'ENGINYERIA INFORMÀTICA

Experiència amb Human based
Computation per a la classificació dels tons
de pell humana

BONNÍN HERNÁNDEZ, JOAN

Tutors

Moyà Alcover, Gabriel

Dr. Jaume-i-Capó, Antoni

Escola Politècnica Superior
Universitat de les Illes Balears
Palma, setembre de 2016

SUMARI

Sumari	I
Índex de figures	III
Índex de taules	V
Acrònims	VII
Abstract	IX
1 Introducció	1
1.1. Motivació	1
1.2. Classificació Fitzpatrick	2
1.2.1. Classificacions alternatives	2
1.3. Human based computation	3
1.3.1. HbC basada en microtasques	4
1.3.2. HbC basada en macrotasques	5
1.4. Objectius	5
2 Sistema	7
2.1. Experiments	7
2.1.1. Human based Computation: Voluntariat	7
2.1.2. Realització de la tasca	8
2.1.3. Plana web	8
2.1.4. Base de dades	11
2.2. Anàlisi de dades i classificador automàtic	12
3 Experiments	13
3.1. Procediment general	13
3.1.1. Difusió	13
3.1.2. Base de dades d'imatges	14
3.1.3. Anàlisi de resultats	14
3.1.4. Nova classificació	15
3.2. Experiment basat en color	16
3.2.1. Preparació i execució	16
3.2.2. Anàlisi de resultats	17
3.2.3. Discussió	21
3.2.4. Nova classificació	22

3.3. Experiment bastat en mostres	23
3.3.1. Preparació i execució	24
3.3.2. Anàlisi de resultats	26
3.3.3. Discussió	30
3.3.4. Nova classificació	30
3.4. Experiment bastat en color II	31
3.4.1. Preparació i execució	31
3.4.2. Anàlisi de resultats	32
3.4.3. Discussió	33
3.5. Comparativa entre experiments	34
3.5.1. Participació	34
3.5.2. Distribució de respostes	34
3.5.3. Desviació de respostes	35
3.5.4. Nova classificació	36
4 Classificació automàtica	39
4.1. Sistema de classificació	39
4.2. Tractament de dades	40
4.2.1. Tractament previ de les mostres	40
4.2.2. Aprenentatge	40
4.2.3. Validació	41
4.3. Anàlisi de resultats	42
4.4. Discussió	42
5 Conclusions	45
Bibliografia	47

ÍNDIX DE FIGURES

1.1. Implementació de la guia Unicode per tonalitat de pell	3
1.2. Selecció de tècniques HbC	6
2.1. Flux de navegació per realitzar l'experiment	8
2.2. Disseny <i>responsive</i> : versió d'escriptori	9
2.3. Disseny <i>responsive</i> : versió mòbil	10
2.4. Implementació reCAPTCHA	11
2.5. Modelat de la base de dades	11
3.1. Selectors de color: exemple d'experiment	17
3.2. Exp. selectors color: distribució de respostes-boxplot	19
3.3. Exp. selectors color: desviació de respostes	20
3.4. Exp. selectors color: desviació respostes classe 3	21
3.5. Exp. selectors color: desviació respostes classe 3 per sota de llindar	21
3.6. Exp. selectors color: redistribució d'imatges	23
3.7. Selectors de mostres: exemple d'experiment	24
3.8. Exp. selectors mostres: nous selectors	25
3.9. Exp. selectors mostres: distribució de respostes-boxplot	28
3.10. Exp. selectors mostres: desviació de respostes	28
3.11. Exp. selectors mostres: desviació respostes classe 3	29
3.12. Exp. selectors mostres: desviació respostes classe 4	30
3.13. Exp. selectors mostres: redistribució d'imatges	31
3.14. Selectors de color II: exemple d'experiment	32
3.15. Exp. selectors color II: mostres classe 2	33
3.16. Exp. selectors color II: mostres classe 3	33
3.17. Comparativa experiments: distribució de respostes	35
3.18. Comparativa experiments: desviacions respostes	36
3.19. Comparativa experiments: noves classes	37
4.1. Classificador automàtica: distribució del GT	43
4.2. Classificador automàtica: distribució de la nova classificació	43

ÍNDIX DE TAULES

1.1.	Atributs de l'escala Fitzpatrick	2
1.2.	Guia Unicode per tonalitat de pell	3
3.1.	Exp. selectors color: anàlisi general	18
3.2.	Exp. selectors color: distribució de respostes	19
3.3.	Exp. selectors color: dades classe 3	20
3.4.	Exp. selectors color: nova classificació	23
3.5.	Exp. selectors mostres: anàlisi general	27
3.6.	Exp. selectors mostres: distribució de respostes	27
3.7.	Exp. selectors mostres: dades classes 3 i 4	29
3.8.	Exp. selectors mostres: nova classificació	30
3.9.	Exp. selectors color II: anàlisi per classes	32
4.1.	Classificadors automàtics: anàlisi	42

ACRÒNIMS

HbC Human based Computation

GT Ground Truth

BD Base de Dades

SGBD Sistema Gestor de Base de Dades

GMM Gaussian Mixture Model

UGiVIA Unitat de Gràfics i Visió per Computador i Intel·ligència Artificial

EPS Escola Politècnica Superior

UIB Universitat de les Illes Balears

CDA Canadian Dermatology Association

ABSTRACT

Es presenta un sistema d'experiments basat en tècniques de Human Based Computation per aconseguir classificar tons de pell segons l'escala Fitzpatrick. Mitjançant diferents experiments, es recopilen respostes d'usuaris que classifiquin imatges de persones segons dita escala. Les variacions d'aquests experiments consisteixen en la mostra que visualitza l'usuari com a referència de cada classe. Dites respostes són analitzades per comprovar quin grau de confiança pot suposar que un usuari inexpert realitzi aquesta tasca.

Es disposa d'un Ground Truth definit per un expert, el qual es compara amb les respostes dels usuaris. En base a les dades validades, es proposa una nova classificació que combina les respostes d'usuaris amb les originals de l'expert. Finalment, tant les noves classificacions, com el conjunt original són avaluat mitjançant un algorisme de classificació automàtica. Aquest classificador automàtic també resulta útil per avaluar la qualitat les diferents classificacions generades.

INTRODUCCIÓ

Aquest projecte tracta la necessitat de classificar les diferents tonalitats de pell a través d'una aproximació a la Human based Computation. Aquest recorregut implica l'estudi de les possibles classificacions, les diferents tècniques per classificar a través de tasques per usuaris, i fins i tot una aproximació a la classificació automàtica.

1.1. Motivació

Arrel del plantejament d'aquest projecte par part dels tutors, i d'un seguit d'articles que em varen facilitar, vaig començar a descobrir conceptes com Human based Computation o algunes nocions de tractament d'imatges que es realitzen a l'Unitat de Gràfics i Visió per Computador i Intel·ligència Artificial (UGiVIA) [1]. Amb aquests nous descobriments, va sorgir una motivació per dur a terme aquest projecte i adquirir coneixements sobre del funcionament de la Human based Computation i, sobre tot, treballar en projectes on la tasca de l'usuari humà no queda desplaçat a un segon pla. A més, em sembla interessant poder veure com responen els usuaris a diferents *inputs* i com les seves respostes varien en funció d'aquests.

Un cop explicada la vessant més humana, no vull deixar passar l'oportunitat d'explicar que també m'atreu la idea d'ampliar els meus coneixements relatius a la intel·ligència artificial amb algorismes de classificació automàtica, així com obtenir nocions bàsiques sobre tractament d'imatges, camp que desconec completament. Esper que, en conjunt, siga una experiència que m'ensenyi com es duen a terme projectes d'un tarannà més científic del que estic habituat, així com a treballar en projectes complexos i de llarga durada.

1.2. Classificació Fitzpatrick

L'escala Fitzpatrick és un esquema de classificació numèric pel color de la pell humana. Desenvolupada pel dermatòleg Thomas B. Fitzpatrick l'any 1975, és un mètode per estimar la resposta de diferents tipus de pell a la llum ultraviolada UV[2]. L'escala tracta de classificar els diferents tipus de pell en sis grans grups. Aquests s'agrupen en funció de com responen al sol i com es bronzegen. Les diferents classes es poden consultar en la [Taula 1.1](#)

Tipus	Es crema la pell?	Es bronzeja?	Altres atributs
#1	Sempre es crema	Mai es bronzeja	Cabell pèl-roig
#2	Habitualment es crema	Resulta complicat bronzejar-se	Cabell ros o clar
#3	Es pot cremar lleugerament	Es bronzeja gradualment	Cabell castany o ros fosc
#4	Normalment no es crema	Es bronzeja amb facilitat	Cabell castany fosc
#5	No es crema gairebé mai	Es bronzeja amb molta facilitat	Cabell negre o molt fosc
#6	No es crema	Es bronzeja sempre	Cabell negre

Taula 1.1: Propietats segons l'escala Fitzpatrick [3]

Cal destacar que és la classificació més usada des de la seva aparició i que a més d'en àmbits mèdics, també s'usa en cosmètica per recomanar diferents castes de protectors solars en funció del tipus de pell.

Tot i això, el que resulta especialment interessant per aquest projecte, és que a més d'analitzar la resposta de la pell als UV, també es poden identificar els tipus de pell en funció del color. Tot i que la classificació recau directament en la resposta als UV, es pot observar com cada tipus de pell habitualment ve acompanyat d'altres característiques com el color de cabell o d'ulls, el qual també pot ser útil a l'hora de classificar algunes mostres en cas d'insuficiència de dades.











Tan estesa és aquesta classificació que és usada per diferents organismes oficials, com ara la Canadian Dermatology Association (CDA) les tenen presents als seus treballs [4] i, fins i tot, proposen als usuaris de la seva plana web un senzill test per, a través d'una sèrie de preguntes amb resposta tancada, poder saber a quin tipus de pell s'assemblen més segons l'escala Fitzpatrick. Dit test conté preguntes relatives a la pell, color d'ulls, de pèl, bronzejat, etc. El test es pot trobar a la plana web de la CDA, i concretament a l'adreça <http://www.dermatology.ca/skin-hair-nails/skin/photoaging/know-your-skin-type/>.

1.2.1. Classificacions alternatives

Existeixen classificacions alternatives a la proposada per Fitzpatrick, com per exemple la **Von Luschan**, de la qual va sorgir la Fitzpatrick. El principal problema de l'escala Von Luschan respecte Fitzpatrick és el gran nombre de classes que presenta, ja que

proposa una organització en **36 classes diferents** [5], les quals són evidentment acla-
paradores per persones sense coneixements específics en dermatologia. Un possible
avantatge respecte Fitzpatrick seria que Von Luschan classifica directament sobre el
color de pell, no sobre com aquesta reacciona al sol. Aquest mateix aspecte, que a sim-
ple vista només sembla avantatjós pel tractament d'imatges, també té els seus punts
febles, com que la pell a classificar no ha d'estar ni bronzejada ni cremada. Amb tot
això, cal aclarir que l'escala Von Luschan pretenia realitzar una classificació racial de
poblacions, mentre que l'objectiu de Fitzpatrick és la classificació de cada individu
segons les característiques cutànies.

Emoji Modifiers

Code	Name	Samples
U+1F3FB	EMOJI MODIFIER FITZPATRICK TYPE-1-2	 
U+1F3FC	EMOJI MODIFIER FITZPATRICK TYPE-3	 
U+1F3FD	EMOJI MODIFIER FITZPATRICK TYPE-4	 
U+1F3FE	EMOJI MODIFIER FITZPATRICK TYPE-5	 
U+1F3FF	EMOJI MODIFIER FITZPATRICK TYPE-6	 

Taula 1.2: Definició de l'escala de color de pell segons Unicode[6]

Una classificació sense pretensions científiques o mèdiques és l'adoptada per
l'estàndard *unicode* [6] per a la representació dels emoji. Aquesta classificació beu
directament de l'escala Fitzpatrick i assigna un color de pell a cada classe. Tot i això,
en lloc de definir 6 grups, només en defineix 5, agrupant les classes 1 i 2 a causa de la
semblança entre els colors que les definirien. La **Taula 1.2** serveix com a guia dels colors
representatius de les classes, per a poder fer diferents representacions més lliures, com
per exemple la mostrada a la **Figura 1.1**

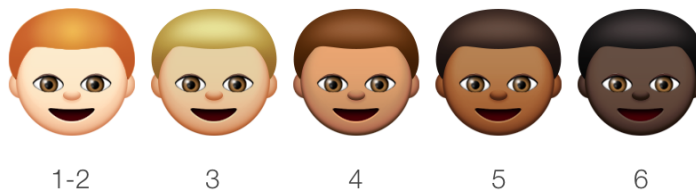


Figura 1.1: Implementació de la guia unicode [7]

1.3. Human based computation

La Human based Computation (**HbC**) (*computació basada en humans*) és una
tècnica informàtica en la que el treball és externalitzat a humans, habitualment en
format de microtasca. Aquest plantejament aprofita les diferències d'habilitat i cost per
treball entre humans i màquines per maximitzar-ne el profit [8].

La computació tradicional consisteix en l'ús de la màquina per part de l'humà, que planteja un problema i un algorisme per resoldre-lo, per rebre una solució de la màquina. La **HbC** realitza un plantejament invers, ja que és l'ordinador qui sistemàticament sol·licita a l'usuari (o a un gran conjunt d'aquests) realitzar una tasca per obtenir-ne els resultats.

1.3.1. HbC basada en microtasques

Donat un problema que necessita múltiples solucions relativament ràpides o senzilles, es recorre a la **HbC** de microtasques. Aquest tipus de tasques són independents entre elles, i gairebé qualsevol persona sense coneixements específics les pot resoldre amb facilitat [8]. Tot i senzillesa d'aquestes tasques, l'usuari pot millorar el seu rendiment mitjançant l'aprenentatge i entrenament. Una altra característica d'aquest tipus de tasques és la complexitat que comporten completar-les mitjançant una màquina pel seu **elevat cost computacional**. Enfront aquest cost, la alternativa de que els humans realitzin el treball resulta encara més interessant, ja que en general les microtasques **no comporten dificultat per un usuari** comú.

Pel que fa a aquest projecte, aquesta diferència entre costos de realització de tasca es compleix, ja que per els humans no és una tasca costosa la comparació de colors dins una imatge, mentre que la màquina no disposa d'eines tan eficients per poder distingir o classificar regions concretes de pell.

Independentment de la problemàtica a tractar, la principal dificultat a l'hora de realitzar **HbC** no és la generació de tasques útils, sinó més aviat la dificultat per què aquestes es realitzin. Com a resum, podem exposar les diferents vies a través les quals les tasques arriben a usuaris.

Voluntariat També conegut com a *ciència ciutadana*. Consisteix a sol·licitar la realització de forma voluntària, sense esperar cap recompensa. És, sorprenentment un mètode efectiu, com es pot comprovar gràcies a l'experiència del projecte.

Videojocs casuals A més de l'al·licient altruista, en aquest cas s'insereix un factor de *gamificació* a les tasques, fent que la participació sigui més elevada per factors com la diversió, la competitivitat o l'autosuperació.

Mercat de microtasques És el mètode més evident i conegut des de la perspectiva de l'usuari per a la realització de microtasques. Consisteix en el pagament directe per la realització d'un treball o un paquet d'aquests.

Tasques forçoses Aquest mètode, a diferència dels anteriors, no cerca estimular a l'usuari, sinó tot el contrari. Consisteix en establir una sèrie de barreres per l'usuari a l'hora de realitzar una tasca informàtica, i la barrera es supera realitzant la tasca requerida pel sistema de **HbC**. L'exemple més conegut és el CAPTCHA.

Educatives Mètode especialment útil per a tasques de disciplines concretes o més avançades. Consisteix en assignar tasques a usuaris en procés d'aprenentatge

relacionat amb la tasca a realitzar, assolint dues fites: la completesa de la tasca **HbC** i el progrés educatiu de l'alumne.

1.3.2. HbC basada en macrotasques

A diferència de les tècniques anteriors, les macrotasques necessiten d'usuaris experimentats per resoldre tasques de major volum, dificultat i complexitat. Això es pot aconseguir bé buscant usuaris avançats (més costós) o bé formant a l'usuari amb feines d'aprenentatge.

Videojocs complexos A diferència dels videojocs *casuals* per realitzar microtasques, aquests no assignen un gran volum de tasques a l'usuari, sinó que n'assignen unes poques de molta major complexitat i durada. Igualment, aquest mètode compta amb els beneficis de la *gamificació* exposats anteriorment.

Concursos obert d'innovació Es planteja un problema complex i es convida obertament a qui vulgui plantejar-ne una solució. Un cop avaluades les respostes a la tasca, es premien les millors tant econòmicament, com amb el prestigi que comporta haver resolt un problema d'aquesta magnitud.

Per maximitzar l'eficiència del projecte cal triar la via de comunicació amb l'usuari que realitzarà les tasques. Hom pot fer-ho seguint el diagrama de la **Figura 1.2**.

1.4. Objectius

Els objectius d'aquest projecte estan relacionats directament amb els conceptes presentats de classificació de pell i de **HbC**. Aquestes vénen determinats al següent llistat:

- Comprensió de diferents tècniques de **HbC** i selecció de la més adequada per a realitzar una classificació de to de pell.
- Realitzar experiments per tal que els usuaris classifiquin un conjunt d'imatges segons l'escala Fitzpatrick.
- Contrastar la validesa del Ground Truth mitjançant les noves dades recollides a partir d'usuaris.
- Generar noves classificacions per tal de millorar l'etiquetat original.
- Realitzar un procés d'etiquetatge automàtic per comparar-lo amb l'habilitat dels humans i qualificar les noves classificacions.

Allarg del document es pot trobar resposta als objectius plantejats en aquest capítol introductori i un extens desenvolupament de les metodologies seguides.

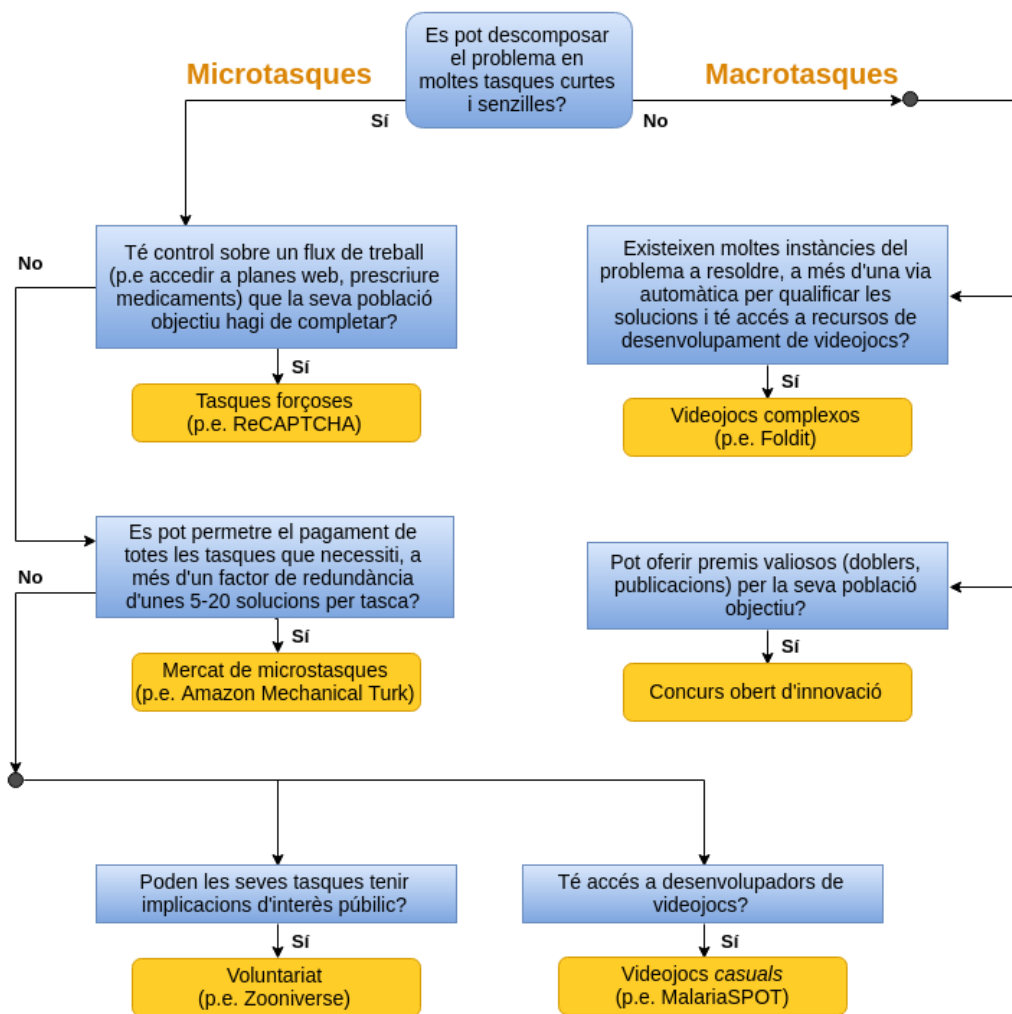


Figura 1.2: Diagrama de selecció de tècniques HbC [8]



SISTEMA

Per dur a terme el projecte s'han fet servir diferents eines informàtiques, cadascuna adequada a les necessitats que han sorgit durant el plantejament i desenvolupament de les diferents etapes del mateix. Les eines que s'han usat es poden agrupar fàcilment en dos grans blocs: eines per a la realització d'una aplicació web i eines per l'anàlisi i tractament de dades amb una vessant més científica.

2.1. Experiments

Un dels fonaments del projecte són els experiments per la captació d'informació dels usuaris, que són la font de dades a tractar. Els experiments s'han dut a terme mitjançant una plana web donat que, és la via idònia per fer accessible de forma oberta els experiments, a més de ser un sistema de fàcil desenvolupament.

2.1.1. Human based Computation: Voluntariat

La manera més senzilla d'obtenir aquestes dades, ja que es necessita un gran volum de respostes, és mitjançant les eines de **HbC**. Més concretament, donat el caràcter educatiu i amb pocs recursos del projecte, s'ha recorregut a **HbC** de microtasques, i més específicament, de voluntariat.

El plantejament per obtenir les dades consisteix en la realització d'uns petits experiments on, mitjançant diferents mètodes, els usuaris ens proporcionin informació nova de com interpreten les imatges, és a dir, com classificarien cada imatge segons l'escala Fitzpatrick.

2.1.2. Realització de la tasca

Pel que fa a la funció dels usuaris, han de realitzar una tasca que on se'ls mostra una sèries d'imatges i han d'intentar classificar-les segons l'escala Fitzpatrick. El flux de navegació per realitzar l'experiment ve donat pel diagrama present a la [Figura 2.1](#).

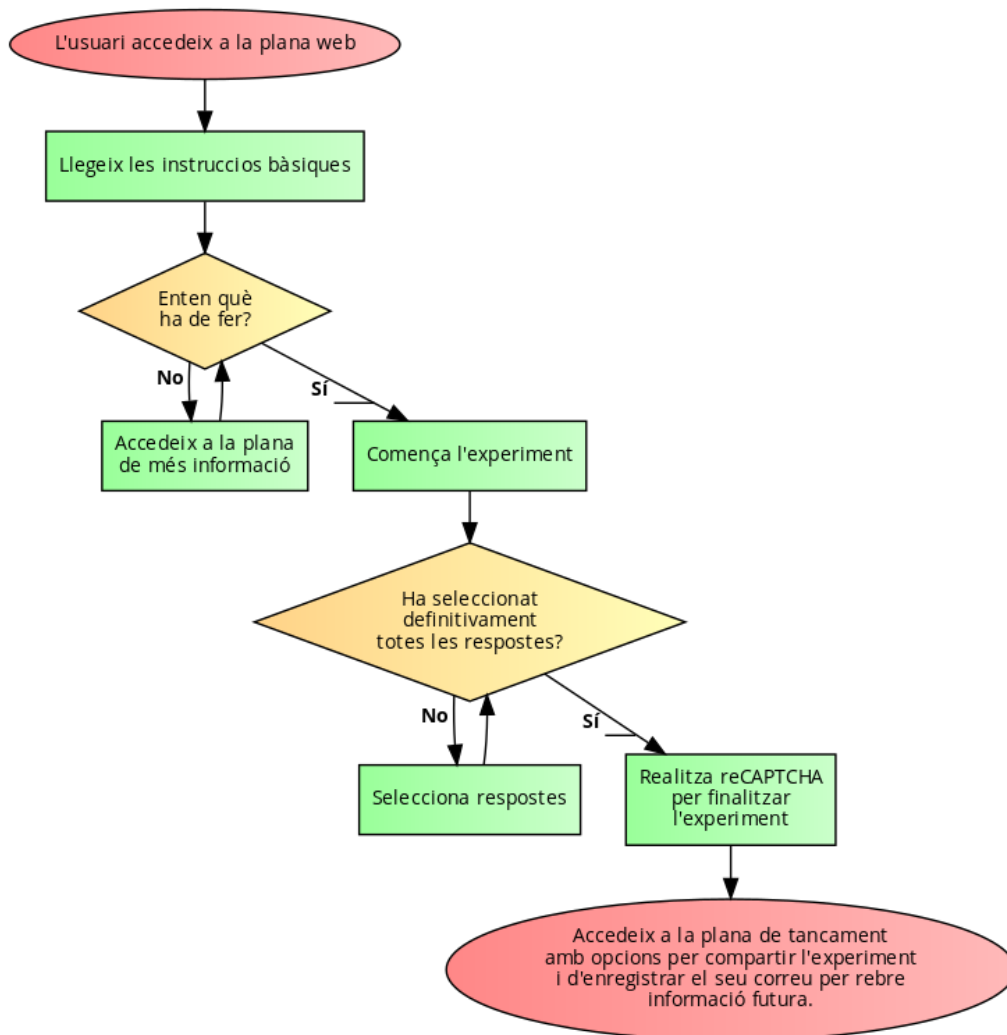


Figura 2.1: Flux de navegació per realitzar l'experiment

Amb això el que es pretén és que la tasca sigui el més senzilla possible, minimitzant la fuga d'usuaris abans d'haver realitzat la tasca, ja que tota la tasca es realitza en una única plana, sense navegació farragosa.

2.1.3. Plana web

La plana web, allotjada al domini humania.uib.es, s'ha desenvolupat fent ús de diferents eines, totes elles ben conegudes i amb una gran presència a les aplicacions d'entorns professionals.

Pel que fa al front-end, s'ha fet servir el framework d'HTML, CSS i JS **Bootstrap**. Aquest framework permet el desenvolupament àgil de planes web, especialment pel que fa a les diferents necessitats *responsive*. Aquest aspecte va resultar clau a l'hora de dissenyar els experiments, ja que la web mòbil és fonamental per interaccions ràpides amb usuaris. Es poden comprovar les diferències entre les versions d'escriptori (Figura 2.2) i mòbil (Figura 2.3).

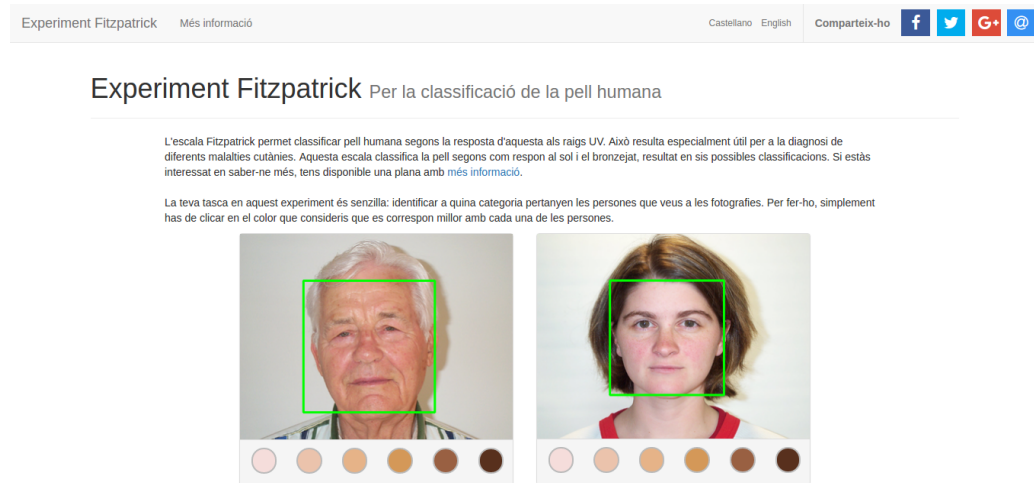


Figura 2.2: Versió d'escriptori de la plana web

Com és evident, alguns aspectes s'han hagut d'adaptar específicament al projecte, essent necessari l'ús de CSS i JavaScript a un nivell més baix. En aquest cas, s'han usat LESSCSS i jQuery respectivament per mantenir un desenvolupament més eficient. Donat el gran ajut que han suposat aquestes eines, es considera apropiat fer-ne una petita explicació:

LESSCSS LESSCSS, també conegut com Less, és un preprocessador de CSS, el qual vol dir que és una extensió del llenguatge CSS, afegint-hi característiques que permeten l'ús de variables, funcions i moltes altres tècniques que permeten crear CSS més mantenible i extensible [9]. El principal avantatge d'usar Less o algun altre llenguatge de preprocessat enlloc de CSS és la capacitat de desenvolupant amb les facilitats i automatització que comporta un llenguatge de programació per sobre d'un llenguatge de marcat.

jQuery jQuery és una ràpida, petita i rica en característiques llibreria de JavaScript. Gràcies a jQuery i la seva senzilla API que funciona a través d'una gran multitud de navegadors, es faciliten accions com el recorregut o manipulació de documents HTML, la gestió d'esdeveniments, animacions i comunicació Ajax [10]. A més, la gran comunitat al voltant d'aquest popular *framework* simplifica la resolució de problemes o dubtes.

Un altre aspecte important per arribar a més usuaris, és que la plana web estigui disponible en **vàries llengües**. Per satisfer aquest requeriment, es va fer ús de les eines

2. SISTEMA

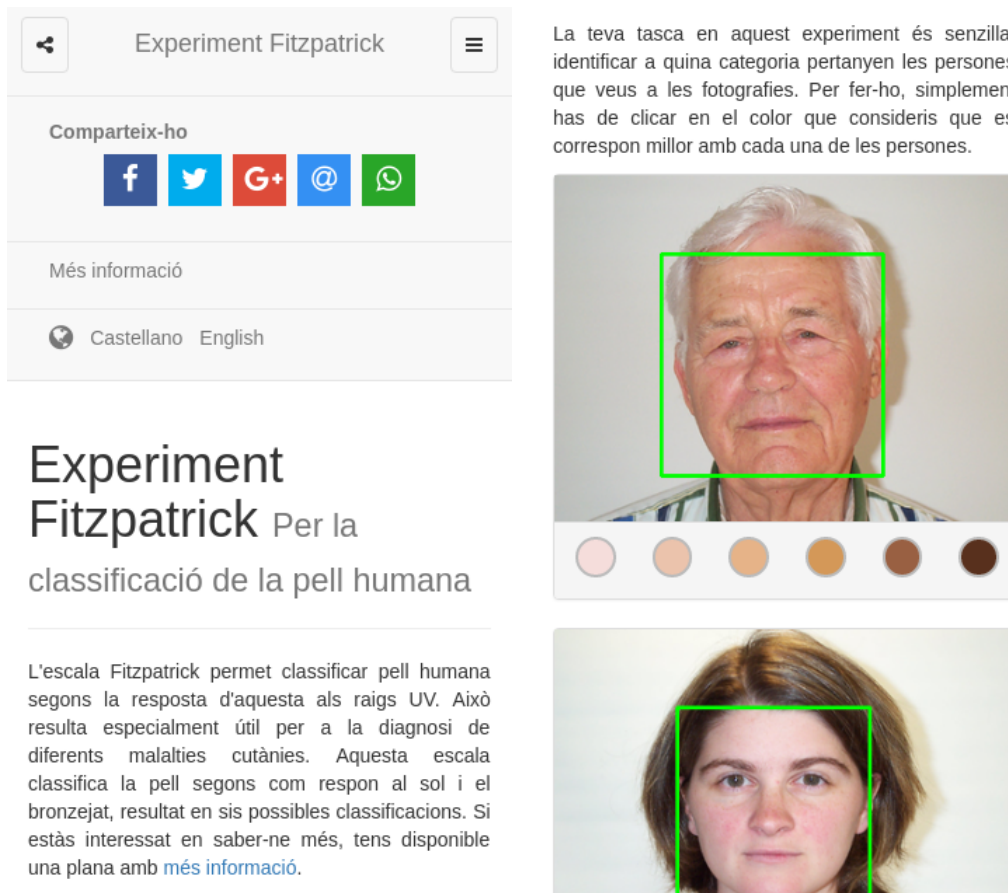


Figura 2.3: Versió mòbil de la plana web

gettext de GNU [11]. Aquest sistema es basa la identificació de les cadenes de text que es troben a l'aplicació per, mitjançant una sèrie de fitxers creats pel desenvolupador, traduir cada una de les cadenes a la llengua que pertorqui. En aquest cas, la plana es trobava disponible en català, castellà i anglès.

Per aconseguir més difusió, es va assignar metainformació a la plana per tal que els usuaris abans d'accedir-hi poguessin veure'n una imatge relacionada i un petit resum del què s'hi podia fer. Per inserir aquestes dades s'han hagut d'adaptar diferents paràmetres generals i específics per cada plataforma de compartició, com Twitter (*TwitterCards*), Facebook (*OpenGraph*), Google+ i cercadors en general (*itemprop*).

En l'aspecte de la seguretat, com es pot observar a la [Figura 2.4](#), per completar l'experiment cal resoldre un **ReCAPTCHA** [12] per evitar la possible automatització de respostes per part de possibles usuaris malintencionats. A més, també es comprova mitjançant una cookie i l'estat de la sessió PHP si l'usuari ha realitzat algun experiment prèviament. Aquest mètode pot semblar molt rudimentari, però és la manera menys invasiva d'enregistrar la informació sobre les repeticions d'experiments.

Finalment, indicar que s'ha usat PHP com a llenguatge de servidor. Des de diferents

Si no vols realitzar cap modificació sobre les opcions elegides, **finalitza** validant que ets humà mitjançant el següent reCAPTCHA.

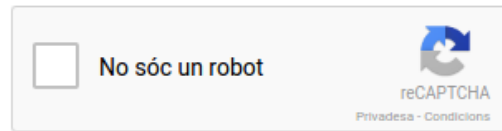


Figura 2.4: Sol·licitud de reCAPTCHA a l'hora de finalitzar l'experiment

scripts PHP s'ha desenvolupat tant les opcions de multi-idoma, la gestió de l'aleatorietat de mostres o les diferents connexions amb la base de dades per emmagatzemar els resultats dels experiments.

2.1.4. Base de dades

Tota la informació relativa als experiments s'ha d'emmagatzemar, com resulta evident, en una Base de Dades (BD). En aquest cas, s'ha utilitzat el Sistema Gestor de Base de Dades (SGBD) MySQL, ja que és el gestor integrat en el domini i té una gran comunitat de la qual obtenir suport.

Donat el caràcter dels experiments, per gestionar les dades s'ha creat un modelat relativament senzill, del qual es pot consultar el diagrama UML a la [Figura 2.5](#). Tot i això, algunes de les seves característiques mereixen ser explicades.

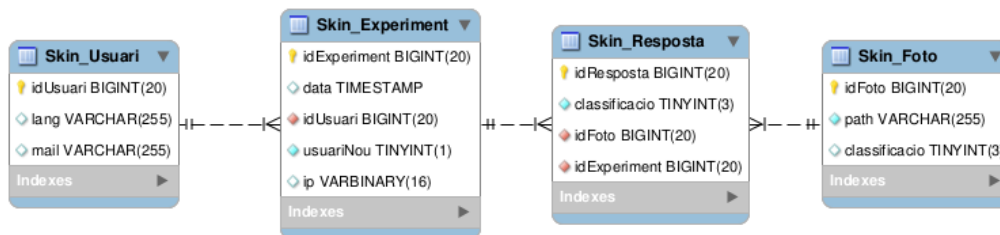


Figura 2.5: Modelat de la BD

Skin_Usuari Tot i que els experiments es poden realitzar de forma anònima i no requereixen cap tipus de registre, un cop finalitzat es dona la opció a l'usuari d'enregistrar un correu electrònic com a via de contacte, a més de la llengua en la que vol rebre la comunicació.

Skin_Experiment Es deixa constància de en quin moment s'ha realitzat, a més d'un valor booleà que identifica l'experiment com a realitzat per un usuari que no n'ha realitzat cap altre. A més a més, també s'emmagatzema la IP pública de l'usuari, per tal de poder limitar en certa mesura l'accés a usuaris malintencionats.

Skin_Experiment S'enregistra dins el camp *classificació* quina resposta ha donat l'usuari a una determinada fotografia.

Skin_Experiment S'enregistra el nom d'arxiu de la imatge en qüestió per poder-hi accedir des dels scripts de servidor, a més de la seva classificació segons el Ground Truth (GT)

Així doncs, gràcies a aquestes eines, s'ha desenvolupat una plana web totalment funcional que satisfà les necessitats dels experiments de microtasques.

2.2. Anàlisi de dades i classificador automàtic

Un cop recollides les dades mitjançant els experiments a la plana web, aquestes s'han recuperat de la base de dades per així poder-les analitzar i s'han emmagatzemat en arxius en format CSV.

L'anàlisi de dades s'ha realitzat mitjançant el llenguatge de programació **Python** gràcies a la gran comunitat i quantitat de llibreries usades en diferents àrees d'investigació. Aquest llenguatge s'ha usat en diferents versions, concretament les versions 2.7 i 3.4, a causa d'algunes necessitats i algunes dependències concretes. D'entre les llibreries que s'han usat, les més destacables són les següents:

numpy Llibreria que agrega major suport per al tractament de vectors i matrius a través d'una biblioteca de funcions matemàtiques d'alt nivell per tractar dites estructures de dades. S'usa a gairebé tots els scripts realitzat gràcies a les facilitats que atorga per tractar les dades.

matplotlib Llibreria per la generació de gràfiques a partir de les estructures bàsiques de Python i les extensions de NumPy. Usada per generar totes les gràfiques de sortida del projecte.

colorsys Mòdul que defineix conversions de color bidireccionals entre diferents espais de color, com ara RGB, YIQ, HLS, HSV, etc. Usat bàsicament pel tractament d'imatges juntament amb els algorismes de classificació automàtica.

OpenCv Llibreria enfocada a la visió per computador en temps real. Tot i això, en el projecte s'usa per a realitzar detecció de la zona d'interès del conjunt d'imatges a tractar. En concret, la seva funció principal és el reconeixement facial per descartar totes les regions on no apareix una persona.

sklearn Llibreria enfocada a diferents tècniques de *machine learning* implementat diferents classificadors, algorismes de regressió, *clustering*, etc. Concretament, en aquest projecte es fa servir per generar *Gaussian mixture model* que serviran per realitzar classificació automàtica.

EXPERIMENTS

L'eix central i punt de partida d'aquest experiment són els experiments. A través de diferents variants d'experiments, s'analitzen les respostes d'usuaris front imatges de cares. La seva tasca és classificar aquestes imatges segons l'escala Fitzpatrick mitjançant una sèrie de selectors que es veuen modificats a cada experiment.

3.1. Procediment general

Per obtenir noves dades, s'han plantejat una sèrie d'experiments amb usuaris, amb una sèrie de característiques comunes entre ells. La idea bàsica és obtenir nova informació sobre com es classificaria segons l'escala Fitzpatrick un conjunt d'imatges del qual ja tenim una classificació original.

3.1.1. Difusió

Una vegada es va planificar l'execució d'experiments basats en la **HbC**, fou necessari establir una sèrie de canals de comunicació amb diferents usuaris, per tal de fer-los arribar les tasques. Per arribar al màxim nombre d'usuaris possibles, la difusió es va realitzar a través dels següents mitjans:

Xarxes socials Facebook, Twitter, Google+. L'enllaç als experiments es va compartir tant a títol personal per diferents usuaris de les xarxes, com a perfils d'institucions com l'Escola Politècnica Superior (**EPS**), la Universitat de les Illes Balears (**UIB**) o el programa de ràdio Balears fa ciència.

Fòrums Per aprofitar eines amb resposta dels usuaris, es va compartir també a diferents fòrums com elotrolado i Reddit, específicament als subreddits */Participants*, */CitizenScience* i */volunteer*.

Missatges directes Es va facilitar la compartició directa entre usuaris per qualsevol via (com correu electrònic o aplicacions de missatgeria mòbil) gràcies a la correcta configuració de metainformació de la plana web.

Mailing Correus electrònics amb gran volum de destinataris. Per al primer experiment, la **UIB** va incloure en el butlletí #SOMUIB. Donat que els experiments permetien als usuaris deixar un correu de contacte, al llançament de la resta dels experiments se'ls va enviar

Ciència per a tothom 2016 Aprofitant el gran volum de participació i assistència a l'esdeveniment *Ciència per a tothom*, es va publicitar l'experiment en aquesta diada, tant presencialment com a través de ràdio al programa Balears fa Ciència.

3.1.2. Base de dades d'imatges

Disposam d'un Ground Truth **GT** que consisteix en un conjunt de 524 imatges classificades sota el criteri d'un **expert en dermatologia**. Tot i que no s'han usat totes per al desenvolupament dels experiments, sí que s'han utilitzat les mateixes entre experiments, per així poder realitzar un millor estudi estadístic.

Ground Truth és un terme usat per fer referència a tota aquella informació que no s'obté per inferència, sinó que s'ha obtingut per observació directa. A més, en àmbits de *machine learning*, el **GT** és tota aquella informació usada per un conjunt d'entrenament de tal manera que aquesta informació és considerada **vàlida i precisa**.

A més, les imatges mostrades a cada experiment són seleccionades de forma aleatòria, per evitar que un mateix usuari pugui realitzar dues vegades experiments sobre les mateixes mostres, ja que ens interessa maximitzar la diversitat d'opinions a cada mostra.

L'objectiu de cada experiment és obtenir un nombre de respostes mínim per cada imatge, en concret 5 mostres, per poder-ne analitzar les respostes i intentar classificar la imatge més acuradament.

3.1.3. Anàlisi de resultats

Els resultats dels experiments seran tractats de la mateixa manera, analitzant en tot cas a les següents mètriques sobre les respostes dels usuaris i la relació d'aquestes amb el **GT**:

Nombre de respostes Quantes respostes sobre les imatges s'han aconseguit gràcies al total de l'experiment. Donat que als experiments els usuaris etiqueten dues imatges, sempre serà el doble al nombre d'experiments.

Nombre de coincidències Quantes etiquetes sobre una imatge coincideixen amb la seva classificació al **GT**.

Percentatge de coincidència Percentatge d'etiquetes coincidents amb la seva classificació al **GT**.

Desplaçament mitjà Per entendre aquesta mètrica cal considerar que les classes es poden organitzar seqüencialment, és a dir establir una relació que segueix l'ordre cardinal. Amb això, el desplaçament mitjà es calcula mitjançant la Fórmula 3.1 i és un valor que serveix per tenir una visió molt general de si existeix un desplaçament generalitzat de les respostes respecte el GT.

$$DespMitjà = \frac{\sum_{i=1}^n Resposta_i - classeGT_i}{nRespostes} \quad (3.1)$$

Desplaçament mitjà absolut Molt similar a la mètrica anterior, pretén mostrar l'error absolut per resposta al conjunt de l'experiment. Ve donat per la Fórmula 3.2.

$$DespMitjàAbsolut = \frac{\sum_{i=1}^n |Resposta_i - classeGT_i|}{nRespostes} \quad (3.2)$$

A més de totes aquestes mètriques, que són útils per poder tenir una vista general de l'experiment, també es defineixen altres mètriques específiques per realitzar un anàlisi basat en les diferències entre classes. Totes elles es calculen usant només les **imatges d'una donada classe**.

Nombre d'imatges Quantes imatges s'han etiquetat de la classe a tractar.

Nombre de respostes Quantes respostes s'han donat a imatges de la classe a tractar.

Mitjana Mitjana aritmètica de totes les respostes donades a imatges de la classe a tractar.

Desviació típica Desviació típica de totes les respostes donades a imatges de la classe a tractar.

Desplaçament mitjà D'igual manera que en l'apartat anterior, mètrica calculada mitjançant la fórmula 3.1, només tenint en compte les respostes donades a imatges de la classe a tractar.

Juntament amb aquestes mètriques de classe, sempre es mostrarà una graella on apareix la distribució de les respostes sobre imatges de cada classe. Gràcies a aquesta taula es pot observar fàcilment el percentatge de respostes entre classes i les desviacions segons l'esperat.

3.1.4. Nova classificació

Un dels objectius del projecte és, a partir de les respostes dels experiments, intentar obtenir una nova classificació per cada imatge amb la intenció de perfeccionar l'etiquetat original. Aquesta millora de la classificació pot servir a realitzar diferents tasques com:

- **Validar el GT** comparant-ne les classes, especialment en les imatges de les quals se'n dubta el correcte etiquetat.
- Comparar la **resposta dels usuaris**, per veure en quines mostres s'obtenen classes unànimes i en quines altres una major diferència.
- Crear diferents conjunts d'entrenament per un **classificador automàtic** i estudiar com es comporta davant diferents classificacions.

Aquestes noves classificacions es generen a través de l'**Algorisme 3.1**, que pondera les respostes de les imatges *ad hoc* per aquests experiments. Donat que es preveuen 5 etiquetes/imatge, es poden definir pesos fixats per les respostes dels usuaris i per la classificació original al **GT**. En un cas general, es pot dir que s'estaria assignant aprop d'un 30% a la classificació original, i la resta al *feedback* que ens donen els usuaris.

```
1 function calculaNovaClasse(classificacio_GT, respostesUsuaris){
2   PES_GT = 2
3   PES_RESPOSTA = 1
4   puntuacio = [0, 0, 0, 0, 0]
5
6   puntuacio[classificacio_GT] += PES_GT
7   foreach resposta in respostesUsuaris do{
8     puntuacio[resposta] += PES_RESPOSTA
9   }
10
11   millorClasse = -1
12   maxPuntuacio = 0
13   foreach idx, punt in puntuacio do{
14     if punt > maxPuntuacio then{
15       maxPuntuacio = punt
16       millorClasse = idx
17     }
18   }
19   return millorClasse
20 }
```

Algorisme 3.1: Càlcul de noves classes

3.2. Experiment basat en color

El primer experiment realitzat es basa en la selecció per color. La tasca de l'usuari consisteix en identificar quin color dels quals es mostren als selectors es sembla més al to de pell de la imatge.

3.2.1. Preparació i execució

Preparació

En aquest primer experiment els selectors de color estan ordenats de forma ascendent seguint l'escala Fitzpatrick, per tal que l'usuari tingui la màxima facilitat a l'hora de comparar dues classes properes. A la **Figura 3.1** es pot observar un exemple de tasca per un usuari, on s'aprecia l'ordre i estil dels selectors.

Les imatges apareixen de forma aleatòria, sempre que la imatge en qüestió no s'hagi respost en un mínim d'ocasions. En concret, per aquest experiment el límit de respostes

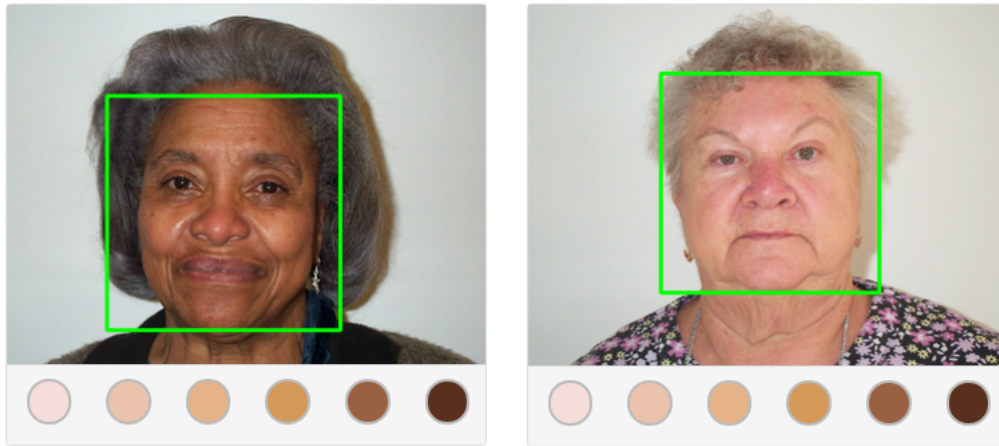


Figura 3.1: Exemple de mostres/selector

per imatge es situa en 5 respostes. Amb això es pretén obtenir el màxim nombre de imatges amb mostres suficients com per poder veure la dispersió de respostes i a més realitzar-ne una nova classificació.

Execució

Les expectatives de participació eren bastant elevades, ja que per contrastar totes les imatges amb 5 etiquetades, calien 1285 experiments realitzats completament. Donat que durant l'execució de l'experiment es va comprovar que el nombre de participants era més baix de l'esperat, i no es podia assegurar arribar al llindar, es va preparar una solució per obtenir dades interessants amb menys participació.

Donat que la baixa participació comportava la possibilitat de tenir una gran dispersió de respostes per imatge, es va plantejar una solució incremental. Aquest mètode va consistir en agrupar les imatges en blocs, de tal manera que les respostes estiguessin més agrupades. A les imatges es va assignar un bloc, formant blocs de 100 imatges les classes dels quals estaven uniformement distribuïdes.

La funció d'aquests blocs era que les imatges es responguessin forma controlada. En un precís instant, els experiments només oferien als usuaris imatges d'un bloc, mantenint-se actiu mentre no s'haguessin obtingut el mínim de respostes per les imatges d'aquell bloc. Un cop totes les imatges d'un bloc havien estat etiquetades 5 vegades, s'obria el següent bloc i es tancava l'anterior.

Gràcies a aquesta tècnica es van poder agrupar les respostes el suficient com per tenir valors significants sobre les imatges, sense perdre el factor d'aleatorietat vers les imatges mostrades.

3.2.2. Anàlisi de resultats

Un cop finalitzat l'experiment, amb un total de **617 participants** (el qual suposa 1234 respostes), es va procedir a l'anàlisi de les respostes. Cal tenir en compte que no

3. EXPERIMENTS

totes les respostes dels experiments entraren dins l'anàlisi, ja que només es tractaren les imatges amb un **mínim de 3 respostes**, descartant de l'estudi aquelles amb informació suficientment rellevant.

Ground Truth

Per fer una observació general i inicial, es varen comparar les respostes dels usuaris amb el que s'esperava en relació al **GT**.

Nombre de respostes: 1107

Nombre de coincidències: 505

Percentatge de coincidència: 45.61%

Desplaçament mitjà: -0.100

Desplaçament mitjà absolut: 0.604

Es pot comprovar que el percentatge de coincidències no és tan elevat com es podria esperar, ja que ni 1/2 de les respostes coincideixen amb el **GT**. A més, els valors de desplaçament mitjà i mitjà absolut indiquen quan llunyanes són les respostes del que s'esperaria al **GT**. El desplaçament mitjà és molt proper a 0, el que vol dir que les desviacions donades sobre les imatges es troben disperses tant superiorment com inferiorment (és a dir, cap a les classes veïnes en ambdues direccions). Amb el desplaçament mitjà absolut es pot comprovar que hi ha un desplaçament significatiu entre les respostes i el **GT**.

Per comprovar l'origen del baix nombre de coincidències, es realitza un anàlisi basat en les classes de cada imatge. Aquesta metodologia ens permet comprovar si les imatges difícils de classificar es troben agrupades en alguna classe en concret o és un problema generalitzat.

Anàlisi per classes

Donada la necessitat d'entendre el comportament dels usuaris front cada tipus d'imatge, s'analitzen les respostes en funció de la classe original de la imatge. Així es podrà veure si el problema és general o es concentra en un subconjunt del **GT**.

Classe	1	2	3	4	5	6
#imatges	1	111	53	19	30	7
#respostes	6	556	268	91	153	33
Mitjana	1.6667	2.0737	2.444	4.1648	4.9804	5.4242
Desviació típica	0.8165	0.8468	0.8395	0.7343	0.7115	0.5019
Dif. absoluta	0.6667	0.0737	-0.556	0.1648	-0.0196	-0.5758

Taula 3.1: Anàlisi d'etiquetatge del primer experiment respecte el **GT**

Classe	Respostes d'usuaris (%)						
	1	2	3	4	5	6	
GT	1	50.0	33.33	16.67	0.00	0.00	0.00
	2	26.08	45.32	24.82	3.06	0.36	0.36
	3	10.07	47.01	32.46	9.33	1.12	0.00
	4	0.00	2.20	12.09	53.85	30.77	1.10
	5	0.65	0.00	1.96	13.73	65.36	18.30
	6	0.00	0.00	0.00	0.00	57.58	42.42

Taula 3.2: Distribució de respostes en relació al GT

Gràcies a la [Taula 3.1](#) i la [Taula 3.2](#) es pot veure numèricament la distribució de respostes i la resposta dels usuaris davant cada classe de les mostres. Tot i això, per analitzar les respostes d'una forma més visual i senzilla, s'ha recorregut a un gràfic boxplot (present a la [Figura 3.2](#)), on es pot comprovar fàcilment la dispersió i mitjana de respostes sobre cada classe.

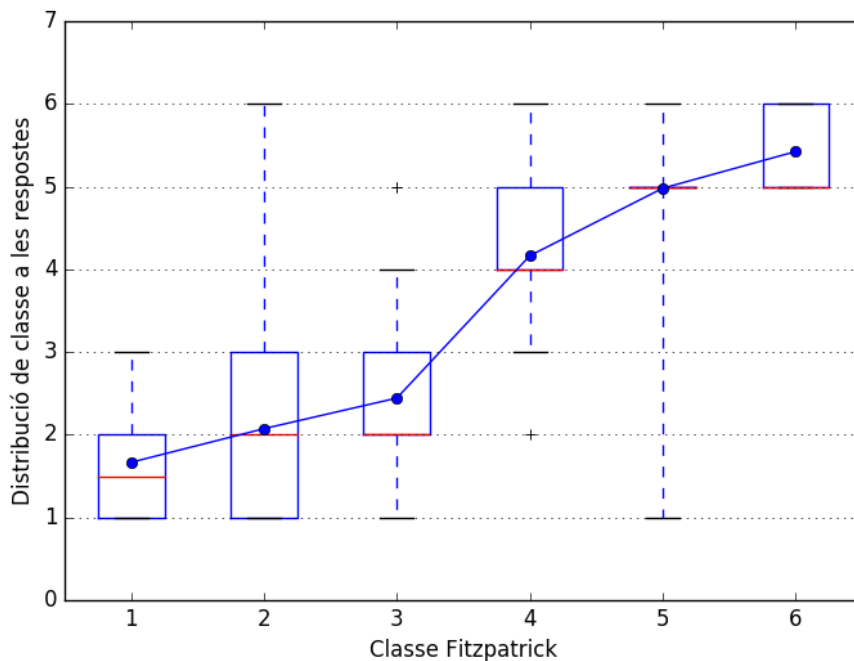


Figura 3.2: Boxplot sobre la distribució de respostes en l'experiment basat en color

Per analitzar els valors més dispersos o desplaçats de l'esperat, es cerca quina classe té més respostes etiquetades a valors llunyans a l'esperat. També es considera que les classes dels extrems no poden ser tractades amb normalitat, ja que només tenen una classe veïna. Això comporta que les respostes no es poden equilibrar superior i inferiorment, sinó que una resposta diferent a l'esperat sempre desviarà la classe. Un altre motiu per rebutjar la classe 1 és que dins el **GT** només es disposa d'una imatge

etiquetada amb aquest classe i, per tant no es pot analitzar acuradament.

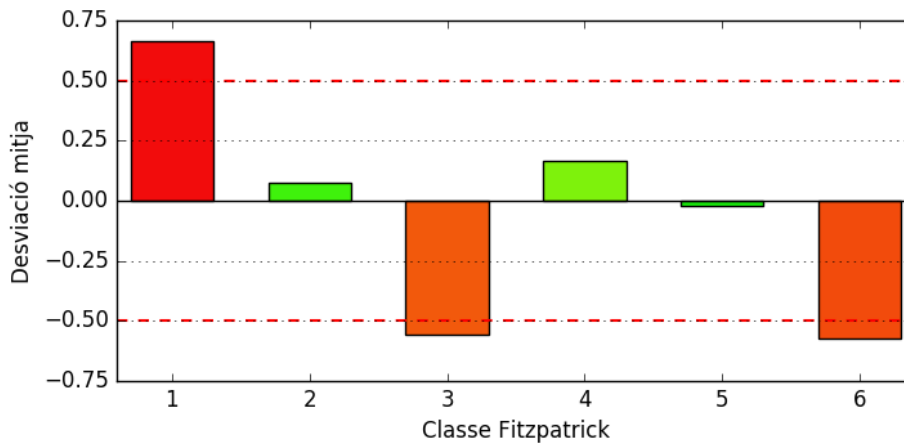


Figura 3.3: Desviació mitjana de les etiquetes respecte el GT

En la [Figura 3.3](#) es poden observar quines classes tenen respostes més desviades. Superar el llindar de 0.5 suposa que es les imatges s'han classificat més properes a una classe diferent a l'esperada. Així doncs, podem veure clarament que els extrems són els més afectats, com és lògic, ja que no tenen valors amb els que compensar les desviacions. A més, són les classes amb menys imatges i respostes, el qual dificulta el seu estudi. El cas la classificació 1 resulta encara més extrem, doncs només té una imatge associada i per tant no pot ser analitzada.

Així com es pot veure que la mitjana de respostes de les classes 2, 4 i 5 són bastant properes a l'esperat, es veu clarament com la **classe 3 té una gran desviació inferior**, és a dir, es confon amb la classe 2.

Per tal de cercar una explicació a aquesta anomalia, s'estudien amb més detall les respostes relatives a la classe 3.

Mostres	Etiquetes	Coincidència	Mitjana	SD	Dif. absoluta
53	268	32.5%	2.444	0.8395	-0.556

Taula 3.3: Dades relatives a la classe 3

Classe 3 A la [Taula 3.3](#) es pot observar que la desviació típica no és especialment elevada en comparació a la resta de classes, així que les etiquetes proposades pels usuaris no són més disperses per la classificació 3. Donat que la dispersió no és elevada, cal descobrir per quin motius la mitjana es troba desplaçada cap a la classe 2.

Si observam la distància respecte la classificació desitjada, a la [Figura 3.4](#) veim que les hi ha un gran volum de mostres etiquetades fora del rang desitjable $[2,5-3,5]$, i en concret un 52,83% de les imatges es troben fora d'aquest rang.

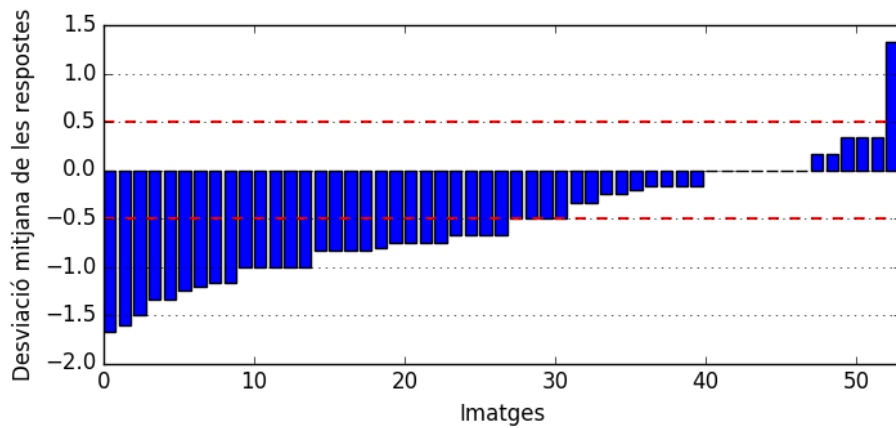


Figura 3.4: Desviació a les respostes d'imatge de tipus 3 en relació a l'esperat

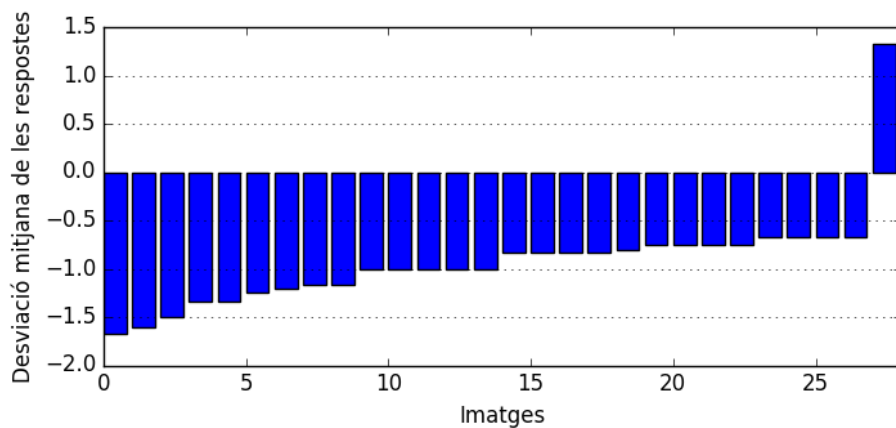


Figura 3.5: Imatges amb major desviació del tolerable

Si ens centram amb les imatges que no encaixen en el llindar desitjat, com es pot veure a la [Figura 3.5](#), el principal problema de desplaçament és en direcció a la classe 2, i fins i tot en cas extrem, a la classe 1.

Així doncs, queda clar que existeix una problemàtica a l'hora de diferenciar les classes 2 i 3 per part dels usuaris amb aquest sistema.

3.2.3. Discussió

Donat que la dispersió de les respostes és relativament baixa, amb una **desviació típica inferior a 1.0** es pot afirmar que l'etiquetat dels usuaris ha resultat consistent, tot i haver-se allunyat del **GT** en la classe 3

Com ja s'ha exposat anteriorment, no és viable realitzar una discussió de les classes 1 i 6, a causa dels següents motius:

- Un baix nombre de mostres impedeixen l'estudi d'aquestes classes. Únicament disposam d'una imatge classificada com a classe 1 i només 7 mostres per la classe 6.
- Les classes extremeres no tenen dues classes veïnes, com si ho fan la resta. Això suposa que tota desviació d'aquestes classes serà negativa, ja que no és possible compensar les respostes diferents a l'esperat. Això pot implicar una dificultat pels usuaris a l'hora de seleccionar mostres amb valors extrems, ja que és complex identificar si la pell es troba en el valor més extrem.

A més, pel que fa a les classes 4 i 5 observam una mínima desviació general, així que podem dir que els usuaris coincideixen en gran mesura amb la classificació de l'expert.

Amb tot això, les classes que més interès han resultat tenir són les 2 i 3, a causa de la seva aparent proximitat. La distància entre mitges d'aquestes classes és tan sols de **0,3703**, quan el desitjable és que sigui 1,0. Crida especialment l'atenció que la majoria d'imatges de classe 2 estiguin ben classificades (mitjana d'etiquetat: 2.0737) i que les de classe 3 no. Analitzant aquest fet, s'ha arribat a la conclusió de que hi ha dos possibles factors per aquests resultats:

1. Les classes 2 i 3 podrien difícilment diferenciables per usuaris sense experiència o coneixements sobre com classificar en l'escala Fitzpatrick. Els usuaris no entrenats podrien percebre les classes 2 i 3 com una única unitat que identificarien com a tipus 2, i per això aquesta ha obtingut bons resultats.
2. Un mal plantejament de l'experiment, concretament dels selectors de classe, ha desencadenat aquest resultat. Donat que les tasques dels experiments es realitzen molt ràpidament, l'ordre dels selectors pot ser un factor clau a l'hora de seleccionar la resposta. Si l'usuari sempre trobar abans el selector de classe 2 que el de 3, i a més considera que el de classe 2 s'assembla suficientment a la mostra, probablement en cas de dubte seleccioni aquesta com a resposta vàlida. Com conseqüència d'aquesta teoria, s'ha desenvolupat un **nou experiment** per veure en quina mesura pot afectar l'ordre dels selectors a imatges de classe 2 i 3.

3.2.4. Nova classificació

El principal objectiu dels experiments és obtenir informació addicional sobre cada mostra i, si escau, una nova classificació. En aquest apartat s'exposa la metodologia seguida per obtenir una nova classificació i quines diferències presenta respecte el **GT**

Ponderació de respostes Així doncs, una vegada aplicat l'**Algorisme 3.1** a les mostres analitzades en l'apartat anterior, es genera una nova classificació cada mostra. La relació d'**imatges/classe** es pot consultar a la **Taula 3.4**.

A la **Taula 3.4** hom pot observar com el volum de les classes 1 i 3 s'ha modificat considerablement, pel qual sembla que s'ha produït una reestructuració vers les classes 1-3.

	C1	C2	C3	C4	C5	C6
GroundTruth	1	111	53	19	30	7
NovaClassificacio	20	109	35	19	32	6

Taula 3.4: Nombre de mostres per classe

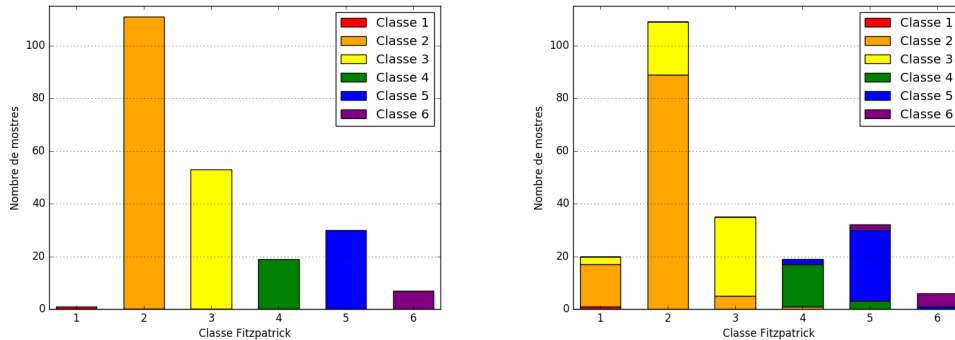


Figura 3.6: Redistribució del volum de mostres/classe

Per analitzar el comportament de les classes 1-3 es pot consultar la [Figura 3.6](#), que mostra com s'han desplaçat les imatges cap a noves classes. Hom pot comprovar com les imatges de les classes 1-3 es reestructuren desplaçant-se cap a classes inferiors, destacant algunes mostres de classe 3, que s'han desplaçat tant cap a la 1 com a la 2. També es pot comprovar com algunes mostres les classes 4-6 es desplacen, però en menor mesura que les 1-3.

Aquesta observació vers les classes 1-3 no només no resolen la problemàtica abans detectada per classificar correctament les classes 2 i 3, sinó que també s'hi afegeix la classe 1. Com es pot comprovar, fins i tot mostres classificades al **GT** com a classe 3 es col·loquen a la classe 1. Això reforça la idea de que els usuaris troben dificultats per diferenciar les classes 2 i 3.

Tot i això, si es determina la validesa de les noves classes, s'obté una possible millora a la classificació de les mostres. Això afecta especialment a la classe 1, que a partir d'aquest moment, gràcies l'augment de volum, podria ser estudiada.

3.3. Experiment bastat en mostres

Aquest experiment basat en la comparativa entre mostres planteja una alternativa a l'anterior, basat en color. La tasca de l'usuari consisteix en identificar quina mostra (*en forma d'imatge*) de les quals es visualitzen als selectors es sembla més al to de pell de la imatge.

3.3.1. Preparació i execució

Preparació

En aquest experiment l'ordre de les classes als selectors és totalment aleatori. Amb això es pretén evitar que l'usuari identifiqui l'ordre de les mostres amb una representació lineal de les classes. Així doncs, l'usuari en lloc de comparar els selectors entre ells, només compara cada un dels selectors amb la imatge de la mostra. Com es pot comprovar a la [Figura 3.7](#), a més de l'ordre, respecte a l'experiment basat en color també ha canviat la mesura dels selectors, que ha augmentat.

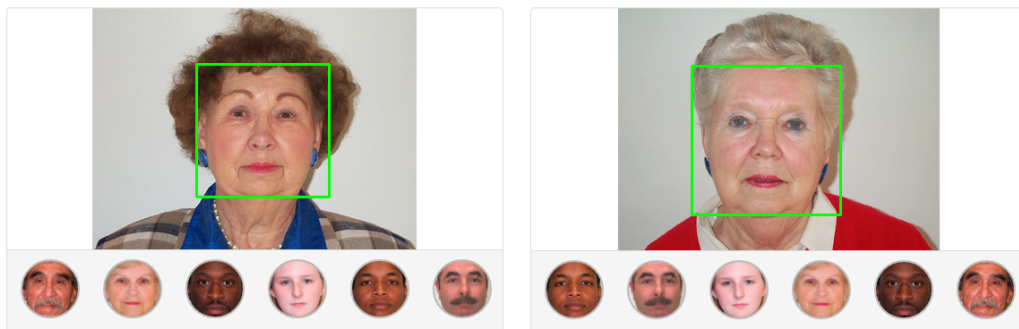


Figura 3.7: Exemple de mostres-selectors en un experiment basat en mostres

Una de les claus de l'èxit d'aquest experiment és la bona selecció de les imatges que apareixen als selectors com a referència. Principalment existeixen dos plantejaments possibles per assignar les imatges als selectors:

Mostres estàtiques Els selectors mostren les mateixes imatges per totes les mostres, tasques i usuaris. Aquest mètode suposa obtenir dades més estables i fàcils d'analitzar, ja que tots els usuaris estan responenent al mateix *input*.

Mostres dinàmiques Els selectors van canviant les imatges que mostren com a referència. Aquest mètode podria modificar les imatges en base a diferents paràmetres com el nombre de tasques realitzades per un usuari, la data, etc. En una etapa inicial d'un gran projecte podria resultar interessant modificar les imatges dels selectors per rangs de dates i així poder analitzar com responen els usuaris als canvis i a cada imatge assignada als selectors.

Per aquest experiment s'han utilitzat selectors amb mostres estàtiques, ja que per obtenir dades útils de selectors amb mostres dinàmiques caldria realitzar experiments amb major durada i volum de participació.

Com resulta evident, la tria de bons selectors és clau per l'aplicació d'aquest mètode. Per seleccionar les mostres a usar, s'ha creat criteri de selecció basat tant en respostes d'usuaris al primer experiment, com del **GT**. Aquest és un senzill algorisme que cerca el màxim nombre de respostes comunes sobre una imatge per tal d'etiquetar-la com a vàlida per representar una classe.

```

1 function calculaMillorMostra (conjuntImatges) {
2     PES_GT = 2
3     PES_RESPOSTA = 1
4
5     maxPuntuacions = [-1, -1, -1, -1, -1, -1]
6     idsSeleccionats = [-1, -1, -1, -1, -1, -1]
7     foreach imatge in conjuntImatges do{
8         puntuacioActual = [0, 0, 0, 0, 0, 0]
9
10        puntuacioActual[imatge["classeGT"]] += PES_GT = 2
11        foreach resposta in imatge["respostesUsuaris"] do{
12            puntuacioActual[resposta] += PES_RESPOSTA
13        }
14
15        foreach idx, maxPunt in maxPuntuacions do{
16            if puntuacioActual[idx] > maxPunt then{
17                maxPunt = puntuacioActual[idx]
18                idsSeleccionats[idx] = imatge["ID"]
19            }
20        }
21    }
22    return idsSeleccionats
23 }

```

Algorisme 3.2: Definició de mostres per els selectors

Una vegada s'aplica l'**Algorisme 3.2** s'obtenen imatges que representen bé la seva classe. Caldria indicar que en cas d'empat a puntuació, les mostres es seleccionen aleatòriament, i més concretament, en l'ordre que es trobin emmagatzemades dins l'estructura de dades *conjuntImatges*. A la **Figura 3.8** es poden observar les imatges resultants del càlcul anterior i la classe que representen.

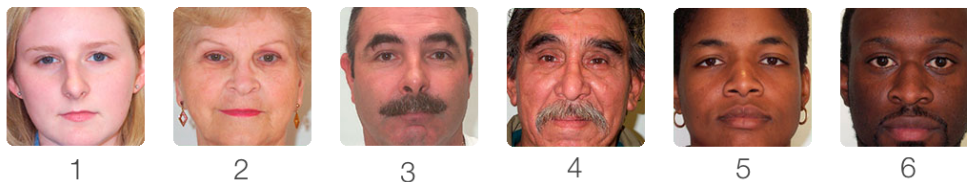


Figura 3.8: Mostres seleccionades com a selectors per cada classe

Novament, les imatges apareixen de forma aleatòria i cada mostra només apareix fins haver-se etiquetat en 5 ocasions. Tot i que el comportament de la selecció de tasques és idèntic al de l'experiment bastat en color, les imatges que poden aparèixer no ho són. Això és degut a que en aquest experiment només es treballa amb imatges ja etiquetades en l'experiment anterior per tal de tenir el màxim volum d'informació sobre un subconjunt del **GT** i poder comparar els resultats entre experiments.

En aquesta ocasió s'ha seguit des de l'inici de l'experiment l'estructura de blocs utilitzada durant l'experiment bastat en color. Aquesta estructura permet tenir un major control sobre el nombre de mostres totalment etiquetades, per tal de minimitzar el nombre de mostres amb respostes insuficients un cop hagi finalitzat l'experiment.

Execució

Durant el període d'execució no varen sorgir gaire problemes inesperats i gràcies a la distribució de mostres per blocs, en aquest cas de 37 mostres cadascun, l'assignació de tasques no va suposar un problema.

La única anomalia que es va observar va ser una petita errada a l'assignació de mostres, que en alguns casos permetien l'etiquetat en més de 5 ocasions per algunes mostres. Això és degut a que el sistema de selecció de mostres només té en compte les mostres amb tasques completades, però no les assignades a un treball en un determinat instant. A causa d'aquesta mancança, una mateixa mostra es podia assignar a la tasca de més d'un usuari alhora i, sota algunes circumstàncies, això va desencadenar mostres amb més de 5 respostes a causa de la realització de tasques simultànies. Tot i esser conscients de l'error, es va estudiar i es va decidir no modificar el sistema a causa de dos motius:

- Aquesta anomalia, lluny de ser un problema greu vers l'experiment, només atorgava més informació a algunes mostres. És evident que aquestes tasques haurien d'haver sigut assignades a altres mostres per maximitzar el nombre de imatges respostes, però un major volum d'etiquetat a una imatge en cap cas és un fet negatiu.
- Les modificacions necessàries al sistema per tal que gestionés aquest comportament eren massa costoses pel benefici que comportarien.

Tot i el poc impacte final d'aquesta mancança, s'ha entès l'origen de l'error i com solucionar-lo, quedant així com una lliçó apresada per a futurs sistemes d'experiments o **HbC** distribuïts per la xarxa.

3.3.2. Anàlisi de resultats

Un cop finalitzat l'experiment, amb un total de **374 participants** (el qual suposa 748 respostes), es va procedir a l'anàlisi de les respostes. Novament no totes les respostes dels experiments entraren dins l'anàlisi, ja que es mantengué el llinard d'un **mínim de 3 respostes** per mostra, descartant de l'estudi aquelles amb informació suficientment rellevant.

Ground Truth

Per fer una observació general i inicial, es varen comparar les respostes dels usuaris amb el que s'esperava en relació al **GT**.

Nombre de respostes: 713

Nombre de coincidències: 322

Percentatge de coincidència: 45.162%

Desplaçament mitjà: -0.0336

Desplaçament mitjà absolut: 0.6031

En aquestes dades es pot observar que el rati d'encert és elevat, gairebé la meitat dels usuaris coincideixen amb el **GT** i el desplaçament mitjà sembla baix. Per poder confirmar aquestes aproximacions inicials, es realitza l'anàlisi per classes.

Anàlisi per classes

Classe	1	2	3	4	5	6
#imatges	1	67	36	13	15	2
#respostes	6	360	186	66	83	12
Mitjana	1.6667	2.1472	2.5914	4.5909	5.0482	6.0
Desviació típica	0.5164	0.8163	0.8666	0.7011	0.714	0.0
Dif. absoluta	0.6667	0.1472	-0.4086	0.5909	0.0482	0.0

Taula 3.5: Anàlisi d'etiquetatge de l'experiment basat en mostres respecte el **GT**

Per analitzar els valors més dispersos o desplaçats, es cerca quina classe té més respostes llunyanes a l'esperat. Cal tenir en compte que les classes 1 i 6, a causa del seu baix nombre d'imatges i respostes, poden no aportar resultats significatius.

Resulten especialment preocupants els percentatges de coincidència i les mitjanes allunyades a l'esperat de les classes 3 i 4. A més de la distància amb el **GT**, també cal analitzar la distància entre les seves mitjanes, que essent de 1.995 es situa en gairebé **el doble del desitjable**. Aquest espai pot ajudar a discernir millor les classes 3 i 4, però dificulta molt la relació 2-3 i 4-5.

Com aspecte positiu es pot comprovar que els valors relacionats amb les classes 5 i 6 són molt propers al **GT**. Un 100% de coincidència en la classe 6 és, tot i el baix nombre de respostes, un gran èxit. Pel que fa a la classe 5, el que destaca de forma especialment positiva és el valor de la mitjana, que es situa només a 0.0482 de l'esperat.

Analitzant la **Figura 3.10** hom pot veure clarament que les classes 1, 3 i 4 estan molt allunyades de l'esperat, i fins i tot les classes 1 i 4 superen el llindar de 0.5. Així

Classe	Respostes d'usuaris (%)					
	1	2	3	4	5	6
1	33.33	66.67	0.0	0.0	0.0	0.0
2	19.72	51.94	23.06	4.72	0.28	0.28
3	5.91	47.85	27.96	17.74	0.54	0.00
4	0.00	1.52	7.58	21.21	69.70	0.00
5	0.00	1.20	2.41	8.43	66.27	21.69
6	0.00	0.00	0.00	0.00	0.00	100.00

Taula 3.6: Distribució de respostes en relació al **GT**

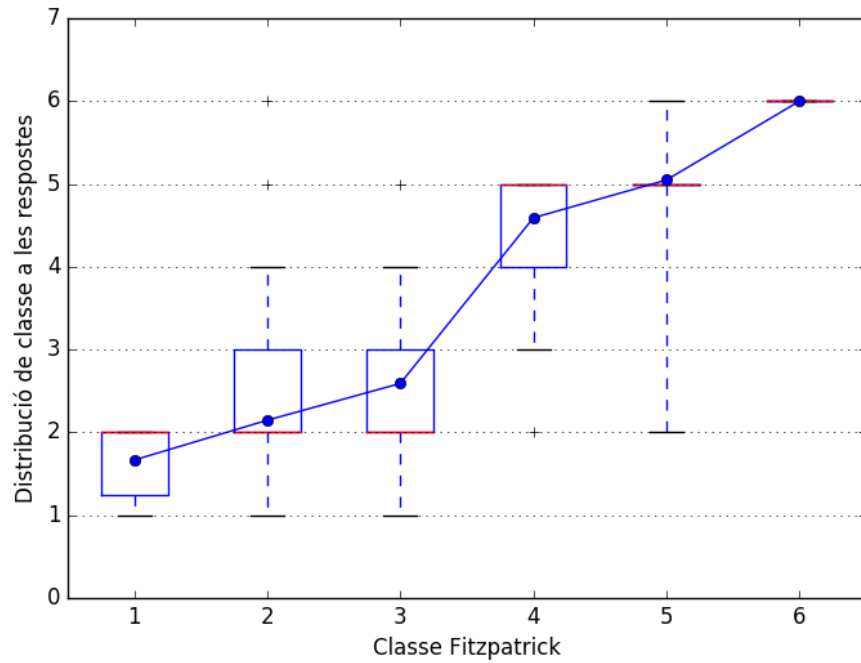


Figura 3.9: Boxplot sobre la distribució de respostes en l'experiment basat en mostres

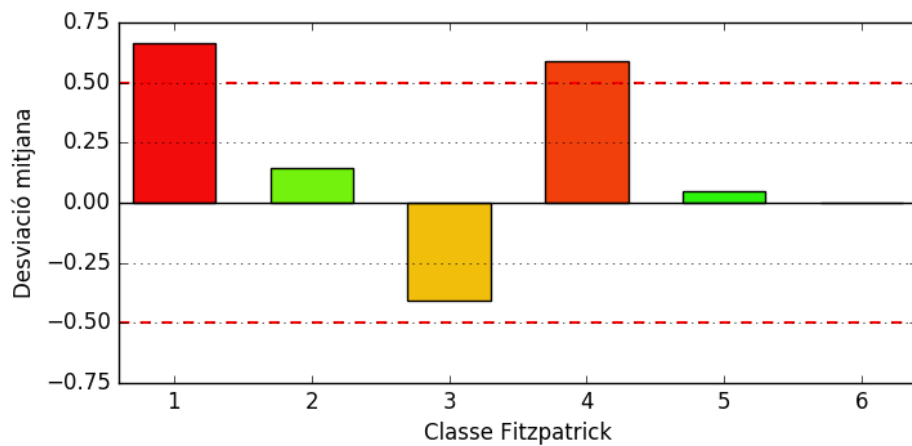


Figura 3.10: Desviació mitjana de les etiquetes respecte el GT

doncs, caldria analitzar-les, però com s'ha exposat, les mostres per les classes 1 i 6 són insuficients, així que es tractaran les classes 3 i 4.

Classe	Mostres	Etiquetes	Coincid.	Mitjana	SD	Dif. absoluta
3	36	186	27.96%	2.5914	0.8666	-0.4086
4	13	66	21.21%	4.5909	0.7011	0.5909

Taula 3.7: Dades relatives a la classes 3 i 4

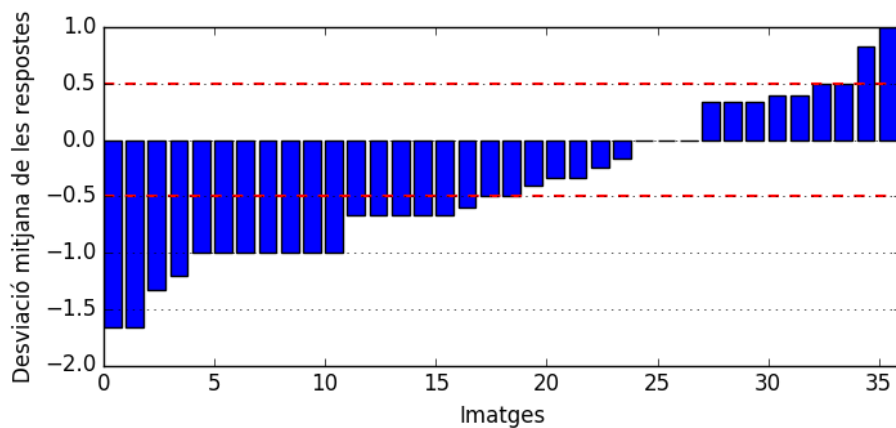


Figura 3.11: Desviació a les respostes d'imatge de tipus 3 en relació a l'esperat

Classe 3 Si observam la distància respecte la classificació desitjada, veim que les hi ha un gran volum de mostres etiquetades fora del rang desitjable $[2,5-3,5]$, i en concret un 52,77% de les imatges es troben fora d'aquest rang.

Si ens centram amb les imatges que no encaixen en el llinar desitjat, podem veure que el principal problema de desplaçament és en direcció a la classe 2, i en dos casos la distància és tan gran que s'acosten a la classe 1. Només dues mostres surten del llinar cap a la classe 4 i en cap moment la superen (la distància mai és superior a 1.0)

Així doncs, les imatges classificades al **GT** com a classe 3 tenen un problema desviació a l'hora de ser classificades, ja que el **44% de les respostes** dels usuaris s'han dirigit cap a la classe 2.

Classe 4 A la classe 4, sembla que el problema és semblant però en sentit oposat, ja que la seva mitjana (4.59) s'acosta a la classe 5. Gràcies a la **Figura 3.12** es pot a visualitzar més fàcilment la desviació de de les respostes cap a la classe superior.

Al diagrama podem observar clarament com el nombre de mostres dins el rang acceptable, $[3,5-4,5]$, és baixíssim. Concretament, un 61.54% de les imatges es troben fora d'aquest rang, essent aquest cas encara més distant del **GT** que la classe 3.

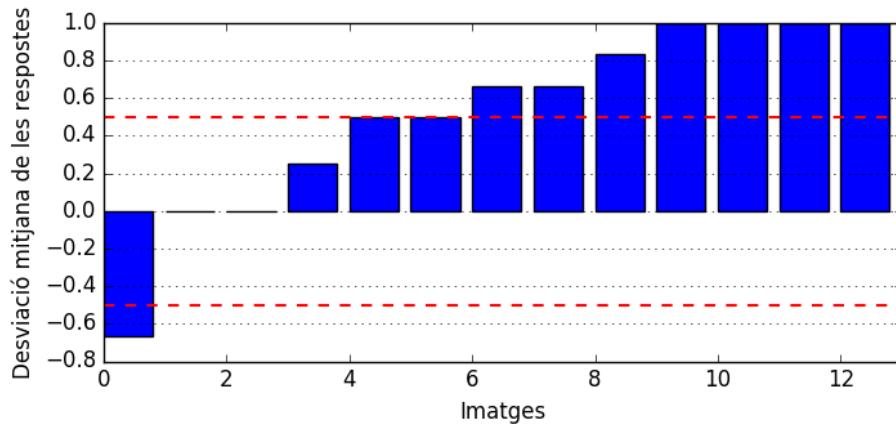


Figura 3.12: Desviació a les respostes d'imatge de tipus 4 en relació a l'esperat

3.3.3. Discussió

Com ja s'ha exposat anteriorment, no és viable realitzar una discussió de la 1 a causa del baix nombre de mostres (una única mostra amb 6 respostes).

Llevat del cas de la classe 1, a la [Figura 3.9](#) es pot observar mitjançant el digrama de caixes com, tot i haver desviacions a les classes 3 i 4, les respostes dels usuaris són bastant consistents entre elles.

Aquesta poca dispersió indica que els usuaris reconeixen similituds entre les mostres i els selectors. Tal vegada amb una millora de les mostres dels selectors es podria aconseguir d'igual manera una baixa dispersió, i una reducció a la desviació respecte la classe objectiu. Així doncs, sembla que aquest mètode amb unes mostres que que definissin prou bé les classes als selectors podria resultar molt efectiu.

A això cal afegir-hi el bons resultats de mitjana de les classes 5 i 6, que gairebé han resultat immillorables, ja que les distàncies cap a la classe del **GT** són mínimes.

3.3.4. Nova classificació

Aplicant l'[Algorisme 3.1](#) pel càlcul de noves classes basat en les respostes de l'experiment, s'obté una nova classificació per cada mostra de la **BD**. Es pot consultar el nombre d'imatges per classe a la [Taula 3.8](#).

	C1	C2	C3	C4	C5	C6
GroundTruth	1	67	36	13	15	2
NovaClassificacio	10	66	24	10	20	4

Taula 3.8: Nombre de mostres per classe

Per analitzar el comportament de la nova classificació es pot consultar la [Figura 3.13](#), que mostra com s'han desplaçat les imatges cap a noves classes. Resulta especialment

sorprenent com s'han recol·locat les imatges que originalment eren de classe 2 i 3, ja que s'han distribuït entre les classes 1-4, essent destacables els canvis que comporten una distància major que 1 (com per exemple els moviments $2 \rightarrow 4$ i $3 \rightarrow 1$).

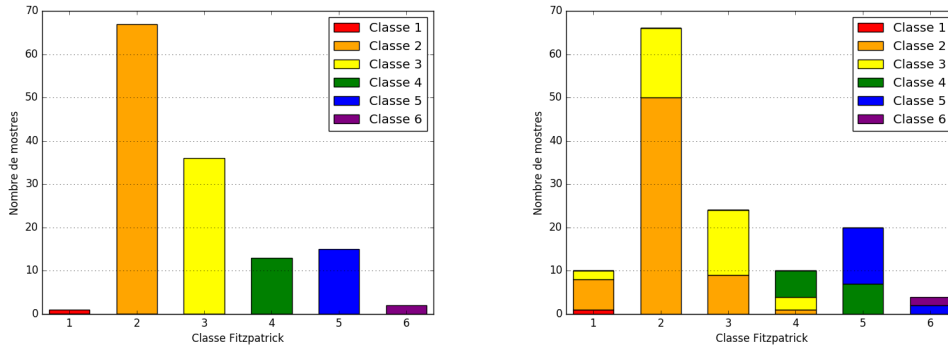


Figura 3.13: Redistribució del volum de mostres/classe

Com a totes les mostres tractades al llarg de l'estudi, podem observar com existeix una clara dificultat per definir correctament l'espai de les classes 2-3, ja que les imatges intercanvien l'etiquetat entre elles.

De cara a futurs usos d'aquesta nova classificació, com ara el sistema classificador automàtic, les classes 1 i 6 es veuen reforçades gràcies a que inicialment tenien molt poques mostres i han augmentat en volum.

3.4. Experiment bastat en color II

En resposta als resultats observats a l'experiment basat en color (Secció 3.2) vers les classes 2 i 3, es va preparar un experiment de curta durada per comprovar si, com s'exposa a la discussió de l'experiment basat en color, l'ordre dels selectors pot influir en l'usuari a l'hora de classificar.

3.4.1. Preparació i execució

Les principals novetats respecte l'experiment original són dues: la selecció d'imatges a etiquetar i els canvis als selectors.

Imatges Pel que fa a les imatges, en aquest cas es va cercar minimitzar l'aleatorietat. Donat que la planificació de l'experiment es va realitzar pensant en un volum baix de participació, es seleccionaren únicament 6 imatges sobre les quals recaptar la màxima informació possible. Com és evident, aquestes imatges són de les classes 2 i 3, ja que és la característica final a estudiar en aquest experiment. El mètode de selecció de les mostres es va basar en l'experiment original. Es seleccionaren tres imatges de la classe 2, i tres més de la classe 3. Les imatges seleccionades tenen el màxim nombre de respostes possibles al primer experiment, a més de ser diverses en gènere i edat de les persones que hi apareixen.

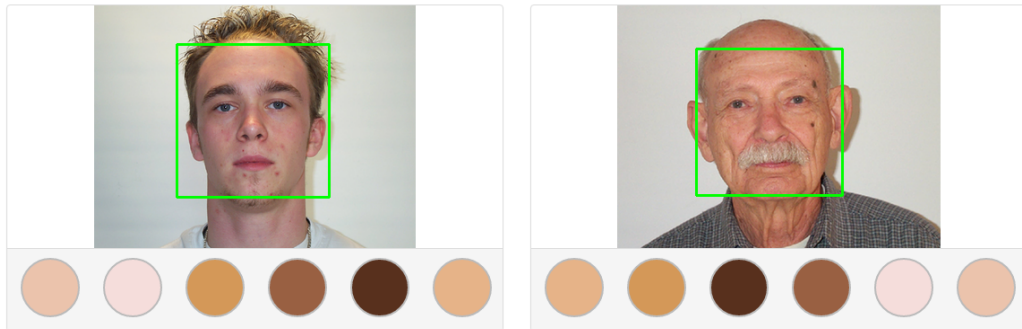


Figura 3.14: Exemple de la reorganització dels selectors basats en color

Selectors Com es pot comprovar a la [Figura 3.14](#), els selectors es mantenen amb el mateix color per identificar cada classe. Tot i això, se n'augmenta la mida i, el que és més important, es reorganitzen. Tot i que la distribució pot semblar aleatòria, realment no ho és. La intenció és fer pensar a l'usuari que no segueix cap ordre lògic, de tal manera que hagi de comprovar un per un els colors en comparació a la mostra. Amb això, l'ordre real ve donat per que als selectors sempre es troben les classes 2 i 3 als extrems, mentre que la resta s'organitzen, ara si, de forma aleatòria.

Aquests són els únics canvis respecte l'experiment original, ja que es mantenen tant les vies de comunicació com el flux de navegació per realitzar-lo completament.

3.4.2. Anàlisi de resultats

Un cop finalitzat l'experiment, amb un total de **105 participants** (el qual suposa 210 respostes), es va procedir a l'anàlisi de les respostes. L'experiment inicialment es va plantejar per una participació i durada molt curta, amb l'expectativa de que es complissin uns 20-30 experiments. Tot i això, la comunitat d'usuaris va realitzar difusió i finalment el nombre de tasques realitzades va ser superior. Aquest excés de dades, lluny de suposar un problema, permetrà realitzar un anàlisi més precís del comportament dels usuaris vers els factors d'aleatorietat.

	Exp. original		Exp. alternatiu	
	Classe 2	Classe 3	Classe 2	Classe 3
#respostes	18	18	105	105
Coincidència	22.22%	44.44%	59.05%	43.81%
Mitjana	2.1111	2.5	2.1429	2.8286
SD	1.0786	0.7071	0.8817	0.8929
Dif. absoluta	0.1111	-0.5	0.1429	-0.1714

Taula 3.9: Anàlisi per classe del segon experiment basat en color

A la [Taula 3.9](#) es pot observar com amb el nou experiment s'han obtingut dades més semblants a l'esperat (*GT*), ja que les mitjanes són més properes a la classe objectiu, especialment destacable el cas de la classe 3. També resulta notori l'augment de coincidències amb classe 2, gairebé **triplicant (x2.69) el rati d'encerts**.

3.4.3. Discussió

Donat el caràcter de l'experiment i el baix volum de mostres implicades, és factible realitzar una discussió per mostra, podent així observar la imatge per realitzar-ne una valoració qualitativa.

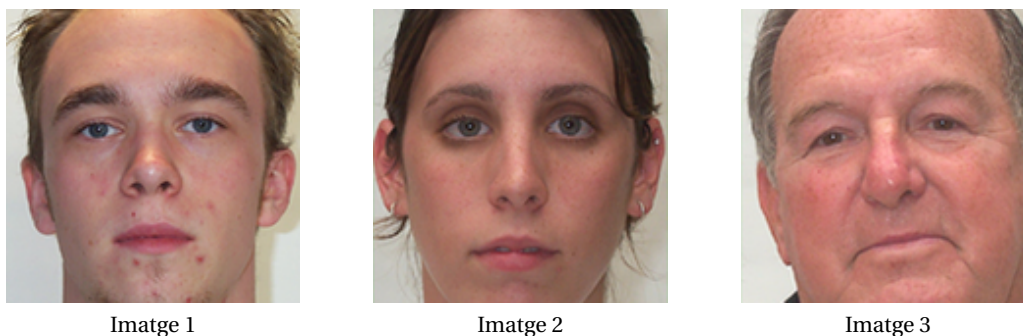


Figura 3.15: Mostres de la classe 2



Figura 3.16: Mostres de la classe 3

Com es pot comprovar tant a la [Figura 3.15](#) com a la [Figura 3.16](#), les mostres de les classes 2 i 3 són molt semblants a primera ullada. Podria semblar que les de tipus 3 tenen un to més taronjós, mentre que les de tipus 2 semblen més rosades.

Tot i això, cal tenir en compte un factor que no havia aparegut fins ara: les marques de la pell, així com l'edat semblen afectar molt a la pigmentació. Un exemple molt clar d'aquest cas és la Imatge 5, on la dona que hi apareix té una sèrie de taques a la pell que dificulten molt la identificació de la seva classe dins l'escala Fitzpatrick. Així doncs, possiblement l'edat de la persona de la mostra hauria de ser un factor a tenir en compte a l'hora de realitzar aquests tipus de classificacions. Una alternativa seria aplicar algun tipus de filtre suavitzant del color a la mostra abans de realitzar la tasca per tal de minimitzar aquestes imperfeccions.

Un altre factor que afecta molt a la imatge, és la il·luminació, doncs es pot veure clarament que les imatges 2 i 3 no comparteixen les mateixes condicions lumíniques. Això afecta al to que es representa a la imatge, i tot i que aquest factor pot esser corregit per la interpretació d'un humà, una màquina no corregeix aquestes desviacions si les

dades no es preparen amb anterioritat, el qual complica encara més la classificació automàtica.

Amb aquest experiment, doncs, lluny de poder aclarir la problemàtica entre les classes 2 i 3, queda encara més patent que les mostres són molt similars i que és comprensible la dispersió de respostes. Caldria analitzar per un futur si aplicant alguna modificació a la imatge, com ara un suavitzat de color o una correcció d'il·luminació aquesta tasca es simplificaria.

3.5. Comparativa entre experiments

Una vegada estudiats tots els resultats dels experiments, cal comparar les dades i conclusions obtingudes de cada un d'ells per assabentar-se dels avantatges i inconvenients que pot suposar l'ús de cada tipus de selectors.

Primer de tot cal tenir en compte que l'experiment basat en color II no és present en aquesta comparativa a causa que es va realitzar amb un objectiu molt concret, com a part de la discussió de l'experiment basat en color I. Entre altres, els principals motius per no tractar-lo en aquesta comparativa són:

- Baix nombre de mostres: Només es tracten 6 mostres en tot l'experiment. En comparativa a la resta d'experiments (221 i 134) és un valor baixíssim.
- Classes analitzades: Donat el caràcter de l'experiment, només es tractaren mostres de les classes 2 i 3, així que no té sentit realitzar una comparativa amb els altres, que tracten l'espectre de classes complet (1-6).

Així doncs, en aquesta secció analitzen les diferències i semblances entre els resultats dels diferents mètodes de selecció que s'han posat a disposició dels usuaris.

3.5.1. Participació

El volum de participació en els dos experiments ha sigut significativament diferent ja que l'experiment basat en color va ésser realitzat en 617 ocasions, mentre que la participació del basat en mostres es redueix a 374. Tot i semblar una diferència insalvable, les quantitats de mostres són suficients com per poder realitzar un anàlisi paral·lel.

3.5.2. Distribució de respostes

Per realitzar la comparativa s'analitzen les respostes per classes de cara experiment, així com s'ha realitzat a l'anàlisi de cada experiment.

Com es pot comprovar a la [Figura 3.17](#), la distribució de respostes té algunes diferències però segueix un patró relativament semblant als dos experiments. A la figura es poden observar un seguit de característiques interessants:

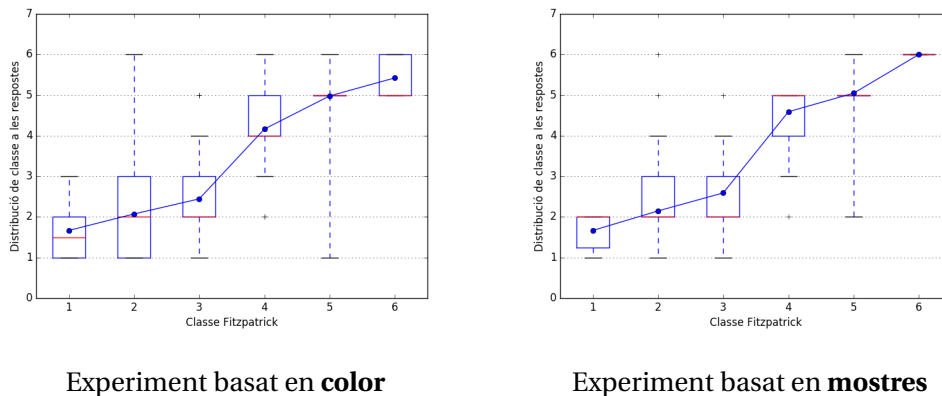


Figura 3.17: Comparativa de distribucions de les respostes

- Les etiquetes de l'experiment basat en color mostren una major precisió respecte al **GT** pel que fa a la mitjana de respostes. Aquesta característica seria útil en algun tipus d'experiment amb un gran volum d'etiquetes per imatge, així es podria reduir l'efecte de la dispersió i mantenir una bona mitjana de coincidència.
- Les etiquetes de l'experiment basat en mostres, tot i tenir majors desviacions de mitjanes respecte el **GT** són molt menys disperses, així que són dades útils quan es cerquin respostes vàlides amb pocs experiments. També s'hauria d'estudiar l'efecte de diferents mostres als selectors, per veure si els usuaris es comporten de la mateixa forma.
- Sembla que existeix una constatació de dificultat a l'hora de classificar les classes 1-3, ja que els usuaris les confonen amb facilitat.
- En ambdós casos les classes 5 i 6 semblen ben definides, ja que es mouen dins rangs acceptables. Especialment destacable el cas de la classe 5, ja que en ambdós experiments aconseguix una mitjana de respostes gairebé perfecta i tots els quartils es troben a la mateixa classe 5.
- La pendent més pronunciada en els dos casos és entre les classes 3 i 4, el qual suposa que són més fàcilment diferenciables per als usuaris, però també que es confonen amb les classes 2 i 5 respectivament. El cas ideal seria que la diferència entre totes les classes fora 1, però aquesta distància més pronunciada també podria ser útil si només es volguessin classificar imatges dels tipus 3 i 4 en concret.

Una vegada comprovades les diferències respectives a la distribució de respostes, anam a analitzar les diferències relacionades amb la desviació.

3.5.3. Desviació de respostes

A més de com disperses i encertades són les respostes dels usuaris, és necessari conèixer en quina mesura aquestes s'allunyen de l'esperat al **GT**.

3. EXPERIMENTS

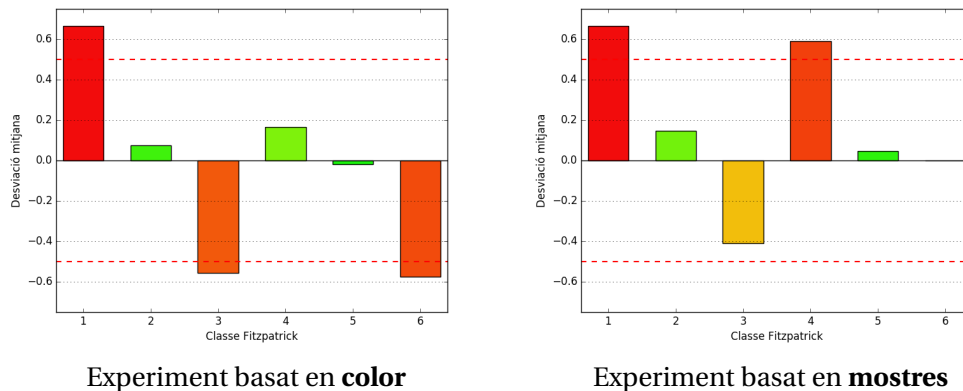


Figura 3.18: Comparativa de desviacions per classe

A la **Figura 3.18** es poden observar els desplaçaments de les respostes respecte la classe que s'esperava.

Com ja s'ha explicat a les pertinents anàlisis de cada experiment, les classes dels extrems (1 i 6) no poden ser avaluades correctament pel poc volum de mostres que se'n tenen. Tot i això, cal aclarir que la raó més probable de que la classe 1 tinguí tanta desviació a ambdós és que només es té una mostra de dita classe i no se'n pot fer cap estudi generalitzat per la classe, només per dita imatge.

A dita figura es pot observar com una de les classes en ambdós casos supera el llindar tolerable de la desviació ± 0.5 . També resulta interessant veure que en ambdós experiments les classes 1-4 es desvien la mateixa direcció, el qual pot suposar que els usuaris interpreten que el **GT** té errors i els identifiquen independentment del tipus de selector.

Resulten especialment sorprenents els canvis de desplaçament a les classes 4 i 6 entre experiments. Caldria tenir en compte que en el cas de la classe 6, al segon experiment el nombre de mostres imatges tractades (2) és molt més baix que al primer experiment (7), i les mostres del segon podrien ser més clares per als usuaris.

3.5.4. Nova classificació

Finalment, cal comparar la nova distribució de les mostres generada per l'**Algorisme 3.1** als dos experiments.

Com es pot observar a la **Figura 3.19**, les classes 1-3 s'han redistribuït de forma semblant als dos experiments. En ambdós casos es veu com la frontera 2-3 no està ben definida i molts d'usuaris dubten a l'hora de classificar. Diferent és el cas de les classes 5 i 6, on es troben diferències clares entre experiments. A l'experiment basat en color, les mostres de dites classes es mantingueren gairebé en la seva totalitat coincidents amb el **GT**. Això significa que els usuaris relacionaren molt bé el color del selector amb les seves mostres. Per altra banda, a l'experiment basat en mostres, aquestes classes

3.5. Comparativa entre experiments

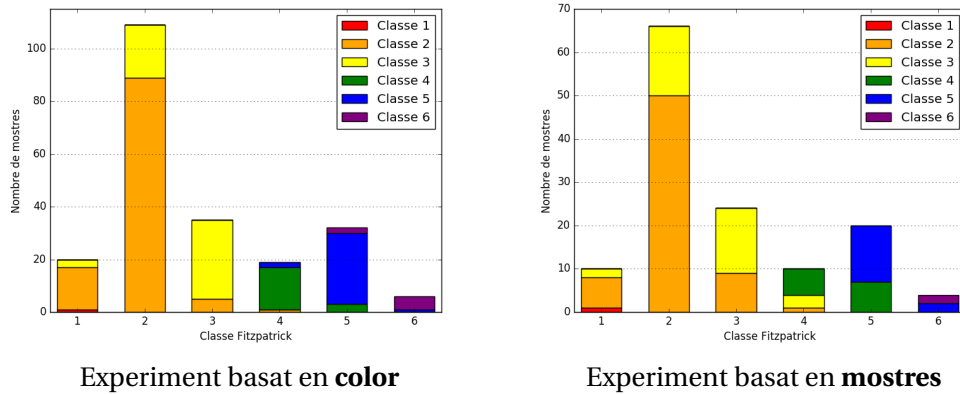


Figura 3.19: Comparativa de les noves classificacions

són molt més permeables. Tant és així, que més de la meitat de les imatges de classe 4 es desviaren cap a la classe 5.

Gràcies a aquest anàlisi queda clar que les principals diferències entre els experiments es donen a les classes 4-6, i per tant caldria analitzar quina de les dues classificacions és realment més efectiva per etiquetar imatges de les classes 4-6.

CLASSIFICACIÓ AUTOMÀTICA

Un dels objectius del projecte és comprovar la diferència d'efectivitat entre les classificacions d'imatges realitzades per humans (a través dels experiments i la **HbC**) i un algorisme de classificació automàtica. A més, també s'usen es classificadors automàtics per analitzar diferències entre les noves classificacions generades al capítol d'experiments.

4.1. Sistema de classificació

Per realitzar la classificació automàtica s'han realitzat diferents proves amb múltiples eines. Finalment, s'han utilitzat Gaussian Mixture Model (**GMM**) (*models de mixelures gaussianes*) donats els bons resultats que ens ha atorgat.

Un **GMM** és un model probabilístic que suposa que tots els punts de dades són generats a partir d'una combinació d'un nombre finit de distribucions Gaussianes amb paràmetres desconeguts. Hom pot interpretar la combinació de models com una generalització del *clustering* k-means per incorporar informació relativa a l'estructura de covariància de les dades, així com els centres de les Gaussianes latents [13].

L'objectiu d'aquest classificador automàtic no és crear una aplicació capaç de classificar definitivament una mostra qualsevol segons el seu tipus de pell. La intenció és més aviat fer una comparativa entre els diferents models de classificació que s'han aconseguit arrel dels experiments. Amb això s'aconseguirà veure quina via d'etiquetat dona millors mostres per construir una futura eina automàtica.

4.2. Tractament de dades

Per procedir a la classificació automàtica es treballa en dues etapes ben diferenciades:

1. La preparació de les dades i procés ensenyament del **GMM**.
2. La classificació automàtica i obtenció de resultats.

Per a la preparació de les dades, es generen dos subconjunts disjunts d'imatges a partir de les **mostres classificades**. El primer d'aquest subconjunts és el conjunt d'aprenentatge, mentre que el segon és el de validació. Per poder comparar els resultats del **GMM** és imprescindible que els conjunts de validació siguin els mateixos entre tots els classificadors diferents. Amb això el que s'aconsegueix és que, independentment de com hagi sigut la fase d'aprenentatge, la tasca a classificar pel **GMM** sempre és la mateixa.

4.2.1. Tractament previ de les mostres

Tot i que les passes definides són l'aprenentatge i classificació automàtica, cal tenir en compte que les imatges de mostra han de ser tractades per que el **GMM** les pugui fer servir. Donat que és un procediment comú tant per la passa d'aprenentatge com per la classificació, cal que es defineixi amb claredat abans d'explicar el procediment complet. Aquesta metodologia ve definida per les següents passes:

1. Mitjançant la llibreria OpenCV es **localitza la cara** dins l'imatge a tractar.
2. La **cara es subdivideix** mitjançant una graella de dimensions $N \times M$. Més concretament, per aquesta experimentació s'ha subdividit en una graella de 6×8 àrees uniformes.
3. Observant les mostres, s'arriba a la conclusió de que una bona regió per realitzar l'anàlisi són les **galtes**, ja que en el conjunt d'imatges que tractam són zones lliures d'imperficcions i reflexes de llum. A més, així també tenim la certesa de que no estam incloent zones sense interès, com ara zones capil·lars o oculars.
4. Es recullen tots els píxels de la regió d'interès de cada imatge i s'etiqueten amb la classe de la imatge. Aquests píxels es transformen a l'**esquema de color HSV**, ja que les components de saturació i valor són els que millor defineixen les característiques de cada classe.

4.2.2. Aprenentatge

A més, a l'hora de preparar les dades cal definir una metodologia de filtrat de les mostres, ja que només s'ha de tractar una regió d'interès, no tota la imatge. En aquest cas, la regió d'interès s'ha definit de la següent manera:

1. Es recullen les imatges a tractar a partir de la definició **conjunt d'aprenentatge**.
2. Es realitza el tractament d'imatges definit a la **Subsecció 4.2.1** per l'obtenció de dades d'interès. A més, s'agrupen totes les dades en una mateixa estructura, realitzant un tractament per píxel, i no per imatge.
3. S'itera sobre el nombre de classes desitjades. En un cas general aquesta iteració seria des de la classe 1 fins la 6, però donades les poques mostres relatives a la classe 1, només s'intentarà classificar en base a les classes 2-6.
 - 3.1.) Per la classe actual es genera una **GMM** amb **5 gaussianes independents** amb una **covariància de tipus complet** a partir de les dades obtingudes en el tractament anterior.
 - 3.2.) S'emmagatzema la **GMM** generada per classificar la classe actual

Seguint aquestes passes, aconseguim les **GMM** capaces de classificar noves imatges, concretament les definides al conjunt de validació.

4.2.3. Validació

Una vegada el **GMM** ha après a classificar mostres, cal validar-ne les seves habilitats. Novament, cal recordar que la intenció d'aquest experiment no és obtenir un sistema de classificació automàtica efectiu, sinó usar-lo com una eina de comparació entre les diferents classificacions que s'han generat.

Els conjunts de validació seran comuns a tots els **GMM** amb diferents aprenentatges. Aquest està compost per un total de 98 imatges, acompanyades evidentment de la seva classificació. El procés de classificació automàtica ve donat per les següents passes:

1. Es recullen les imatges a tractar a partir de la definició **conjunt de validació**.
2. Per cada imatge del conjunt, s'itera:
 - 2.1.) Es realitza el tractament d'imatges definit a la **Subsecció 4.2.1** per l'obtenció de dades d'interès.
 - 2.2.) S'itera sobre el nombre de classes desitjades. En consistència amb l'apartat d'aprenentatge, només es classifica en base a les classes 2-6.
 - 2.2..1.) Per les dades d'interès de la imatge actual, se'n consulta al **GMM** de la classe desitjada, quants de píxels es consideren de dita classe. Per comprovar si un píxel pertany a una classe, es consulta al **GMM** quina probabilitat té donat píxel de pertànyer a la classe.
 - 2.2..2.) Si aquesta probabilitat supera un llindar determinat, es considera que el píxel és vàlid per la classe donada. En el cas d'aquests experiments el llindar, o *threshold* en anglès, és de $8e - 6$, ja que realitzant diferents proves ha resultat ser un paràmetre que retorna valors prou acurats.
 - 2.3.) Es realitza el recompte de quants de píxels s'han assignat a cada classe.

- 2.4.) La classe amb més píxels vàlids és considerada la classe del total de la imatge.

Aquests procediments d'aprenentatge i classificació automàtica s'han aplicat a dos conjunts de mostres: **tot el conjunt del GT** i les resultants de la nova classificació de **l'experiment basat en color**. Les dades de l'experiment basat en mostres no s'han usat ja que s'ha considerat que el nombre de mostres és massa baixa com per poder-ne generar un classificador representatiu.

4.3. Anàlisi de resultats

Una vegada exposada, es pot aplicar la metodologia de classificació automàtica sobre els nostres conjunts d'imatges, un basat en el **GT** i l'altre en **l'experiment amb selectors de color**. Aquesta tècnica ens indica de quina classe preveu que és cada imatge del conjunt de validació. Així doncs, el procediment actual és tan senzill com calcular quantes de les previsions coincideixen amb la classificació real, i calcular-ne el percentatge.

Classe	Ground truth	Experiment colors
Global	0.4082	0.6224
2	0.3878	0.6607
3	0.2500	0.4375
4	0.5556	0.6364
5	0.6667	0.6154
6	0.0000	0.5000

Taula 4.1: Resultats d'ambdós classificadors automàtics

Una vegada executat l'algorisme pertinent, s'ha calculat el percentatge d'encert sobre cada classe en funció de l'etiquetat original. Aquests percentatges es poden consultar a la **Taula 4.1**. Com es pot comprovar, els resultats no són propers a una aplicació final, ja que es tracta d'un percentatge relativament baix en ambdós casos. Tot i això, es pot observar fàcilment que en general la classificació generada per usuaris resulta més útil usant el classificador automàtic. Això és degut a que les classes estan millor diferenciades, i per tant, els píxels assignats a cada classe també.

4.4. Discussió

Per veure més clarament com el classificador automàtic basat en les respostes dels usuaris és més precís que el que basat en el **GT**, podem analitzar les figures **Figura 4.1** i **Figura 4.2**. Es pot observar clarament com, tot i mantenir la seguir la mateixa per gairebé totes les classes, el classificador basat en respostes d'usuaris defineix molt millor les zones principals de les distribucions, visualitzant els pics d'intensitat al centre de cada classe.

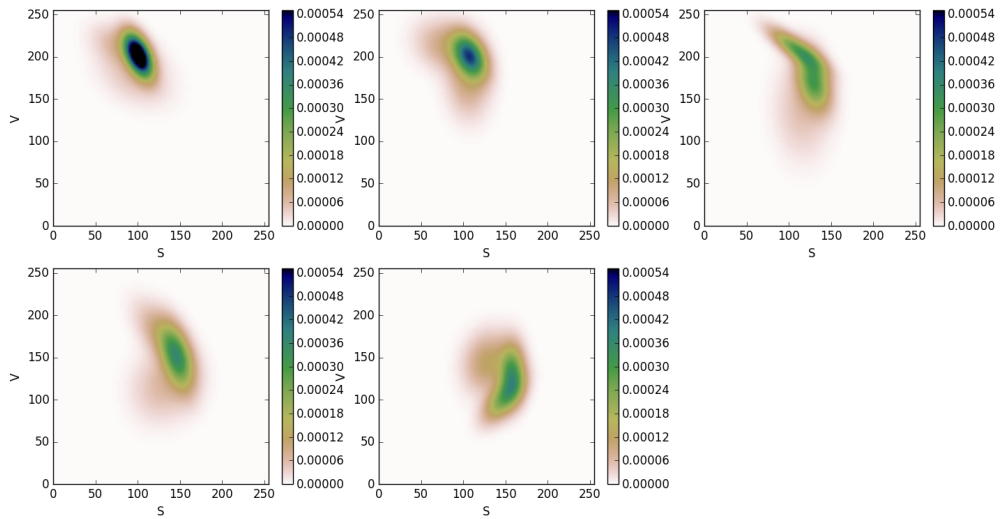


Figura 4.1: Representació del classificador basat en GT

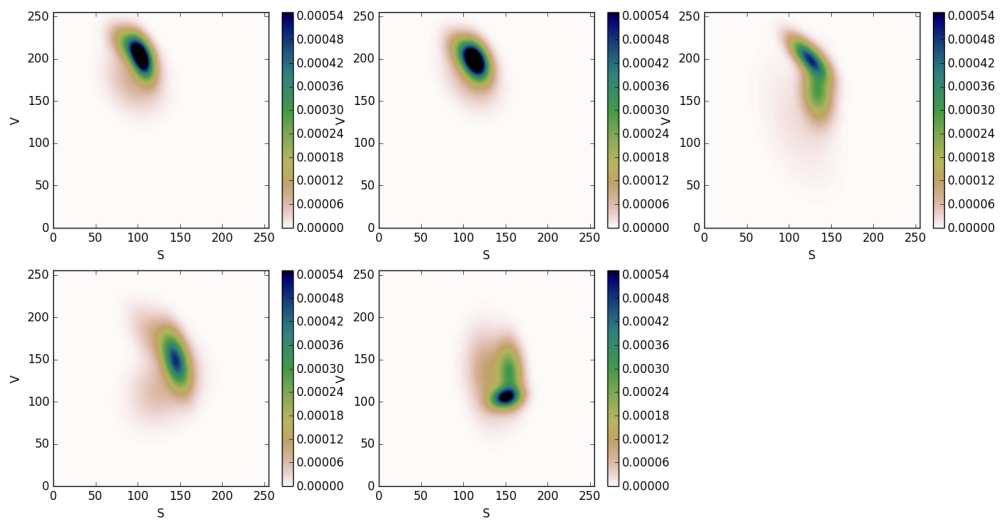


Figura 4.2: Representació del classificador basat en respostes d'usuaris

La principal diferència s'observa a la classe 6, que presenta no presenta la mateixa forma als dos classificadors, però com s'ha exposat abans, el nombre de mostres de les classes 1 i 6 són baixos, així que la diferència pot venir donada per aquest motiu. És possible que amb un major nombre de mostres les formes també fossin similars, però amb una major definició en el segon classificador.

A mode de cloenda, i en vista de que els classificadors basats en respostes d'usuaris defineixen millor els llindars entre classes, queda clar que cal continuar investigant i realitzar experiments que puguin donar més informació sobre les classes. A més, així també es podria realitzar un estudi de la classe 1, que s'ha descartat per falta de mostres, tan per l'aprenentatge com per la validació.

CONCLUSIONS

Una vegada realitzats tots els experiments amb usuaris, així com el posterior tractament de dades i una aproximació a la classificació automàtica, es pot donar el projecte per tancat.

Mitjançant la planificació i realització del projecte s'han assolit una sèrie de coneixements relacionats amb la **HbC** i com realitzar experiments amb usuaris. Aquest aprenentatge no ha sigut només relatiu a aspectes tècnics, com podria ser la preparació de l'experiment amb diferents tecnologies, o el llançament d'una aplicació web. Aquest aprenentatge, a més de tot això, també ha vingut acompanyat d'una sèrie de coneixements sobre com fer arribar a la gent el teu contingut, com analitzar el *feedback*, les suggerències que et retornen, etc.

Durant el recorregut del projecte s'han assolit els diferents objectius relacionats amb la **HbC** ja que s'ha entès l'objectiu de realitzar tasques mitjançant aquest mètode, i quines són les vies correctes per fer-ho. A més, els experiments realitzats basant-nos en aquestes tècniques semblen haver funcionat de la forma esperada, llevat d'alguns aspectes de participació, tot i que controlar el volum d'usuaris no estava dins l'abast del projecte.

En un altre ordre de coses, s'han realitzat tècniques de classificació automàtica amb un resultat positiu, tot i que les prediccions oferides pel sistema no són les pròpies d'una aplicació en estat final.

Finalment, fent una observació general a les tasques assolides i pensant en una possible continuació futura, hi ha una sèrie d'apartats que semblen viables per continuar investigant:

5. CONCLUSIONS

- Realitzar una aplicació per un entorn real on les tasques de classificació de pell es realitzassin continuadament. Un dels mètodes que sembla més viable és la realització de les tasques forçoses. El procediment és suficientment ràpid com per que els usuaris el realitzin de forma semblant als captcha, i de moment, per les proves que s'han realitzat, no sembla que els algorismes automàtics tinguin la mateixa capacitat d'encert que els humans. En cas de crear un projecte com aquest, caldria tenir en compte que els selectors no només han de ser senzills per als humans, sinó que també han de ser de difícil resolució per a les màquines.
- Provar l'ús de diferents eines de classificació automàtica per tal de maximitzar-ne l'eficàcia. Les tasques de classificació automàtica d'aquest projecte només són una primera aproximació al que podria ser un projecte totalment independent, comprovant l'eficàcia i eficiència de diferents mètodes, no només de **GMM**. Aquest projecte podria tenir futur també en un entorn d'empresa privada, allunyat de la investigació, ja que podria ser útil per a determinades empreses de cosmètics i protectors solars.

Amb tot això, i per tancar aquest projecte, m'agradaria exposar le meu agraïment i la meva satisfacció personal per haver realitzat tasques totalment noves per a mi, i adornar-me'n de que ha estat un procés d'aprenentatge útil, no només un tràmit, ja que a cada passa que realitzava en relació al projecte, entenia una mica més i millor el que s'estava realitzant en conjunt.

BIBLIOGRAFIA

- [1] B. M.-A. Antoni Jaume-i Capó, Carlos Mena Barco, “Análisis de la morfología de glóbulos rojos en dispositivos táctiles a través de un captcha,” 2016. 1.1
- [2] Wikipedia, “Fitzpatrick scale — wikipedia: The free encyclopedia,” 2016. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Fitzpatrick_scale&oldid=737228246 1.2
- [3] F. . Flawless, “Melanin rich skin tones.” [Online]. Available: <https://fairandflawless.com/pages/skin-tone> 1.1
- [4] C. D. Association, “Skin.” [Online]. Available: <http://www.dermatology.ca/skin-hair-nails/skin/> 1.2
- [5] Wikipedia, “Von luschan chromatic scale — wikipedia: The free encyclopedia,” 2016. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Von_Luschan_chromatic_scale&oldid=721878043 1.2.1
- [6] Unicode, “Unicode emoji,” 2015. [Online]. Available: <http://www.unicode.org/reports/tr51/tr51-2.html> 1.2, 1.2.1
- [7] J. Burge, “The trouble with redheads, a possible solution,” 2015. [Online]. Available: <http://blog.emojipedia.org/the-trouble-with-redheads> 1.1
- [8] A. I. S. Benjamin M. Good, “Crowdsourcing for bioinformatics,” 2013. 1.3, 1.3.1, 1.2
- [9] “Less - getting started.” [Online]. Available: <http://lesscss.org> 2.1.3
- [10] T. jQuery Foundation, “Query.” [Online]. Available: <https://jquery.com/> 2.1.3
- [11] G. Project, “gettext - gnu project.” [Online]. Available: <https://www.gnu.org/software/gettext/> 2.1.3
- [12] Google, “recaptcha.” [Online]. Available: <https://www.google.com/recaptcha/intro/index.html> 2.1.3
- [13] Scikit-learn, “Gaussian mixture models,” 2014. [Online]. Available: <http://scikit-learn.org/stable/modules/mixture.html> 4.1