Brief research report: Exploring the Diversity of Our Skin

Lizzy Vimbisai Pswarai, James Farrell, Mahfuzur Rahman Khokan, Soumya Paria, Andrea Giardina, Kaustubh Adhikari

School of Mathematics and Statistics, The Open University, Milton Keynes, UK.

Abstract

This report will discuss the process of collecting skin tone measurements from a diverse group of participants and analyse the accuracies of machine-based readings through comparing with reference measurements. The machines usually take the readings of the skin tone level by generating the images and analysing them with various statistical techniques to read the information received. Most methods used today can be completed using machines and technology and we have taken advantage of this by obtaining image samples of the participants faces or/and parts of their arms and have the image processed by self-created algorithm to get the colour values of each image sample for every sample given. Instead of the results received being based on societies ethnicities the colours are received in RGB (red, green, and blue) reflectance values with a median number between 0 and 255 for each image sample processed. Using this information, we can observe the accuracy that is assumed for the machine to have when reading the skin tone of a range of skin tones and use the results as a basis to improve accuracy in machines when it comes to recognising those of a darker skin tone removing the Caucasian bias within them.

Figure.01: A diverse group of people with different appearances, ethnicities, and religions.

1 Introduction

1.1 Background Research

Appearance can be a difficult topic to talk about or investigate because appearance is something people mostly base their judgment on someone. Whether that be through posture, physique, clothing style down to how much jewellery we wear. People have tried to abide by the societal rule of not "judging a book by the cover" however judgement has since been confused with moral judgement. Many misconceptions or misunderstandings have been made since. Moral judgments refer to judgments that have moral content; they are used to evaluate situations, courses of action, persons, behaviour, etc. (Seven Pillar Institute, 2017). Moral judgment has been a philosophical discussion for quite a time as it is a very subjective matter. When it comes to appearance the best moral compass to align this too is "a theory of moral sentiment" (Seven Pillar Institute, 2017).

"Proponents of this theory point to the human tendency to sometimes favour, or privilege, family and friends in moral decisions to substantiate their claim." (Seven Pillar Institute, 2017). In other words, people tend to prefer to use or bring up their friends of family in moral choices/decisions for support and agreement to further push for their claim to sound true. This can happen both ways and tends to happen in communities to peer pressure the receiving party into thinking one form of just reasoning due to unified agreement of their fellow peers/community. Coming back to appearance, this then causes people to create "appearance norms" (Andrew Mason, 2016).

Appearance norms are where, in a community or society, there are unspoken standards as to how one should look or come across as and that goes as far as race, gender, and sexuality. And this has been in place for as long as anyone can remember making it difficult to eradicate. Appearance norms have contributed to many downfalls in society such as Racism and Discrimination and have also been the cause of many disorders such as Eating Disorders, Body Dysmorphia, Dissociative Identity Disorder (DID), anxiety disorders and stress disorders.

Appearance norms tend to be biased and wrongfully forced upon by those of a similar appearance in a community/society. These biases can also be seen in technology for example in aspects relating to the cosmetics industry where you want to get an accurate foundation colour for your skin tone. Or when genetics research takes place, and they are used to understand human skin traits throughout history. They can be most recently seen in the use of Oximeters during COVID. Where the machines would overestimate those of a darker skin tone to have more oxygen than they state. Preventing them from getting the treatment they require (JAMA Internal Medicine, 2022). So hopefully with my research we can start bringing these issues to light and begin correcting these machines and remove any biases presently affecting those of an ethnic minority in multitudes of ways.

1.2 Project Aims

The aim of this project is to use statistical analyses to assess the biases in these pigmentation measurement instruments, starting off by finding how inaccurate the machines are when reading images and comparing it to the real digital colour. The digital RGB colour measurements are converted by the machine into measurements of skin colour and blood saturation, and we will focus on assessing the biases of these numerical measurements, since these represent biologically relevant factors that are used in various assessments.

2 Methodology

2.1 Design of The Project

This project involved processing photographs of various body parts, such as inner and outer arms, etc. They are a diverse group of people according to appearance/ethnicities. Correspondingly, machine measurements from the same body parts were available too, for comparison.

These images were edited them in a software called Gimp, if the images were too bright or too dark, to make them seem natural. After the appropriate changes, the images are cropped to find a suitable region where there is the least amount of reflection and a smooth surface, and use those to represent the skin colour.

An in-house Python code was developed to process the images to extract median red, median green and median blue values of each image. The extracted information is be placed into an excel file.

Once that was done, the data is visualized with multiple scatter graph(s). They compare the machines reading to the real colour results of the image to see how accurate the machine is in comparison.

2.2 Image Handling

The skin colour photographs all had images taken with a credit card sized photography scale for reference. There would be a white paper on the arm with a hole in the middle which indicates the place or section of skin that is assessed, as displayed below in Fig2.

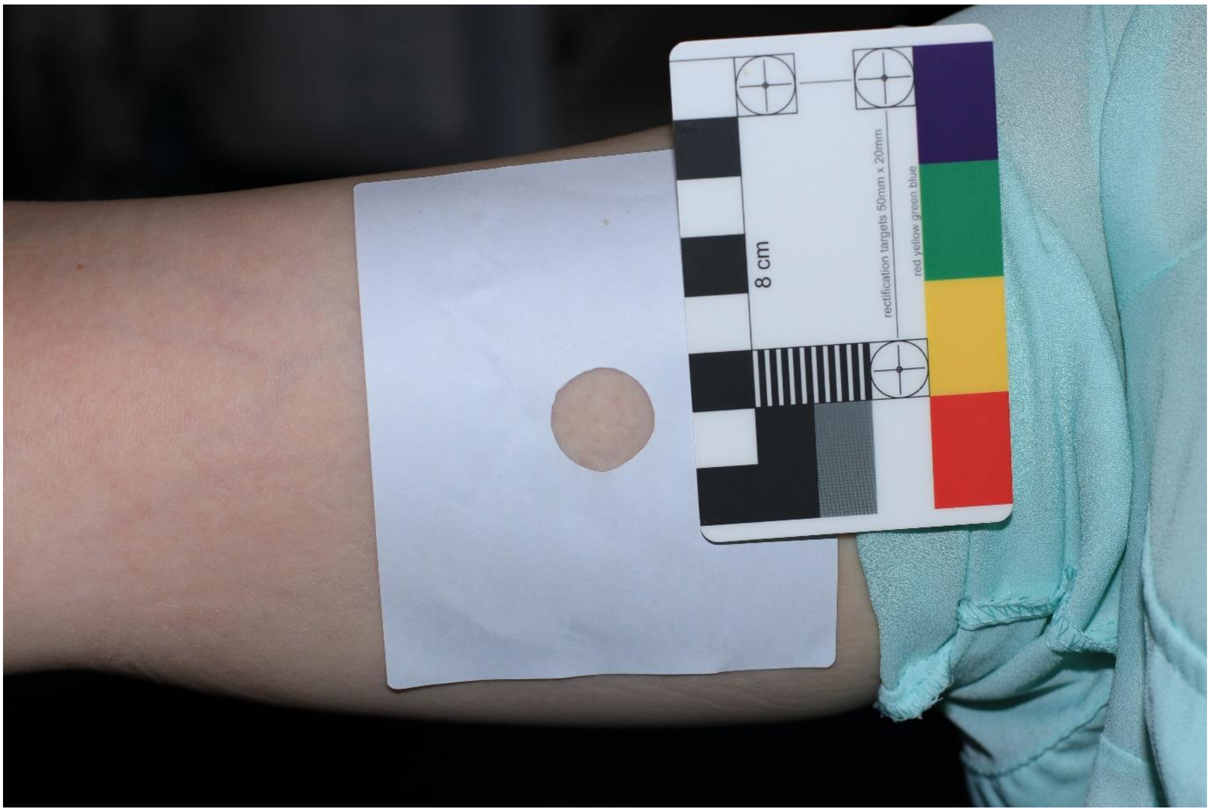

Fig2. An unedited image of a participant's inner upper arm with a reference card and a white piece of paper with a circle in it.

Annotation information of the images is stored in an excel file, which indicate if the image is useful or not quality wise, Problems, and which part of the body that image contained.

File Name	Useful	Problems	Body Part
IMG_5845	Yes	None	Inner Forearm
IMG_5847	No	When cropped the image is fuzzy	Inner Upperarm
IMG_5848	Yes	None	Outer Upperarm
IMG_5849	Partialy	There were a lot of surfaces with flash which couldn't be removed	Face
IMG_5850	Yes	None	Face
IMG_5852	Yes	None	Face
IMG_5824	Partialy	A lot of reflection	Inner Wrist
IMG_5825	No	Blurry/Unclear/Fuzzy	Inner Wrist

Fig3. Image annotations

All the information, having a total of 86 images, has been updated and kept only suitable images, which included:

- Those that were not very blurry
- Those without an excessive amount of reflection
- Those without too many beauties marks
- And any that were not deemed as N/A, Unused or Unidentifiabl

The following functions of GIMP has been applied to the selected images to maintain consistency in editing for all and create a natural state that will help the machines to read them accurately.

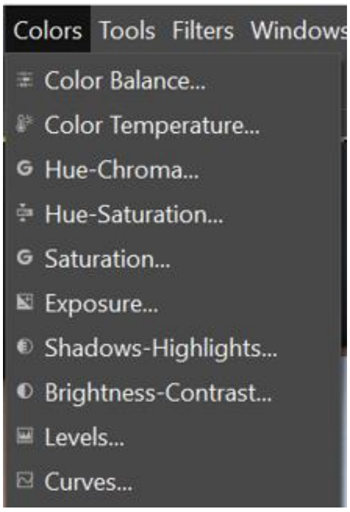

Fig4. Tools used to edit the images

All the edited images have been saved keeping the name similar with the original image file ended with underscore ‘adj` (_adj), to be more specific, for the cropped version of any parts of the body saved as IMG_[number]_adj_cr(.JPG).

2.3 Algorithm: Design

This section describes the algorithm through which the cropped images has been transformed into median level of read, blue and green colours, and saved into an excel sheet keeping the file and folder name as their source names.

2.4 Algorithm: Automation

An algorithm has been created with a loop function, shown in the following figure.05, which will help to read all the cropped images from the respective files and folders, the only things to be specified manually is the file path name.

```python
# Python program to read
# image using PIL module
import os
import PIL
import re

# importing PIL
from PIL import Image
from os import listdir
from os import walk

def load(mypath):
    filenames = next(walk(mypath), (None, None, []))[2]
    subdir = next(walk(mypath), (None, None, []))[1]

    for file in filenames:
        if re.search('cr', file):
            print(mypath + '/' + file)

    print(subdir)

    for folder in subdir:
        filenames = next(walk(mypath + '/' + folder), (None, None, []))[2]
        for file in filenames:
            if re.search('cr', file):
                print(mypath + '/' + file)
```

Fig.5. A function to load all the cropped images file names into a user display window

This forms the start of the process of automation as all that is needed now is to give the code the functionality to read the colours within the images requested (which we confirmed we can request the relevant images and their locations earlier).

The following mathematical complex function (in Fig. 7) helps to read the images requested and retrieve the median red, green and blue from them.

```python
def geometric_median(red, green, blue, eps=1e-5):
    X = np.stack((red, green, blue), axis=1)
    y = np.mean(X, 0)

    while True:
        D = cdist(X, [y])
        nonzeros = (D != 0)[:, 0]

        Dinv = 1 / D[nonzeros]
        Dinvs = np.sum(Dinv)
        W = Dinv / Dinvs
        T = np.sum(W * X[nonzeros], 0)

        num_zeros = len(X) - np.sum(nonzeros)
        if num_zeros == 0:
            y1 = T
        elif num_zeros == len(X):
            return y
        else:
            R = (T - y) * Dinvs
            r = np.linalg.norm(R)
            rinv = 0 if r == 0 else num_zeros/r
            y1 = max(0, 1-rinv)*T + min(1, rinv)*y

        if euclidean(y, y1) < eps:
            return y1

        y = y1
```

Fig.7. A mathematical function coded to read an images median red, green and blue levels

The following simple python code indicates the way of getting the median red, green and blue value levels from a phot realistic image.

```python
red = np.array(img[:,:,2]).flatten()
green = np.array(img[:,:,1]).flatten()
blue = np.array(img[:,:,0]).flatten()
```

Fig.8. A simple python code to produce a photo-realistic-image

Both of the above codes, showed in Fig.7 and Fig.8, has been used to fetch all the required images automatically, read them, and display each median red, green and blue colour levels found in each one in sequence. An example of the results is displayed below in Fig.9.

```
=========== RESTART: C:\Users\lvpsw\OneD
[218.54691623 228.00979046 242.33925297]
[199.79677085 210.42081528 230.95710518]
[199.54880243 179.63360719 174.05732019]
[197.1639739  162.94826788 159.58862832]
[175.7432592  135.06008742 127.76649438]
[157.7329581  110.69066918 103.18267231]
[163.49223793 124.68901193 115.45565076]
[162.28936596 117.06161427 107.0876669 ]
[166.94428115 119.97635875 111.13759069]
[163.47517917 114.58292115 108.44728392]
[191.22815687 155.36217274 152.04917683]
[219.41628499 186.49338948 185.33732245]
[181.86083554 140.10847649 129.65278641]
[212.40515784 174.57010177 170.63599569]
[142.10500975 103.65513395  95.54704803]
[149.08923984  95.20844836  91.66072858]
[166.80642126 127.87314048 124.95458303]
```

Fig.9. An example of the output results for all the cropped images for an individual [ordered followed as median red, green, and blue color values]

Once the median value levels of all the images retrieved successfully, the data has been stored into an excel file for the future applicability.

2.5 Algorithm: Scripting to Excel

The excel file needs to be created in the same place where the code belongs, and the cell names need to be specified along with corresponding row and column. The following picture (Fig. 10) shows an example of creating the workbook/excel sheet using the format described above:

```
your_workbook = Workbook()
sheet = your_workbook.active
sheet["A1"] = "Folder Name"
sheet["B1"] = "File Name"
sheet["C1"] = "Red Median"
sheet["D1"] = "Green Median"
sheet["E1"] = "Blue Median"
```

This included all the information required for any future referencing or searches that needed to be done to find any trends in the data presented if needed.

Now, the information of all the participants have been collected using the coding described earlier and stored in the excel sheet following the correct order. One thing to make sure that the row numbers were accumulating accordingly and not starting from one, but based on the number set from the start, and every time it would loop through the images of another folder. However, this can be fixed by making sure that a new number (representing a row number) would be sent into the main loop every time it starts going through and reading the images of another folder as displayed in Fig.11 below.

```
for subd in directory:

    c = load('C:/Users/lvpsw/OneDrive/Documents/Nuffield Research Placement/Mela
    print(c)
    z = c
    print(z)
```

Fig.11. Displays the field number checks (the prints) and update that made sure the data being written into the excel file would continue to the row below every time

The above code would then take the number stored in a variable (a digital storage box with a label) called “z” overwriting or rewriting the row number the code should start writing the data read from. Fig12 displays the code used to do so below.

```
sheet["A{0}".format(z)] = subd
sheet["B{0}".format(z)] = file
sheet["C{0}".format(z)] = gm[0]
sheet["D{0}".format(z)] = gm[1]
sheet["E{0}".format(z)] = gm[2]
```

For context “.format(z)” indicates that the “{0}” should be the number inside the variable “z”. For example, if z = 10 the code above will write the information assigned to it in A10, B10, C10, D10 and E10. Keeping it all in the same row and changing when accumulated in the main loop for the same folder or when it moves on to another folder variable “z” will be overwritten to start accumulating from the previous number if stopped when it finished reading all the images in the previous folder. The variables displayed in Fig.12 will be explained below:

subd: This variable holds the Folder name used to find which folder the code is reading images in.

file: This variable holds the file name of the images read by the code.

gm[]: This holds the list of colours/as we see it number levels or all three colours red, blue and green. To put each one in the correct column we use “[number position in the list]” to retrieve the correct number and put it in the correct column. In python lists the first item of the list is number 0 and onwards. These are also called Indexes. In all the lists (refer to Fig.9 for an example) the positions are the same with red being 0, blue being 1 and green being 2.

sheet []: This just reads the specific cell requested, for example “A1” and whatever it is assigned to (= variable/number/words otherwise known as a string) will be written in the cell requested which in this case is cell “A1”.

All the data would then be added accordingly to an excel file named “Image Data” as partially displayed in Fig.13 below.

File Name	Red Median	Green Median	Blue Median
IMG_5845_adj_cr.JPG	218.5469162	228.0097905	242.339253
IMG_5847_adj_cr.JPG	199.7967708	210.4208153	230.9571052
IMG_5848_adj_cr.JPG	199.5488024	179.6336072	174.0573202
IMG_5849_1_adj_cr.JPG	197.1639739	162.9482679	159.5886283
IMG_5849_2_adj_cr.JPG	175.7432592	135.0600874	127.7664944
IMG_5849_3_adj_cr.JPG	157.7329581	110.6906692	103.1826723
IMG_5849_4_adj_cr.JPG	163.4922379	124.6890119	115.4556508
IMG_5849_5_adj_cr.JPG	162.289366	117.0616143	107.0876669
IMG_5849_6_adj_cr.JPG	166.9442811	119.9763588	111.1375907
IMG_5849_7_adj_cr.JPG	163.4751792	114.5829212	108.4472839
IMG_5849_adj_cr.JPG	191.2281569	155.3621727	152.0491768
IMG_5850_1_adj_cr.JPG	219.416285	186.4933895	185.3373224
IMG_5850_2_adj_cr.JPG	181.8608355	140.1084765	129.6527864
IMG_5850_adj_cr.JPG	212.4051578	174.5701018	170.6359957

Fig.13. Overview of a sample collected data into an excel file

3 Results and Discussion

3.1 Results

The visual results of all the cropped images and their median colour value levels have been plotted on a few scatter graphs to show the relationships among the median RGB color values as well as the pigmentation indices such as skin colour i.e. melanin index (MI), and blood saturation i.e. erythema index (EI). The following scatter diagram shows correlation between melanin index (putting on the x axis) and erythema index (EI) (putting on the y axis) as shown below in Fig14.

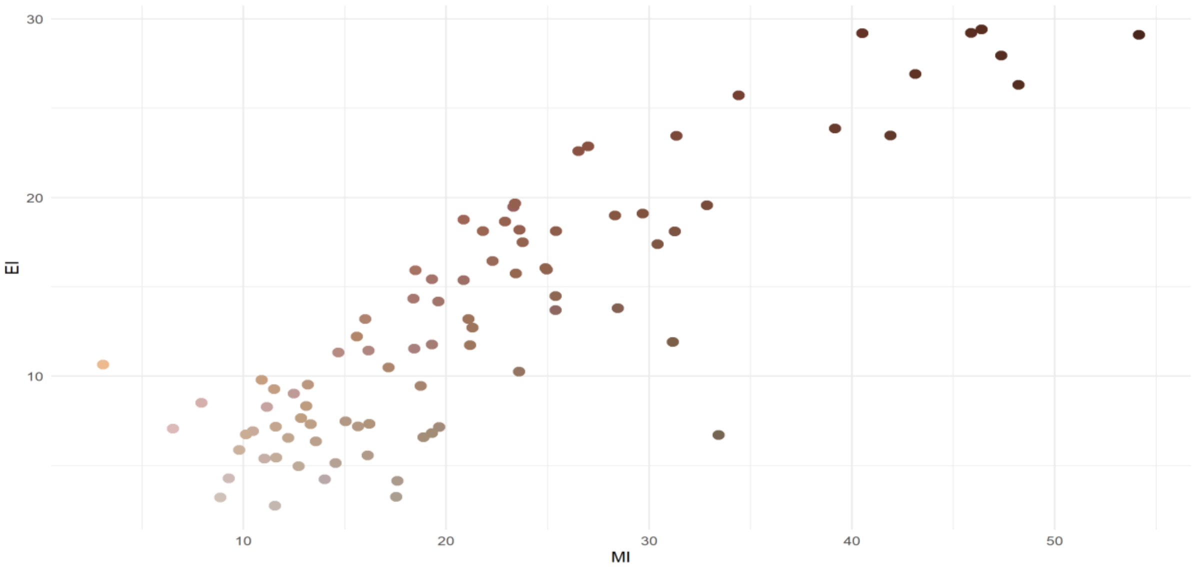

Fig14. A scatter diagram shows positive correlation between melanin and erythema index. (Based on reference measurements or real image)

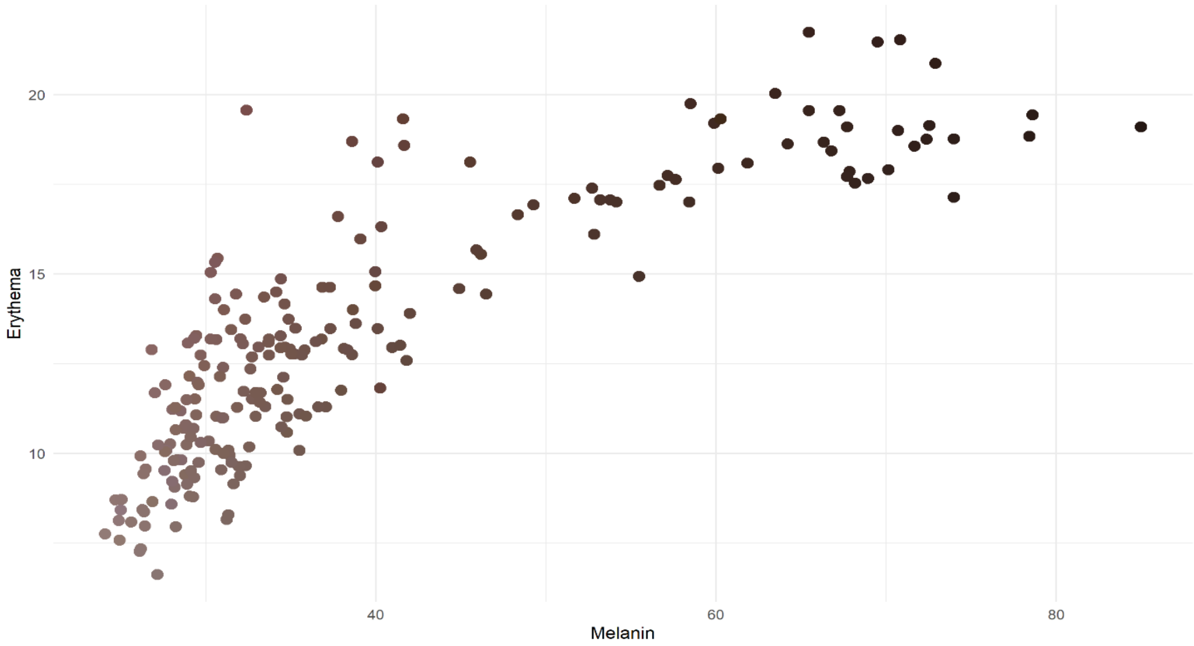

Fig.15. Another graph representing positive correlation between melanin and erythema index. (Based on the measurements obtained from the participants in this study)

Erythema is indicates the extent of blood flow and/or blood saturation under the skin, so higher erythema increases the red colour levels in the skin. In Fig15 above we can see the levels of erythema are matching the Melanin levels for those of a lighter skin however as the skin tone gets darker the levels start to lose accuracy, showing that machine readings are getting inaccurate due to darker skin people.

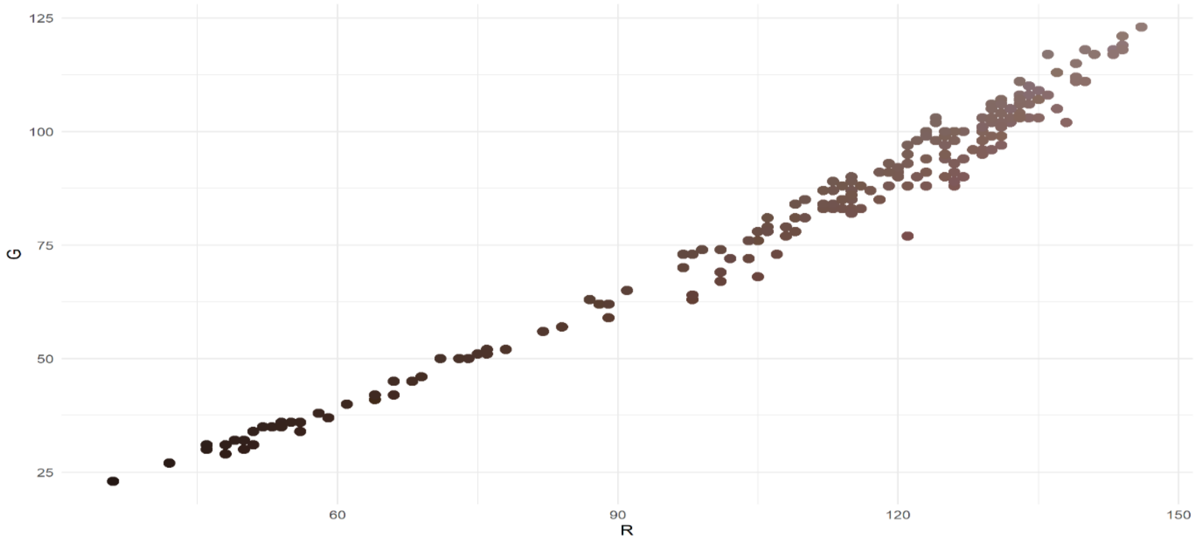

Fig.16. Scatter graph shows positive correlation between median red vs median green, also supporting the relationships showed in Fig.15 (Based on reference measurements)

Here is also a scatter graph displaying the correlation between the green and red median colours retrieved by the machine below in Fig17.

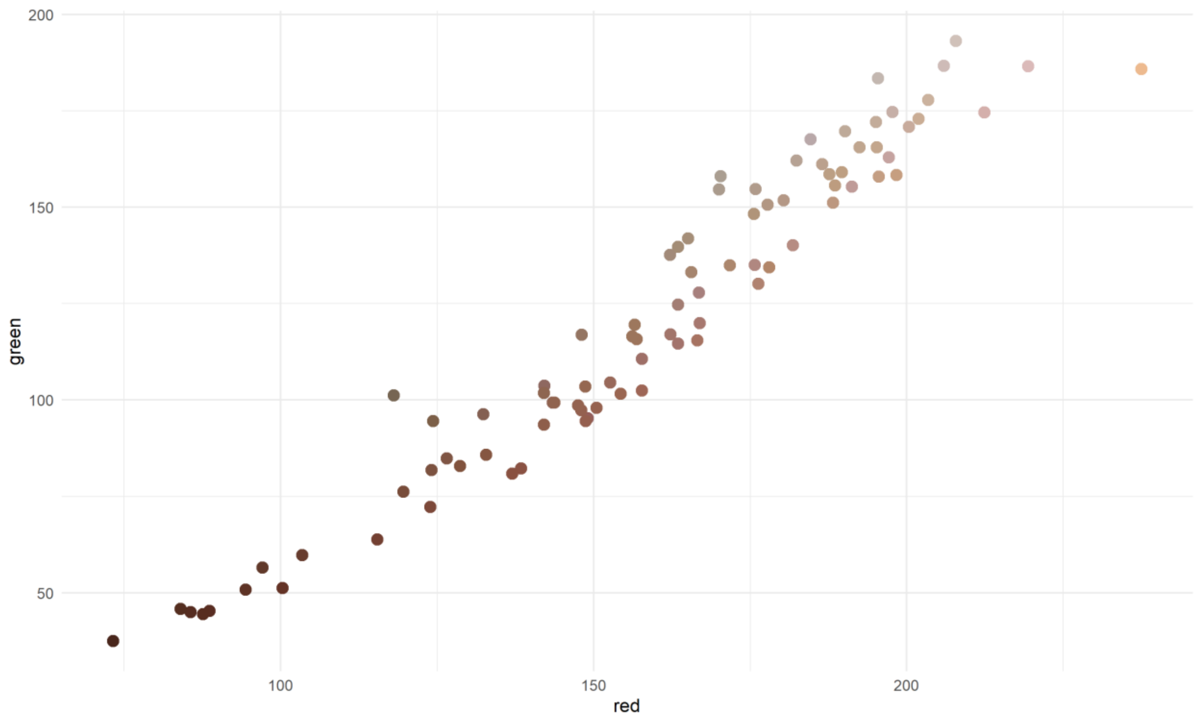

Fig.17. A scatter graph showing positive correlation between the medians red and green produced by the machines.

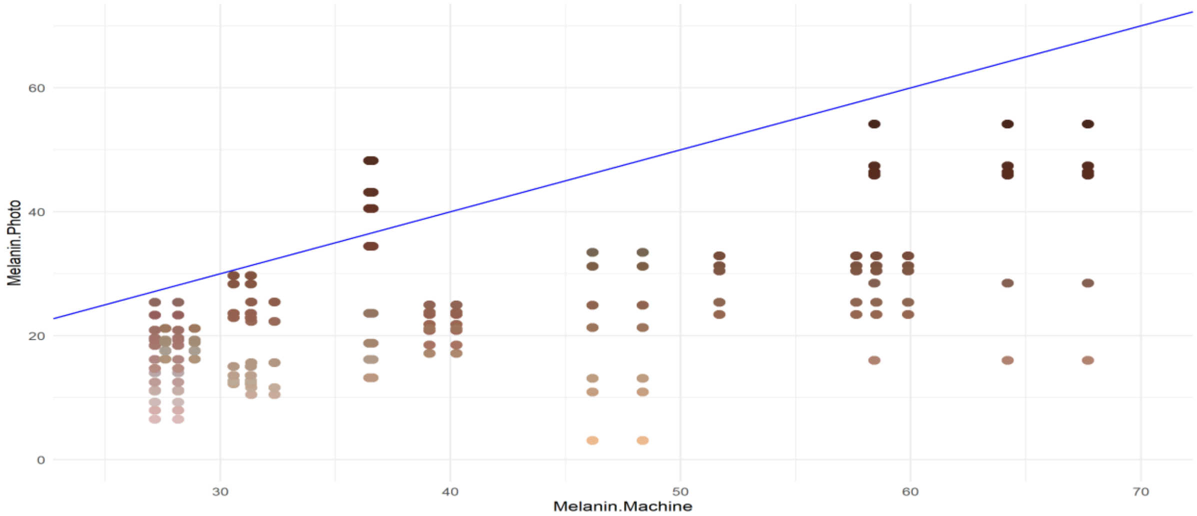

Fig.18. Comparison between skin color readings taken from machine and digital photos. The point's deviation from the blue line is an indication of biases in machine readings. This picture proves that machines giving more biases when the melanin level increases i.e., for the darker skinned people.

3.2 Limitations And Improvements

One of the key limitations of this research is human error, since there are a chance of possible misjudgements during the image editing. It is difficult to adjust an image to what the skin colour of person would normally look like if you have never seen the person before in their natural light. Which made making the image look more natural more difficult.

Though the colour reference cards are used to provide an objective template, human subjectivity might still remain. Another limitation was keeping consistency among the adjustments, since they were not taken in the same format.

An improvement that could be made to this experiment only if it is possible to meet the participants beforehand and taking their images amid in natural light to be better suit with their natural appearance.

3.3 Supporting Research

In a previous publication on a large cohort of Latin American participants showed the extent of quantitative bias in the measurements of skin color [Adhikari et al. (2022)], which is consistent with our preliminary results.

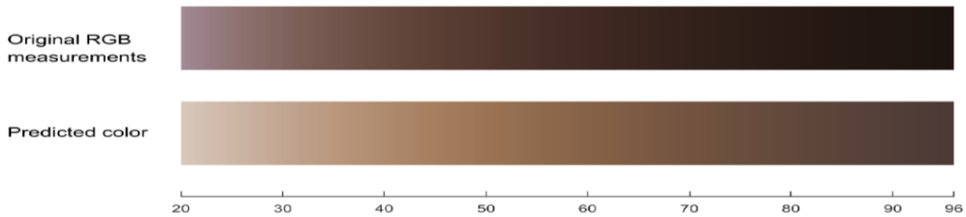

Fig.19. Differences in two scales represent the reflection of measurement errors (biases), occurred due to using machines.

The above situation is coinciding with our research cases (Fig.18) i.e., the machines are producing more measurement biases when the melanin level of the participants gets higher.

As shown also in Fig.14 it shows that you go across what is expected to be a relatively strong correlation to represent colour accuracy between what the machine read and the real image, begins to curve as the skin tone gets darker. Eventually curving greatly when approaching the darkest skin tone. With that we can confirm that there is a clear bias in these machines that need to be adjusted/fixed so that it is all inclusive preventing any form of expulsion from occurring due to technology we did not modify to be all inclusive.

4 Conclusion

With all the undoubtable evidence gathered we can conclude that there is a large bias when machines read skin tone. The bias presented in Fig.14, Fig.15 and Fig.16 through the machine's readings are just a small window into the amount of technology that holds this bias. Technology may evolve but this issue has been there since equipment that used to read skin tones such as Oximeters have been with us since the 1940's and have still not been developed to fit all races since the bias had been inputted due to appearance norms in the 1940's.

These types of issues need to be solved as soon as possible because people that use such technology hinder themselves in various ways as much as they hinder others that use it through them, such as the cosmetics industry. Being all inclusive may start with being humane with each other, overcoming the appearance norms and biases through ourselves but we need to reach further. We need to make sure that anything that has been made with an appearance norm or bias in mind, needs to be updated so that they are made for everyone and not a specific set of people. Making anything with biases like these, to be made for everyone just by starting with removing appearance biases from technology can make a larger impact on the world than we think. From perfect matches of foundation to skin, to saving lives and giving care to those who need it in hospitals.

Our research may be a small step in proving that something needs to be done, but if this issue can be brought to light rather than be locked up and away from public view, we can make a difference. We can change not only ourselves but everything around us for the betterment of all humans, no matter the appearance.

5 Appendix

We have recently discovered an article by The Guardian that explains, AI being developed to diagnose skin cancer. This is an incredible development for the medical industry however the full title discloses the fact that it may be less accurate for those of a darker skin. They are trying to rectify this by adding sample images for the AI to analyse and learn from. Those that already do this however have very few samples to work from. They have attempted looking through data bases to find more samples of darker skinned people however “the team found just 2,436 of a total of 106,950 images within the 21 databases had skin type recorded. Of these, only 10 images were from people recorded as having brown skin and one was from an individual recorded as having dark brown or black skin.” Showing there is a massive issue with darker skin recognition, variety wise which “could bring other problems including risking avoidable surgery, missing treatable cancers and causing unnecessary anxiety” according to the team involved.

A similar paper on AI driven dermatology by The Atlantic highlights the risks in numerical values just to bring it further down to reality. “While fair-skinned people are at the highest risk for contracting skin cancer, the mortality rate for African Americans is considerably higher: Their five-year survival rate is 73 percent, compared with 90 percent for white Americans, according to the American Academy of Dermatology.” This just shows the urgency that needs to be brought upon bringing darker skinned to be updated into technology as we rely on it more than ever for such life defining issues such as cancer.

6 References

1. University of Virginia, College and Graduate School of Arts and Sciences, 2022. *The Individual Society*. [image] Available at: <https://gened.as.virginia.edu/sites/gened.as.virginia.edu/files/iStock-1202344480.jpg> [Accessed 1 August 2022].
2. Institute, S., 2017. *Moral Judgment*. [online] Seven Pillars Institute. Available at: <https://sevenpillarsinstitute.org/glossary/moral-judgment/> [Accessed 23 August 2022].

3. Mason, A. (2017), Appearance, Discrimination, and Reaction Qualifications. Journal of Political Philosophy. [Journal] Andrew Mason. Available at: <https://doi.org/10.1111/jopp.12099> [Accessed 23 August 2022].
4. Gottlieb, E., Ziegler, J., Morley, K., Rush, B. and Celi, L., 2022. *Assessment of Racial and Ethnic Differences in Oxygen Supplementation Among Patients in the Intensive Care Unit*. JAMA Internal Medicine.
5. Giardina, A., 2022. *stylegan2-directions/color_skin.py at main · agiardina/stylegan2-directions*. [online] GitHub. Available at: <https://github.com/agiardina/stylegan2-directions/blob/main/color_skin.py> [Accessed 2 August 2022].
6. Adhikari, K., 2022. PigmentationGWAS_paper_Supplement. [Accessed 30 August 2022]
7. Davis, N., 2022. AI skin cancer diagnoses risk being less accurate for dark skin – study. [online] The Guardian. Available at: <https://www.theguardian.com/society/2021/nov/09/ai-skin-cancer-diagnoses-risk-being-less-accurate-for-dark-skin-study> [Accessed 26 August 2022].
8. Lashbrook, A., 2022. AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind. [online] The Atlantic. Available at: <https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/> [Accessed 25 August 2022].
9. Negrin, L. (2008). Appearance and Identity. In: Appearance and Identity. Palgrave Macmillan, New York. https://doi.org/10.1057/9780230617186_2
10. Giorgio Sirugo, Scott M. Williams, Sarah A. Tishkoff The Missing Diversity in Human Genetic Studies Cell, Volume 177, Issue 4, 2 May 2019, Pages 1080