Application of Data Science in Skin Care Formulation

Farkhanda Dalal   
Department of Computer Engineering  
Pune Institute of Computer TechnologyPune, India  
7farkhanda@gmail.com  
  
  
Misbah Bagwan  
Department of Computer Engineering  
Pune Institute of Computer TechnologyPune, India  
@gmail.com  
  
  
Mansi Jangle  
Department of Computer Engineering  
Pune Institute of Computer TechnologyPune, India  
@gmail.com

*Abstract*— Skincare formulations have predominantly been based on research and data derived from Caucasian skin, leading to a lack of comprehensive studies on skin color. This oversight has resulted in limited product efficacy and relevance for individuals with diverse skin tones and conditions. To create better skincare products, developing formulations that are tailored to the skin type, age, and skin condition is essential.

By leveraging the power of Data Science, we can develop products with higher efficacy for skin of color, by analyzing factors like melanin concentration, collagen amount, epidermis thickness, skin type, etc. This paper discusses how Data Science can facilitate advancement in Skincare development and provide a more inclusive approach to skincare formulations.

*Keywords*— Data Science, Data Collection, Data Visualization, Statistical Analysis, Correlational Analysis.

# Introduction

In recent years, the skincare industry has faced significant challenges in addressing the diverse needs of consumers with different skin types. Historically, the formulation of skincare products has been guided by limited datasets, predominantly focusing on Caucasian skin types but marketed to women of all skin colors [3]. As a result, many skincare solutions are ineffective for individuals with different ethnic backgrounds, skin tones, and conditions. This lack of diversity in data collection and analysis has led to the development of one-size-fits-all products that fail to cater to the specific needs of underrepresented demographic groups [3] The pressing issue is the skincare industry’s inability to address the full spectrum of skin types, which has left a gap in efficacy and consumer satisfaction, particularly for individuals with non-Caucasian skin.

With the growing awareness of this diversity, there is an urgent need for a more comprehensive, data-driven approach that uses large and diverse datasets. By collecting and analyzing data on multiple skin types, including factors like melanin content, collagen, skin elasticity, pigmentation, sensitivity, and environmental impact, the skincare industry can better understand the needs of all consumers. This shift toward a data-centric methodology can lead to the creation of more inclusive and effective products that cater to a wider audience, thereby reducing the disparities in skincare outcomes.

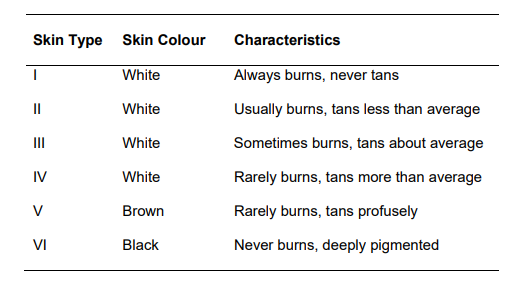
# need of inclusive formulations

## Classification of Skin Types:

Sun-reactive Typing:

The concept of sun-reactive "skin typing" was created in 1975 to classify persons *with white skin* for treatment of psoriasis. A clinical research was conducted by T.B. Fitzpatrick from 1988. This paper which categorizes skin types based on how they react to sun exposure. The Fitzpatrick classification system classifies women of color as type IV, V, and VI in the table below [5].

Even though this is the most well-known and generally accepted in the somatology industry, it has limitations as not all women of color can be accommodated in this scale [3].



## Structural and Physiological Differences in Skin Characteristics Based on Ethnic Grounds:

Human skin has structural differences based on four major characteristics namely, melanin content, TEWL, skin irritation and sebum levels [6]. These parameters have a profound impact on how skin behaves and reacts to environmental factors, treatments, and products. Hence, these differences are of notable significance when we think of inclusive skin care formulation.

The primary short coming of the current skincare industry is that these differences are neglected during the Research and Development phase of manufacturing. As a result, we see an array of formulations that are unfit for use, for non-Caucasian skin.

1. Table II. II Structural and Physiological Differences in Skin Characteristics Based on Ethnic Grounds

|  |  |  |  |
| --- | --- | --- | --- |
| Property | Caucasian | Brown | Black |
| Melanosomes | Small,  approx.  (0.5nm \* 0.3 nm) [4] | Larger,  approx. (0.6nm \* 0.3 nm) [4] | Largest,  approx. (1.0nm \* 0.5 nm) [4] |
| Higher Melanin Concentration is seen at | Chin and Neck [3] | Neck and  Periauricular space [3] | Fore Head and Chin  [3] |
| Dermis | Thinner,  between  1mm to 2mm | Intermediate,  between  1mm to 3mm | Thicker,  between  1.5mm to 4mm |
| Aging Results In | Photodamage, Loss of Collagen [1] | Dyspigmentation [1] | Frequent xerosis,  Hyperpigmentation [1] |
| Irritation Causes Following Reaction | Erythema | Erythema, Hyperpigmentation | Results in darkening of effected area instead of Erythema |
| Sebum  Production | More on Fore Head,  Periauricular space, Chin [3] | More on Fore Head,  Periauricular space, Chin [3] | More on Fore Head,  Periauricular space, Cheek [3] |

Identify applicable funding agency here. If none, delete this text box.

## Under Representation of Skin of Color:

The traditional method of formulation is deep rooted to cater to the European or Caucasian skin type. Hence, previously products were developed to cater to their needs and wants, whereas the same products were also marketed towards non-Caucasian women with less consideration to how these products would address their needs and how their skin would react to these formulations.

This problem is particularly prevalent in dermatological and cosmetic literature, including dermatology textbooks, where a racial bias is reflected in the underrepresentation of images of minorities compared to the general population, which can lead to inequalities in skin health care. [1]

In an article published by King (1998) in the USA, in which 21 cosmetic company executives were interviewed, it was found that most cosmetic companies do not market their products differently to women of color, neither do they have special cosmetic collections for these diverse ethnic groups [6].

This biasness has been prevalent even with the rise of Artificial Intelligence and Machine Learning and its implementation in the Cosmetics and Personal Care Industry.

AI has proven to inherit and adopt human biases and this is clearly evident as there is lack of focus on diverse skin types and skin concerns. Here are a few examples:

**1. Sampling Bias:**

Sampling bias arises when the input data lacks diversity or is skewed. In cosmetic skincare research, panelists with light skin tones are often overrepresented, while those with darker skin tones are underrepresented. This imbalance in datasets can lead to algorithms that are non-generalizable across race, ethnicity, sex, gender, or age. [1]

**2. Confounding Bias:**

Confounding bias occurs when the relationship between variables is obscured by the presence of additional variables not considered in the model. [1] An example related to skin of color is when predicting hyperpigmentation using UV exposure data but failing to account for melanin levels. Melanin significantly influences how skin reacts to UV light and excluding it from the model can lead to inaccurate predictions, particularly for individuals with darker skin tones.

**3. Measurement Bias:**

Measurement bias occurs due to discrepancies in how skin properties are measured across different devices and methods. For instance, skin color measurements from spectrometers might vary depending on ethnicity and the device used, leading to inconsistencies. Additionally, post-processing features in cameras can introduce biases, particularly in representing different skin tones. [1]

**4. Label Bias:**

Label bias arises when numerical values obtained for skin properties do not account for individual differences. For instance, redness values may vary based on pigmentation, leading to inaccurate classifications. Inconsistent human visual assessments for features like age prediction can also introduce bias, as cultural and individual beliefs may affect judgment. [1]

# METHODOLOGY

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. A. Georgievskaya, T. Tlyachev, D. Danko, K. Chekanov, and H. Corstjens, "How artificial intelligence adopts human biases: The case of cosmetic skincare industry," *AI and Ethics*, vol. 4, no. 2, pp. 85-95, 2023.
2. E. Markiewicz and O. C. Idowu, "Personalized skincare: from molecular basis to clinical and commercial applications," 2018.
3. C. D. Kaur and S. Saraf, "Skin care assessment on the basis of skin hydration, melanin, erythema, and sebum at various body sites," *Asian Journal of Pharmaceutical and Clinical Research*, vol. 4, no. 2, pp. 40-45, 2011.
4. L. J. Teixeira, "Specific cosmetic and skincare needs of women of color in South Africa,", *South African Journal of Dermatology*, vol. 12, no. 1, pp. 60-72, 2006.
5. FITZPATRICK, T.B. , “The Validity and Practicality of Sun Reactive Skin Types I Through VI.” *Archives of Dermatology*, 124: 869-871, 1988.
6. GRIMES, P.E. , “Skin and Hair Cosmetic Issues in Women of Colour. Dermatological Clinics”, 18(4), Oct: 659-665, 2000