IFSCC 2025 full paper ***IFSCC2025-922***

Rethinking Hair Color Classification: A Data-Driven Framework for Innovation

Olfa Bchir [1], Gautier Deconinck [1]

[1] L'Oréal Research and Innovation, Paris, France

Abstract

The current hair color classification uses a numerical system based on the tone level and reflects to describe the color result. This system relies on subjective visual evaluation and varies between brands and companies, limiting therefore the objective analysis of shade portfolios and strategic decision-making. The technically complex classification is understood by the experts of the industry but lacking a consumer-centric approach. Therefore, an objective, data-driven and consumer-centric approach is essential. We introduce a new universal classification system for permanent hair color based on color science.

The hair color classification approach leverages objective L*a*b* colorimetric data measured with a hair color imaging system to classify shades into three distinct categories, further divided into color families and sub-families based on undertone, chroma, and lightness. Over 700 physical swatches have been first classified by experienced colorists under controlled illumination (D65). Then a data analysis was conducted by defining minimum overlapping convex hulls each representing the gamut of a color family and a nearest convex hull classification algorithm was developed. This algorithm calculates the distance of a shade to predefined convex hulls in comparison with a threshold to determine its classification and confidence. Future research will focus on developing and refining machine learning models by incorporating image data to capture nuanced color variations, further enhancing the system's accuracy and applicability.

Keywords: Hair Color Classification, Algorithms, Color Science, Data Science, Digital applications, Colorimetry

1. Introduction

The selection of hair color products remains a complex and often frustrating process for consumers. Current classification systems, primarily based on numerical codes varying across brands, often lead to confusion, mismatched expectations, and dissatisfaction. This inconsistency prevents objective shade comparison, limiting the potential for personalized recommendations and data-driven product development.

In the rapidly evolving beauty industry, personalization is important. Consumers increasingly seek tailored products and services that cater to their individual needs and preferences. A standardized, objective, and universally understood color classification system is crucial for enabling data-driven personalization in hair color. Such a system would empower consumers to navigate with confidence the vast array of hair color options and facilitate the development of advanced recommendation tools.

This research introduces a novel, data-driven framework designed to address these challenges. Our method utilizes precise L*a*b* colorimetric measurements from standardized hair swatches and employs a Nearest Convex Hull algorithm to classify shades objectively. We define three consumer-centric families (Natural, Natural Fashion, and Fashion) further categorized into sub-families and sub-categories. This comprehensive approach enables accurate, scalable, and consumer-friendly classification, paving the way for enhanced product development, portfolio management, competitor analysis, and personalized consumer experiences.

2. Materials and Methods

Workshops with experts: To establish a robust foundation for hair color classification, a series of workshops were conducted with six expert colorists. These experts were tasked with visually classifying a set of 700 real hair swatches representing a diverse range of shades. The goal was to achieve a consensus on the classification of each swatch. To ensure objective color evaluation, observations were made under standardized D65 lighting using LED light boxes. The workshops took place in a controlled observation environment (Figure 1), guaranteeing optimal conditions for color assessment and comparison.
Each swatch was classified visually in the different families, sub-families and sub-categories which we defined to be the most consumer-centric.

(**a**)

(**b**)

Figure 1. (a) Grey room with D65 lighting; (b) Example of workshop with experts to classify red hair swatches in the grey room under D65 lighting.

Hair Color Families, Sub-Families and sub-Categories: We have established a new hair color classification system consisting of three distinct families and several sub-families, centered around a consumer-centric approach (Figure 2):

- ***Natural Color Family:*** This family encompasses all hair shades naturally occurring in nature. Included in this family are Natural Red shades, which are often excluded from other classification systems but are fundamental for an inclusive approach to color.
  Sub-Families: Blonde, Brown, Dark Brown/Black and Natural Red

- ***Natural Fashion Color Family:*** This second family comprises natural shades with a fashionable reflect. This family includes for example Beige or Brown shades with red, gold, ashy or copper reflects.
  Sub-Families: Beige, Natural fashion Brown, Natural Fashion Dark Brown/Black

- ***Fashion Color Family:*** This final family consists of more expressive and vibrant colors that deviate further from natural shades and are thus bolder. This includes intense Reds, Blues, Pinks and other deep, vibrant shades.
  Sub-Families: Blue, Violet, Pink, Red, Copper

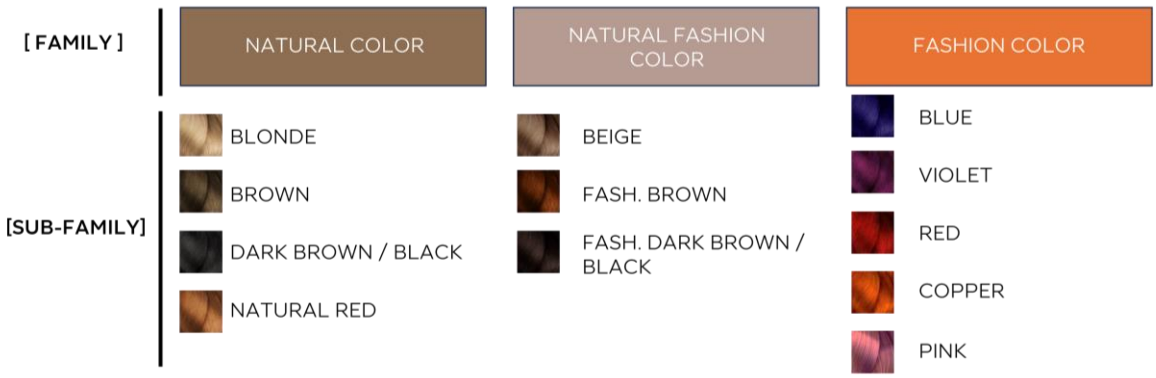


Figure 2. Representation of the different hair color families and sub-families.

Each sub-family is further divided into several sub-categories. These color descriptors provide more detailed information describing the shade:

- **Sub-Lightness:** Describes the lightness of the shade. This is categorized into three levels or a more granular scale (e.g., 1 to 10).
- **Sub-Hue:** Cool/Warm and Sub-reflect. This descriptor characterizes the undertone of the shade. For some sub-families, it also specifies the dominant reflect (e.g., Natural Fashion Brown with a Red Sub-reflect, Beige with Gold Sub-reflect…).
- **Sub-Chroma:** Low/Medium/High. This describes the saturation and intensity of the color.

Our initial focus is on the classification of families and sub-families.

Hair Swatches: Standardized hair swatches were used for hair color evaluation. They contain 90% white hair and 10% brown hair blended together. They have all similar homogeneity and a target weight of about 2g.

Color Application: Hair color products were applied to swatches following a standardized protocol. Products were chosen to cover a large gamut of hair color from different brands for the 700 swatches evaluated during the workshops.

Color Measurement: Color measurements were performed using a multispectral imaging system. The system illuminates hair swatches with specific wavelengths of light from 400nm to 700nm at 10nm intervals. Reflected light is captured by a multispectral CCD with a resolution of 2000x2000 pixels over a 15x15mm^2 area. The spectral data are then used to calculate L*a*b* values for each pixel, using the D65 standard illuminant as a reference. The reported L*a*b* value for each swatch represents the average L*a*b* value across all pixels within the swatch area. Image acquisition time was less than 10 seconds per swatch.

Convex Hull Algorithm: The convex hull of a set of points in a Euclidean space is defined as the smallest convex polygon (in 2D) or polyhedron (in 3D) containing all points within the set [1]. In this study, the convex hull algorithm is employed to delineate sub-family boundaries within a dataset. A convex hull is constructed for each sub-family, effectively defining the smallest convex region encompassing all data points belonging to that sub-family.

A total of 700 physical hair swatches were measured using a multispectral imaging system. The resulting L*a*b* colorimetric data were used to define the color gamut of each sub-family within a three-dimensional (3D) color space. Optimal convex hulls were generated to represent these gamuts following data analysis and cleaning. The goal was to achieve distinct convex hulls with minimal overlap, ensuring clear separation between sub-families in the 3D L*a*b* space. Overlap, in this context, refers to significant entanglement of the convex hulls, lacking distinct, separable 3D regions for each sub-family. It is important to note that a sub-family can be entirely encompassed within a larger sub-family, occupying a central region within the larger sub-family's gamut.
Analysis of the convex hulls and their degree of overlap played a critical role in refining sub-family classifications. For example, the initial Natural Fashion Blonde sub-family exhibited important overlap with the Natural Blonde sub-family, causing a re-examination and reassignment of swatches to either the Natural Blonde or Beige sub-families. Initially, a Natural Fashion Gray sub-family was included; however, due to extensive overlap with the Beige sub-family, it was removed and its members integrated into the Beige sub-family, distinguished by an ashy Sub-reflect. Furthermore, the initial Natural Red sub-family comprised only 13 swatches. To enhance the precision of our algorithm and the representativeness of the Natural Red gamut, 51 Natural Red hair swatches were acquired, measured, and incorporated into the Natural Red sub-family. Figure 3 summarizes our workflow methodology.

Figue 3. Chart representing the different steps for the Hair Color Classification.

Nearest Convex Hull classification (NCH): Accurately classifying hair color shades presents a challenge due to the subtle variations and richness of possible colors. The Nearest Convex Hull classification method offers a geometrically intuitive and robust approach to address this complexity, leveraging the uniform nature of L*a*b* color space and its compatibility with Euclidean distance calculations. This method demonstrates robustness to outliers, as highlighted in [2], which compares NCH favorably to state-of-the-art techniques.

Each hair color sub-family is represented by a convex hull in 3D L*a*b* space, generated using the Convex Hull algorithm. Classification of a new shade involves determining its position relative to these convex hulls, using a parameter called "Count True." "Count True" calculates the number of sub-families a given shade falls within.

If Count True = 1, the shade is directly assigned to that single sub-family. However, shades can fall within multiple sub-families or outside all of them, requiring further analysis. A distance threshold, optimized for visual coherence, is then employed. The distance between the shade's L*a*b* coordinates and the boundary of each convex hull it falls within is calculated.

This approach results in three classification scenarios:

(a) Inside (Count True = 1): The shade falls within a single sub-family's convex hull. If the shade is near the boundary of another sub-family (within the distance threshold), the confidence is 80%, and the neighboring sub-family is indicated. Otherwise, the confidence is 100%.

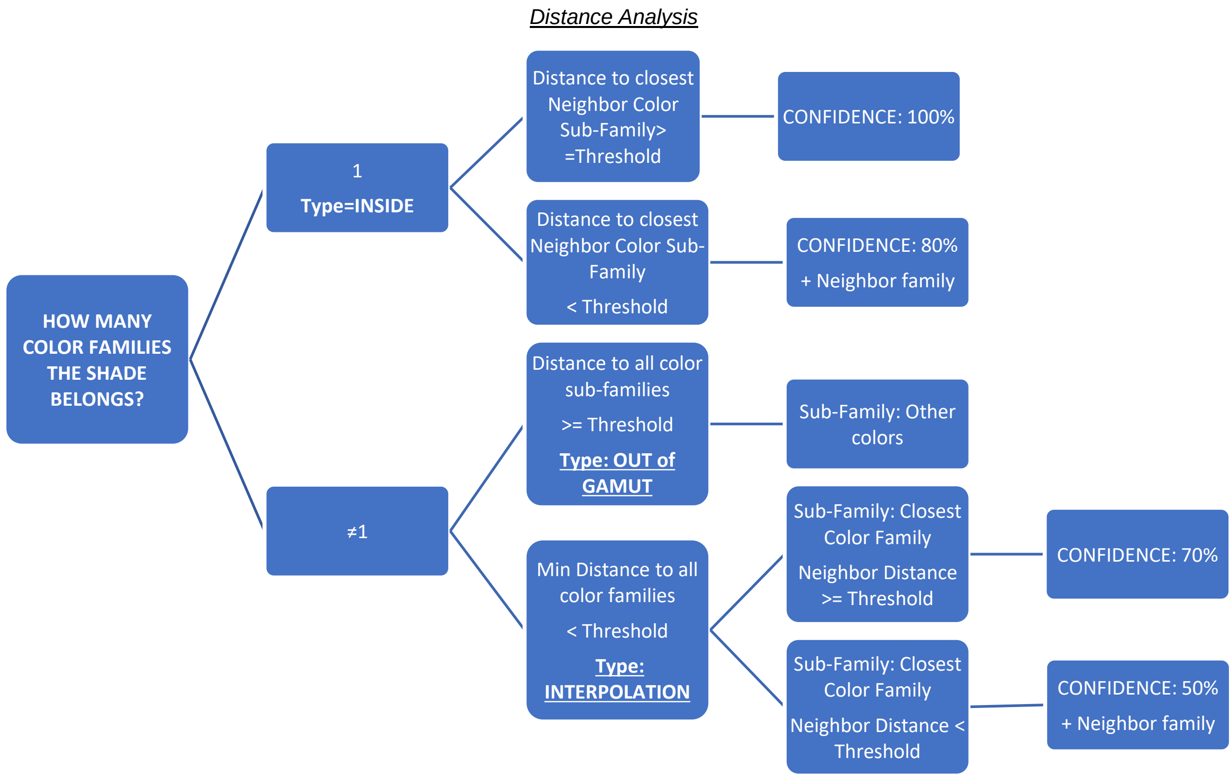
(b) Interpolation (Count True ≠1 and distance to the closest sub-family is within the threshold): The shade may fall within multiple sub-families' convex hulls or outside all of them. The shade is assigned to the closest sub-family. The confidence level and neighboring sub-family are then determined:

- 50% confidence: The distances to both the closest and second-closest sub-families are within the threshold. The shade is assigned to the closest sub-family, but the second-closest is also very close, hence the reduced confidence.
- 70% confidence: The distance to the closest sub-family is within the threshold, while the distance to the second-closest is beyond it. The second-closest sub-family (even if outside the threshold) is still identified as the neighbor to provide additional context.

(c) Out of Gamut (distance to all sub-families is beyond the threshold): The shade falls outside all sub-family convex hulls. The shade is considered “Other colors”. The closest sub-family is still indicated for context.

This refined approach, using the "Count True" parameter and distance calculations, allows for accurate classification of complex shades, providing a confidence level and neighboring sub-family information. Figure 4 illustrates the algorithm's classification steps.

Figure 5 illustrates examples of shades classified based on their positioning in the L*a*b* color space.



Figue 4: Diagram illustrating the methodology.

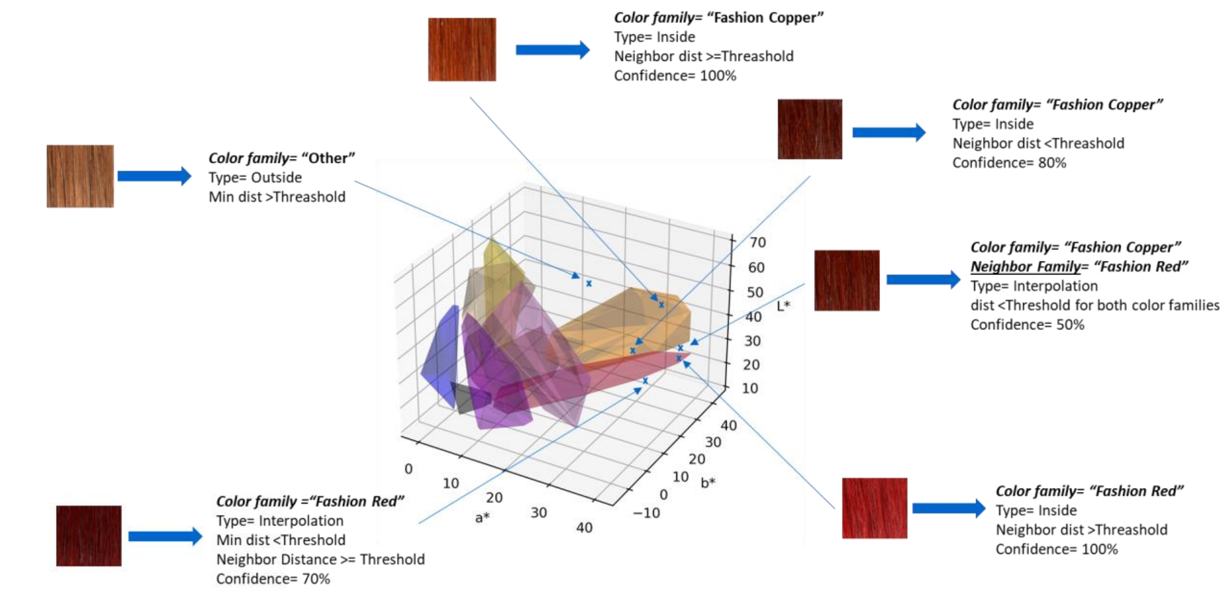


Figure 5: Examples of hair swatches classified based on their L*a*b* positions regarding all the convexhuls.

3. Results

Classification of hair color gamut :

Our initial objective was to map the landscape of existing hair color shades available on the market and understand their relationships within the L*a*b* color space. We compiled a dataset of 2578 shades, all measured using our standardized in vitro protocol. Applying the Nearest Convex Hull algorithm to this dataset successfully classified the shades into distinct families and sub-families, as detailed in Tables I and II.

Natural	Natural Fashion	Fashion
803	1216	559

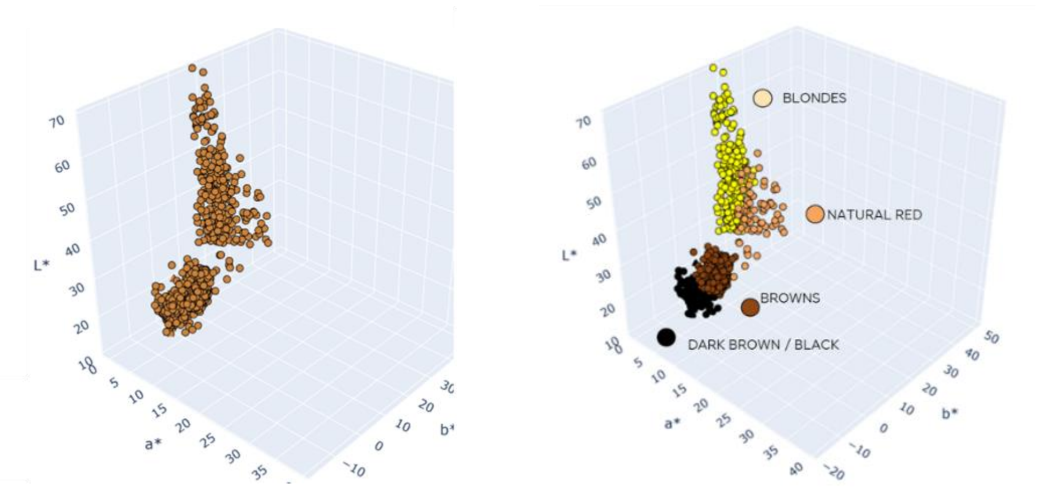
Table I. Distribution of hair color shades across Families.

Family	Sub-Family	Number of shades
Natural	Blonde	225
	Brown	248
	Dark Brown / Black	254
	Natural Red	76
Natural Fashion	Beige	551
	Natural Fashion Brown	584
	Natural Fashion Dark Brown/ Black	81
Fashion	Blue	39
	Violet	105
	Red	201
	Pink	47
	Copper	167

Table II. Distribution of hair color shades across Sub-Families.

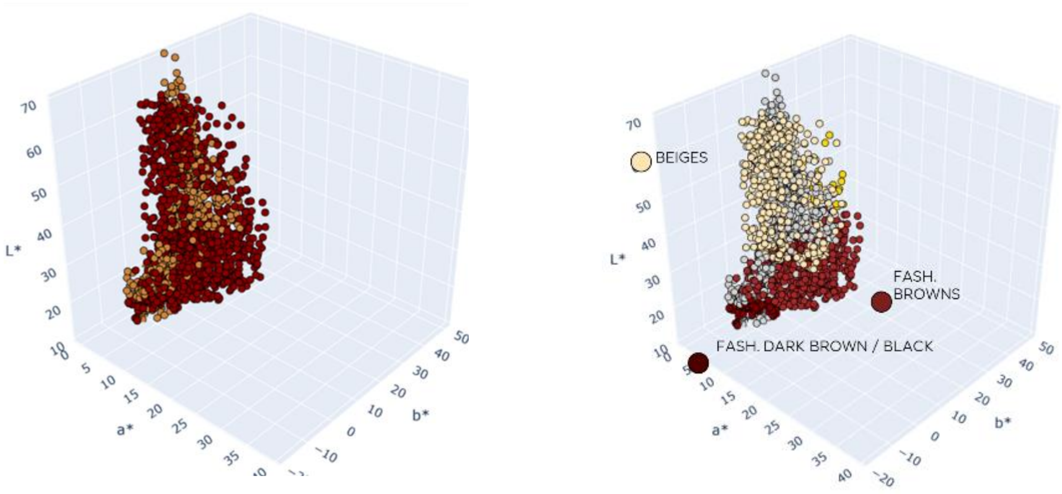
Table I reveals that the Natural Fashion family (1216 shades) is the most represented in our dataset, followed by the Natural family (803 shades) and then the Fashion family (559 shades). This distribution suggests a market emphasis on shades that retain a degree of naturalness while offering fashionable variations. Table II provides a more granular view, showing that within the Natural Fashion family, Beige (551 shades) and Natural Fashion Brown (584 shades) are the most prevalent sub-families. This further emphasizes the popularity of brown and beige tones with fashionable reflects in the current market landscape. In the Natural family, the distribution across Blonde (225), Brown (248), and Dark Brown/Black (254) sub-families is relatively even, indicating balanced representation across the natural tone spectrum.

This classification, visualized in Figures 6-8, demonstrates a clear hierarchical structure. Natural colors form a core palette across the range of tone levels, as expected (Figure 6). Natural Fashion shades extend this core towards more expressive colors (Figure 7). For instance, the Natural Fashion Beige sub-family expands upon the Natural Blonde sub-family by incorporating reflects like ashy, red, copper, and gold, showcasing the algorithm's ability to capture nuanced variations. Similarly, Natural Fashion Brown introduces similar reflect variations to the Natural Brown range. Finally, Fashion colors occupy the outermost layer, representing the most vibrant and saturated hues (Figure 8), with each sub-family clustering in distinct L*a*b* zones.



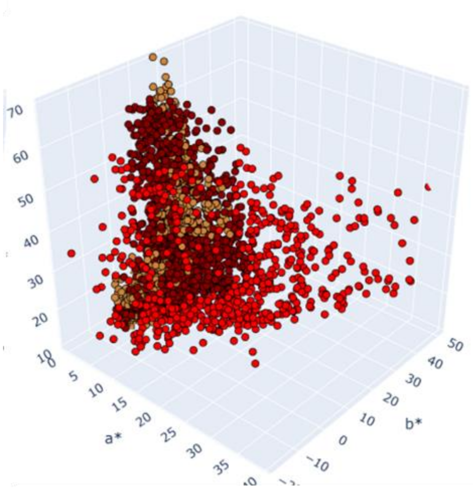
(**a**) (**b**)

Figure 6: Distribution of Natural shades in the L*a*b* color space (a) and their classification by sub-families (b).

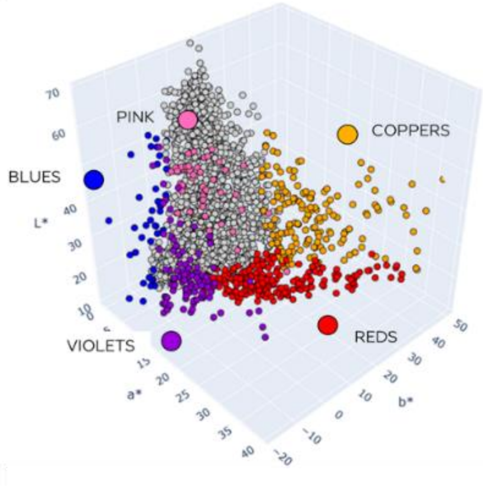


(**a**) (**b**)

Figure 7: Distribution of Natural Fashion shades in the L*a*b* color space (a) and their classification by sub-families (b).



(**a**)



(**b**)

Figure 8: Distribution of Fashion shades in the L*a*b* color space (a) and their classification by sub-families (b).

4. Discussion

Our initial approach using Nearest Convex Hull classification, while effective for broad family and sub-family distinctions, encountered limitations in the finer-grained sub-categorization by undertone, chroma, and lightness likely due to the complexity of these intertwined visual attributes. The difficulty in achieving minimally overlapping convex hulls for these sub-categories suggests that these visual attributes are more complex than initially hypothesized. This complexity necessitates a more sophisticated machine learning approach.

As a next step, we plan to investigate the application of various machine learning algorithms directly on the L*a*b* data for sub-category classification. If this proves insufficient, we will leverage computer vision techniques to extract relevant features from the images to capture nuanced color variations and texture information and use these as input for the machine learning models. This combined approach has the potential to overcome the limitations observed with the convex hull method and achieve higher accuracy in classifying hair color shades according to the full range of consumer-relevant attributes. Future research could explore the effectiveness of different machine learning models for this task, as well as the development of adaptive thresholding techniques to further optimize classification performance.

5. Conclusion

This research introduces a novel, data-driven framework for hair color classification that directly addresses the limitations of existing systems. By leveraging objective L*a*b* measurements and a consumer-centric approach, we have developed a system that offers improved accuracy, scalability, and clarity compared to traditional methods. The hierarchical structure of families, sub-families, and sub-categories, based on expert knowledge and data analysis, provides a robust foundation for personalized recommendations, portfolio management, and future product development. While the Nearest Convex Hull approach proved valuable for initial classification, the transition to a machine learning approach incorporating L*a*b* data and image data offers significant potential for enhancing the classification of nuanced color attributes. Future work will focus on refining these machine learning models and exploring the integration of additional visual features to further improve its accuracy and applicability. This framework holds promise for empowering consumers to navigate the complex landscape of hair color options with greater confidence and for driving innovation within the beauty industry toward more personalized and effective products and services.

References:

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.ConvexHull.html

[2] Georgi I. Nalbantov, Patrick J. F. Groenen, and Jan C. Bioch: Nearest Convex Hull Classification.