IFSCC 35TH
CONGRESS
FRANCE_CANNES 15-18 SEPT 25
THE FUTURE IS SCIENCE
ifscc2025.com
IFSCC
SOCIÉTÉ FRANÇAISE DE COSMÉTOLOGIE

IFSCC 2025 full paper (IFSCC2025-1602)

"AI-Powered Lipstick Color Prediction: Designing Inclusive Shade Ranges for Diverse Skin Tones"

Haiting GU [1,2], Nathalie GEFFROY HYLAND [1,*] , Salma SALHI [1,3] , Sheldon WU [2] , Julie COUTET [1] , and Alexander JASPERS [4]

[1] L'Oréal Research and Innovation, Chevilly-Larue, France;
[2] L'Oréal Research and Innovation, Shanghai, China;
[3] Meritis Technologies, Paris, France;
[4] L'Oréal Research and Innovation, Kawasaki, Japan

1. Introduction

Beauty, a multifaceted symbol deeply rooted in human culture and psychology, plays a pivotal role in self-expression, social perception, and confidence [1]. Cosmetics, particularly lipstick, serves as a powerful icon for enhancing aesthetic appeal and conveying individuality [2]. Lipstick color holds a certain allure, with research indicating that specific shades can subtly enhance perceptions of a woman's attractiveness and femininity [3-4]. However, selecting a flattering shade remains a persistent challenge. The cosmetic industry has historically underserved diverse skin tones, particularly deeper complexions [5]. This results in limited products that fail to include and fully embrace deep skin tone, while simultaneously offering overwhelming lipstick shades designed for light skin tone. Without clear guidance on undertone compatibility, consumers risk purchasing colors that is frustrating or mismatches with their skin tones. All these challenges highlight the need for more inclusive product development and user-friendly tools to simplify the selection.

Prior studies have explored consumer preference and perceived impression of lipstick colors under multiple lighting conditions [6-7]. However, existing methodology is constrained by both limited participant diversity (often < 65 subjects) and narrow color gamut of tested shades (typically 15-20 colors). Additionally, the cross-cultural comparisons of color preference patterns across geographical markets are understudied. Suitability to skin tone has been proved as one of the most critical attributes in lipstick selection [8]. However, the mechanism of quantifying suitability remains poorly understood.

To address these challenges, this research pioneers the development of the first AI-driven lipstick model in predicting the suitability of shades on various skin tones. This model was trained on 204,000 consumer evaluations from four geographically distinct markets, representing the largest and most diverse dataset ever used for lipstick color analysis. The predicted results empower cosmetic companies to create the most inclusive and comprehensive shade ranges, while simultaneously providing consumers with a more streamlined and personalized method for selecting colors.

2. Materials and Methods

2.1. Data Acquisition

Data collection was conducted through a global study across four countries (France, China, US, India) to capture the diverse needs of consumers with varying skin tones, ages, and cultural backgrounds [9]. 2040 Participants were evenly recruited among 6 skin tone clusters in each country, with 120 consumers per cluster, ensuring balanced representation of skin tone diversity. The 6 skin tone clusters were defined based on spectral measurements taken from around 3000 women worldwide, consisting of C1-very light, C2-light cool, C3-light warm, C4-medium, C5-deep, and C6-very deep [10]. A cutting-edge virtual try-on algorithm [9] was employed to simulate 100 lipstick shades on each participant, dynamically adapting to individual bare lip tone to ensure a realistic and faithful representation of how shades will appear. Participants evaluated each simulated shade in randomized order, rating suitability, wearability and other relevant attributes on a scale of 1-5. Ratings of 1-2 indicate disagreement or dislike, 3 reflects neutrality, and 4-5 signify agreement or preference. In total, 204,000 evaluations were accumulated to establish a lipstick color database linking demographic data, skin tone, age and preference rating.

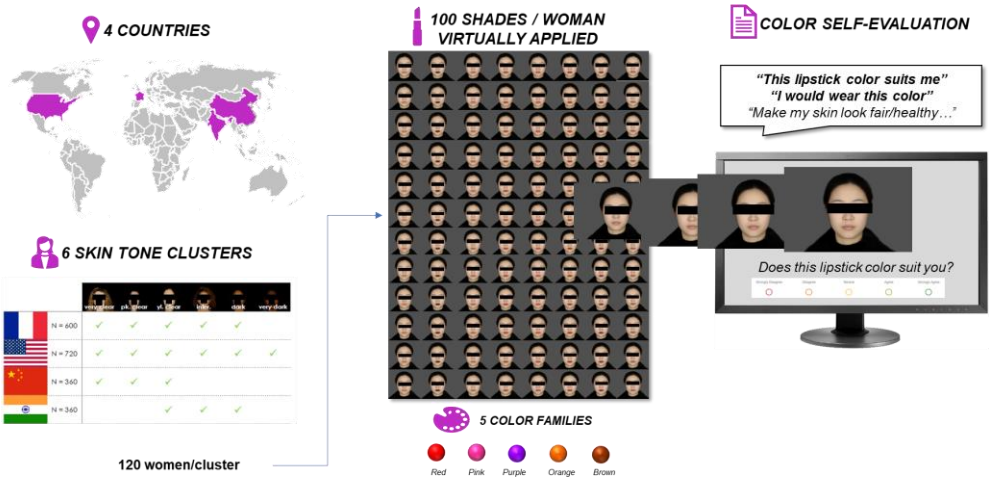


Figure 1. Data acquisition flow from a global lipstick color study

2.2. Data Quality

Data quality is foundational to AI model training because it directly determines the model's accuracy, fairness, and generalizability. High-quality data guarantee that learned patterns reflect real-world diversity rather than artifacts of biased and incomplete datasets. The quality of dataset in this research is ensured through large-scale participant recruitment, comprehensive testing of a wide lipstick color gamut, and repeated evaluations to confirm the rating consistency.

In US and India studies, three shades were evaluated twice per participant in randomized sequence to access the intra-participant variation. If the first and second answers were both 'agree' (i.e., 4 or 5), or both 'disagree' (i.e., 1 or 2), it was defined as a consistent rating. Otherwise, it would be an inconsistent answer, suggesting instable participant and unreliable data. As in Table 1, all the three shades showed high repeatability above 70%, indicating that more than 70% of participants gave similar ratings on the same color. It verified the consistency of rating throughout the whole evaluation, and the high quality of acquired database, establishing the robustness for downstream analytical and AI-training applications.

Table 1. Intra-participant consistency based on ratings of repeated evaluations

Repeated Shade	US	India
Shade A	76.8%	87.4%
Shade B	83.9%	74.8%
Shade C	80.9%	73.3%

2.3. Data Pre-process

The database, as previously described, was constructed by integrating participant ratings and their facial image data. To ensure accurate color analysis, lip and skin color values were extracted from images using facial landmarks from existing functions like Python Dlib [11]. For each evaluated image, the lip region was detected, and all lip color pixels were identified. The median pixel color was ultimately selected as the representative color for the lip region. This approach captures the central area of the lip, which is less affected by edge artifacts or gloss finish. Furthermore, the median pixel exhibited the least mean color difference when compared to the original lip pixels, ensuring accurate and consistent representation (Figure 2). Similarly, by excluding facial features defined by facial landmarks, a facial skin mask was generated, containing only skin pixels. Unlike the lip region, the facial skin exhibits greater inhomogeneity, characterized by tonal variations such as lighter cheek areas and darker jawline zones. This variability makes it challenging to use the median value to represent the entire face accurately. A clustering algorithm, such as k-means clustering [12], was applied to group the skin pixels into two distinct clusters, effectively separating the lighter and shadowed regions of the face. The centroid of the lighter cluster, referred to as 'Dom2,' was validated as a representative color for overall facial perception through a small psychophysical experiment with internal color experts.

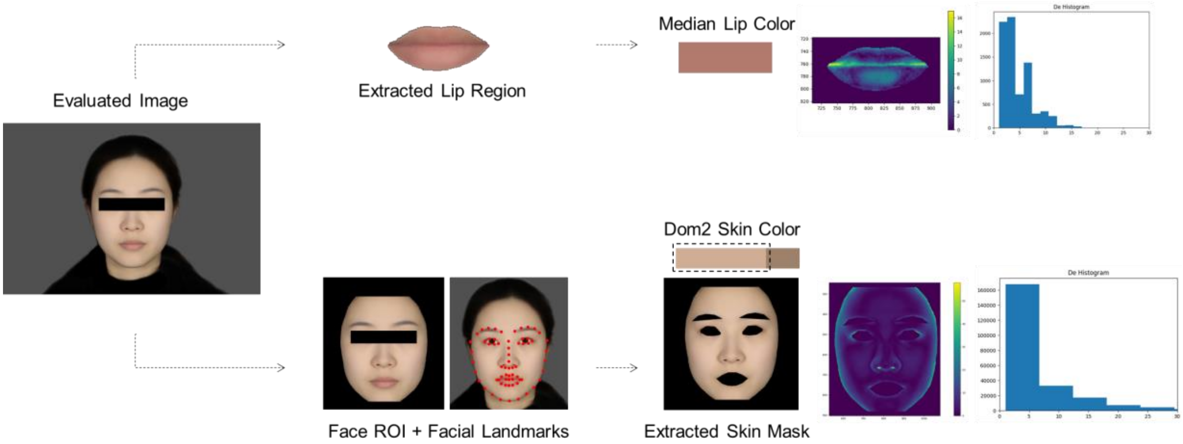


Figure 2. Workflow of image color extraction. The process includes several steps like detected lip/facial skin region, defined representative color, calculated color difference heatmap and histogram

To enhance color analysis aligned with human perception, the extracted median and Dom2 RGB values were converted into CIELAB color coordinates [13]. Figure 3 illustrates examples of the lip color database mapped in the CIELAB space. Key attributes, including L* (lightness), a* (reddishness-greenness), b* (yellowness-blueness), C* (Chroma), and h (hue angle), derived from the extracted skin and lip colors, were utilized as core input parameters for model training.

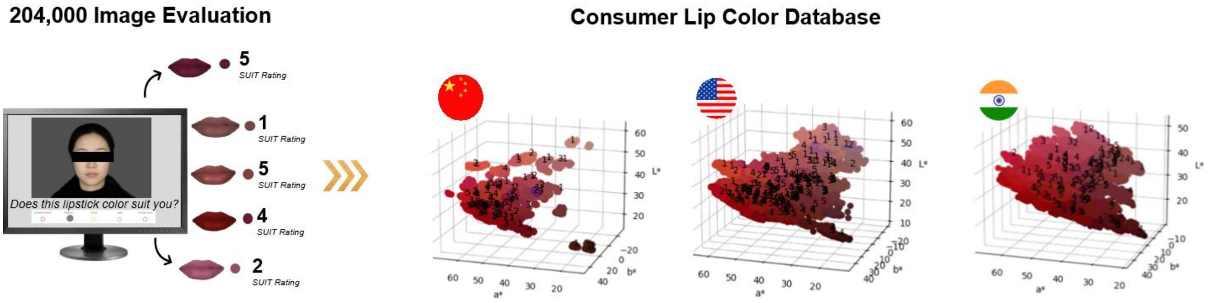


Figure 3. Examples of extracted lip color database in CIELAB color space

2.4. AI Predictive Model

As a first step, all input parameters were standardized. A random forest classifier [14-15] was employed to evaluate feature importance and identify the most relevant inputs for predicting color suitability. The multivariate analysis confirmed the significant cross-cultural disparities in lipstick color preference in four studied countries. Skin tone characteristics were identified as significant predictors in this model, demonstrating a meaningful improvement in predictive accuracy. This observed relationship may be mechanistically explained through two main factors: perceptual harmony and simultaneous contrast effect. Aligned skin undertones contributed to the harmony with lip colors, elevating aesthetic cohesion. Meanwhile, skin and lip colors mutually influence each other's perceived color appearance, impacting overall color suitability. Age was also found to be a key input feature. In contrast, other input parameters such as lipstick coverage, makeup purpose, and lip color selection reasons, were determined to show lower relevance in predicting outcomes.

Following the selection of relevant input features, various AI methodologies were evaluated to identify the optimal model for predicting lipstick suitability. Support Vector Regression (SVR) [16] was determined to be the superior approach, exhibiting statistically high prediction accuracy compared to alternatives such as neural networks and decision trees. This advantage can be attributed to its kernel-based architecture, which efficiently captures nonlinear relationships between chromatic color features and suitability ratings. Additionally, SVR was chosen for its ability to balance high performance with computational efficiency, offering rapid runtimes and reduced resource consumption. To enhance the model further, hyperparameters such as the kernel, regularization parameter, and epsilon were optimized using k-fold cross-validation coupled with a grid search method [17]. This combined approach ensured robust generalization by minimizing overfitting risks. Once the optimal hyperparameters was identified, the final SVR model was trained on the complete training dataset to ensure robust and reliable predictions.

2.5. Inclusive Color Recommendations

An analytical approach, Total Unduplicated Reach and Frequency (TURF) analysis [18-19], was employed to identify the optimal combination of unique colors that maximizes the overall reach rate. Reach rate refers to the number or percentage of targeted participants who find at least one color in the combination suitable for their skin tones. Unique colors were derived from the previously mentioned lip color database by grouping similar shades with a

CIEDE2000 [20] color difference ≤ 1.0, a threshold supported by literature indicating that the human eye cannot distinguish colors with a ∆E2000 value below 1.1 [21-22]. Each grouped color is characterized by attributes such as color family, participant IDs, the number of participants evaluated, and the number of participants reached (rated 4 or 5). The TURF analysis begins with the color exhibiting the highest reach rate and iteratively evaluates all possible combinations to identify subsequent colors that maximize reach among targeted participants. This recommendation process can be fully customized to meet specific constraints, such as defined color count, desired color family ratios, or targeted skin tones. By leveraging this data-driven approach, the resulting color combinations reflect consumer preferences across diverse demographics, promoting inclusivity and enhancing satisfaction.

3. Results

3.1. Prediction Accuracy

Model prediction accuracy was evaluated using k-fold cross-validation, where the dataset was partitioned into 80% training and 20% testing subsets iteratively to ensure robust generalization. For comparison with ordinal rating, continuous suitability scores predicted by the SVR model were discretized into three performance levels: < 2.8 (Not Suit), 2.8-3.6 (Neutral), >3.6 (Suit). These thresholds were derived by analyzing the distribution of predicted suitability scores and identifying natural breaks within the database.

This predictive model achieved high accuracy in classifying lipstick color suitability, with 73.7% precision for “Not Suit” and 78.0% for “Suit” predictions on US dataset (Table 2). Nonetheless, lower accuracy was observed for 'Neutral' predictions, which is expected given the inherent challenges of classifying ambiguous cases. These metrics indicated the robust discrimination between clearly suit and incompatible colors, offering valuable insights in reliably pre-screening performing colors in shade development.

Table 2. Prediction accuracy of SVR model by cross-validation on US dataset

	Prediction: Not Suit	**Prediction: Neutral**	**Prediction: Suit**
Truth: Not Suit (1,2)	14059 (**73.7%**)	4591	4819
Truth: Neutral (3)	1398	1524	3187
Truth: Suit (4,5)	3609	5633	28315 (**78.0%**)

3.2. Business Application: AI-Powered Shade Analysis

The power of AI predictive model in shade analysis lies in its ability to transform subjective aesthetics into scalable, data-driven decisions. As illustrated in Figure 4, for each test shade, by inputting its application color and representative skin colors, this AI model can generate the suitability predictions across diverse skin clusters.

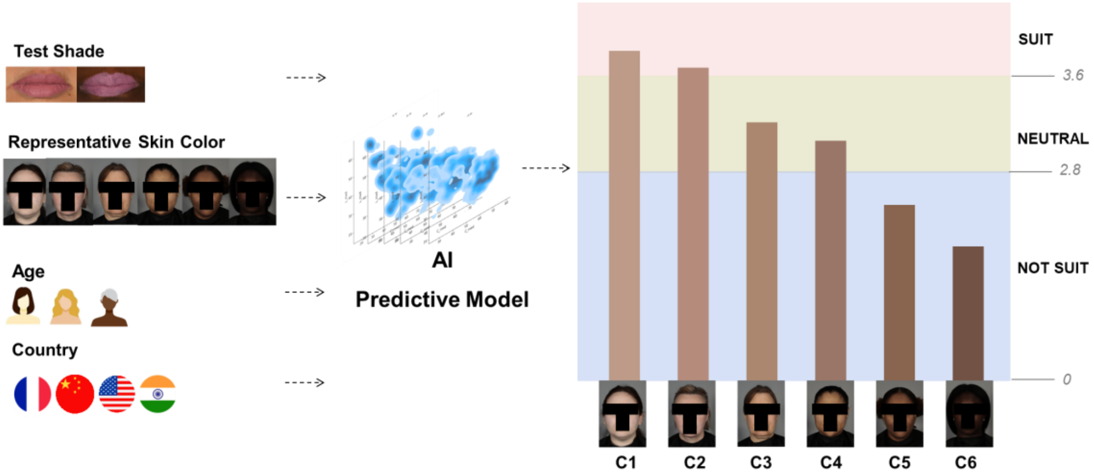
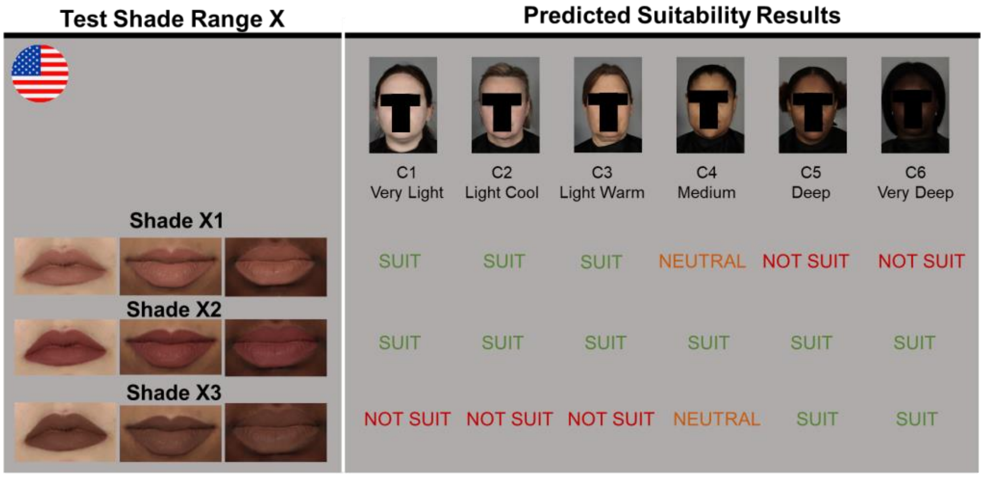


Figure 4. Workflow for predicting shade suitability using AI Model

As an example, the AI-powered shade analysis of specific shade range X in US market is shown in Figure 5. Shade X2 achieved “Suit” predictions on all skin clusters from C1 to C6, which can be strategically positioned as a universal “star shade”. Shade X1, validated for clusters C1 to C3, is particularly relevant for light skin tones. In contrast, Shade X3 demonstrated strong suitability for clusters C5 to C6, making it an ideal choice for deep skin tones and addressing the inclusive need gaps. Similar analysis can be applied to other markets like China, France and India, uncovering nuanced cultural beauty preferences and informing a comprehensive global shade strategy.



Test Shade Range X

Predicted Suitability Results

Test Shade Range X	C1 Very Light	C2 Light Cool	C3 Light Warm	C4 Medium	C5 Deep	C6 Very Deep
Shade X1	SUIT	SUIT	SUIT	NEUTRAL	NOT SUIT	NOT SUIT
Shade X2	SUIT	SUIT	SUIT	SUIT	SUIT	SUIT
Shade X3	NOT SUIT	NOT SUIT	NOT SUIT	NEUTRAL	SUIT	SUIT

Figure 5. Example of AI-powered shade analysis for shade range X in US market

Another major strength is the ability to translate predictions into consumer-friendly shade guidance (Figure 6). By using this mapping, consumers can effortlessly identify suitable shade options based on their skin tones, minimizing the risk of dissatisfaction from online purchases.

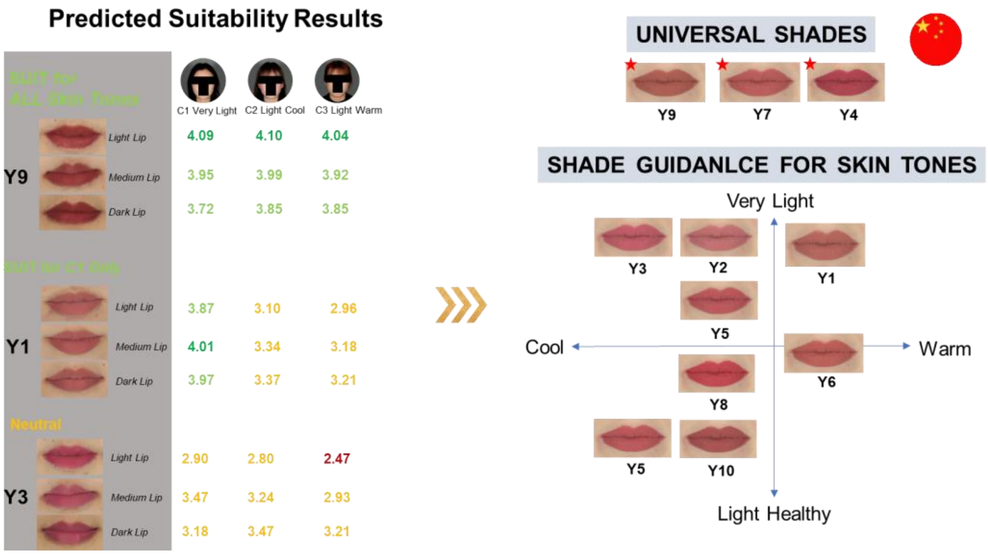


Figure 6. Example of shade guidance based on suitability predictions for China

3.3. Business Application: AI-Powered Shade Inclusivity

The integration of AI predictive model with TURF analysis offers an innovative framework for designing truly inclusive shade ranges that maximize consumer satisfaction. A practical application is illustrated in the optimization of shade range Z for Indian market (Figure 7). Initially, the suitability of existing shades for Indian skin clusters (C3, C4 and C5) was assessed through model predictions. The optimized shade range retained high-performing existing shades (Z1-Z5) while strategically incorporating new colors. TURF analysis identified the optimal combination of additions, including deep brown, universal purple, and pink colors. Following this AI-driven optimization, the shade range Z was transformed into a truly inclusive range with 98% reach rate, indicating 98% of Indian consumers found at least one suitable shade within the range.

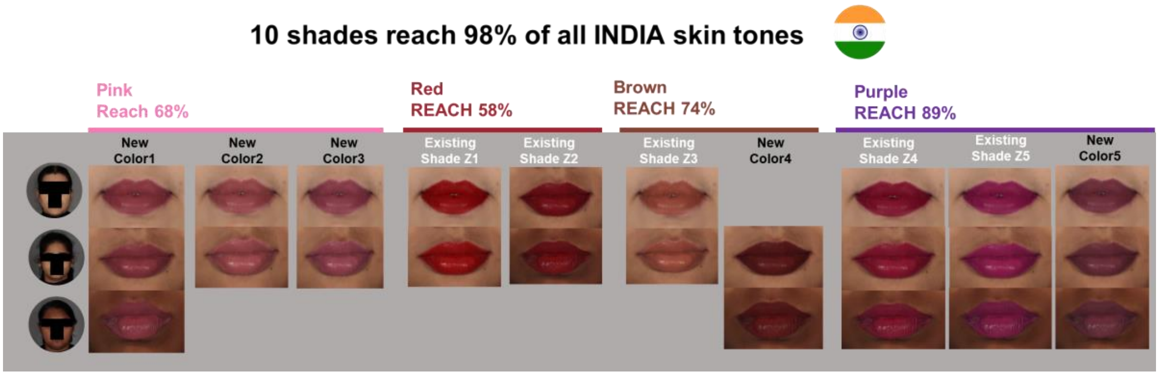


Figure 7. Example of AI-powered inclusive shade range for India

Leveraging these insights, the whole digital process empowers cosmetic companies to analyze shade performance, optimize existing offerings, and create highly satisfying lipstick color ranges to meet the diverse needs of global consumers.

4. Discussion

This research introduces a groundbreaking data-driven framework for predicting lipstick color suitability and developing inclusive shade ranges. Leveraging a global database of 204,000 consumer images and evaluations, this AI predictive model decodes intricate relationship between skin tones, lip colors and cultural preferences.

This novel framework represents a revolution from traditional methods, bringing a significant advancement in inclusivity and efficiency. Traditionally, shade development has depended on subjective inputs from small focus groups, designer intuition, or limited market testing, often resulting in the unintentional prioritization of certain skin undertones or prevailing trends. Additionally, existing methods for validating lipstick shade suitability often rely on small-scale consumer tests involving approximately 60 participants over a two-week period. This is resource-intensive and time-consuming, while also suffering from subjective biases and narrow sample sizes. In contrast, this AI-powered analysis can predict shade suitability on diverse skin tones and various countries in just hours. The extensive number of participants enhances the generalizability of predictions. The high accuracy, exceeding 70%, serves as evidence of the robustness and reliability of the results. Furthermore, by reducing the development timeline from months to days, this framework delivers inclusive shade ranges with maximum satisfaction. The fast, data-driven results are statistically validated to meet the diverse beauty needs of every individual worldwide.

Existing research on lipstick color suitability remains limited, with most studies focusing on correlation between lipstick color and perceived impressions [6-7]. A preference for orange-red lipstick colors among Chinese female was identified in [6], aligning with findings in this research: Chinese participants in database rated orange and red color highest, while US participants favored pink and brown, and India participants prioritized purple and brown. Few research explored the compatibility between skin and lip color. This research is the first comprehensive and global study to bridge this research gap, providing accurate predictions of color suitability in relation to skin tone.

This research may face limitations in reflecting the evolution of lipstick color preferences. It relies on static database collected in a specific year, which risks becoming outdated over time. Future research could explore an agile, high-quality way of acquiring consumer evaluations, enabling dynamic database updates while preserving high accuracy. Such advancements would ensure that this AI-powered process remains relevant and continues to resonate with consumers across diverse demographics, promoting inclusivity and enhancing customer satisfaction in the fast-paced beauty industry.

5. Conclusion

Leveraging a global database of 204,000 consumer evaluations, this research pioneered an AI-powered framework for predicting lipstick color suitability and designing inclusive shade ranges. The SVR predictive model achieved high accuracy over 70% in classifying “Suit” and “Not Suit” shades, streamlining the development process with data-driven insights. Integrating the AI model with TURF analysis further enables the design of optimized shade ranges, maximizing inclusivity and consumer satisfaction.

The whole approach redefines inclusivity as a scientific, scalable process, not a marketing buzzword. By replacing subjective bias with data, and inefficiency with speed, it empowers cosmetic companies to analyze shade performance, refine existing products, and develop truly inclusive collections, ultimately resonating the beauty needs from every unique skin undertone.

References:

1. Dutton, D. (2009). *The art instinct: Beauty, pleasure, & human evolution.* Oxford University Press, USA.
2. Gurrieri, L., & Drenten, J. (2021). The feminist politics of choice: lipstick as a marketplace icon. *Consumption Markets & Culture*, *24*(3), 225-240.
3. Stephen, I. D., & McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception*, 39(8), 1104-1110.
4. Guéguen, N., & Jacob, C. (2012). Lipstick and tipping behavior: when red lipstick enhance waitresses tips. *International Journal of Hospitality Management*, *31*(4), 1333-1335.
5. Van, A. Y. (2017). *More than Skin Deep: An Analysis of Black Women's Experiences with Race, Skin Tone, and Cosmetics*. Minnesota State University, Mankato.
6. Tian, B., Gong, H., Chen, Z., Yu, X., Pointer, M. R., Yu, J., ... & Liu, Q. (2023). Assessment of color preference, purchase intention and sexual attractiveness of lipstick colors under multiple lighting conditions. *Frontiers in Neuroscience*, *17*, 1280270.
7. Wu, Y. A., Gong, S. M., & Lee, W. Y. (2024). A study on the impressions induced by lipstick colors. *Color Research & Application*, *49*(4), 374-383.
8. Wartaka, M. (2016). Analysis of the consumers preferences of lipstick product and its relationship with the segmentation of the lipstick products. *The Management Journal of Binaniaga*, *1*(2).
9. Geffroy-Hyland N, Qiao Y, Phan Van Song T, Gu H, et al (2023) Augmented lip color evaluation for multicultural consumers. *IFSCC Congress 2023*: Poster – PRB -139.
10. Qiao Y, Atis B, Cointereau-Chardon S, et al (2020) Developing the most inclusive and relevant liquid foundation ranges for multicultural consumers. *IFSCC Congress 2020*: Poster – 328
11. King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, *10*, 1755-1758.
12. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (Vol. 5, pp. 281-298).
13. CIE 15: 2004 Colorimetry, Vienna, Austria: *Commission Internationale de l'Eclairage*; 2004.
14. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
15. Huang, B. F., & Boutros, P. C. (2016). The parameter sensitivity of random forests. *BMC bioinformatics*, *17*, 1-13.
16. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*, 199-222.
17. Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American statistical association*, *78*(382), 316-331.
18. Marti ML, Jain DC (1994) Optimal product positioning: A TURF analysis approach. *Journal of Advertising Research*, 34(5): 27-33.
19. Koenigsberg O, Czepiel JA (1996) Improving media selection decisions: A TURF analysis approach. *Journal of Advertising Research*, 36(6): 21-28.
20. Luo, M. R., Cui, G., & Rigg, B. (2001). The development of the CIE 2000 colour‐difference formula: CIEDE2000*. Color Research & Application*, 26(5), 340-350.
21. Luo, M. R., & Rigg, B. (1986). Chromaticity‐discrimination ellipses for surface colours. *Color Research & Application*, 11(1), 25-42.
22. Huang, M., Liu, H., Cui, G., Luo, M. R., & Melgosa, M. (2012). Evaluation of threshold color differences using printed samples. *Journal of the Optical Society of America A*, *29*(6), 883-891.