UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Unlocking the Face Code: How Facial Characteristics Drive Social Biases

Permalink
https://escholarship.org/uc/item/25g3s339

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Gurkan, Necdet
Rathod, Vismay

Publication Date
2024

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Unlocking the Face Code: How Facial Characteristics Drive Social Biases

Necdet Gürkan (ngurkan@stevens.edu)
School of Business, 1 Castle Point Terrace
Hoboken, NJ 07030, USA

Vismay Rathod (vrathod4@stevens.edu)
Computer Science, 1 Castle Point Terrace
Hoboken, NJ 07030, USA

Abstract

People rapidly form first impressions based on facial appearances, which have significant real-life consequences. While various computational models have been developed to analyze how facial characteristics influence these impressions, they often have limitations, such as focusing on limited trait impressions, restricted facial characteristics, reliance on black-box machine learning methods, and dependency on manual annotations. In this study, we address these shortcomings by utilizing recent advancements in computer vision to extract human-interpretable, quantitative measures of facial characteristics (e.g., facial morphological features and skin color) and emotional attributes from face images. Using machine learning techniques, we modeled 34 first impressions and validated our model's generalizability and predictive accuracy with out-of-sample faces. Our model demonstrates the relative importance of facial characteristics and emotional attributes in shaping these 34 first impressions. Our results provide a comprehensive understanding of how various facial characteristics and emotional attributes collectively influence social biases.

Keywords: first impressions, computational model, face perception

Introduction

The human face is a primary source of visual information for recognizing individuals and interpreting their emotions and mental states. Despite common advice against judging a book by its cover, people frequently infer various attributes, such as trustworthiness, dominance, or intelligence, based on the facial characteristics of strangers. These judgments are formed quickly and have a salient impact on human behavior (Olivola, Funk, & Todorov, 2014), ranging from mate selection (Langlois et al., 2000), sentencing decisions (Funk & Todorov, 2013), employment opportunities (Olivola, Eubanks, & Lovelace, 2014) to voting patterns (Ballew & Todorov, 2007). Given the importance of first impressions in everyday decisions, including the increasingly pervasive use of images of faces on social media, an important question arises: Which facial characteristics do people rely on when forming first impressions from faces?

Previous research has identified numerous facial features associated with the perception of trait impressions (Hehman et al., 2019; Todorov et al., 2015). These insights form the basis for more comprehensive theories of social perception, seeking to elucidate the accuracy and practical importance of these trait impressions (Todorov, Said, Engell, & Oosterhof, 2008; Zebrowitz, 2017). One key area of focus is the structural similarity between a person's facial features and emotional expressions. Faces that resemble expressions of happiness are often perceived as trustworthy, while those resembling anger are seen as dominant (Adams Jr, Nelson, Soto, Hess, & Kleck, 2012; Said, Sebe, & Todorov, 2009). Overgeneralization Theory explains this phenomenon, suggesting that people are highly attuned to emotional expressions in faces, leading them to perceive emotions and associated traits even in neutral faces (Todorov et al., 2008; Zebrowitz, 2017). This theory indicates that these perceived emotion significantly influence impression formation, influenced by an overly responsive emotion detection system.

Moreover, various theories in impression formation emphasize different facial attributes. For example, facial morphological features influence perceptions of trustworthiness and dominance (Geniole, Molnar, Carré, & McCormick, 2014; Ormiston, Wong, & Haselhuhn, 2017). While some studies argue that these features are markers of behavioral tendencies due to the influence of biological factors on both facial morphology and behavior (Carré, McCormick, & Mondloch, 2009), others contend that there is no clear relationship between facial morphological features and behavioral tendencies (D. Wang, Nair, Kouchaki, Zajac, & Zhao, 2019). Jaeger and Jones (2022) found that some facial morphology features plays a role in impression formation, suggesting a potential link between facial structure and trait dispositions.

Beyond emotional expressions and facial morphological features, skin color significantly influences first impressions of faces (Stepanova & Strube, 2009; Strom, Zebrowitz, Zhang, Bronstad, & Lee, 2012). Darker skin tones are often associated with more stereotyped trait impressions (Blair, Judd, Sadler, & Jenkins, 2002), negative associations (Livingston & Brewer, 2002), and harsher penalties in the criminal justice system (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Much of the research on race perception in facial features has pointed to variations in skin tone as a key factor. Studies where participants were asked to rate the importance of various facial features and skin color in determining a person's race have found that skin color is perceived as the most significant indicator (Brown, Dane, & Durham, 1998).

Many recent studies have employed modern deep learning methods to develop a more comprehensive understanding of trait impressions and to generate face stimuli for face percep-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*. ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

tion experiments (Liu et al., 2014; Liang, Jin, & Li, 2014; Peterson, Uddenberg, Griffiths, Todorov, & Suchow, 2022). For instance, Peterson et al. (2022) demonstrated that deep features obtained from StyleGAN2 (Karras et al., 2020), a synthetic face generator, can be used to predict first impressions of faces. Gurkan and Suchow (2022a) further developed this approach by utilizing these deep features as a latent construct within the Cultural Consensus Theory (Romney, Batchelder, & Weller, 1987) to align these features with culturally constructed beliefs. Although these methods demonstrate high predictive capabilities, they lack interpretability regarding which specific features contribute to various traits.

While there has been significant effort to develop computational models for understanding the contribution of facial characteristics to first impressions, previous studies have often focused on the effects of only one or a few features. These studies commonly used non-photorealistic face images and a limited set of stimuli. To address this gap, some researchers have begun building computational models that combine the effects of multiple facial characteristics using larger sets of photorealistic face images (Jaeger & Jones, 2022; Vernon, Sutherland, Young, & Hartley, 2014; McCurrie et al., 2017). However, these studies face several challenges. First, models often restrict their focus to a narrow range of trait impressions, although the dominance and valence model is argued to generalize findings (Todorov et al., 2008). Second, there has been a limited variety of facial morphological features studied, with a heavy reliance on measures like the Facial Width-to-Height Ratio (FWHR) and emotion resemblance. Lastly, these methods often depend on manual annotations of facial characteristics, which can limit the scalability and efficiency of the research process.

To address the shortcomings of current computational models, we modeled first impressions of faces by analyzing facial characteristics and emotional attributes to identify the underlying factors contributing to these impressions. We utilized recent advancements in computer vision to extract human-interpretable, quantitative measures from craniofacial features—drawing on anthropometric studies of facial morphology (Farkas, 1995; Farkas, Katic, & Forrest, 2005)-, skin color (Chardon, Cretois, & Hourseau, 1991), and emotion analysis (Serengil & Ozpinar, 2021). Given the high correlation among some features, we trained our model using Lasso Regression (Tibshirani, 1996). We validated our model on 34 different first impressions of faces using out-of-sample faces. Our results reveal which facial characteristics drive social biases and contribute to the field of downstream applications, such as training people to overcome stereotypes (Bohil, Kleider-Offutt, Killingsworth, & Meacham, 2021).

Methods

Facial Characteristics Coding

In this section, we explain the implementation of five different facial characteristics coding schemes, chosen for their quantitative viability, interpretability, and validity in face perception and anthropometric studies. These schemes offer diverse methods, facilitating a comprehensive evaluation of statistical parameters for facial characteristics (Merler, Ratha, Feris, & Smith, 2019). The five coding schemes capture multiple modalities of facial features, including emotional attributes, craniofacial areas, ratios, and distances, and skin color predicitons. These measurements can be reliably estimated from photos of frontal faces using landmark points of the face (Farkas, 1995) and computer vision algorithms (Serengil & Ozpinar, 2021; Jafari et al., 2016). We used 18 facial landmarks for creating three craniofacial feature coding (Table 1), and adopt the abbreviations from Farkas (1995) when referring to these facial landmark points instead of using full anatonmical terms.

In order to extract the 18 facial landmark points, we employed DLIB's facial key-point extraction tools (King, 2009), which provide a set of 68 key-points for each face. As shown in Figure 1, we mapped the 68 DLIB key-points to the 18 facial landmarks. These landmarks were then used for extracting craniofacial features. All craniofacial features and skin color measurements were z-standardized prior to analysis.

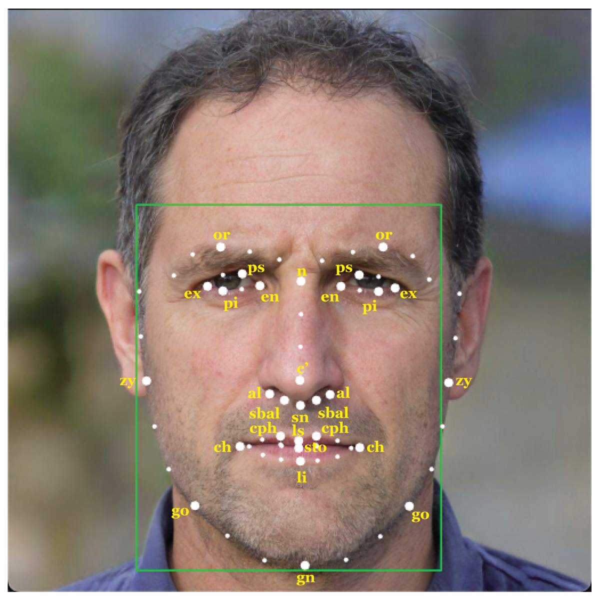

Figure 1: We utilized 68 key-points extracted using DLIB from each face (represented as small dots) to localize 18 facial landmarks (indicated by large, labeled dots). These 18 landmarks served as the basis for extracting craniofacial measures for our coding schemes. The face image is synthetically generated using the StyleGAN2 model and does not correspond to any known individual.

Craniofacial Areas: The first facial coding scheme, derived from a subsequent work by Farkas et al. (2005), includes measurements related to various regions of the face. The ten dimensions of these craniofacial area features are outlined in Table 2.

Anatomy	Abbr.	Anatomy	Abbr.
alare	al	subalare	$sbal$
orbitale	or	subnasale	sn
palpebrale superius	ps	crista philtre	cph
palpebrale inferius	pi	labiale superius	ls
endocanthion	en	stornion	sto
exocanthion	ex	labiale inferius	li
nasion	n	chelion	ch
pronasale	c'	gonion	go
zygion	zy	gnathion	gn

Table 1: Anatomical terms and their abbreviations obtained from Farkas (1995) for the set of facial landmarks utilized to compute the craniofacial measurements.

Craniofacial Ratios: The second facial coding scheme consists of measurements that represent various facial ratios. These features were utilized to assess age progression in individuals aged 0 to 18, as detailed in Ramanathan and Chellappa (2006). The eight dimensions of these craniofacial ratio features are summarized in Table 3.

Craniofacial area	Measure
Face height	$n - gn$
Face height	$sn - gn$
Face width	$zy - zy$
Face width	$go - go$
Orbits intercanthal width	$en - en$
Orbits fissure length (left and right)	$en - ex$
Orbits biocular width	$ex - ex$
Nose height	$n - sn$
Nose width	$al - al$
Labio-oral region	$ch - ch$

Table 2: Ten craniofacial areas representing various proportions of the faces (Farkas, 1995).

Craniofacial Distances: The third facial coding scheme for measuring craniofacial distances (Farkas, 1995) consists of eight metrics (Table 4). These metrics are used to determine the vertical distances between various facial elements, such as the top of the forehead, eyes, nose, mouth, and chin.

Skin Color: Researchers have been using different methods to characterize skin, including skin reflectance (Weyrich et al., 2006), skin type (Chardon et al., 1991), and skin color (Chardon et al., 1991; M. Wang, Xiao, Wuerger, Cheung, & Luo, 2015). Although some recent studies have adopted the Fitzpatrick Skin Type (FST) to categorize sun-reactive skin types (Buolamwini & Gebru, 2018), there is no universal measurement for skin color, even in the dermatology field (Ware, Dawson, Shinohara, & Taylor, 2020).

The Melanin Index (MI), used to assign FST, is found to be highly correlated with the Individual Typology Angle (ITA) (Wilkes, Wright, du Plessis, & Reeder, 2015). Given the high

Craniofacial ratio	Measure
Facial index	$(n - gn)/(zy - zy)$
Mandibular index	$(sto - gn)/(go - go)$
Intercanthal index	$(en - en)/(ex - ex)$
Orbital width index (left and right)	$(ex - en)/(en - en)$
Eye fissure index (left and right)	$(ps - pi)/(ex - en)$
Nasal index	$(al - al)/(n - sn)$
Vermilion height index	$(Is - sto)/(sto - li)$
Mouth-face width index	$(ch - ch)/(zy - zy)$

Table 3: Eight craniofacial ratios representing various ratios of the face (Farkas et al., 2005).

Craniofacial distance	Measure
Intercanthal face height	$n - sto$
Eye fissure height (left and right)	$ps - pi$
Orbit and brow height (left and right)	$or - pi$
Columella length	$sn - c'$
Upper lip height	$sn - sto$
Lower vermilion height	$sto - li$
Philtrum width	$cph - cph$
Lateral upper lip heights (left and right)	$sbal - ls'$

Table 4: Craniofacial distances corresponding to different vertical regions on the face, as defined in Farkas' anthropometric measurements (Farkas, 1995).

correlation between MI and ITA, and the simplicity of calculating ITA, this measure could be practical for assessing skin color (Chardon et al., 1991). This is particularly relevant since ITA can be calculated from an image, making it a more accessible method. As described in (Chardon et al., 1991), we implemented ITA in the CIE-Lab color space. To measure ITA directly from an image, we converted the *RGB* image to CIE-Lab space using standard image processing techniques.

As ITA is a point measurement, enabling every pixel corresponding to skin to have an ITA measurement, we extracted ITA values for pixels within a masked face region. The skin mask in the face (pixels corresponding to skin) was extracted using a deep neural network (Jafari et al., 2016).

Emotion Analysis: We utilized the DeepFace library (Serengil & Ozpinar, 2021), a deep learning framework designed for facial recognition and emotion analysis, to obtain emotion attributes. This library is widely used in various fields for emotion analysis (Zielonka, Bolkart, & Thies, 2022; Li, Yeh, & Huang, 2023; Park, Lee, Doosti, & Tan, 2023) and provides a platform for employing pre-trained deep learning models in image and video analysis. DeepFace generates probabilities for seven emotional attributes (angry, disgust, fear, happy, sad, surprise, and neutral) based on facial expressions. For our study, we selected the most dominant emotion identified for each face image.

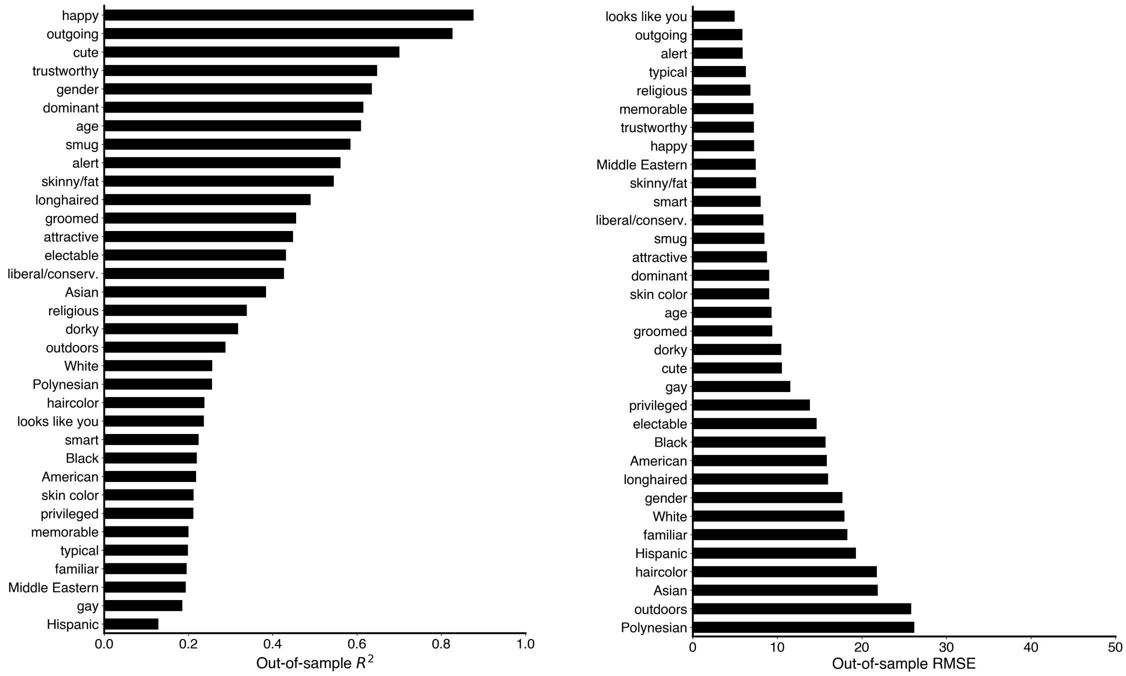

Figure 2: **Left:** The R^2 (Coefficient of Determination) indicating the performance of our model. **Right:** The Root Mean Square Error (RMSE) illustrating the prediction accuracy of our model.

Data

We applied the facial coding scheme and used it to model a large dataset of people's first impressions of faces (Peterson et al., 2022). The dataset contains over 1 million judgments of 34 trait inferences for 1,000 face images. Each face was rated by 30 unique participants for each trait using a scale from 1 to 100. In this dataset, the face images were generated using a synthetic photorealistic image generator (Karras et al., 2020). The DLIB toolkit (King, 2009) failed to detect two faces; therefore, we had a total of 998 face images in our dataset. For each trait, we used the ratings of 798 faces for training and the remainder (200 faces) for validation, an 80-20 split.

Results

We employed machine learning techniques to predict first impressions of faces based on facial characteristics and emotional attributes. Ideally, we would not only examine these face features in isolation but also consider their interactions (e.g., the combined effect of facial width-to-height ratio and intercanthal face height). However, focusing solely on two-way interactions would result in an additional 703 predictors in the model. Given that we had only a limited number of images for each trait to train the model, we decided not to include interaction terms in our analysis.

We examined the influence of all 38 facial characteristics by incorporating them into a single regression model. We used Lasso Regression (Tibshirani, 1996), which achieves two key objectives: (a) it minimizes overfitting through regularization by shrinking predictors, and (b) it conducts variable selection by reducing the coefficients of uninformative parameters to zero. The model includes one essential hyperparameter, α, which needs fine-tuning. This hyperparameter governs the degree of shrinkage. To determine the best-fitted α, we employed 10-fold cross-validation and a grid search approach.

Model Accuracy Evaluation

While our model offers valuable insights into the relationship between facial characteristics and first impressions, it also demonstrates predictive accuracy and generalizability. The model's performance, as indicated by the R^2 and RMSE values for different traits, suggests varying levels of predictive accuracy and explanatory power (Figure 2). It demonstrated average RMSE and R^2 values of 11.84 and 0.40, respectively. For traits like 'happy' and 'outgoing,' where we observed high R^2 values and relatively low RMSE, the model not only predicts these traits accurately but also explains a substantial portion of the variance in these impressions. This indicates that the extracted facial characteristics are strong predictors for these particular traits, effectively capturing most of the nuances that contribute to how these traits are perceived based on facial characteristics.

In contrast, traits such as 'gender' and 'looks like you' present a more complex scenario due to inconsistencies in RMSE and R^2 values. Specifically, lower R^2 values for these traits indicate that while the model is quite accurate in its predictions (as evidenced by low RMSE), it captures only a small portion of the variability in trait impressions. This discrepancy suggests that there may be key factors influencing

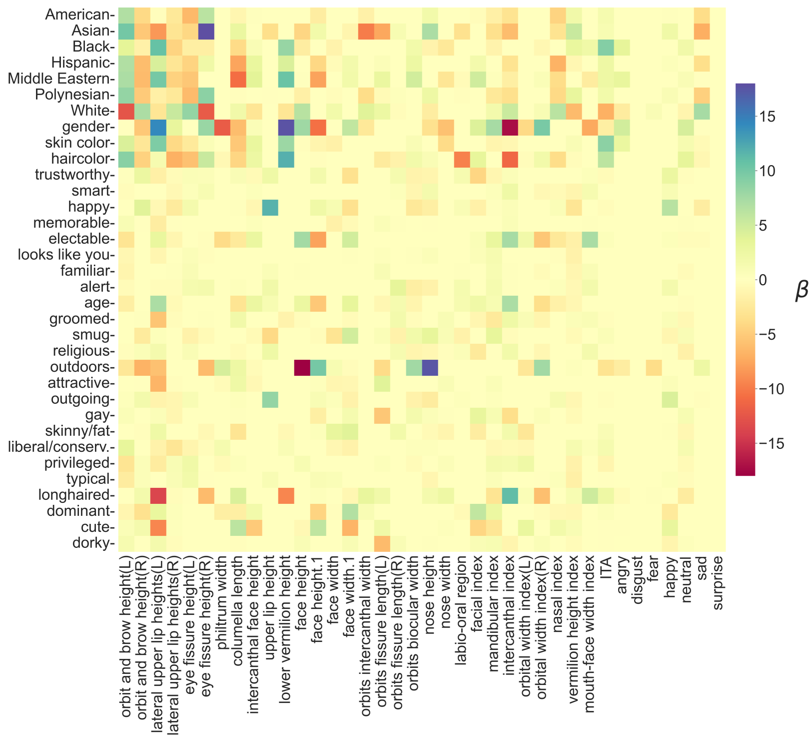

Figure 3: The relationships between facial characteristics and first impressions. Coefficients were derived from Lasso Regression models with cross-validation.

these traits that are not captured by the facial characteristics included in our model. For 'gender' trait in particular, the higher RMSE points to a greater variability in predictions, highlighting the complexity of gender perception, which may be influenced by a wider array of factors beyond simple facial characteristics. The 'looks like you' trait, on the other hand, poses a unique challenge due to its highly subjective nature. It is closely tied to an individual's personal perception of resemblance, which is difficult to quantify and may not be fully captured by the facial characteristics we have extracted. This reveals the limitations of our model in dealing with highly subjective and intricate traits, suggesting a need to incorporate more diverse and possibly non-visual factors to enhance predictive accuracy and explanatory power.

Analysis of Facial Characteristics

In our Lasso Regression model, the coefficient weights assigned to each facial characteristic offer insights into their relative importance in shaping first impressions. These coefficients indicate the strength and direction of the relationship between each facial feature and the perceived traits. A positive coefficient suggests that an increase in the feature is associated with an increase in the trait impression, while a negative coefficient indicates an inverse relationship.

For traits reflecting race inference from faces, such as American, Asian, Black, Hispanic, Middle Eastern, Polynesian, and White, orbit and brow height had significant effects on these judgments, suggesting their strong influence in creating impressions of race ($M_{|\bar{\beta}|} = 8.12$, $SD_{|\bar{\beta}|} = 2.36$). Previous studies have found that orbit and brow height vary among different races (Glass et al., 2014; Price, Gupta, Woodward, Stinnett, & Murchison, 2009). Our findings validate that human face perception is sensitive to these measurements when making inferences about a stranger's race.

Our study represents a pioneering effort in deriving skin color measurements directly from facial images and incorporating them into models for first impression analysis. Utilizing the Individual Typology Angle (ITA) as our metric for skin color, we discovered a significant relationship between this measurement and the inference of the 'skin color' trait. Although the 'skin color' trait reflects people's subjective perception of facial skin color, the strong relationship between our ITA measurement and 'skin color' impressions validates our methodology. The ITA metric demonstrates that people often rely on skin color as a key factor when forming impressions related to traits such as 'Black' or 'White'. Additionally, our research indicates a correlation between first impressions of hair color and our ITA measurements.

In the formation of first impressions, it appears that some traits are inferred from similar facial characteristics, while others are discerned from contrasting ones. For example, traits like 'happy' and 'outgoing', 'gay' and 'dorky', and 'American' and 'Polynesian' may be perceived based on similar facial features. In contrast, traits such as 'dominant' and 'trustworthy', 'cute' and 'dominant', 'White' and 'Polynesian', and 'age' and 'cute' seem to be based on contrasting facial characteristics. Drawing on the understanding that some first impressions are formulated based on similar and contrasting features, it is evident that facial cues significantly influence the trait inferences.

While our model offers valuable insights into which facial characteristics contribute to first impressions of faces, it falls short in identifying these characteristics for traits such as 'memorable', 'typical', 'familiar', and 'looks like you'. This limitation may stem from the inherently subjective nature of these traits. Unlike more objective characteristics, these traits are highly influenced by individual experiences and perceptions, making them difficult to quantify and model. Consequently, the model's inability to capture the subtle variations of these personal and subjective aspects could account for its failure to accurately identify the facial characteristics associated with these particular traits.

Discussion

In this study, we introduced a novel multimodal approach that combines various methods for analyzing facial characteristics. By leveraging recent advancements in computer vision, we extracted these characteristics and used them as inputs for our machine learning model. We specifically applied Lasso Regression to model first impressions and tested the model's generalizability and accuracy using faces not previously seen during training. Our model demonstrates a remarkable capability in predicting unseen faces and significantly contributes to the understanding of how facial characteristics and emotional attributes collectively influence first impressions. The impact of each facial characteristic and emotion attribute is further elucidated by analyzing the coefficient weights learned by our model.

While comparing the predictive capabilities and generalizability of our model across various first impressions, we observed that certain traits displayed high R^2 values yet also had high RMSE. This paradoxical outcome might be attributed to the complex nature of these traits and the variability in how they are perceived by different individuals. High R^2 values suggest that our model successfully explains a large proportion of the variance in these trait impressions. However, the corresponding high RMSE indicates significant absolute prediction errors. This discrepancy could stem from the inherent subjectivity and diversity in human perception of these traits, leading to greater variability in the data. Consequently, while the model is generally effective in predicting the overall trend or direction of these impressions, it faces challenges in accurately predicting the precise level or intensity of the trait for each individual. It highlights the difficulties in modeling subjective human perceptions, which are influenced by a range of subtle factors, including individual differences and cultural variations.

We found that certain facial characteristics significantly influence multiple aspects of impression formation for faces. For example, orbit and brow height were found to have a substantial impact on racial impression inferences. This finding aligns with previous cross-cultural research (Glass et al., 2014; Price et al., 2009), which has noted variations in orbit and brow height across different races, suggesting an involuntary sensitivity to these facial characteristics among observers. Despite prior studies highlighting the importance of emotional attributes in shaping first impressions of faces, we observed that these attributes did not centrally explain the formation of all first impressions. This may be attributed to the specific measurements derived from face images in our study and the diversity of emotional expressions present in the faces of our dataset.

Instead of relying on high-dimensional feature representations of faces, utilizing interpretable measurements of facial characteristics can elucidate variations in first impressions among different sub-groups (Gurkan & Suchow, 2022a). Future research could incorporate these derived facial characteristics as a latent construct within the framework of Cultural Consensus Theory (Romney et al., 1987). Moreover, this approach could illuminate how intersubjective norms influence perception of facial features (Over & Cook, 2018; Gurkan & Suchow, 2022b).

Despite the relatively good performance of some of our models, the results also suggest that our list of facial characteristics was not exhaustive for all first impressions of faces. Other characteristics (e.g., perceived weight, age) might demonstrate significant contributions to first impressions. To enhance model performance and examine the relative importance of facial characteristics, it is crucial to model faces considering all potentially meaningful characteristics. As we relied on recent advancements in computer vision to derive facial characteristics from face images, our study did not account for pose estimation. Although most images in the dataset were portraits, the derived measurements could still be affected by the orientation of the face images. Future work should focus on expanding the list of facial characteristics and incorporating pose estimation from face images.

Ethical implications Our research, while advancing our understanding of how facial characteristics and emotional attributes influence first impressions, raises significant ethical concerns. The models, which demonstrate predictive accuracy and generalizability, may inadvertently perpetuate and solidify societal prejudices, particularly if the underlying data mirrors existing biases. These biases in first impressions about the group under examination do not accurately represent the actual identities, beliefs, or abilities of the individuals depicted in the images. This misrepresentation is a critical ethical issue, as it can reinforce stereotypical thinking and oversimplify complex human traits. Furthermore, the use of advanced machine learning techniques to analyze human facial features and emotions raises the potential for misuse of this technology. The comprehensive analysis of how various facial characteristics influence first impressions, while valuable academically, risks oversimplifying human interactions and undervaluing personal attributes beyond physical appearance.

References

Adams Jr, R. B., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition & Emotion*, *26*(3), 431–441.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*(46), 17948–17953.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of afrocentric features in person perception: judging by features and categories. *Journal of Personality and Social psychology*, *83*(1), 5.

Bohil, C. J., Kleider-Offutt, H. M., Killingsworth, C., & Meacham, A. M. (2021). Training away face-type bias: Perception and decisions about emotional expression in stereotypically black faces. *Psychological Research*, *85*, 2727–2741.

Brown, T. D., Dane, F. C., & Durham, M. D. (1998). Perception of race and ethnicity. *Journal of Social Behavior and Personality*, *13*(2), 295.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77–91).

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*(10), 1194–1198.

Chardon, A., Cretois, I., & Hourseau, C. (1991). Skin colour typology and suntanning pathways. *International Journal of Cosmetic Science*, *13*(4), 191–208.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386.

Farkas, L. G. (1995). Anthropometry of the head and face. *Raven Press*.

Farkas, L. G., Katic, M. J., & Forrest, C. R. (2005). International anthropometric study of facial morphology in various ethnic groups/races. *Journal of Craniofacial Surgery*, *16*(4), 615–646.

Funk, F., & Todorov, A. (2013). Criminal stereotypes in the courtroom: Facial tattoos affect guilt and punishment differently. *Psychology, Public Policy, and Law*, *19*(4), 466.

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1526.

Glass, L. R. D., Lira, J., Enkhbold, E., Dimont, E., Scofield, S., Sherwood, P. R., & Winn, B. J. (2014). The lateral brow: position in relation to age, gender and ethnicity. *Ophthalmic Plastic and Reconstructive Surgery*, *30*(4), 295.

Gurkan, N., & Suchow, J. (2022a). Cultural alignment of machine-vision representations. In *SVRHM 2022 Workshop@ NeurIPS*.

Gurkan, N., & Suchow, J. W. (2022b). Learning and enforcing a cultural consensus in online communities. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44).

Jaeger, B., & Jones, A. L. (2022). Which facial features are central in impression formation? *Social Psychological and Personality Science*, *13*(2), 553–561.

Jafari, M. H., Karimi, N., Nasr-Esfahani, E., Samavi, S., Soroushmehr, S. M. R., Ward, K., & Najarian, K. (2016). Skin lesion segmentation in clinical images using deep learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 337–342).

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8110–8119).

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, *10*, 1755–1758.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? a meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390.

Li, Y.-T., Yeh, S.-L., & Huang, T.-R. (2023). The cross-race effect in automatic facial expression recognition violates measurement invariance. *Frontiers in Psychology*, *14*.

Liang, L., Jin, L., & Li, X. (2014). Facial skin beautification using adaptive region-aware masks. *IEEE Transactions on Cybernetics*, *44*(12), 2600–2612.

Liu, L., Xing, J., Liu, S., Xu, H., Zhou, X., & Yan, S. (2014). Wow! you are so beautiful today! *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *11*(1s), 1–22.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*(1), 5.

McCurrie, M., Beletti, F., Parzianello, L., Westendorp, A., Anthony, S., & Scheirer, W. J. (2017). Predicting first impressions with deep learning. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 518–525).

Merler, M., Ratha, N., Feris, R. S., & Smith, J. R. (2019). Diversity in faces. *arXiv preprint arXiv:1901.10436*.

Olivola, C. Y., Eubanks, D. L., & Lovelace, J. B. (2014). The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly*, *25*(5), 817–834.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive*

Sciences, *18*(11), 566–570.

Ormiston, M. E., Wong, E. M., & Haselhuhn, M. P. (2017). Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual Differences*, *105*, 40–42.

Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from? *Cognition*, *170*, 190–200.

Park, K., Lee, S., Doosti, S., & Tan, Y. (2023). Provision of helpful review videos: Effects of video characteristics on perceived helpfulness. *Production and Operations Management*.

Peterson, J. C., Uddenberg, S., Griffiths, T. L., Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences*, *119*(17), e2115228119.

Price, K. M., Gupta, P. K., Woodward, J. A., Stinnett, S. S., & Murchison, A. P. (2009). Eyebrow and eyelid dimensions: an anthropometric analysis of african americans and caucasians. *Plastic and reconstructive surgery*, *124*(2), 615–623.

Ramanathan, N., & Chellappa, R. (2006). Modeling age progression in young faces. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 1, pp. 387–394).

Romney, A. K., Batchelder, W. H., & Weller, S. C. (1987). Recent applications of cultural consensus theory. *American Behavioral Scientist*, *31*(2), 163–177.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260.

Serengil, S. I., & Ozpinar, A. (2021). Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)* (p. 1-4). Retrieved from `https://doi.org/10.1109/ICEET53442.2021.9659697` doi: 10.1109/ICEET53442.2021.9659697

Stepanova, E. V., & Strube, M. J. (2009). Making of a face: Role of facial physiognomy, skin tone, and color presentation mode in evaluations of racial typicality. *The Journal of Social Psychology*, *149*(1), 66.

Strom, M. A., Zebrowitz, L. A., Zhang, S., Bronstad, P. M., & Lee, H. K. (2012). Skin and bones: The contribution of skin tone and facial structure to racial prototypicality ratings. *PLoS One*, *7*(7), e41193.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460.

Vernon, R. J., Sutherland, C. A., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, *111*(32), E3353–E3361.

Wang, D., Nair, K., Kouchaki, M., Zajac, E. J., & Zhao, X. (2019). A case of evolutionary mismatch? why facial width-to-height ratio may not predict behavioral tendencies. *Psychological Science*, *30*(7), 1074–1081.

Wang, M., Xiao, K., Wuerger, S., Cheung, V., & Luo, M. R. (2015). Measuring human skin colour. In *Color and imaging conference* (Vol. 23, pp. 230–234).

Ware, O. R., Dawson, J. E., Shinohara, M. M., & Taylor, S. C. (2020). Racial limitations of fitzpatrick skin type. *Cutis*, *105*(2), 77–80.

Weyrich, T., Matusik, W., Pfister, H., Bickel, B., Donner, C., Tu, C., … others (2006). Analysis of human faces using a measurement-based skin reflectance model. *ACM Transactions on Graphics (ToG)*, *25*(3), 1013–1024.

Wilkes, M., Wright, C. Y., du Plessis, J. L., & Reeder, A. (2015). Fitzpatrick skin type, individual typology angle, and melanin index in an african population: steps toward universally applicable skin photosensitivity assessments. *JAMA Dermatology*, *151*(8), 902–903.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*(3), 237–242.

Zielonka, W., Bolkart, T., & Thies, J. (2022). Towards metrical reconstruction of human faces. In *European Conference on Computer Vision* (pp. 250–269).