A Computational Analysis of "Dog-like" and "Cat-like" Impressions in Human Faces Using PCA and SVM

Inase Kondo, Kouichiro Ueno, Yuzhou Jin, and Kiryu Miura*

Abstract

People often describe human faces with subjective impressions of animal similarity, such as "dog-like" or "cat-like." While prevalent in popular culture, the computational basis for this perception remains unexplored. This study aims to computationally model this subjective impression. To achieve this, we developed a classification model using Principal Component Analysis (PCA) for feature extraction and a Support Vector Machine (SVM) for classification. A dataset of 8,000 dog and cat facial images was used to build a feature space defining "canine" and "feline" characteristics. The trained model was then tested on real human faces from the Labeled Faces in the Wild (LFW) dataset. For validation, these human faces were also annotated as "dog-like" or "cat-like" by two human annotators and the Gemini large language model. Our primary results indicate that human faces, when projected onto the PCA feature space, distribute between the distinct clusters formed by dog and cat faces. A comparative analysis reveals a moderate agreement between our model's classifications and human/LLM annotations, but also highlights a significant "dog-like" classification bias in both the model and the LLM. Furthermore, by visualizing the model's decision-making process using LIME, we qualitatively compared its feature importance against the reasoning provided by human and LLM annotators. This revealed that the model's "dog-like" bias is strongly linked to facial expressions, while its "cat-like" classifications are consistently driven by features in the nasal region. This study concludes that the subjective impression of animal-like features can be effectively modeled, providing a computational framework for analyzing facial morphology beyond simple identification.

Keywords: *PCA, SVM, human faces, animal similarity, eigenfaces*

1. Introduction

It is a common phenomenon for humans to perceive and describe the faces of others using animal archetypes, such as remarking that someone has a "cat-like" or "dog-like" face. This subjective impression, though anecdotal, is pervasive in social communication and has potential applications in entertainment, character design, and personalized avatar generation. However, the underlying facial characteristics that trigger these specific impressions are not well understood from a computational perspective.

This ambiguity raises a fundamental question: is it possible to computationally model and classify these subjective impressions based on objective facial features? Existing facial analysis research has largely focused on identity recognition, demographic estimation, or emotion detection, leaving the domain of subjective, cross-species impression analysis relatively untouched.

To address this gap, this paper proposes a computational model to classify human faces as either "dog-like" or "cat-like." Our approach utilizes Principal Component Analysis (PCA) to extract a low-dimensional set of core facial features from a collection of dog and cat images. These features are then used to train a Support Vector Machine (SVM) classifier. The primary objective is to build a model that can determine which animal category a given human face more closely resembles and to investigate the features that drive this classification.

This leads to our primary research question:

RQ: How does a classifier trained on dog and cat facial images classify real human faces, and what are the key facial features that inform its decisions?

2. Related Work

Our research is situated at the intersection of classical face recognition, image classification, and the computational analysis of subjective impressions.

The field of automated face analysis has a rich history, with early foundational work focusing on holistic, appearance-based methods. One of the most influential techniques is "Eigenfaces," which uses Principal Component Analysis (PCA) to represent faces in a low-dimensional feature space [1]. While primarily developed for recognition tasks, the ability of PCA to capture the principal modes of variation in a set of faces is highly relevant to our goal of extracting defining features.

Support Vector Machines (SVMs) have been widely and successfully applied to various image classification tasks due to their effectiveness in high-dimensional spaces and their robustness, particularly when training data is limited [2]. SVMs have been used for everything from object detection to medical image analysis, demonstrating their versatility as a powerful classification tool, making them a suitable choice for our binary classification problem.

While computational analysis of facial attractiveness and emotion is an active area of research [3], few studies have ventured into modeling more abstract or metaphorical impressions, such as animal resemblances. Most existing work is centered on intra-species analysis (i.e., human-to-human).

*The University of Osaka, Osaka 565-0871, Japan. (e-mail: i-kondou@ics.osaka-u.ac.jp; k-ueno@ist.osaka-u.ac.jp; jin.yuzhou@ist.osaka-u.ac.jp; kiryu.miura@ist.osaka-u.ac.jp)

The novelty of our study lies in its explicit focus on quantifying a subjective, cross-species facial impression. Whereas most face analysis systems aim to identify who a person is, our work aims to classify what a person's face looks like in an abstract, analogical sense. By applying established techniques like PCA and SVM to this unconventional problem and comparing the model's rationale to that of humans and LLMs, we aim to provide a new perspective on facial feature analysis.

3. Methodology

Our methodology consists of four main stages: (1) data preparation, (2) image preprocessing, (3) feature extraction using PCA, and (4) classification using an SVM.

3.1. Datasets

Source Data: The initial data was sourced from the "cats vs dogs" dataset provided by Microsoft [4], available via the Hugging Face Datasets repository. We utilized 4,000 images for each class (4,000 cats, 4,000 dogs), creating a base dataset of 8,000 images.

Training and Test Split: The combined animal dataset was partitioned into a training set (80%) and a testing set (20%) using a stratified split to maintain the class balance. This resulted in 6,400 images for training and 1,600 images for testing. A fixed `random_state` was used to ensure the reproducibility of this split.

External Test Data (Human Faces): The human face data was sourced from the Labeled Faces in the Wild (LFW) dataset from the University of Massachusetts Amherst [5]. From the full dataset of 13,233 images, we randomly selected a subset of 50 images for analysis. For validation against our model's output, these 50 images were independently annotated as either "dog-like" or "cat-like" by two human annotators and the Gemini large language model (gemini-2.5-flash). A final label was assigned based on a majority vote among the three annotators.

3.2. Preprocessing

All images in our datasets underwent a uniform preprocessing pipeline to standardize them for model training and evaluation.

1. **Image Resizing:** All dog and cat images were resized to 128x128 pixels.
2. **Background Removal:** The `rembg` tool was used to remove the background from each image. This step helps to eliminate potential noise and allows the model to focus specifically on the features of the subject.
3. **Vectorization:** The human face images underwent a similar preprocessing pipeline, and all final images were converted to grayscale and flattened into 1D vectors for analysis.

3.3. Feature Extraction: PCA

Principal Component Analysis (PCA) was employed to reduce the dimensionality of the image data. Each 128x128 pixel image, initially represented as a 16,384-dimensional vector, was transformed into a lower-dimensional feature vector. The PCA model was configured to retain 85% of the cumulative variance in the data. Crucially, the PCA model was fitted exclusively on the 6,400 images of the training set. This prevents data leakage from the test set. The same fitted model was then used to transform all other datasets.

3.4. Classification: SVM

A Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel was used for the final classification task. An SVM seeks to find an optimal hyperplane that maximizes the margin between the two classes (dogs and cats) in the PCA feature space. The RBF kernel was chosen for its effectiveness in handling non-linear relationships. The SVM classifier was trained on the PCA-transformed feature vectors of the training data, with labels assigned as 1 for "dog" and 0 for "cat."

4. Experiments

This section presents the results of our experiments, beginning with the baseline performance of our classifier, followed by a comparative analysis of its classifications of human faces against those of human and LLM annotators.

4.1. Model Performance on Animal Classification

First, to validate the effectiveness of our PCA-SVM model, we evaluated its performance on the held-out test set of 1,600 animal images. The model achieved an accuracy of 75.125% in distinguishing between cats and dogs. This strong baseline performance confirms that the model successfully learned discriminative features from the training data, providing a solid foundation for the subsequent analysis of human faces.

4.2. Classification of Human Faces: A Comparative Analysis

We applied our trained model to the 50-image human test set and compared its classifications with the labels provided by two human annotators and the Gemini model(gemini-2.5-flash).

4.2.1. Quantitative Comparison of Classifications

To provide a quantitative overview, we first analyzed the label distribution for each of the four annotators. As shown in Table 1, there were significant differences in classification tendencies. The human annotators showed relative balance, with Annotator A leaning slightly towards "cat" (56%) and Annotator B being perfectly balanced (50%). In stark contrast, both the LLM and our ML model exhibited a strong bias towards "dog-like" classifications, at 86% and 78% respectively.

Table 1: Label Distribution by Annotator

Annotator	Dog	Cat	Dog-like Ratio	Cat-like Ratio
Annotator A	22	28	44%	56%
Annotator B	25	25	50%	50%
LLM (Gemini)	43	7	86%	14%
ML Model	39	11	78%	22%

Next, we examined the pairwise agreement rates between the annotators (Table 2). The agreement between the two human annotators (A vs. B) was 62%. Notably, the agreement between the LLM and our ML model was the highest of any pair, at 68%. Our ML model's agreement was higher with Annotator B (60%) than with Annotator A (46%). The four parties reached a unanimous decision on only 14 of the 50 images (28%), highlighting the subjective nature of the task.

Table 2: Pairwise Agreement Matrix (number of matching labels out of 50; percentage in parentheses)

	Annotator B	LLM	ML Model
Annotator A	31 (62%)	29 (58%)	23 (46%)
Annotator B		26 (52%)	30 (60%)
LLM			34 (68%)

4.2.2. Qualitative Analysis of Classification Rationale

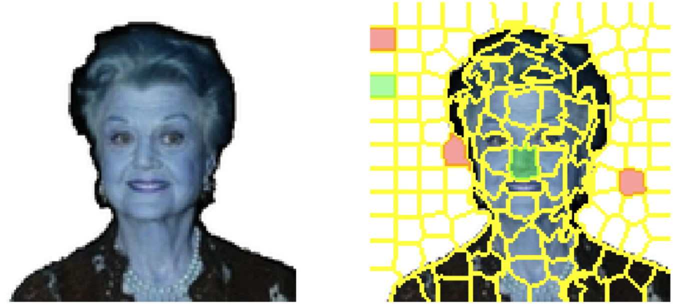
To understand the reasons behind these quantitative differences, we analyzed the justification provided by the human and LLM annotators and compared them to our model's visual focus, as identified by LIME (Local Interpretable Model-agnostic Explanations). This analysis revealed two distinct, and sometimes conflicting, patterns in our model's decision-making process.

First, the "dog-like" classification bias appears strongly linked to facial expressions. As shown in Table 3, the LLM frequently cited "Expression" (66%) as a reason for its labels, a category that Annotator A never used. Our ML model's high agreement with the LLM (68%) suggests it learned a similar association.

Second, a remarkably consistent pattern emerged for "cat-like" classifications: the model's reliance on the nasal region. Our LIME analysis revealed that in all 17 instances where the model highlighted the nose as a significant feature, it was to provide a positive contribution to the "cat" class. This suggests the model learned a simple, powerful heuristic: a certain shape or texture in the nose is a decisive indicator of feline features.

This analysis reveals a key insight: the LLM and, by extension, our ML model, appear to strongly associate positive expressions ("friendly smile," "open expression") with "dog-like" features. This explains their strong bias. Humans, particularly Annotator A, relied more on static structural features like the nose and jawline.

Case Study: Conflict Between Local Feature Importance and Final Classification.



(a) Original Image (b) LIME Visualization

Figure 1: Case Study: Conflict Between Local Feature Importance and Final Classification. (a) Original Image (b) LIME Visualization

Case Study: Conflict Between Local Feature Importance and Final Classification. These two learned patterns can lead to complex and sometimes counter-intuitive results. Figure 1 presents a compelling case where both human annotators classified an image as "cat-like." In contrast, both the LLM and our SVM model classified it as "dog-like."

The LIME visualization for our model (Figure 1b) reveals the internal conflict in its decision. The nasal region is highlighted in green, indicating a strong positive contribution to the "cat" class, consistent with the heuristic we identified. However, the final prediction was "dog." This implies that other features—likely related to the subject's expression, which aligns with the model's "dog-like" bias—provided a stronger, opposing contribution that ultimately tipped the classification. This case study vividly illustrates that the model's final decision is a weighted sum of multiple, sometimes contradictory, local features, and it does not always align with its own single most prominent visual cue.

5. Discussion

The results from our experiments offer several points for discussion regarding the nature of subjective facial perception and the behavior of our computational model.

The quantitative analysis highlights a significant gap between human perception and the judgments of both the LLM and our PCA-SVM model. The strong "dog-like" bias in the automated systems, linked to the interpretation of facial expressions, suggests that these models may have learned a simplistic correla-

Table 3: Frequency of Feature Mention in Reasoning

Feature Category	LLM	Annotator A	Annotator B
Eyes	84%	38%	48%
Nose	68%	44%	4%
Mouth	22%	14%	12%
Contour	78%	40%	44%
Cheeks	12%	8%	6%
Expression	66%	0%	22%

tion where positive affect (e.g., a smile) is associated with "dog-like" traits of friendliness and approachability. Human annotators, conversely, demonstrated a more nuanced approach, capable of separating structural features from transient expressions.

The qualitative analysis, particularly the LIME visualizations, provides a deeper understanding of our model's internal logic. The discovery that the nasal region is a consistent predictor for the "cat" class is a fascinating artifact of the training process. It suggests the model identified a highly discriminative, local feature in the training data. However, the case study in Figure 1 demonstrates that this is not an absolute rule but a weighted factor. The model's final output is a result of competition between various learned features, which explains why a strong "cat-like" signal from the nose can be overridden by other "dog-like" signals from elsewhere in the image.

5.1. Limitations

This study has several limitations that should be acknowledged.

Dataset Bias: Our training data was limited in the diversity of dog and cat breeds, which could affect the generalizability of the learned features.

Methodological Simplicity: We intentionally used a classical approach (PCA+SVM). Modern deep learning models, such as Convolutional Neural Networks (CNNs), could potentially extract more complex and robust features.

Ambiguity of Definition: The very concepts of "dog-face" and "cat-face" are subjective and culturally dependent. Our model provides one possible computational definition, but it is not absolute.

Preprocessing Artifacts: The LIME analysis occasionally highlighted regions corresponding to the background, despite the use of a background removal tool. This indicates that the preprocessing was not perfect and that residual background pixels may have introduced noise into the model's decision-making process.

5.2. Future Work

Our findings suggest several promising directions for future research:

Model Refinement and De-biasing: A key priority is to address the model's "dog-like" bias. This could involve augmenting the training data with a wider variety of animal facial expressions or employing algorithmic de-biasing techniques. Furthermore, refining the preprocessing steps to completely eliminate background artifacts would improve model accuracy.

Exploring Advanced Architectures: A comparative study using a Convolutional Neural Network (CNN) would be a valuable next step. A CNN could learn hierarchical features, potentially capturing more nuanced structural information than our current PCA-based approach and reducing the reliance on simplistic heuristics like the "nose rule."

Enhancing Interpretability: While LIME provided local explanations, future work could explore methods that offer global interpretations of the model. This would help to understand the overall feature space and how different characteristics (e.g., eye shape vs. expression) are weighted against each other.

Large-Scale Human-in-the-Loop Evaluation: To validate our findings on subjectivity, a larger-scale study involving more human participants is crucial. This would allow for a more robust analysis of inter-annotator agreement and a more accurate "ground truth" against which computational models can be benchmarked.

6. Conclusion

In this paper, we successfully developed and evaluated a computational model to classify the subjective impression of "dog-like" and "cat-like" features in human faces. By training a PCA-SVM classifier on a dataset of animal faces, we demonstrated that this task is computationally feasible. Our comparative analysis against human and LLM annotators revealed a significant "dog-like" bias in automated systems, which we attribute to a learned association between positive facial expressions and canine traits. Through the use of LIME, we uncovered specific, and at times conflicting, heuristics learned by our model, such as the consistent association of the nasal region with "cat-like" features. The case study of a conflicting classification highlighted the complexity of the model's decision-making, where the final judgment results from a weighted competition between multiple learned features. This research serves as a proof-of-concept that computational methods can be used to investigate abstract aspects of facial perception, revealing not only the potential of such models but also their inherent biases and the fascinating gap between computational and human perception.

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 01 1991.

[2] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, p. 273–297, Sep. 1995.

[3] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, "Data-driven enhancement of facial attractiveness," *ACM Trans. Graph.*, vol. 27, no. 3, p. 1–9, Aug. 2008.

[4] J. Elson, J. Douceur, J. Howell, and J. Saul, "Asirra: A captcha that exploits interest-aligned manual image categorization," in *Proceedings of 14th ACM Conference on Computer and Communications Security (CCS)*, October 2007.

[5] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.