Implementation of Feed-forward Model In Learning and Behavioral Mode

MohammadHosein Ostadi[1], Setare Roshan[2], and Reza ebrahimpoor[3]

[1]**mh.ostadi.v@gmail.com**
[2]**setareRoshan@gmail.com**
[3]**RezaEbrahimpoor@gmail.com**

ABSTRACT

This study presents a dual investigation into the efficacy of the Hierarchical Model and X (HMAX) for visual categorization and parallels it with human cognitive performance on similar tasks. The HMAX model, structured on the hierarchical processing observed in the primate visual cortex, is examined for its classification capabilities, enhanced through the application of Support Vector Machine (SVM) classifiers. Complementarily, human visual recognition is evaluated through a behavioral task employing the Psychtoolbox in MATLAB, assessing accuracy and reaction times in the categorization of animal and non-animal images. The analysis reveals a notable proficiency in human subjects for categorizing images with pronounced features, while the HMAX model demonstrates robustness across varied image types, with discrepancies highlighting the differences in artificial and natural processing systems. The study further discusses the potential improvement in model accuracy through k-fold cross-validation, which was not utilized, and analyzes the performance impact of different SVM kernels, concluding the linear kernel's comparative adequacy for this specific dataset.

Keywords: Hmax, behavioral, feedforward

INTRODUCTION

In the burgeoning field of computational neuroscience, the quest to understand and emulate the human brain's remarkable ability to process and categorize visual information has led to the development of various computational models. Among these, the Hierarchical Model and X (HMAX) model stands as a pivotal contribution, offering significant insights into the intricate processes underpinning rapid object recognition. This study delves into an in-depth exploration of the HMAX model, specifically investigating its ability to categorize and classify visual stimuli, drawing parallels with human cognitive capabilities in similar tasks.

The HMAX model, inspired by the hierarchical processing in the primate visual cortex, has been instrumental in bridging the gap between biological plausibility and computational efficiency. It mimics the layered processing observed in the visual cortex, from primary visual regions to higher-order inferotemporal cortex, thus offering a robust framework for understanding visual categorization and recognition. Our study aims to not only replicate and analyze this model but also to extend its application through the integration of modern machine learning techniques, particularly the Support Vector Machine (SVM) classifier, to enhance its classification performance.

Parallel to the computational exploration, this study also incorporates a behavioral component, leveraging the Psychtoolbox in MATLAB to gather and analyze visual recognition data from human subjects. This facet of the research focuses on examining how humans categorize and respond to visual stimuli, specifically animal and non-animal images, across varied contexts and perspectives. The inclusion of a Linear Classification toolbox further refines our analysis, allowing for a nuanced understanding of human categorization accuracy and response times.

By juxtaposing the computational efficiency and accuracy of the HMAX model with human performance in similar tasks, this study aims to contribute to the broader understanding of visual processing both in artificial and natural systems. Our research is not just a comparative analysis but also a step towards a more integrated approach in cognitive science, where computational models and human behavioral studies inform and enhance each other.

In the following sections, we will delve into the details of the HMAX model, the methodologies employed in gathering and analyzing both computational and behavioral data, and present a comprehensive analysis of the results. The conclusion will synthesize these findings, offering insights into the implications of this study for future research in computational neuroscience and cognitive psychology.

MATERIALS AND METHODS

Implementation of the HMAX Model

Model Overview: The HMAX model, inspired by the hierarchical structure of the primate visual cortex, is a computational framework designed for rapid categorization and classification of visual stimuli. It mimics the multi-stage processing of visual information, from basic feature extraction to complex pattern recognition.

Model Architecture: The HMAX model comprises four primary layers, each corresponding to different stages of visual processing:

- **S1 (Simple Cells Layer):** This layer models the primary visual cortex (V1) and is responsible for the initial processing of visual stimuli. It involves filtering the input images with Gabor filters of various orientations and scales, capturing basic features like edges and bars. To understand better, note figures 1,2,3.

$$G(x,y) = exp(-\frac{x_0^2+\gamma^2 y_0^2}{2\sigma^2})cos(\frac{2\pi}{\lambda}x_0)$$

$$x_0 = xcos(\theta)+ysin(\theta), \quad y_0 = xsin(\theta)+ycos(\theta)$$

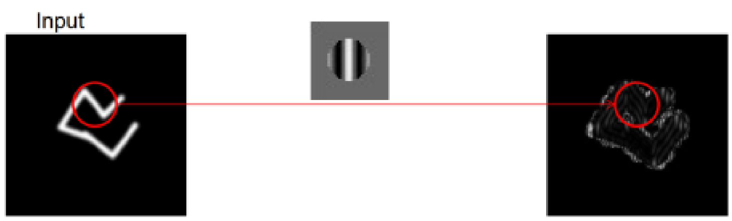


Figure 1. S1 output sample

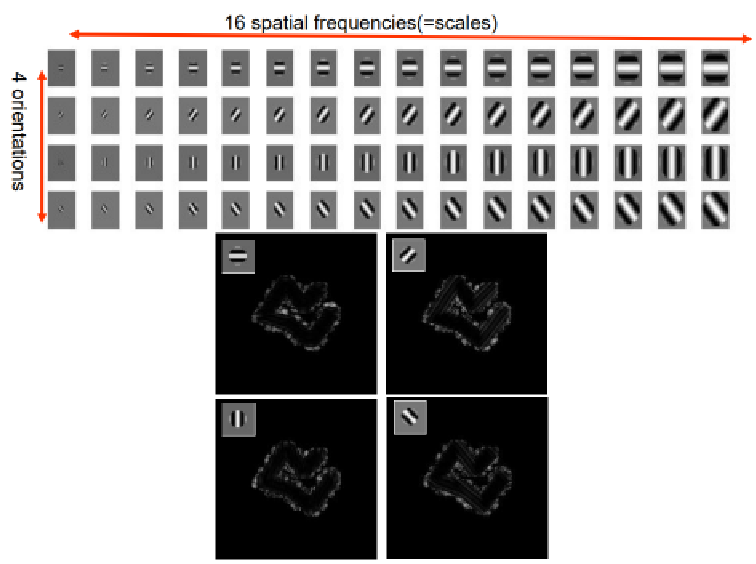


Figure 2. 16 Spatial Frequencies and 4 Orientations

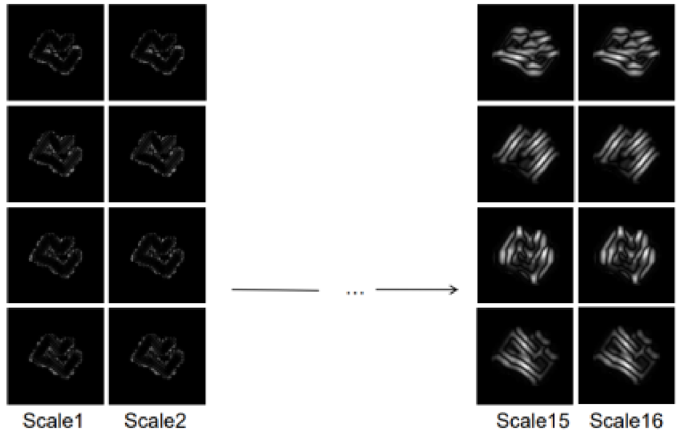


Figure 3. Scale 1 to 16 output of S1

- **C1 (Complex Cells Layer):** Following the S1 layer, the C1 layer simulates the behavior of complex cells in V1 and V2 regions. It performs a max-pooling operation over the outputs of S1 cells. This process introduces invariance to small shifts and distortions in the visual stimuli. To understand better, note figures 4,5.

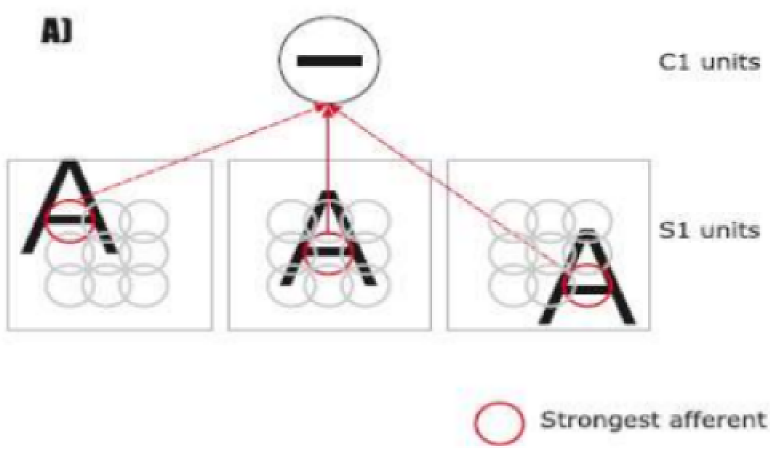


Figure 4. centered at different location thus providing some translation invariance.

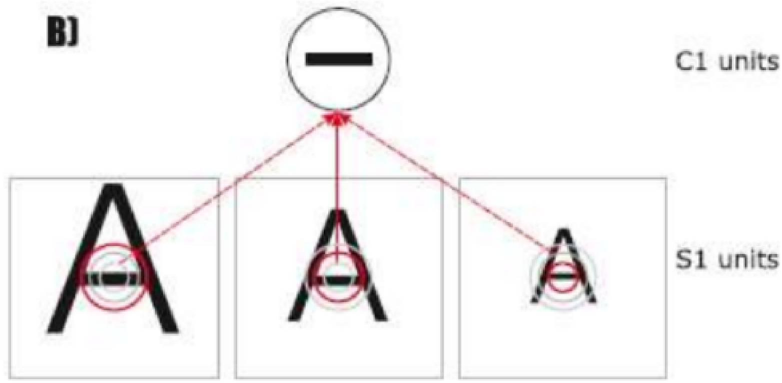


Figure 5. at different scales providing some scale invariance to the complex cell

- **S2 (Intermediate Complex Cells Layer):** The S2 layer further abstracts the visual information by combining the features detected in the C1 layer. This layer uses a set of learned or predetermined templates to detect more complex patterns, such as textures and object parts.

  $$FM_{S2}(x,y)^s_m = exp(-\beta ||X^s - P^m||)$$

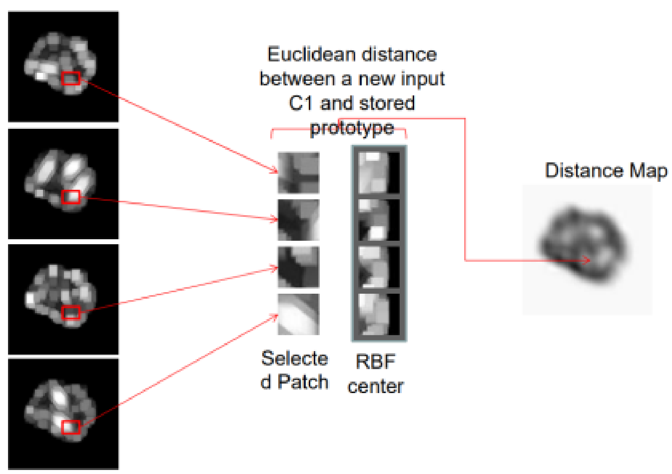


Figure 6. Response of S2 cell's after train

- **C2 (Higher-level Complex Cells Layer):** The final layer, C2, aggregates the information from S2 and provides a global representation of the input image. The C2 layer's output serves as the feature vector for categorization tasks and is highly robust to variations in position and scale.

Model Training and Classification: For training the HMAX model, we employ supervised learning techniques. The model is exposed to a diverse set of training images, allowing it to learn and adapt its S2 templates for effective pattern recognition. Post-training, the C2 layer's output is used as input for a classifier, such as Support Vector Machine (SVM), for the categorization task.

Behavioral Task Implementation

Experimental Setup: The behavioral task is designed to assess human object recognition capabilities under rapid presentation conditions. We used the Psychtoolbox in MATLAB for precise stimulus presentation and response recording.(figure 7)

Subjects: Three male subjects, all aged 22 years, participated in the study. The choice of a homogeneous group in terms of age and gender aimed to control for demographic variables that might influence visual processing.

Task Description: The task involved presenting subjects with a series of images (600 animal and 600 non-animal images, subdivided into categories of head, close body, medium body, and far body). Each subject was exposed to these images in blocks of 120, with the task of rapidly categorizing them as either 'animal' or 'non-animal'.

Data Collection: Accuracy and reaction times were recorded for each subject. The accuracy data encompassed the subject's ability to correctly categorize images, while reaction times provided insights into the speed of cognitive processing.

Analysis Tools: We utilized the Linear Classification toolbox in MATLAB for analyzing the behavioral data. This involved computing the average accuracies and reaction times across different image categories and blocks.

This comprehensive approach, combining computational modeling with human behavioral analysis, allows for a nuanced understanding of visual categorization processes, bridging the gap between artificial and natural intelligence systems.

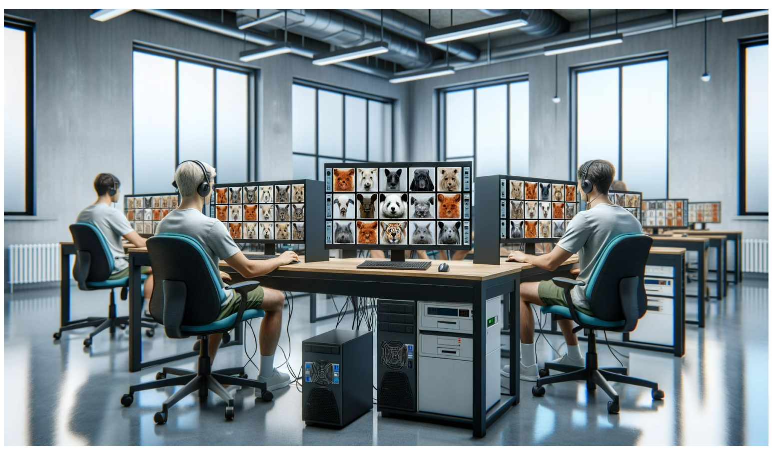


Figure 7. Task Setup

Dataset Description:

The study utilizes a meticulously curated dataset, sourced from the comprehensive collection at MIT's Center for Biological and Computational Learning (CBCL). This dataset is integral to both the computational modeling and behavioral analysis components of our research. It comprises 1,200 images, evenly split into two primary categories: 600 animal and 600 non-animal images. These images are further categorized into four distinct subgroups - head, close body, medium body, and far body, each containing 150 images. This diverse and well-organized collection of visual stimuli enables a thorough examination of object recognition and categorization processes. The dataset's variety in terms of image composition and complexity provides a robust foundation for testing the HMAX model's efficiency in visual categorization and serves as a crucial tool in assessing human performance in the corresponding behavioral tasks. see figures 8 and 9.

(a) Head

(b) Medium Body

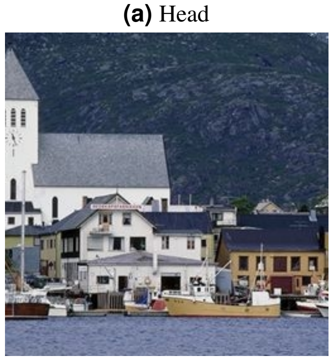
(c) Far Body

(d) Body

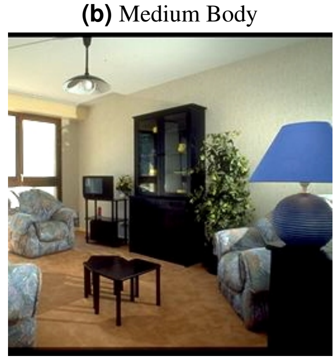
Figure 8. Target images.

(a) Head

(b) Medium Body

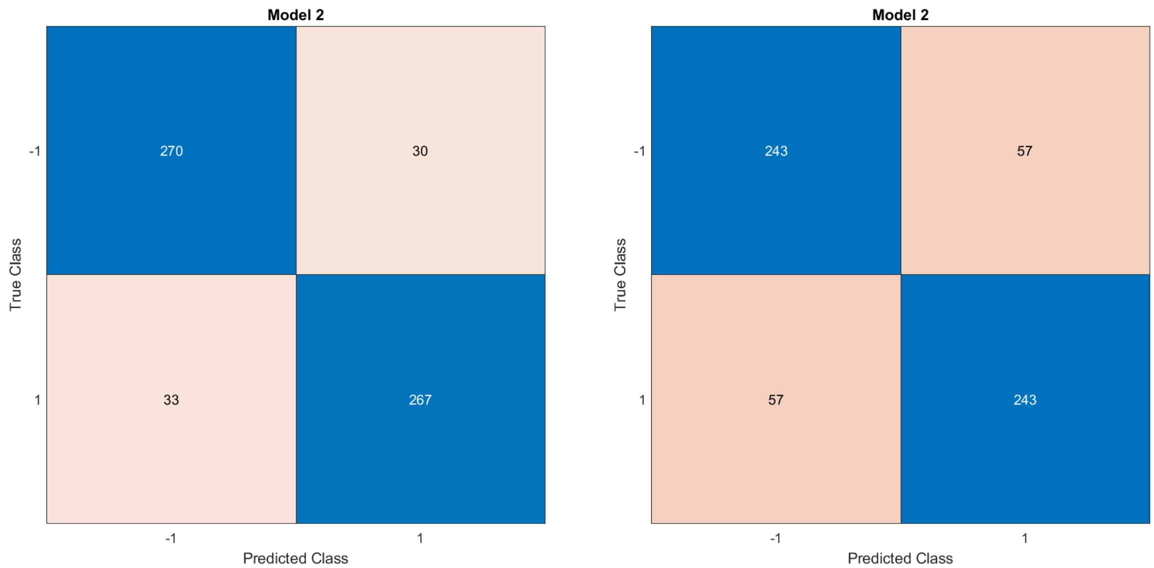


(c) Far Body



(d) Body

Figure 9. Distractor images



(a) Train

(b) Test

Figure 10. Confusion matrices

RESULTS

Model Results

The performance of the hierarchical model was quantitatively assessed using confusion matrix and Receiver Operating Characteristic (ROC) curve analyses on both training and testing datasets.

Confusion Matrix Analysis

The confusion matrix for the training dataset shows that the model achieved a high accuracy rate of 89%. It correctly predicted 270 instances as class '-1' and 267 instances as class '1', with 30 and 33 instances misclassified, respectively. For the testing dataset, the model's accuracy was slightly lower at 81%, correctly classifying 243 instances for both classes '-1' and '1', while misclassifying 57 instances for each class. See figures 10a and 10b.

ROC Curve Analysis

The model's discriminative capability is further elucidated by the ROC curves. The training dataset's ROC curve yielded an AUC of 0.95, suggesting excellent model performance, whereas the testing dataset's ROC curve had an AUC of 0.89. These AUC values indicate a strong predictive ability within the model's classification function.

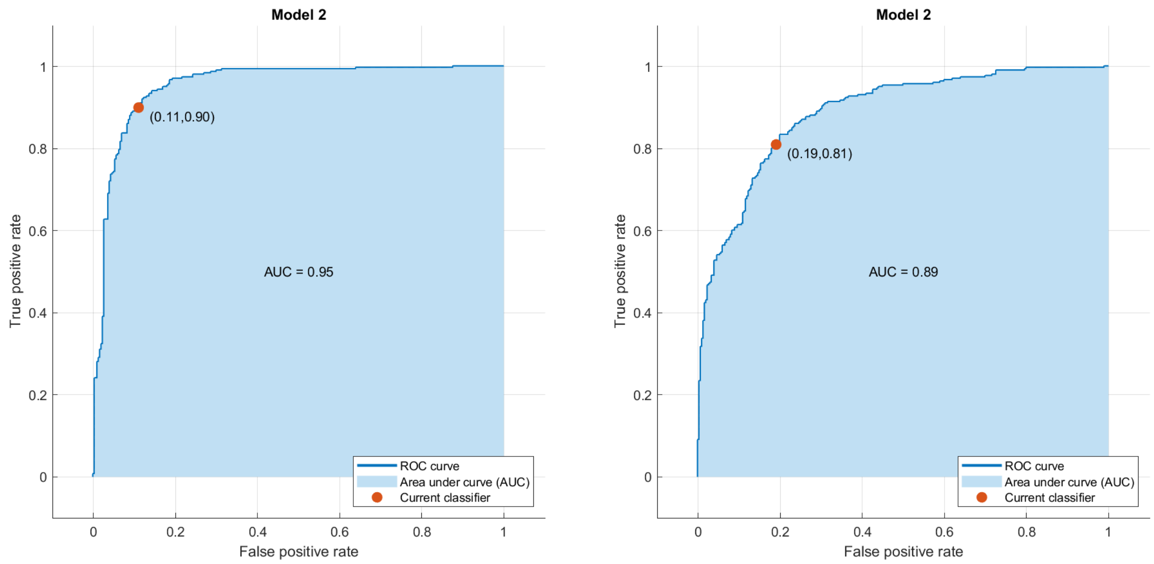
The ROC curve for the testing dataset places the current classifier's operating point at (0.19, 0.81), reflecting a true positive rate of 81% in contrast to a false positive rate of 19%. In comparison, the training dataset's operating point stands at (0.11, 0.90), signaling a true positive rate of 90% for a false positive rate of 11%. See figures 11a and 11b.

These metrics demonstrate the model's effective learning from the training data and its subsequent proficiency in categorizing new, unseen visual stimuli. The accuracy on the test data, while lower than the training, still indicates a robust generalization capability. These results underscore the potential of hierarchical models in replicating aspects of visual categorization processes.

Behavioural Results

In the rapid categorization task, human participants were tasked with categorizing images across four distinct subcategories: head, close body, medium body, and far body. The task was designed to mimic the conditions under which computational models such as the HMAX operate, allowing for a direct comparison of performance between human and machine vision.

- **Subject1** demonstrated remarkable acuity, achieving the highest average accuracy across all subcategories: 96% for head, 97.33% for close body, 87.67% for medium body, and 94.33% for far body images. Notably, Subject1's reaction time averaged at 0.362 seconds, indicating not only precision but also expediency in categorization.
- **Subject2** also displayed high levels of accuracy, with average scores only marginally trailing Subject1: 94.67% for head, 95.67% for close body, 86.33% for medium body, and 94.33% for far body images. The average reaction time for Subject2 was recorded at 0.440 seconds, slightly slower than Subject1 but still within the rapid response range.



(a) Train **(b)** Test

Figure 11. ROC Curves

- **Subject3** presented consistent accuracy, particularly for the head and close body categories, with scores of 94% and 94.67% respectively. However, a notable drop was observed in the medium body category at 79.33%, suggesting a potential area for targeted improvement. Subject3's accuracy for far body images was strong at 93.33%, and the reaction time averaged at 0.380 seconds, reflecting swift processing similar to Subject1.

Overall, the human participants showed a tendency to excel in the head and close body categorization tasks, likely due to the pronounced features and cues present in these subcategories. The medium body images posed a greater challenge, as evidenced by the lower accuracy rates, particularly for Subject3. These results highlight the proficiency of human visual processing and provide a benchmark against which the performance of computational models can be compared.

Compare Results

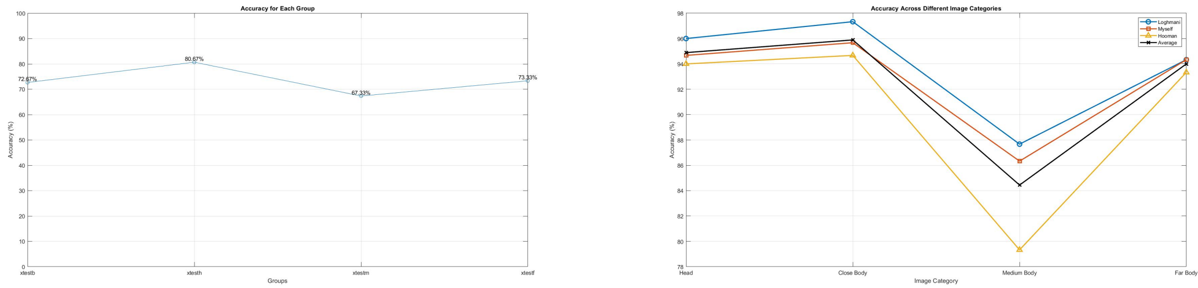
A comparative analysis between the computational model and human subjects reveals insightful distinctions in visual categorization capabilities. The first image(figure 12a), depicting the accuracy for each group, illustrates a uniform trend where all groups exhibit a decline in performance in the medium body category. However, this trend is more pronounced in human subjects, suggesting that this category may inherently possess more complex features that are challenging for both human and computational processing.

The second image(figure 12b), displaying the accuracy across different image categories for each human subject alongside the average, indicates that while human performance varies across individuals, it remains consistently high for certain categories, such as head and close body images. This observation is aligned with the model's performance, which also shows higher accuracies for these categories.

In direct comparison, the human subjects generally outperformed the computational model in the head and close body categories. This could be attributed to the intricate pattern recognition and abstraction capabilities inherent in human cognition. Conversely, the computational model demonstrated a comparable or slightly superior performance in the medium body category, potentially due to its systematic feature extraction and processing techniques, which may be less influenced by the complexity of the stimuli.

The far body category presented a unique challenge, with human subjects showing slightly higher variability in performance, yet maintaining a competitive edge over the model. This could suggest that the human visual system is adept at extrapolating relevant features even from less detailed stimuli, a trait that computational models are progressively striving to emulate.

Overall, the comparison underscores the complexities of visual processing and the nuances that differentiate human cognition from computational models. The findings advocate for continued advancements in computational modeling to bridge the gap in performance, particularly in categories where human cognition excels.



(a) Model

(b) Subjects

Figure 12. Accuracy

DISCUSSION

The efficacy of the Support Vector Machine (SVM) model, as indicated by the training and testing accuracies, raises important considerations about model validation techniques. The lack of cross-validation in this study's methodology may have resulted in a less accurate estimate of the model's generalization performance. K-fold cross-validation, in particular, is a powerful tool that could potentially enhance the accuracy of the model by providing a more comprehensive assessment of its predictive capabilities.

Kernel Analysis

The comparison of SVM kernels, as depicted in Figure 13, demonstrates that the linear kernel was sufficient for the task at hand. This observation leads to an intriguing conclusion that, despite the complexity of the visual categorization task, a linear decision boundary was adequate to achieve a high level of accuracy.

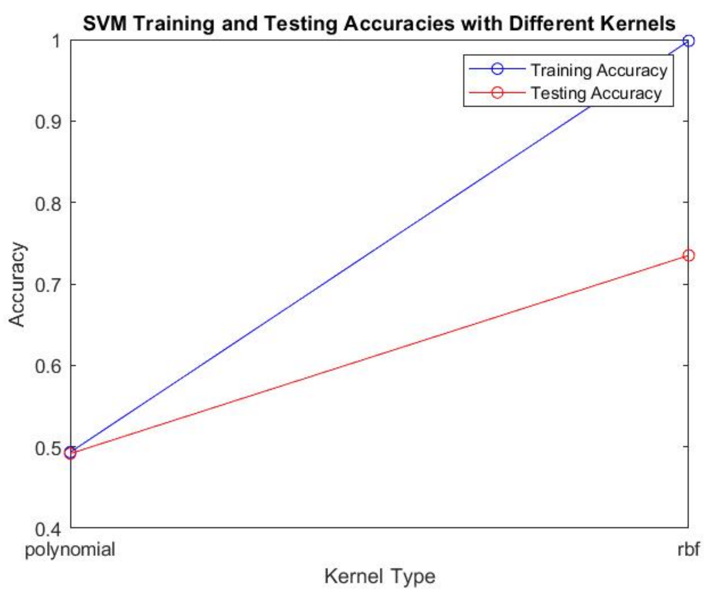


Figure 13. SVM kernels

The polynomial and rbf kernels, often preferred for their ability to model non-linear decision boundaries, did not improve the model's performance. This result suggests that the additional complexity introduced by these kernels did not translate to a practical advantage within the scope of this dataset. One could hypothesize that the feature space was sufficiently expressive such that a non-linear transformation did not provide additional benefits.

In light of these findings, future studies could benefit from a detailed hyperparameter optimization process, including the exploration of different kernel functions and the application of cross-validation techniques. This approach may lead to an enhancement of the model's accuracy and a more reliable interpretation of its ability to generalize.