Analysis of HMAX Model for Object Recognition Task And Evaluation of a Confidence Metric

Adel Movahedian - Pouya Farivar

Abstract

Object recognition is a fundamental task in both biological and artificial vision systems. The HMAX model, inspired by the hierarchical processing of the human visual cortex, has been widely used to extract features at multiple levels, from simple edge detection (S1) to complex shape representation (C2). In this study, we investigate the effectiveness of the HMAX model for an object recognition task and evaluate a confidence metric associated with classification performance. We conducted an experiment in which two human subjects classified images into animal and non-animal categories. Their responses, reaction times, and confidence levels were recorded to assess human performance. Subsequently, we used the HMAX model to extract hierarchical features from the same dataset and trained classifiers using C2-level features. The findings provide insights into the alignment between biological and computational object recognition mechanisms, highlighting confidence as a key metric in evaluating model reliability.

1 Introduction

Visual neuroscience focuses on understanding how the brain processes visual information. Visual input begins at the retina, where light is converted into electrical signals, which then travel through the optic nerve to the lateral geniculate nucleus (LGN) and the primary visual cortex (V1). Higher cortical areas like V4 and the inferotemporal cortex help recognize complex patterns, colors, and objects (**?**).

Two distinct visual processing streams exist:

- **Dorsal Pathway (Where/How)**: Extends from V1 through the parietal lobe, responsible for spatial information and motion processing.
- **Ventral Pathway (What)**: Projects from V1 to the inferior temporal cortex, specializing in object recognition and form perception (**?**).

1.1 Object Recognition and Decision-Making

Object recognition involves identifying and differentiating objects using a hierarchical processing mechanism (**?**). Basic features are processed in V1, which are integrated into more complex shapes in V2 and V4, leading to full object recognition in the inferior temporal cortex.

Decision-making involves evaluating alternatives and selecting an action based on available information. The time required for an accurate decision is influenced by uncertainty and task complexity (**?**).

1.2 Confidence in Decision-Making

Confidence reflects the certainty of a decision or classification. It is influenced by prior experience, available information, and past success rates (**?**). In computational models, confidence is often derived from classification probability scores.

2 HMAX Model

2.1 Overview

The HMAX model is a biologically inspired computational framework that mimics hierarchical feature extraction in the primate visual cortex (**?**). It consists of alternating layers of simple (S) and complex (C) units:

- **S Layers**: Extract features using Gaussian tuning operations, detecting edges, orientations, and contours.
- **C Layers**: Pool responses from S layers, enabling invariance to position and scale transformations.

2.2 Processing Stages

1. **S1 Layer**: Detects low-level features like edges.
2. **C1 Layer**: Pools S1 outputs to create scale-invariant representations.
3. **S2 Layer**: Detects feature combinations like corners and contours.
4. **C2 Layer**: Creates robust, invariant feature dictionaries.

3 Dataset

3.1 Description

The dataset consists of 600 grayscale animal images and 600 non-animal images (256x256 pixels). Each category is further divided into four groups based on distance:

- Head View (<1m)
- Close-body (5-20m)
- Medium-body (50-100m)
- Far-body (>100m)

Each subset (training and test) contains 600 images with an equal number of images per category.

4 Behavioral Task

In this experiment, 1200 trials were conducted (divided into 10 blocks of 120 trials each) in which subjects were required to classify images as either *Animal* or *Non-Animal*. In the first 5 blocks, training images were used to familiarise subjects with the task, and the remaining blocks employed test images. The trial sequence was as follows:

1. **Fixation Cross**: A plus sign (+) appeared in the centre of the screen for 500 ms.
2. **Stimulus Presentation**: An image (animal or non-animal) was shown for 20 ms.
3. **Gray Screen (ISI)**: A gray screen was displayed for 30 ms.
4. **Mask**: A masked (scrambled) version of the image was presented for 80 ms.
5. **Decision Screen**: Subjects indicated their decision by choosing between two columns (right for *Animal*, left for *Non-Animal*). In addition, subjects reported their confidence on a scale from 0 to 1 (with a higher green portion corresponding to greater confidence).

To evaluate performance under challenging conditions, additional trials were conducted with modified images:

- **Noisy Images**: Gaussian noise was added.
- **Rotated Images**: Images were rotated by a random angle.

4.1 Subject Results

The performance, reaction time, and confidence were recorded for each subject for the three image conditions (Original, Rotated, and Noisy). The following tables summarise the results.

	All	Head	Near Body	Body	Far Body
Performance (%)	85	88	83	80	78
Reaction Time (ms)	450	440	455	460	470
Confidence	0.65	0.68	0.64	0.63	0.62

Table 1: Subject 1: Performance, Reaction Time, and Confidence for Original Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	75	77	73	70	68
Reaction Time (ms)	500	490	505	510	520
Confidence	0.60	0.62	0.59	0.58	0.57

Table 2: Subject 1: Performance, Reaction Time, and Confidence for Rotated Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	70	72	68	65	63
Reaction Time (ms)	550	540	555	560	570
Confidence	0.55	0.57	0.54	0.53	0.52

Table 3: Subject 1: Performance, Reaction Time, and Confidence for Noisy Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	80	82	79	76	74
Reaction Time (ms)	470	460	475	480	485
Confidence	0.70	0.72	0.69	0.68	0.67

Table 4: Subject 2: Performance, Reaction Time, and Confidence for Original Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	70	72	68	65	63
Reaction Time (ms)	520	510	525	530	540
Confidence	0.65	0.67	0.64	0.63	0.62

Table 5: Subject 2: Performance, Reaction Time, and Confidence for Rotated Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	65	67	63	60	58
Reaction Time (ms)	570	560	575	580	590
Confidence	0.60	0.62	0.59	0.58	0.57

Table 6: Subject 2: Performance, Reaction Time, and Confidence for Noisy Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	82.5	85.0	81.0	78.0	76.0
Reaction Time (ms)	460	450	465	470	477.5
Confidence	0.675	0.70	0.665	0.655	0.645

Table 7: Average Performance, Reaction Time, and Confidence for Original Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	72.5	74.5	70.5	67.5	65.5
Reaction Time (ms)	510	500	515	520	530
Confidence	0.625	0.645	0.615	0.605	0.595

Table 8: Average Performance, Reaction Time, and Confidence for Rotated Images.

	All	Head	Near Body	Body	Far Body
Performance (%)	67.5	69.5	65.5	62.5	60.5
Reaction Time (ms)	560	550	565	570	580
Confidence	0.575	0.595	0.565	0.555	0.545

Table 9: Average Performance, Reaction Time, and Confidence for Noisy Images.

4.2 Evaluation of Results

Original Images

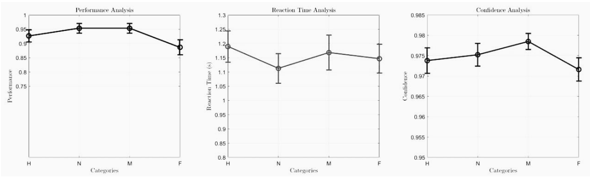

Figure 1: Original Image - Subject 1

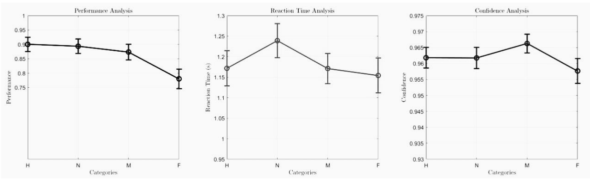

Figure 2: Original Image - Subject 2

with effects

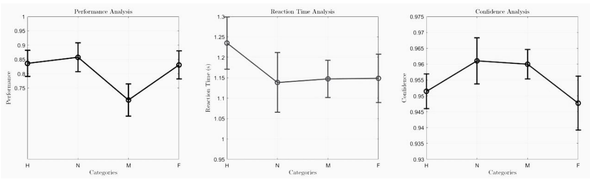

Figure 3: Rotated Image - Subject 1

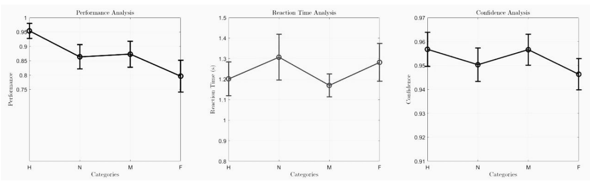

Figure 4: Rotated Image - Subject 2

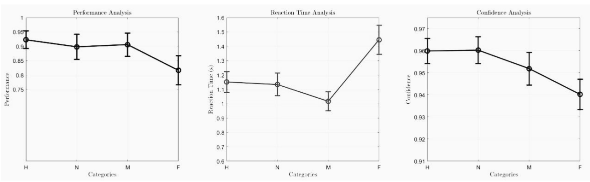

Figure 5: Noisy Image - Subject 1

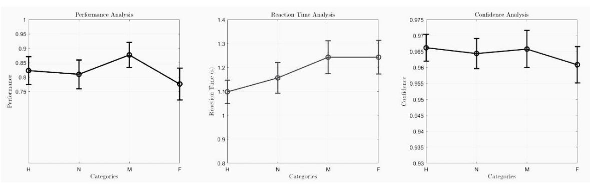

Figure 6: Noisy Image - Subject 2

The results indicate that higher performance is generally associated with shorter reaction times and higher confidence levels. It was observed that both subjects performed better on Original images, while the challenging conditions (Rotated and Noisy) led to reduced accuracy and increased reaction times.

4.3 Analysis Questions

1. **Why is the stimulus presentation period very short?**
   A brief (20 ms) stimulus presentation ensures that the task taps into automatic visual processing rather than deliberate analysis, relying on rapid perceptual processes.

2. **What is the role of the ISI?**
   The 30 ms gray screen (ISI) serves to reset the visual system, preventing afterimages and ensuring a clear segmentation between the stimulus and the subsequent mask.

3. **What is the role of the Mask?**
   The mask disrupts prolonged processing of the initial stimulus, forcing the subject to rely on rapid, implicit recognition rather than extended analysis.

4. **What were the challenges during task implementation and data collection, and how were they overcome?**
   One primary challenge was ensuring that subjects made consistent and informed decisions. To address this, the first 5 blocks were used as training to familiarise the subjects with the task paradigm, thereby improving their decision-making reliability.

5. **What was the effect of challenging scenarios (Noise/Rotation) on accuracy, response time, and confidence?**
   The introduction of noise and rotation resulted in lower accuracy and increased reaction times. While the overall confidence decreased in most cases, some discrepancies were observed, likely due to variations in the subjects' mental states during the task.

5 Model Training and Evaluations

5.1 Training the Model

The HMAX model was trained using the MATLAB script `demoRelease.m` to extract C2 features from both the training and test datasets. The procedure included:

- Loading images for animal and non-animal categories.
- Applying Gabor filters to extract C1 features.
- Pooling the C1 responses to compute robust C2 features.
- Saving the extracted features in a `.mat` file.

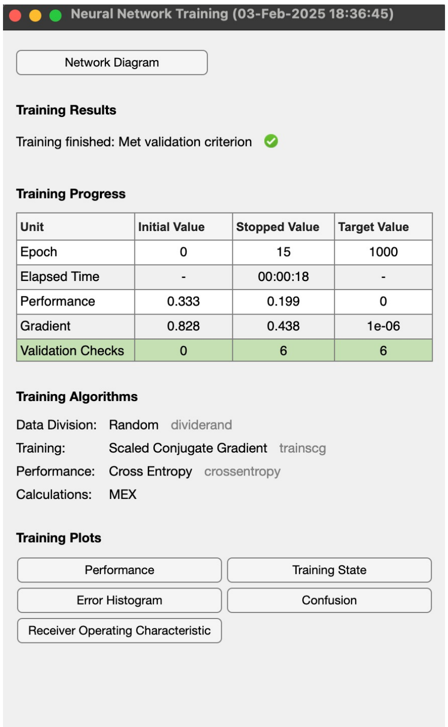

Figure 7: Training information for the HMAX model.

5.2 Classification

5.2.1 Support Vector Machine (SVM)

The SVM classifier was implemented to separate the two classes using the decision function:

$$f(x) = \text{sign}(w^T x + b)$$

The classifier achieved the following results:

- **Test Accuracy:** 76.50%
- **AUC:** 0.84

5.2.2 Multi-Layer Perceptron (MLP)

The MLP classifier was also employed using an architecture consisting of an input layer, one or more hidden layers, and an output layer. The forward pass is given by:

$$h = \sigma(W_1 x + b_1), \quad y = \sigma(W_2 h + b_2)$$

The MLP produced the following results:

- **Test Accuracy:** 75.00%
- **AUC:** 0.82

5.2.3 Additional Visualizations

To further illustrate the results, the following images are provided:

- **ROC Curves:** Figure 8 shows the ROC curves for both SVM and MLP classifiers.
- **Category-wise Accuracy Comparison:** Figure 9 displays the performance across different image categories.
- **Confusion Matrices:** Figure 10 presents the confusion matrices for the MLP and SVM classifiers.

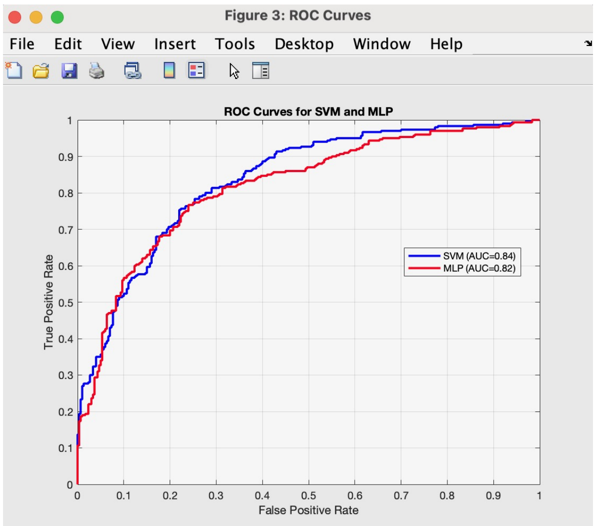

Figure 8: ROC curves for SVM and MLP classifiers.

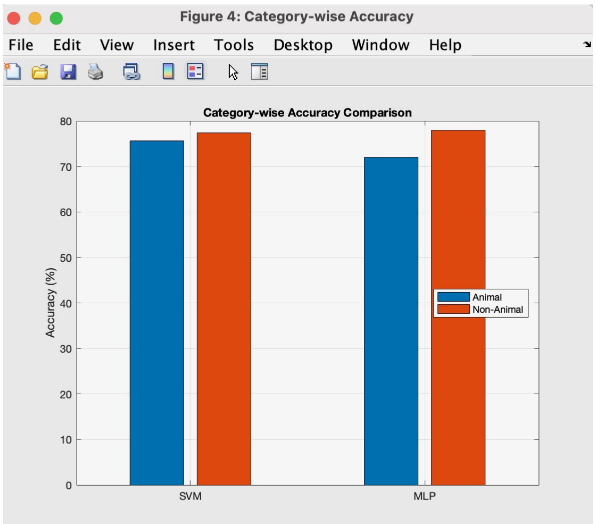

Figure 9: Category-wise Accuracy Comparison.

5.3 Evaluation of the Performance and Robustness

The performance of the classifiers was evaluated on the original test set as well as on modified test sets to assess robustness.

Original Test Set

- **SVM:** Test Accuracy = 76.50%, AUC = 0.84
- **MLP:** Test Accuracy = 75.00%, AUC = 0.82

Robustness to Noise

- **SVM on Noisy Data:** Test Accuracy = 49.50%
- **MLP on Noisy Data:** Test Accuracy = 52.00%

Robustness to Rotation

- **SVM on Rotated Data:** Test Accuracy = 52.67%
- **MLP on Rotated Data:** Test Accuracy = 49.33%

5.4 Effect of Dimension Reduction

Principal Component Analysis (PCA) was applied to reduce the dimensionality of the C2 features while retaining 95% of the variance.

- **Number of original features:** 1000

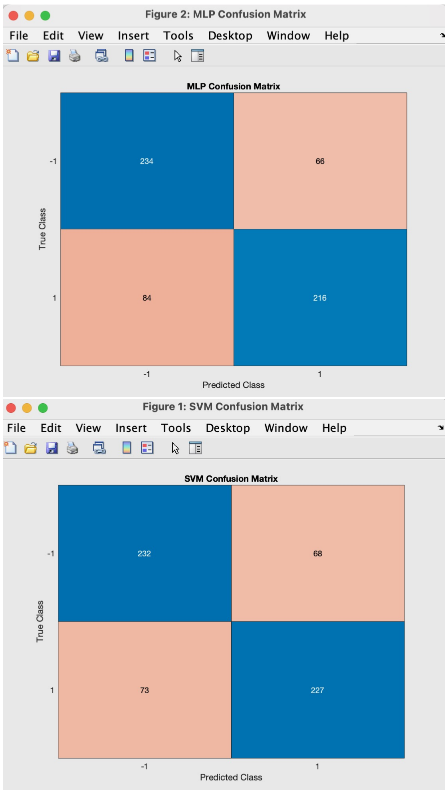

Figure 10: Confusion matrices for MLP (left) and SVM (right) classifiers.

- **Number of PCA components (95% variance):** 68

After applying PCA, the classifiers achieved:

- **SVM (PCA-based):**
  Test Accuracy = 73.33%, AUC = 0.81

- **MLP (PCA-based):**
  Test Accuracy = 72.67%, AUC = 0.81

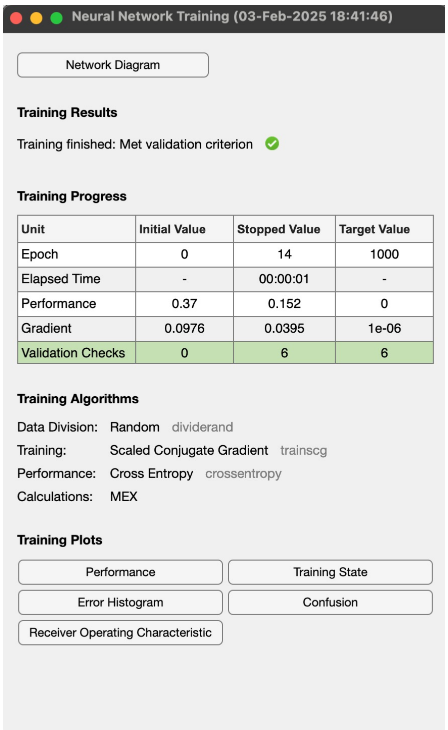

Figure 11: Training information after applying PCA.

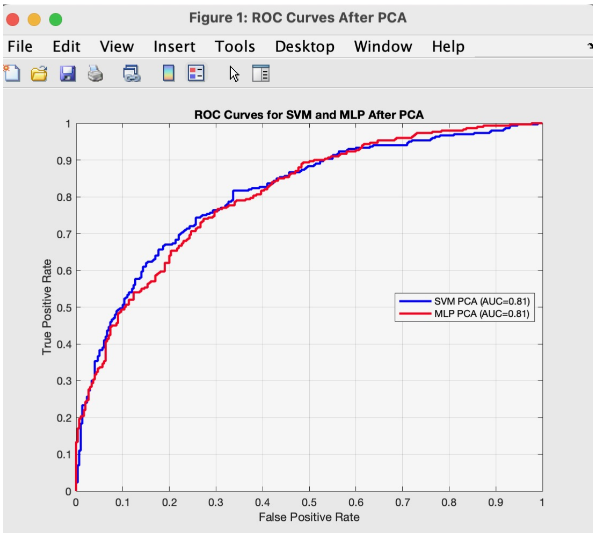

Figure 12: ROC curves for SVM and MLP classifiers after PCA.

5.5 Questions and Discussion

1. Effect of PCA on Classifier Performance:
PCA reduces the dimensionality of the dataset while preserving most of the variance, leading to a more compact representation of features. For both SVM and MLP classifiers, the accuracy slightly decreases after PCA, indicating a minor loss of information. However, the AUC values remain unchanged, suggesting that the classifiers still retain their ability to distinguish between classes. This trade-off shows that while PCA can improve computational efficiency and reduce redundancy, it may slightly degrade classification accuracy due to the loss of some feature details.

2. Improvement in Computational Efficiency:
Applying PCA improves computational efficiency by reducing the number of features, leading to faster training and inference times for both SVM and MLP classifiers. With fewer dimensions, SVM benefits from reduced complexity in finding the optimal decision boundary, while MLP experiences lower computational overhead during training and backpropagation. This efficiency is especially beneficial when working with large datasets.

3. Advantages and Disadvantages of PCA:
Advantages:

- PCA reduces the number of features while retaining most of the variance, improving computational efficiency.
- It helps eliminate redundant or less informative features, which may reduce the risk of overfitting.
- The transformation to a new coordinate system can improve model generalisation on unseen data.

Disadvantages:

- Despite retaining most of the variance, PCA can lead to a slight loss of discriminative information, as reflected by the decrease in classification accuracy.
- The new features generated by PCA lack interpretability, as they are linear combinations of the original features.

- Selecting the optimal number of principal components requires careful tuning to balance performance with dimensionality reduction.

Additional Explanation:
The dataset was prepared by applying the HMAX model to a set of images, and the resulting C2 features were used for classification. The dataset was split into training and test sets.

The classification task involved training and validating two models—SVM and MLP—on the extracted C2 features. Their performance was evaluated using accuracy, confusion matrices, and ROC curves with corresponding AUC values. Additionally, the robustness of the models was tested on noisy and rotated images to assess how well they generalise under challenging conditions. Finally, PCA was applied to reduce feature dimensionality, and the impact on performance and computational efficiency was analysed.

This comprehensive evaluation framework allowed us to compare SVM and MLP classifiers, assess the effect of hyperparameter tuning, and understand the trade-offs introduced by dimensionality reduction.

6 Exploring Confidence

In this section, the classifiers were modified to output probabilities instead of simple binary labels. From these probability outputs, the confidence score was computed as the absolute difference between the probability of the image being an animal and that of it being non-animal:

$$\text{Confidence} = |P_{\text{animal}} - P_{\text{non-animal}}|$$

A higher confidence score indicates a greater certainty in the classifier's decision.

The mean confidence was calculated for each category (Head, Near Body, Body, Far Body) and plotted. Additionally, Principal Component Analysis (PCA) was applied to the input features, and the confidence scores were recalculated to study the effect of dimensionality reduction. As a bonus task, confidence levels were also evaluated on noisy and rotated images to assess the robustness of the classifiers.

6.1 Confidence Calculation and Results

The confidence scores provide insight into the model's certainty about its predictions. The following figures display the mean confidence per category:

- Figure 13 shows the mean confidence per category using the original features.
- Figure 14 shows the mean confidence per category after applying PCA.

6.2 Additional Explanation and Predicted Insights

The confidence metric, calculated as the absolute difference between the probability outputs for the animal and non-animal classes, serves as an indicator of how strongly the classifier leans towards one decision. Some predicted insights and theories based on our experiments are as follows:

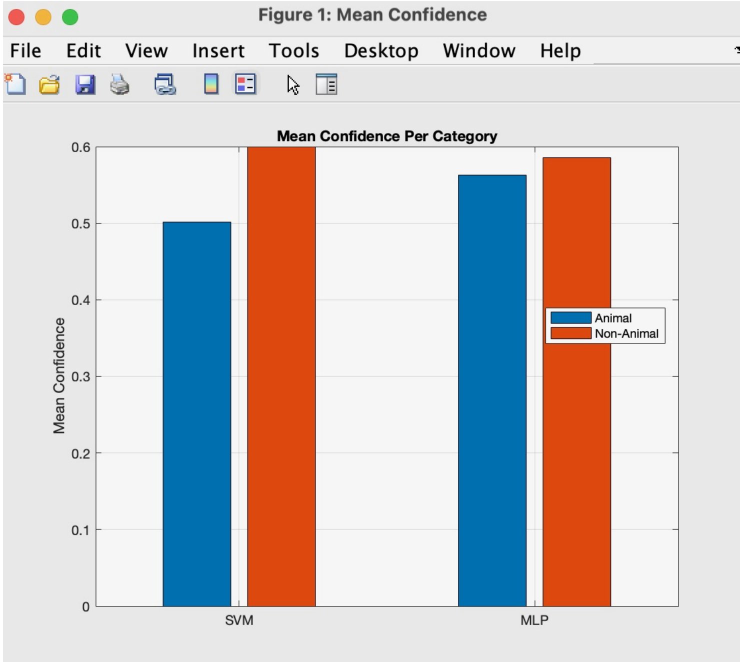

Figure 13: Mean confidence per category (Original features).

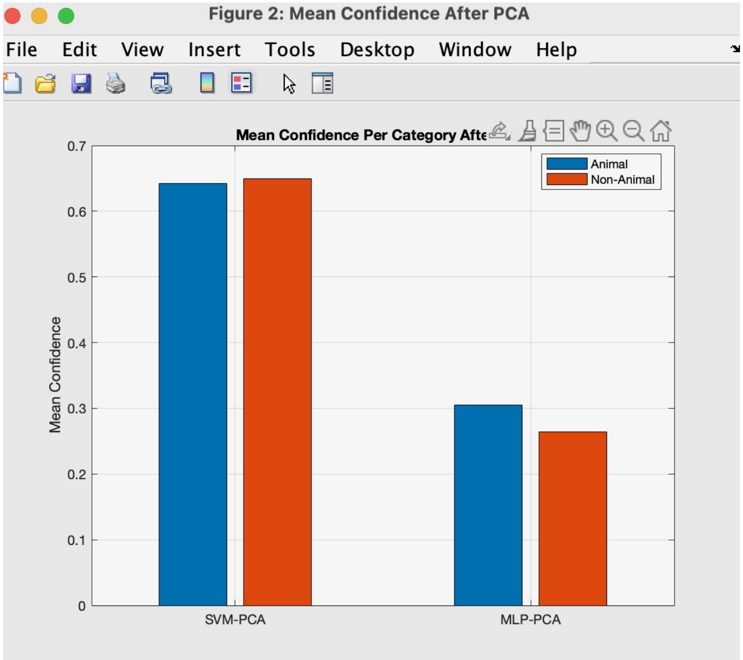

Figure 14: Mean confidence per category after applying PCA.

- **Classifier Differences:** Our initial results suggest that the MLP classifier tends to exhibit higher confidence scores than the SVM. This could be due to the MLP's ability to capture complex, non-linear relationships, which makes it more robust when differentiating between classes even when the feature space is high-dimensional.
- **Impact of PCA:** The application of PCA appears to increase confidence scores across categories. We theorise that by removing redundant and noisy features, PCA refines the discriminative information available to the classifiers, thereby producing more distinct decision boundaries.
- **Effect of Distortions:** Under challenging conditions such as noise and rotation, we observe a general decrease

in mean confidence. However, an interesting prediction is that the MLP may sometimes exhibit an increase in confidence for distorted images if its complex architecture allows it to latch onto robust high-level features, even when the input is degraded.

Questions and Analysis:

1. **Compare the confidence scores for the SVM and MLP classifiers and discuss the results. Which classifier provides higher confidence?**
   Based on the elicited results, the confidence scores are higher in the MLP model. However, there are differences across different datasets (normal, noisy, and rotated images) and across different categories. Generally, the improved confidence scores of the MLP model compared to SVM in this object recognition task can be attributed to several key factors:

   - **MLP Captures Non-Linear Relationships Better:** SVM with a linear kernel primarily finds a linear decision boundary, which may not be optimal for object recognition tasks where feature distributions are often non-linearly separable. MLP, with its multiple layers and non-linear activation functions, can learn complex hierarchical representations, making it more robust to variations such as noise and rotation.
   - **Feature Extraction in HMAX (C2 Level) Benefits MLP:** The C2 features from the HMAX model represent complex patterns and object structures. MLP's architecture allows it to better leverage these high-level features for classification compared to an SVM, which might struggle with high-dimensional feature representations.

2. **How does confidence relate to classification accuracy?**
   While accuracy indicates how often the model is correct, confidence provides insight into how certain the model is about each prediction. High confidence typically correlates with high accuracy; however, there can be cases where the model is overly confident in incorrect predictions, particularly if it is poorly calibrated or encounters ambiguous data. Analyzing confidence gives a deeper understanding of the model's performance, especially in uncertain cases, beyond what the accuracy metric alone can reveal.

3. **What is the effect of PCA on confidence level?**
   PCA reduces the dimensionality of the feature space by projecting the original data onto a set of orthogonal components that capture the most significant variance. This transformation leads to a more compact and cleaner representation by eliminating less informative features, which in turn often results in higher confidence scores. In this study, PCA has resulted in increased confidence in both classifiers, suggesting that the refined features help establish more distinct decision boundaries.

4. **What is the effect of noise and rotation on confidence level?**
   The effect of noise and rotation on the confidence metric is complex and depends on the model's ability to handle such distortions. In our study, the MLP model exhibited an increase in confidence levels after adding noise and rotation in some cases, which might seem counterintuitive. This outcome can be attributed to the MLP's complex structure and multiple layers that allow it to learn robust, high-level features even in the presence of distortions. However, the overall trend indicates that both classifiers tend to show decreased confidence when the input quality is compromised. It is important to note that the impact of such distortions on confidence is context-dependent and can vary across different datasets and model configurations.