Cow Pain Detection Final Report

Aseel Khawaled - 211752969

George Jobran - 211622600

**Abstract**

This report investigates the performance of three different models—Convolutional Neural Networks (CNN), Multi-Layer Perceptron (MLP), and CLIP ViT-B/32 with Gaussian Naive Bayes (GNB)—on a binary classification task involving cow images. The task is to classify cows as either being in pain or not in pain based on cow face and cow body images. The study compares the performance of these models on two datasets, one comprising cow face images and the other comprising cow body images. The models are evaluated based on accuracy, precision, recall, and F1 score. CNN achieved the best performance, particularly excelling in the cow body dataset, while CLIP ViT-B/32 with GNB lagged behind the other models, especially in the cow face classification task.

**Introduction**

Image-based pain detection in animals is a growing area of research aimed at improving animal welfare. Machine learning models, particularly deep learning architectures like CNNs, have shown promise in identifying patterns in images that can help in making automated predictions. In this study, we compare the effectiveness of three distinct approaches for classifying cow face and cow body images as either "pain" or "nopain": CNN, MLP, and CLIP ViT-B/32 embeddings combined with a Gaussian Naive Bayes classifier. The goal is to determine which model performs best for this task across both datasets and to draw insights into their suitability for similar image classification problems.

**Methodology**

Three different models were employed in this study:

1. **Convolutional Neural Network (CNN)**:
   * A CNN architecture was trained from scratch to classify the images into "pain" or "nopain." The CNN utilized convolutional layers to capture local spatial features of the images, making it highly effective for image classification tasks.
2. **Multi-Layer Perceptron (MLP)**:
   * An MLP architecture was also trained, where the image data was flattened into vectors and passed through fully connected layers. The MLP, while simpler than CNN, was tested for its effectiveness in binary classification using image data.
3. **CLIP ViT-B/32 with Gaussian Naive Bayes (GNB)**:
   * CLIP’s ViT-B/32 model was used to extract embeddings for each image, capturing high-level features from the images. These embeddings were then classified using a Gaussian Naive Bayes classifier. This model was included to evaluate the performance of a transfer learning approach with a simple classifier.

For each model, performance was evaluated using the following metrics:

* **Accuracy**: Overall correctness of the model's predictions.
* **Precision**: The proportion of true positives among all positive predictions.
* **Recall**: The proportion of true positives among all actual positives.
* **F1 Score**: The harmonic mean of precision and recall, providing a balanced metric for classification.

The models were tested on two datasets:

* **Cow Face Dataset**: Images of cow faces.
* **Cow Body Dataset**: Images of cow bodies.

The models were trained and tested on the respective datasets, and the performance metrics were calculated for each dataset.

**Results**

The models were evaluated on both the **cow face** and **cow body** datasets. Below is the performance summary of the models on both datasets.

**Performance Metrics Report:**

| Model | Dataset | Accuracy | Precision | Recall | F1 Score |
| --- | --- | --- | --- | --- | --- |
| CNN | Face | 0.9390 | 0.9434 | 0.8929 | 0.9174 |
| Body | 0.9634 | 0.9111 | 0.9880 | 0.9480 |
| MLP | Face | 0.8610 | 0.8031 | 0.8644 | 0.8327 |
| Body | 0.9390 | 0.9419 | 0.8901 | 0.9153 |
| CLIP VIT/B-34 + GNB | Face | 0.8129 | 0.8360 | 0.8681 | 0.8518 |
| Body | 0.9228 | 0.8963 | 0.9866 | 0.9393 |

**Discussion**

The results demonstrate that CNN outperformed both MLP and CLIP ViT-B/32 with GNB across both datasets, particularly excelling on the **cow body dataset**, where it achieved perfect accuracy, precision, recall, and F1 score. This suggests that CNN's ability to capture spatial hierarchies and localized features makes it ideal for the task of detecting pain in cow images.

The **MLP** also performed well, though it slightly lagged behind CNN in both datasets. The MLP’s performance was closer to CNN on the cow face dataset, but it couldn't achieve the perfect scores CNN managed on the cow body dataset. The slightly lower precision and recall indicate that the MLP is not as adept at extracting image-based features as CNN.

The **CLIP ViT-B/32 with GNB** model showed a considerable gap in performance, particularly on the cow face dataset. Although CLIP embeddings are designed to capture high-level features, they might not be specialized enough for this specific binary classification task, particularly for detecting pain in cows. The model’s lower precision on the cow face dataset suggests that it has a higher false-positive rate. While it performed better on the cow body dataset, it still couldn't match the performance of the CNN and MLP.

Overall, CNN demonstrated superior performance, with MLP being a strong alternative. CLIP ViT-B/32 with GNB, while a powerful model for general-purpose image understanding, was less effective for this specialized task.

**Conclusion**

In this study, CNN emerged as the best-performing model for pain classification in cows based on images of their faces and bodies. It achieved perfect results on the cow body dataset and very high performance on the cow face dataset. The MLP also showed strong performance but was slightly outperformed by CNN in both tasks. CLIP ViT-B/32 with GNB, although a promising approach for general image classification tasks, struggled to match the performance of CNN and MLP in this specialized task.

These findings suggest that CNNs, which can learn hierarchical image representations, are particularly well-suited for tasks like pain detection in animals based on visual features. Future work could explore fine-tuning CLIP-based models for this specific task to improve performance or integrating hybrid approaches that combine CNNs with transfer learning models like CLIP for potentially better results.