# Comparative Analysis of CNN and Vision Transformer Models on Image Classification and Object Detection Problems

## Business Objective

Explore and identify at least 5 CNN and Vision Transformer architectures/models that are commonly used for image classification and object detection problems. The objective is to understand these architectures by reviewing literature, research papers, and related materials. Conduct a comparative analysis of these models using datasets tailored for image classification and object detection tasks. Additionally, this project aims to provide insights into designing and training deep learning models for these problems, evaluating their strengths and weaknesses to determine the most effective architecture for specific use cases.

## Dataset Details

Choose datasets relevant to image classification and object detection tasks. Suggested options include, but are not limited to:

1. 1. Image Classification Datasets:

* - ImageNet: Large-scale dataset with millions of labeled images across thousands of categories.
* - CIFAR-10/ CIFAR-100: Small datasets containing labeled images in various classes.
* - Google Open Images V7: A large dataset with diverse and real-world labeled images suitable for classification tasks.
* - iNaturalist Dataset: A large-scale dataset with real-world images focused on species classification, useful for fine-grained categorization.
* - Places365: A dataset with over 10 million images for scene recognition and classification.

1. 2. Object Detection Datasets:

* - COCO (Common Objects in Context): Rich dataset with images labeled for detection, segmentation, and captioning.
* - Pascal VOC: Dataset for object detection with bounding boxes and class labels.
* - Open Images Dataset: Large-scale dataset with annotations for object detection tasks.
* - Waymo Open Dataset: A real-world dataset for autonomous driving, useful for object detection and segmentation tasks.

Participants may also select other relevant datasets as per the project requirements.

## Acceptance Criterion

The final output should be a comprehensive report comparing the performance, computational efficiency, and usability of the selected CNN and Vision Transformer architectures/models. The evaluation metrics will include accuracy, mean Average Precision (mAP) for object detection, training/inference time, and computational resource utilization.

## Milestones

|  |  |  |
| --- | --- | --- |
| Milestone | Duration | Task Description |
| Kickoff and Business Objective Discussion | 1 Day | Discuss project objectives, finalize the tasks (image classification and object detection), and select models. |
| Literature Review & Dataset Preparation | 1 Week | Review literature on CNN and Vision Transformer architectures. Download and preprocess datasets for classification and detection. |
| Model Implementation & Training | 1 Week | Implement and train at least 5 CNN and Vision Transformer architectures/models (e.g., AlexNet, VGG, ResNet, MobileNet, Vision Transformer, etc.). |
| Model Evaluation & Comparative Analysis | 1 Week | Evaluate models on test datasets and perform comparative analysis based on defined metrics. |
| Feedback and Refinement | 1 Week | Incorporate feedback, fine-tune models if necessary, and refine documentation. |
| Final Presentation | 1 Day | Present the results, complete documentation, and submit the final report and code. |

## Protocols

- All participants must adhere to the agreed timelines. Extensions will not be granted.  
- The final presentation, along with documentation and Python code, must be submitted before the presentation day.  
- Participation in review meetings is mandatory.

This document outlines the revised business requirements for the CNN and Vision Transformer-based comparative analysis project, tailored to the new objective provided.