**PaAC Open Project: Finetuning Deep Learning Models on Astrophysical Datasets** **Core Assignment - 1** **Literature Review & Model Selection** **Submitted by: Mahfooj Ali**

### **1. Literature Review**

The Galaxy Zoo project provides a large set of labeled galaxy images contributed by citizen scientists, aimed at supporting research in astrophysical morphology classification. For the purpose of this project, the primary goal is to perform galaxy image classification using deep learning models, particularly convolutional neural networks (CNNs).

Prior works using the Galaxy Zoo dataset have shown that CNNs significantly outperform traditional machine learning models. Notable papers include:

* “Galaxy Morphology Classification using Deep Convolutional Neural Networks” – Dieleman et al.
* "Morphological Classification of Galaxies Using Deep Learning" – Huertas-Company et al.

These papers emphasize the importance of transfer learning and data augmentation to improve performance on astrophysical image datasets.

### **2. Pretrained Model Selection**

After reviewing multiple models, I propose using **ResNet-50** as the base pretrained model due to the following reasons:

* **Feasibility**: ResNet-50 is computationally less intensive than larger architectures like ResNet-101 or ViT, making it suitable for training on standard GPUs or cloud platforms (like Google Colab).
* **Performance**: It is known for its stable convergence and strong performance on a wide variety of image classification tasks.
* **Compatibility**: Available via **Hugging Face Transformers and timm**, allowing easy integration and fine-tuning workflows.

**Framework/Tools**:

* **Hugging Face Transformers**
* **PyTorch + torchvision**
* **timm** (PyTorch Image Models library)

### **3. Fine-Tuning Method Selection**

After reviewing various fine-tuning strategies, the following methods were considered:

#### **✅ Selected Method: LoRA (Low-Rank Adaptation)**

* **Reason**: Reduces the number of trainable parameters, making fine-tuning memory efficient while retaining strong performance.
* **Use Case**: Especially effective when working with large models and limited computational resources.
* **Reference**: Hu et al., *"LoRA: Low-Rank Adaptation of Large Language Models"*

#### **🔍 Other Fine-Tuning Methods:**

* **Full Fine-Tuning**: Updates all model parameters. High accuracy but resource-intensive.
* **Adapter Layers**: Inserted into layers to adapt without changing original model weights.
* **Layer Freezing**: Freeze lower layers and train only higher ones; balances efficiency and accuracy.

### **4. Optional Exploration: Ensemble Learning**

Explored ensemble methods that could potentially outperform individual models:

* **Model Averaging**: Combine predictions of ResNet, EfficientNet, and ViT models.
* **Stacked Generalization**: Use meta-models to combine outputs of base models.
* **Bagging and Boosting**: Apply lightweight ensemble techniques to fine-tuned models.

Although ensemble methods introduce extra complexity and computation, they may enhance performance significantly and are worth trying post baseline model evaluation.

### **5. References**

1. Dieleman, Willett, Dambre. *"Rotation-invariant convolutional neural networks for galaxy morphology prediction"*. (2015)
2. Huertas-Company et al. *"Deep Learning for Galaxy Morphology Classification"*. (2018)
3. HuggingFace Model Hub: https://huggingface.co/models
4. Galaxy Zoo Dataset:<https://data.galaxyzoo.org/#section-21>
5. Hu et al. *"LoRA: Low-Rank Adaptation of Large Language Models"*, 2021.
6. Timm GitHub:<https://github.com/huggingface/pytorch-image-models>
7. Blogs: Towards Data Science, PapersWithCode, Medium articles on image classification