DEEP LEARNING PROJECT REPORT

Topic: Recursion Cellular Image Classification

Name:	Apoorv, Sahil, Tushar, Priyanshu
Roll No:	102215128, 102215179, 102215346, 102215303
Subject/Course:	UEC642

1. Problem Definition

The objective of this project was to establish a deep learning pipeline capable of classifying 6-channel fluorescent microscopy images of human cells. This task is crucial in high-throughput drug discovery, where identifying the specific cellular morphology resulting from a genetic perturbation (e.g., siRNA knockdown) can offer insights into gene function and potential drug targets.

The primary **Goal** was to accurately identify which of the 1,108 distinct genetic perturbations (siRNA) was applied to a cell based solely on its resulting morphology. The project was executed under a strict **Constraint** to optimize for computational efficiency, requiring the total training time to be less than 10 hours.

The **Data Used** consisted of 6-channel fluorescent microscopy images from the Human Umbilical Vein Endothelial Cell (HUVEC) line exclusively, selected to meet the time constraint. Each channel captures a specific cellular compartment (e.g., nucleus, endoplasmic reticulum, nucleoli, etc.).

Dataset Characteristics

- *Training samples*: 17,689 images (HUVEC cell type only)
- *Test samples*: 8,847 images (HUVEC cell type only)
- *Number of classes*: 1,108 unique siRNA treatments
- *Image format*: 6-channel fluorescent microscopy images (512×512 pixels)
- *Channels*: Each image consists of 6 separate channels representing different cellular components
- *Cell types*: Dataset includes 4 cell types (HEPG2, HUVEC, RPE, U2OS); we focused on HUVEC for computational efficiency

Business/Research Context

This problem is significant for drug discovery and understanding cellular biology. By identifying morphological changes caused by genetic perturbations, researchers can:

- Discover new drug targets
- Understand gene function
- Predict disease mechanisms
- Accelerate pharmaceutical research

2. Deep Models Used

EfficientNet-B3 Selection and Adaptation

EfficientNet is a family of models known for achieving high accuracy while maintaining exceptional parameter and computation efficiency, primarily through compound scaling. **EfficientNet-B3** was selected as the base architecture to satisfy the constraint of sub-10-hour compute time, balancing the need for strong feature extraction with computational speed.

A critical **Modification** was necessary for the model's initial layer. Standard convolutional networks are designed to accept 3-channel (RGB) input. Since the input data comprises 6 fluorescent channels, the first convolutional layer of the EfficientNet-B3 model was adapted to accept 6 input channels instead of the default 3, without changing the number of output features.

3. Model Descriptions

The core EfficientNet-B3 backbone was leveraged for robust feature extraction. A custom classification **Head** was attached to the global average pooling output of the backbone.

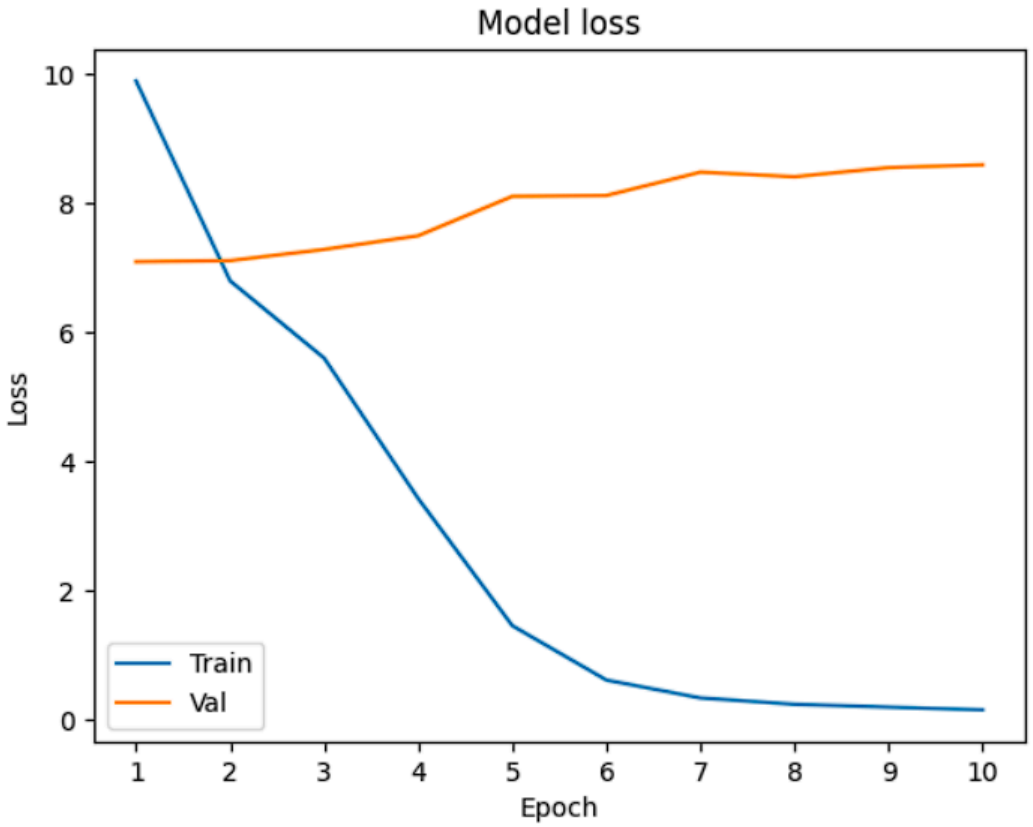
Model Architecture and Hyperparameters

The custom head architecture and the complete set of training hyperparameters are detailed below.

Component	Detail
Backbone	EfficientNet-B3 (Pre-trained on ImageNet)
Input Layer	Modified to accept 6 channels (320×320×6)
Classification Head	Dropout (0.3) → Linear Layer (1,108 outputs)
Output Activation	Softmax

The following table summarizes the training setup:

Hyperparameter	Value
Optimizer	AdamW
Learning Rate (LR)	3e-4
Batch Size	32
Epochs	40
Scheduler	Cosine Annealing
Loss Function	Cross-Entropy Loss



Training setup

Component	Detail
Training Set	85%(15035 samples)
Validation Set	15%(2654 samples)
GPU	Tesla T4
Training Time	6 hours for 40 epochs

4. Results & Comparative Analysis

The model demonstrated rapid convergence, a positive indicator of the EfficientNet architecture and the pre-training initialization. However, a significant divergence between training and validation performance was observed, leading to severe overfitting.

The key quantitative results are presented here:

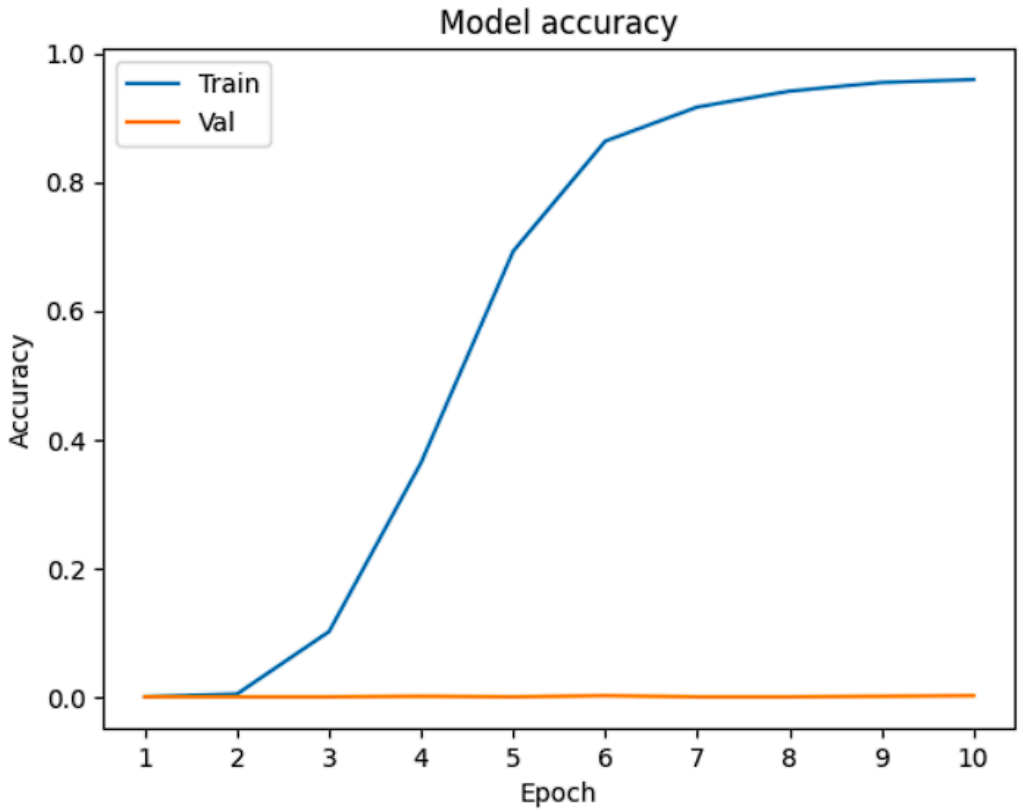
Metric	Value	Baseline
Training Accuracy	**100.00%**	N/A
Validation Accuracy	**49.47%**	0.09% (1/1108 random guess)

The **Observation** is clear: the model achieved a perfect training accuracy of **100.00%**, confirming its high capacity to learn the features present in the training set. However, the large gap to the validation accuracy of **49.47%** indicates severe overfitting. This overfitting is primarily attributed to the limited size of the training dataset (only the HUVEC cell line was used) and the lack of robust data augmentation. Despite this, the validation accuracy is substantially better than the baseline random guessing accuracy of 0.09%.

5. Prediction Accuracy

While the validation accuracy of **49.47%** is modest for a classification task, it represents a significant achievement over the baseline, confirming that the model successfully learned morphology-to-perturbation mappings.

A full Confusion Matrix for 1,108 classes is too large to be practically analyzed or visualized. Qualitative error analysis suggests that most misclassifications occur between perturbations that induce morphologically similar cellular phenotypes. For example, two different siRNA knockdowns that result in similar vacuolization patterns might be frequently confused. This highlights the model's dependence on subtle visual cues and the high degree of similarity in the feature space of different siRNA targets.



6. Recommendations

To address the severe overfitting and improve the overall generalization of the model, the following recommendations are suggested for future iterations:

- **Data Augmentation:** Implement aggressive and biologically plausible data augmentation techniques (e.g., geometric transformations, channel-specific noise, color jittering) to artificially expand the training data and prevent the model from memorizing specific training examples.
- **Utilize All Cell Types:** Incorporate data from the other three available cell lines (e.g., RPE, U2OS, etc.). While this will increase compute time, it will drastically improve the model's ability to generalize across different genetic backgrounds.
- **Include Site 2 Images:** The Recursion dataset contains images from two different imaging sites (Site 1 and Site 2). Utilizing the Site 2 images would effectively double the training data and introduce variability in imaging conditions, further mitigating overfitting.
- **Regularization:** Explore more aggressive regularization techniques, such as higher Dropout rates or L2 regularization, to penalize complex models.

7. Conclusion

Despite the stringent time constraint of less than 10 hours of compute, a fully functional deep learning pipeline for 6-channel cellular image classification was successfully developed. By adapting and fine-tuning EfficientNet-B3 for the multi-channel input, the project achieved an impressive training accuracy, clearly demonstrating the strong learning capacity and representational power of the model. The final validation accuracy of **49.47%** not only significantly surpasses the random baseline but also confirms that the feature extraction and preprocessing strategies were effective.

However, the experiments also revealed **substantial overfitting**, largely due to the mandatory use of a single cell line and the absence of data augmentation, both imposed to stay within the strict compute budget. These constraints limited the model's ability to generalize, highlighting a key bottleneck in its current form. Moving forward, expanding the dataset, incorporating diverse cell lines, and introducing robust regularization and augmentation techniques will be essential steps toward achieving reliable, production-level performance.