# Indian Institute of Technology, Jodhpur

Dependable AI | Assignment 3

Topic : Visualisation and Explainability

**Total Marks :** 60 + 10 Marks(Report) + 30 Marks(Viva)

---

**Submission Policy and Requirements :**

- Any kind of plagiarism is not accepted. We will strictly follow institute policies for plagiarism.
- Recommended programming languages: Python + Keras/TensorFlow/PyTorch.
- You may use any external libraries or GitHub codes. However, the evaluation will test your knowledge of the algorithm and the choice of hyperparameters. Do cite the libraries/codes.
- Submission should include: Working code for each of the parts separately and a report to show the analysis of results in each of the parts.

**Assessment criterion:**

The assessment will be done on the basis of the following components:

- Working codes
- Analysis and clarity of results (drawing comparisons across different parts) & clarity of the report
- Understanding the theoretical concepts and the choice of hyperparameters.

**Guidelines for Submission:**

- A single report(pdf) for all questions.
- Mention all the relevant results, comparisons as asked or wherever required for better understanding for the results.
- A single zip file containing the report, codes and readme if required

---

**Question 1 (Visualisation): [30 Marks]**

**Dataset**: CIFAR-10

**Aim**: Implementation of GradCam.

**Model:** Select a deep learning model (ex. Resnet18, VGG16 etc.) and train it for 10-class classification. (**Note**: You may use your trained model for CIFAR10 dataset from Assignment 1, Question 3)

a. Report the accuracy obtained on the test set.
b. Select 10 samples from each class which were correctly classified by the trained model. Apply GradCam on it and visualise most salient regions being used for prediction.
c. Select 5 samples from each class which were incorrectly classified by the trained model and apply GradCam to visualise most salient regions along with the predicted class and confidence.

**Reference Article:**

https://towardsdatascience.com/demystifying-convolutional-neural-networks-using-gradcam-554a85dd4e48

**Reference Paper:** https://arxiv.org/pdf/1610.02391.pdf

**Question 2 (Explainability): [30 Marks]**

Choose **any one** from the following:
**Aim:** Implementation of LIME model.
**Dataset**: 20newsgroups

1. Train an SVM/Regression Model on this data, visually explain predictions of the model on samples from test data using LIME.(one sample from each class)

**Dataset:** CIFAR10

2. Take any ImageNet Pretrained Model, generate a visual explanation of the model's predictions on 10 samples from CIFAR-10 test data using LIME. (One sample from each class)