

## Project Task

### Skin Cancer Classification

#### General Instructions

1. You may do the project in a group of **maximum 3 students**, or **individually**.
2. You may use any Python libraries like Tensorflow or Pytorch etc.
3. A Viva will be taken for each team after submission.

#### Instructions

1. The task is to classify different categories of skin cancers.
2. A dataset is provided to you for training and validation.
3. The flower dataset contains **7 classes**. 1. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (**akiec**), 2. basal cell carcinoma (**bcc**), 3. benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, **bkl**), 4. dermatofibroma (**df**), 5. melanoma (**mel**), 6. melanocytic nevi (**nv**) and 7. vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, **vasc**).
4. The train set contains grayscale images making a total of **3000** images with dimensions **300 x 300**.
5. The validation set contains **1000** images.
6. The train and validation labels are prefixed to image names like 1\_abc. (Classes 1-7).
7. You need to train **2** multi-layer convolutional neural network architecture (**MobileNet based only**) with any modifications if you want in the architecture, activation functions and optimiser.
8. **PLEASE NOTE THAT YOU CANNOT USE ANY MLP/RNN/LSTM/TRANSFORMER ETC. FOR THIS TASK. YOU HAVE TO USE A MobileNet BASED NEURAL NETWORK ONLY.**
9. You are free to use any pre-trained MobileNet network but not from the categories of networks mentioned in 8.
10. You may use data augmentation for the task.

#### Submissions

1. **CODE SUBMISSION:** You need to provide the model weights of the **2 architectures** which you use in your submission so that the code can be checked on the test dataset which is not shared with you.

2. Please provide a cell in your notebook at the end for classifying test data, where any user can test a set of new images on your trained model. This is **MANDATORY**. If you do not add this cell, then your code will not be evaluated on test data.
  - Submit the notebook with a cell which has a proper code for testing (i.e. with a path to load the data, code for any preprocessing steps etc.), with any supporting functions as well, so that the evaluator just has to change the path of the directory for the testing data.
3. The submission will also include the outputs in the notebook file itself. E.g. The training loss vs epochs plot, **accuracy values for training and validation datasets**.
4. Evaluation metric will be **accuracy** i.e. number of correct predictions out of all predictions. Try to get maximum accuracy (as much as you can) on validation dataset for optimal performance of your model.
5. **PROJECT REPORT:** Every team/individual needs to submit a report (softcopy) in proper format with name and ID of group members, table of contents, results, the training process, and any other details related to your implementation. You must explain all the training or design choices adopted by you.
6. **CONFUSION MATRIX FOR VALIDATION SET IS MANDATORY IN REPORT.**
7. **TOTAL ACCURACY AS WELL ACCURACY FOR EACH CLASS NEEDS TO BE MENTIONED.**
8. **YOUR CODE SHOULD RUN ON GOOGLE COLAB, OTHERWISE WON'T BE EVALUATED.**

### Submission Deadline

1. The submission deadline is **14 April 2025, 11 AM IST**.
2. The submission link will be shared with you in a few days.
3. It is advised that you plan early and try to submit the project prior to the deadline, as **NO EXTENSIONS will be possible after the deadline**.

### Checklist for Submission

1. A single Python notebook file with the code for pre-processing steps, train and validation results, and a cell for testing of your model.
2. Project Report in PDF format.

**These files must be sent in a single zip file with first names of the team members e.g. (Abc\_xyz\_pqr.zip), and without the train and validation datasets.**

-----XXXXXX-----