**Assignment 1**

**Instructions**: Use Jupyter notebook to work on this assignment using Python language. All cell outputs and visualizations should be visible and necessary comments should be put to make the code readable. Once the code is ready, save it as a pdf too. Create a zip file containing the pdf and the .ipynb file obtained from Jupyter Notebook.

This assignment is on image classification problem using shallow supervised machine learning algorithms (**apply only logistic regression (LR) and/or artificial neural networks (ANN)**). The dataset consists of images obtained from pavements. The images can contain either raveling distress or no raveling distress. Your objective in this challenge is to develop a binary image classification model that can classify from a given image if it contains raveling distress or not. Your model performance is expected to depend on the chosen model (LR/ANN), model architecture (for ANN) and the features extracted from the images. To learn more about selection of features, refer to the following paper:

Hoang, N.-D. (2019). Automatic detection of asphalt pavement raveling using image texture based feature extraction and stochastic gradient descent logistic regression. Automation in Construction, 105(102843), 102843. <https://doi.org/10.1016/j.autcon.2019.102843>

Data The dataset consists of 2 folders: 'train' and 'test'. The folder 'train' is a labelled dataset and consists of 2 sub-folders: 'Raveling' and 'Non\_raveling'. The sub-folder 'Raveling' consists of a total of 350 images containing raveling distress, while the 'Non\_raveling' sub-folder contains 350 images with no raveling distress. All images are of the ‘.jpg’ format.

The ‘test’ folder consists of the test dataset images. There are a total of 10 images in this folder (all ‘.jpg’ files). These images are unlabelled i.e., the labels are not provided, whether they contain raveling distress or not. The objective of this challenge is to predict the class of each of these test images. The class can be either ‘Raveling’ or ‘Non\_raveling’. The filenames of these test images are given as 1.jpg, 2.jpg, 3.jpg, …, 10.jpg.

**Submission**

Format Submission files should be in .csv format and contain two columns: filename and class. ‘filename’ should be in string format and will contain the image filename (1.jpg, 2. jpg, etc.). The ‘class’ will also be in string format and contain the corresponding class name. It can take two values: 'Raveling' or 'Non\_raveling'. Since there are 10 images in the test dataset, your submission should have 10 rows (excluding the header), each with 2 columns. A sample submission file is also provided in the dataset. The file should contain a header and have the following format:

|  |  |
| --- | --- |
| filename | class |
| 1.jpg | Raveling |