**Final Report**

[**United Arab Emirates Chapter - AI-Based Road Inspection System | PROJECT MANAGEMENT FOLDER**](https://www.notion.so/omdenadocs/United-Arab-Emirates-Chapter-AI-Based-Road-Inspection-System-PROJECT-MANAGEMENT-FOLDER-db05e310b5a14075a40949f793180882)

**Task1: Data Collection and Preprocessing**

1. **Data Collection**

The required data was categorized into 4 classes:

**Crack Groove**

 

**Rut Subsidence**

 

Being the foundation step in the project, all participants were invited to collect the data, 15+ members endeavored to collect the data belonging to four or less classes. ‘The more, the better’ approach was adopted for this phase, as it was expected that a substantial amount could either be duplicates, irrelevant or insufficiently sized.

Many tools were used for data collection such as Google Search, Scrapping, and [Time Saver Images Collecting Manner (TSICM)](https://drive.google.com/file/d/1mQ-1vwckVcKYtLdDqexWaElSrl2B-D-j/view?usp=share_link).

A total of **33871** images were collected with following distribution. (it is to be noted the collection figures are prior to any duplicate(s) | invalid images removal.

[Table A.1 Collected Images Numbers](https://docs.google.com/spreadsheets/d/1verp-Dn8otQFT5E71vdnePcW2iua3VmERHeawk856nQ/edit#gid=0) Report

The enthusiasm is much appreciated by all the contributors and

1. Pre-Processing

Following data collection phase, the pre-processing of the collection was performed, and the team went through the following steps;

1. Checking for correct class, and corrections as required
2. Removal of duplicates by using various techniques
3. Cropping of images for undesired objects (e.g. vehicles, humans etc.)
4. Checking size and dropping undersized sized of images (< 300x300) as per acceptance criteria as set by the modeling team
5. Converting image format to .jpg
6. Data Exploration

The above steps involved creation of code to process the large amount of collected images, and usage of a variety of tools, and efforts spanning 2 weeks of hard work, the pre-processing was finalized and the finished dataset (images) were handed over to the modeling team in 2 tranches (Crack+Groove and Rut+Subsidence)

**The data exploration charts for all classes are as under:**

**Tools used: spreadsheets, ......**

Link to finished data:

<https://drive.google.com/drive/folders/1V04yMqTuOUdiD4Iid6IYLj964tffFiFh?usp=share_link>

Later on some additional images were added in Crack+Groove dataset by the modeling lead, and link to the same is as follows:

<https://drive.google.com/drive/folders/1V04yMqTuOUdiD4Iid6IYLj964tffFiFh?usp=share_link>