

# **Configuration Manual**

MSc Research Project Data Analytics

Abin Joseph Student ID: 21216312

School of Computing National College of Ireland

Supervisor:

Abdul Shahid

#### National College of Ireland



#### **MSc Project Submission Sheet**

	Scho	ol of	Com	puting
--	------	-------	-----	--------

Student Name:	Abin Joseph
Student ID:	21216312
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Abdul Shahid
Submission Due Date:	31/01/2024
Project Title:	Comparative Analysis between ResNet models on marine oil spill detection.
Word Count:	732
Page Count:	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Abin Joseph
Date:	31/01/2024

#### PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

## **Configuration Manual**

#### Abin Joseph Student ID: 21216312

### **1** Device Specifications

EDA and splitting dataset to create train and test data was done using Jupyter notebook with below specifications.

Processor	12th Gen Intel(R) Core (TM) i7-1260P 2.50
	GHz
Installed RAM	16.0 GB (15.6 GB usable)
System type	64-bit operating system, x64-based processor
Tool used	Jupyter Notebook

Table 1. Laptop specifications used for EDA.

Due to processing complexity preprocessing, data augmentation, model building and evaluation was done using Google Colaboratory.

GPU	A100 GPU with 50GB RAM,
System Ram	System Ram 85.5 GB
Disk Space	166.8 GB
Tool used	Google Colaboratory

Table 2. Device requirements for performing model building, training, validation and testing

### 2 Dataset

The SAR image dataset was downloaded from CSIRO data access portal (Blondeau-Patissier et al., 2022). In Figure 1, the folder structure downloaded dataset is given below. After extracting the zip file there will be two folders called data and Metatdata. The data folder contains two zip folder which should be extracted to get the respective class folders that is, 0 and 1. Since the file name is long when combined with folder path might give issues while running the code. So, the folders 0 and 1 are stored in the ICT code and Artifact folder/Dataset/Complete data. Folder structure of ICT code and artifacts are given below in

Figure 2. The images divided into class 0 and 1 in Complete data folder will be then used for EDA and to create train and test data. All the code files are stored in the Code folder and dataset is stored in Dataset folder. Also, the code folder contains pdf versions of the ipnyb files that was executed using Colaboratory.

Due to the complexity of the task, EDA and train test creation was done in EDA.ipnyb using jupyter. Preprocessing, augmentation, model building, training, and testing was done in Final\_ResNet50.ipnyb, Final\_ResNet50V2.ipnyb, and Final\_ResNet101.ipnyb using Colaboratory<sup>1</sup>. Each filename represents the model that is being trained in that respective code file. It was divided into three since the computational load was high and the memory limit was exceeding.

Blondeau-Patissier\_David\_30\_Sep\_2023.zip

o data

- S1SAR\_UnBalanced\_400by400\_Class\_0.zip
   0
- S1SAR\_UnBalanced\_400by400\_Class\_1.zip

```
• 1
```

Metadata

- collection\_import\_sha256sum.txt
- collection\_import\_sha512sum.txt
- Creative Commons Attribution-Share Alike.html
- dublincore-000057430v001.xml
- dublincore-collection.xml
- readme-help.md

#### Figure 1 Dataset structure

After running the EDA.ipnyb file the resulting training and test data will be saved in the tf\_train and tf\_test folder. This was uploaded to Google Drive<sup>2</sup>. When tried to split and save the dataset manually using colab, it gave unknown errors which led me to use the above method.

#### ICT code and artefacts

Dataset

Complete data
0
1
o tf\_test

<sup>&</sup>lt;sup>1</sup> <u>Welcome to Colaboratory - Colaboratory (google.com)</u>

<sup>&</sup>lt;sup>2</sup> Home - Google Drive

- o tf\_train
- Code
  - o EDA.ipnyb
  - Final\_ResNet50.ipnyb
  - Final\_ResNet50V2.ipnyb
  - Final\_ResNet101.ipnyb
  - Final\_ResNet50.ipnyb Colaboratory.pdf
  - Final\_ResNet50V2.ipnyb Colaboratory.pdf
  - Final\_ResNet101.ipnyb Colaboratory.pdf
- Readme.txt

Figure 2. ICT code and artefacts folder structure

### 3 Model building, training, validation, and testing

After uploading the the tf\_train and tf\_test dataset folders to the Google Drive, upload the respective code files, Final\_ResNet50.ipnyb, Final\_ResNet50V2.ipnyb, and Final\_ResNet101.ipnyb to Colaboratory. Update the respective path for the training\_samples and test\_samples variables to Google Drive folder location for tf\_train and tf\_test in Cell number 3 for all the above-mentioned code files that is, Final\_ResNet50.ipnyb, Final\_ResNet50V2.ipnyb, and Final\_ResNet50V2.ipnyb, and Final\_ResNet50V2.ipnyb, and Final\_ResNet50V2.ipnyb, Final\_ResNet50V2.ipnyb, Final\_ResNet50V2.ipnyb, and Final\_ResNet50V2.ipnyb, Final\_ResNet50V2.i

Also please update the checkpoint path given in Cell number 9 for Final\_ResNet50.ipnyb, Final\_ResNet50V2.ipnyb, and Final\_ResNet101.ipnyb to save the model checkpoint.

Before running the code ensure that below libraries are installed since these are the libraries which is used for model building, training, validation, testing and visualizations.

Table 3. Libraries to be installed.

### **4 Points to remember.**

Below mentioned are important points to remember;

- The EDA code file can be directly run after opening it in the Jupyter notebook since the path is set to be relative.
- Before running the code files on Google Colaboratory, ensure to update the path for training\_samples, test\_files with the respective drive location to tf\_train and tf\_test dataset.

• Also update the checkpoint path to save the best weights of the model during training.

By following the given steps, it would be possible to run the code with out any issues.

### References

Blondeau-Patissier, David; Schroeder, Thomas; Diakogiannis, Foivos; Li, Zhibin (2022): CSIRO Sentinel-1 SAR image dataset of oil- and non-oil features for machine learning ( Deep Learning ). v1. CSIRO. Data Collection. https://doi.org/10.25919/4v55-dn16