National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Ramit Dour
Student ID: x23102764

School of Computing
National College of Ireland

Supervisor: Dr. Anderson Simiscuka

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ramit Dour
Student ID:	x23102764
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr. Anderson Simiscuka
Submission Due Date:	12/08/2024
Project Title:	Configuration Manual
Word Count:	550
Page Count:	9

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.

ALL 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:	
Date:	12th August 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

Ramit Dour
x23102764

1 Section 1: Introduction

This configuration manual provides steps which are required to replicate the research project on any system.

Upload the notebook code artifacts file to Google Drive's **'Colab Notebooks'** Folder.

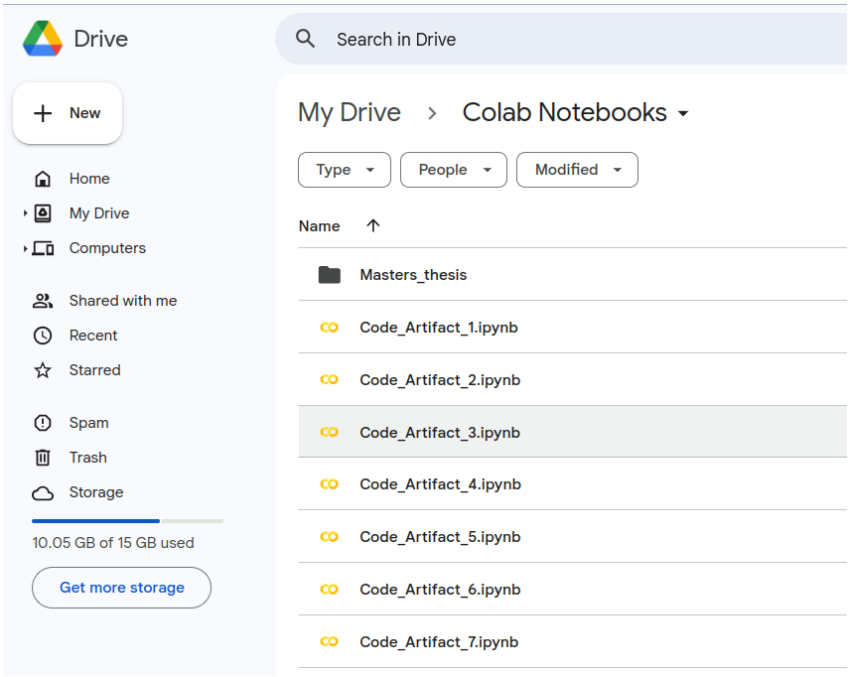

Figure 1: Upload Code Notebooks to Google Drive

2 Section 2: System Configuration

After opening the notebook Go to the 'Runtime' in the top menu.

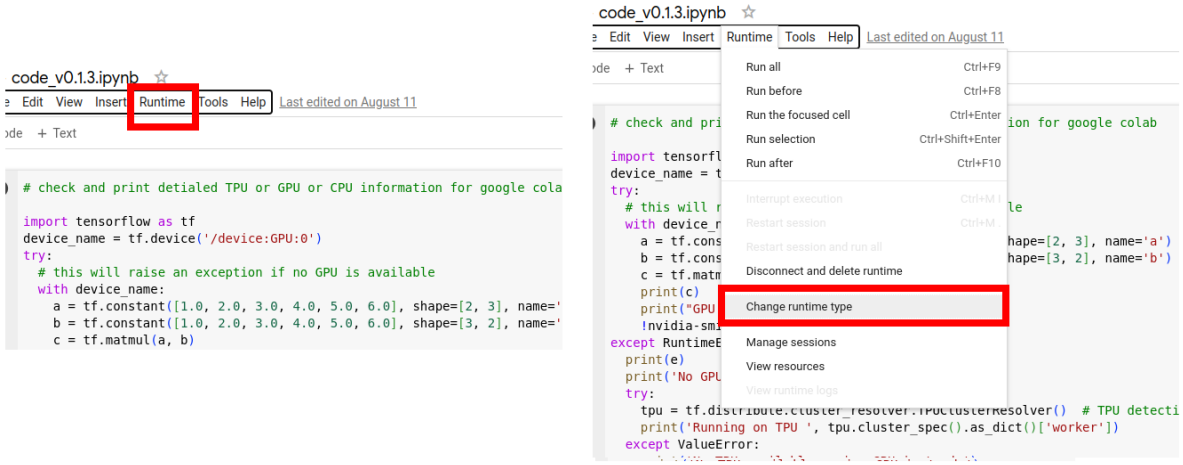

Figure 2: Runtime Select

Select 'Change Runtime Type'

Make sure the Runtime Type is Python3 and the Hardware accelerator is selected as **A100 GPU**

Then in the right upper corner, there is a button named 'Reconnect' click on that to connect to the runtime.

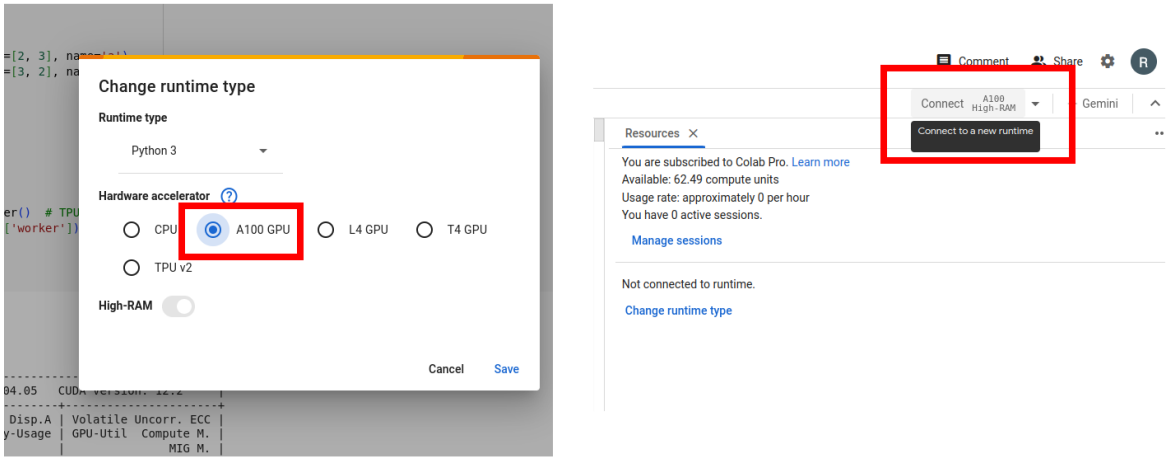

Figure 3: GPU Select

Once Connected Run and execute the first line of the code to Check the allocated resources. Sometimes where there is high traffic then Google Colab might automatically allocate a Lower Grade Graphic Card.

As this project requires image processing which consumes a high level of computational resources its always better to select the best available GPU for training purposes

for batch processing.

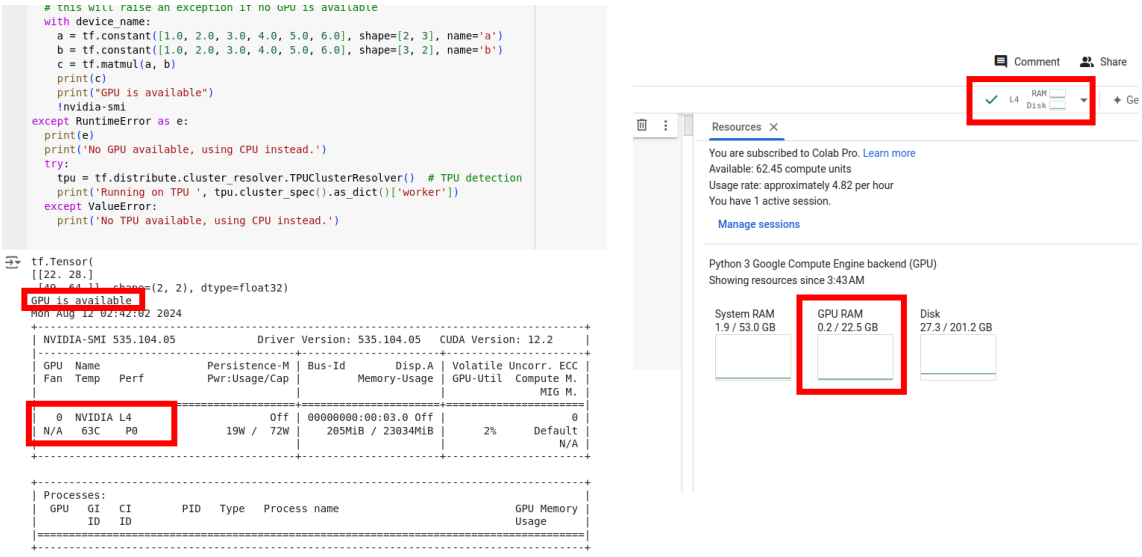

Figure 4: GPU Verification

3 Section 3: Data Collection

Create a folder named **'masters_thesis'** in root of Google Drive. Inside the folder upload the zip files of the all 4 data sets. Database NEU , Database Severstal and KolektorSDD2/KSDD2.

Make sure the code artifacts of all the 9 experiments are uploaded fully. Sometimes while uploading the file is not uploaded fully.

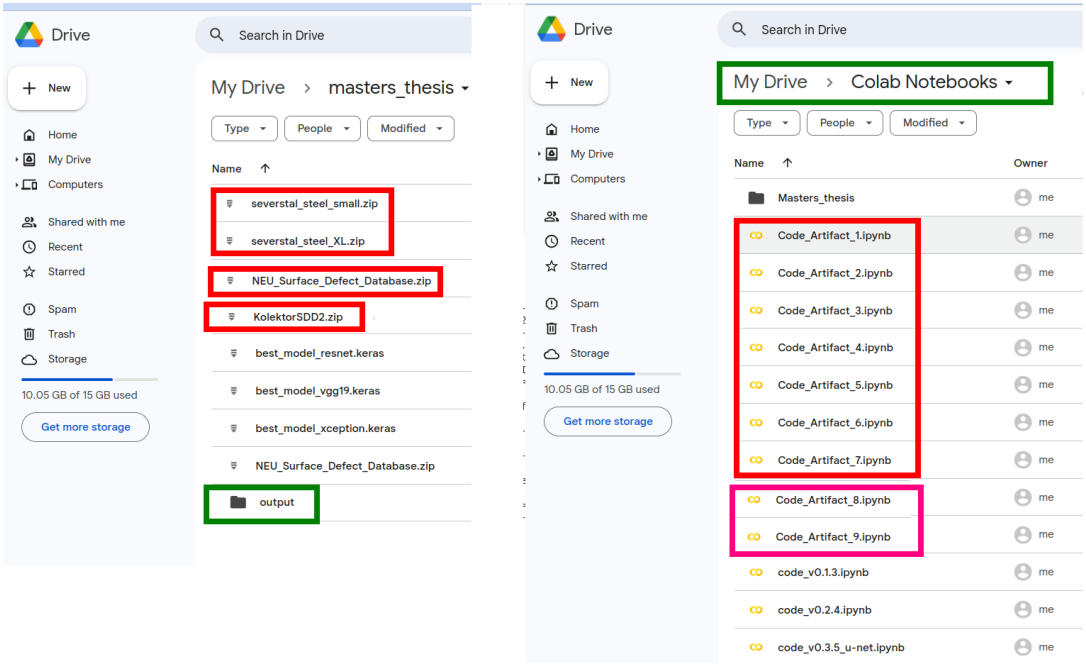

Figure 5: Upload Code Notebooks to Google Drive and Data sets to Drive

The dataset and notebooks can be downloaded from given Drive Link[1]

1. NEUSurfaceDefectDatabase [2]

2. KolektorSDD2 [3]

3. SeverstalSteelDefectDetection [4]

[1]https://drive.google.com/drive/folders/1G7GDF0A4PqYNw-jUXRw43Fph727zvX56?usp=sharing
[2]NEUSurfaceDefectDatabase Link Dikshit (2023)
[3]KolektorSDD2 Link ViCoS Laboratory (2023)
[4]SeverstalSteelDefectDetection Link *Severstal: Steel Defect Detection* (2019)

4 Section 4: Data Preparation and Transformation

Make your when your run the code it is connected to the google drive and the drive is mounted.

Allow the necessary permission and select the account to which the data sets have been uploaded. Once mounted the names of zip files are shown in the output terminal.

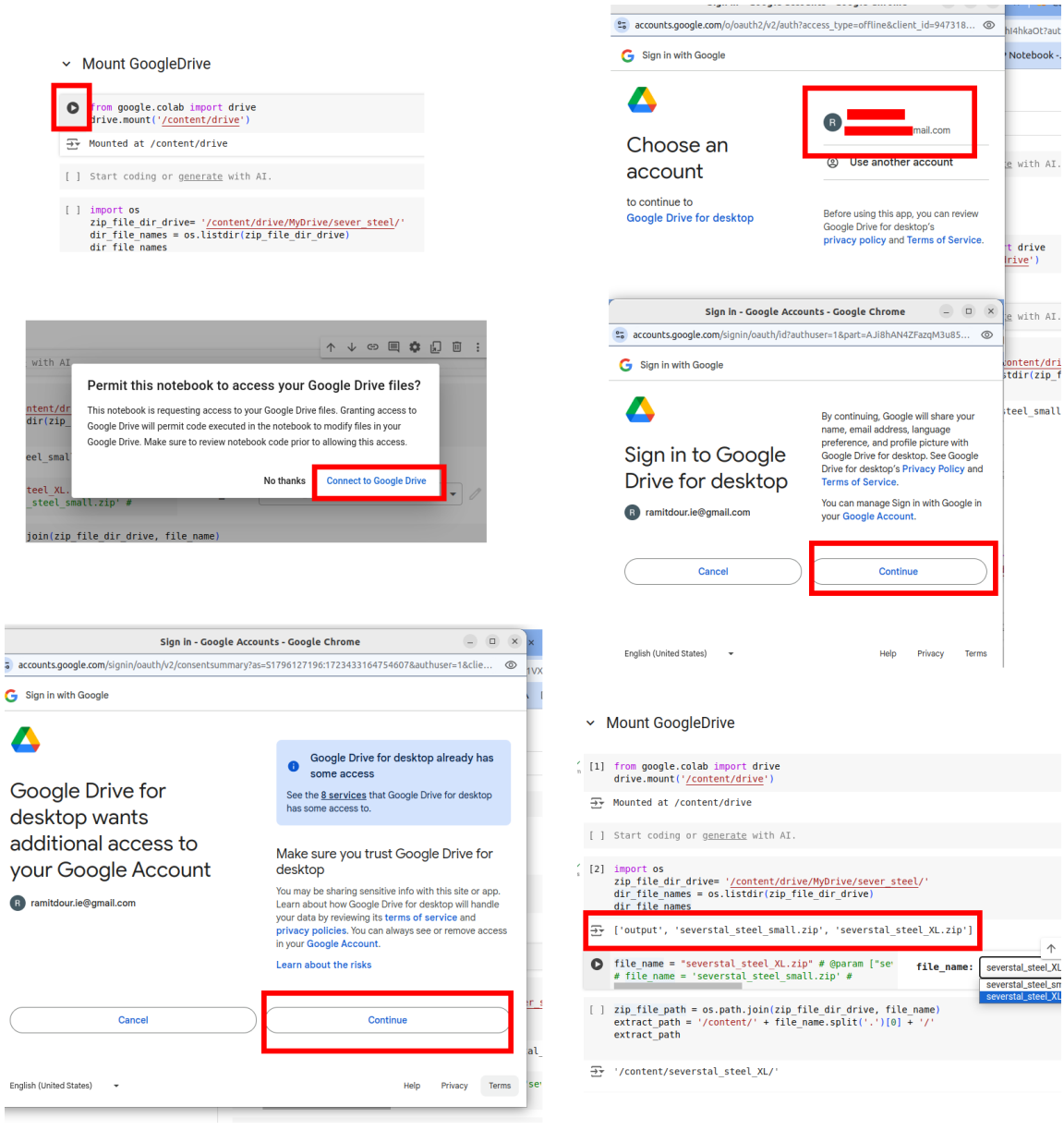

Figure 6: Google Drive Connection

5 Section 5: Model Building/Loading

5.1 Model Building

Select the required split ratio, batch size, and epoch for training of the model by adjusting the sliders or manually putting the numbers.

Figure 7: Hyper parameter tuning and Model Fitting

To start training the model run cell Fit and with for few minutes, the training of first epoch takes time . Once training starts the monitor metrics will start showing up in the output terminal.

5.2 Model Loading

In order to reuse the model for future predictions or deployment it need to be saved. Once the model is traind it can be resued by loading the model weights.

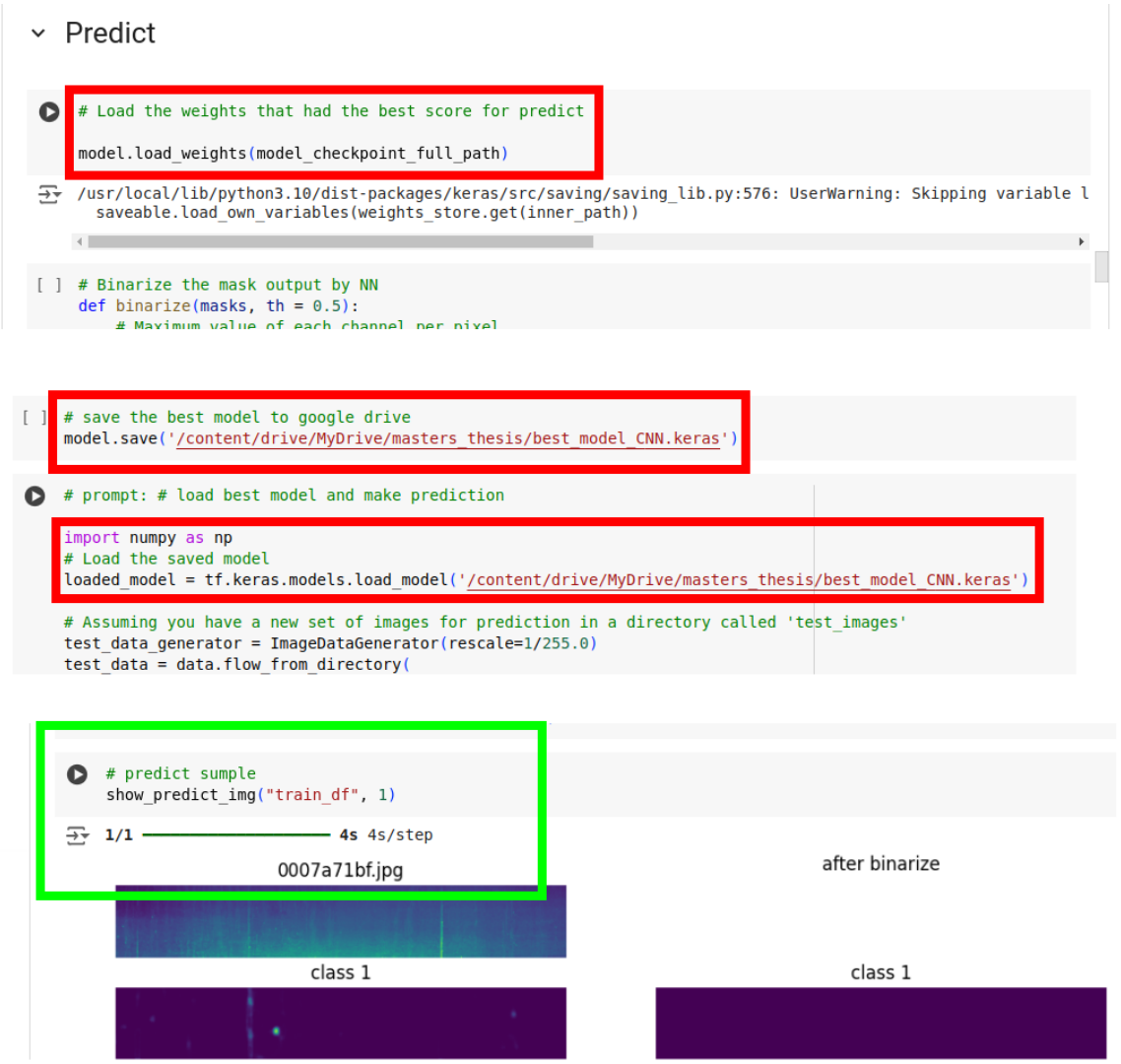

Figure 8: Model Saving and Model Loading

6 Section 6: Model Evaluations

Running the evaluation cells will give the desired outputs for the evaluation of each model.Values life accuracy precision F1-Score ,etc will be printed in the output terminal.

Plots like Confusion Matrix and the ROC curves will also be printed.

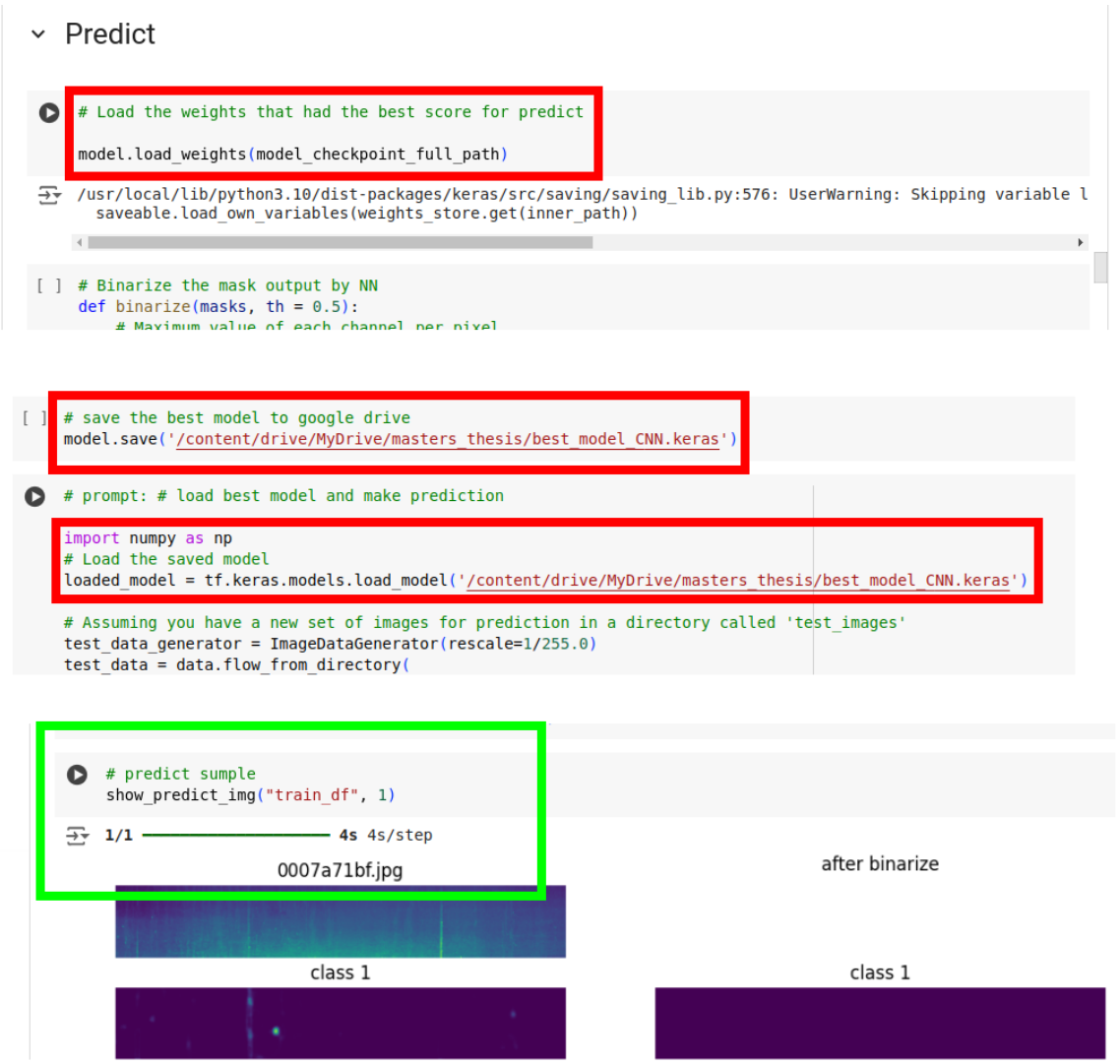

Figure 9: Model Saving and Model Loading

7 Section 7 : Packages and Libraries Used

Altough the installation of required packages are automatically done in Google Colab Notebooks. But for the reference the packages which are used in this project are as below :

- Pandas
- Matplotlib
- NumPy
- Glob
- Seaborn
- Scikit Learn
- TensorFlow
- Pytorch
- Meta-SAM
- Random
- Pandas

References

Dikshit, K. (2023). Neu surface defect database. Accessed: 2024-08-12.
URL: *https://www.kaggle.com/datasets/kaustubhdikshit/neu-surface-defect-database*

Severstal: Steel Defect Detection (2019). Accessed: 2024-08-12.
URL: *https://www.kaggle.com/c/severstal-steel-defect-detection/data*

ViCoS Laboratory, U. o. L. (2023). Kolektor surface-defect dataset 2 (kolektorsdd2/ksdd2). Accessed: 2024-08-12.
URL: *https://www.vicos.si/resources/kolektorsdd2/*