Additional Data:

# Data Motivation:

The reason for potentially introducing a new dataset into our project is the number of classes in our current working model is rather scarce. The number of classes which our model is able to detect currently is 4. These being:

1. Traffic lights
2. Speed Limits
3. Road Crossing
4. Stop Signs

This current classification set is good and works well but only offers a niche solution. The additional included dataset would ideally cover a wider range of signs to allow the program to be used more widely. The speed limit classification is also only able to check if a sign is a speed limit sign but not the speed within the sign itself.

The purpose of additional data would be to cover more road sign topics for example thinks such as hazard purposed signs, as the sign options are only for very limited options which are unable to retrieve much road information from the images/videos.

# Data Preparation:

For the purpose of speed and time the dataset which was used was originally set up for use on a YOLOv3 object detection framework. But the dataset was utilised for use within the YOLOv5 object detection framework. The dataset was repurposed and set up through the use of Roboflow, this split the data into a training set, of 518 images, a validation set, of 148 images and finally a testing set of 74 images. The split was 70, 20, 10 (%) respectively. The classes which the dataset was able to detect were made up of 4:

1. Prohibitor
2. Danger
3. Mandatory
4. Other

This dataset focused on the generalised purpose of the road signs rather than the specific instruction which they present. The datasets link is: <https://www.kaggle.com/valentynsichkar/traffic-signs-dataset-in-yolo-format>  
The dataset was also not integrated into the current model but utilised to produce a new model for the new classes to save time on training and reproducing the same model but modified.

This dataset is suitable for the purpose of our project as it allows the model to be more generalised to be able to detect a larger plethora of road signs without the need of huge amounts of data for training and testing. This dataset allows the generalised format of our project to be explored. This opens up great opportunity of discovery, as if the dataset was integrated with the current model, we can find out if it would be able to detect a larger amount of road signs with great confidence.

The current User Interface which Ben Sisk is building should allow the user to define the location of the model in which they wish to detect road signs in their specified image/video. For this purpose, the creation of a second model with just the new dataset is not an issue for comparison and to comprehend the impact of the change as the models are able to be interchanged with ease.

The new models code etc which was used to create it was used through the YOLOv5 git hub repository and using the roboflow step by step tutorial. This was used to help with the speed and the building process. It also allows the code to retrieve diagrams which are able to output a tensor board off of different metrics such as the model’s precision etc. This code will be within the model folder which is in the Additional Data folder.

# Impact Report:

Due to endless user caused and system-based errors which I encountered the data received was formulated from one model created which utilised a batch size of 16, 100 epochs (before it failed) and an image size of 416. This data displayed the capabilities which the current model could utilise to improve as the general results were clear that many more road signs are recognized in images which include more than just the four classes which are used in the original dataset.

The introduction of the additional dataset displays the future capabilities of our system, and where it could improve. The new data would allow for the current model to be able to identify more than just four specific classes (Traffic Lights, Road Crossings, Stop Signs and Speed Limit), it would be able to identify a much larger majority of signs but in a more generalised format. If the integration were done correctly, it would allow the identification of the original classes which wouldn’t have a large confidence score to be generalised into one of the four new classes to give it some form of detection.

The graphs shown below are based on this partially failed model:  
Chart, line chart

Description automatically generated  
Chart, line chart

Description automatically generated

Finally, the files which are for the new model are not the final form due to last errors which were unable to be corrected or re-modelled in time for the hand in. They have still been included for sake of reference and marking use if needed. Access to this file can be found through <https://github.com/EdDavies1/EdDavies1.git> this link.