MARMARA UNIVERSITY

FACULTY OF ENGINEERING

CSE4288

Introduction to Machine Learning

TERM PROJECT

Model Evaluation Report

Group: 2

Table of Contents

1. Evaluation Metrics and Results 3

1.1 Decision Tree 3

1.2 Naïve Bayes 4

1.3 K-NN 5

1.4 Logistic Regression 6

1.5 YOLO 7

2. Comparison Between Models 9

3. Optimization Steps Taken 10

1. Evaluation Metrics and Results

1.1 Decision Tree

Figure 1 gives the confusion matrix of our model with Decision Tree classification. In this case, the image is classified as either a crosswalk exists or not.

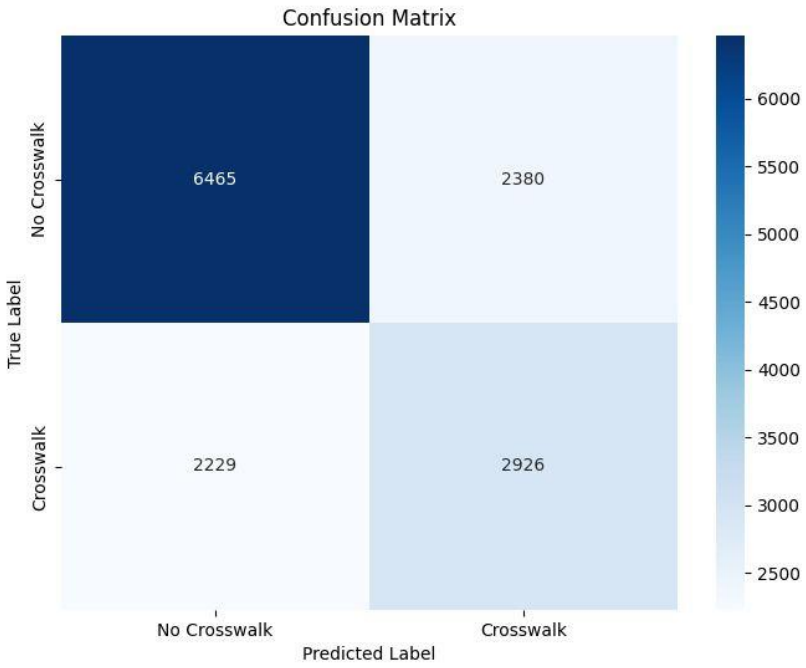

Figure 1: Confusion matrix of Decision Tree

Table 1 gives the evaluation results of our model with Decision Tree classification.

Table 1: Evaluation results of Decision Tree

	Precision	**Recall**	**F1-Score**	**Support**
No Crosswalk	0,74	0,73	0,74	8845
Crosswalk	0,55	0,57	0,56	5155
Macro avg.	0,65	0,65	0,65	14000
Weighted avg.	0,67	0,67	0,67	1400

Accuracy of the model with Decision Tree classification is 67,08%.

1.2 Naïve Bayes

Figure 2 gives the confusion matrix of our model with Naïve Bayes classification. In this case, the image is classified as either a crosswalk exists or not.

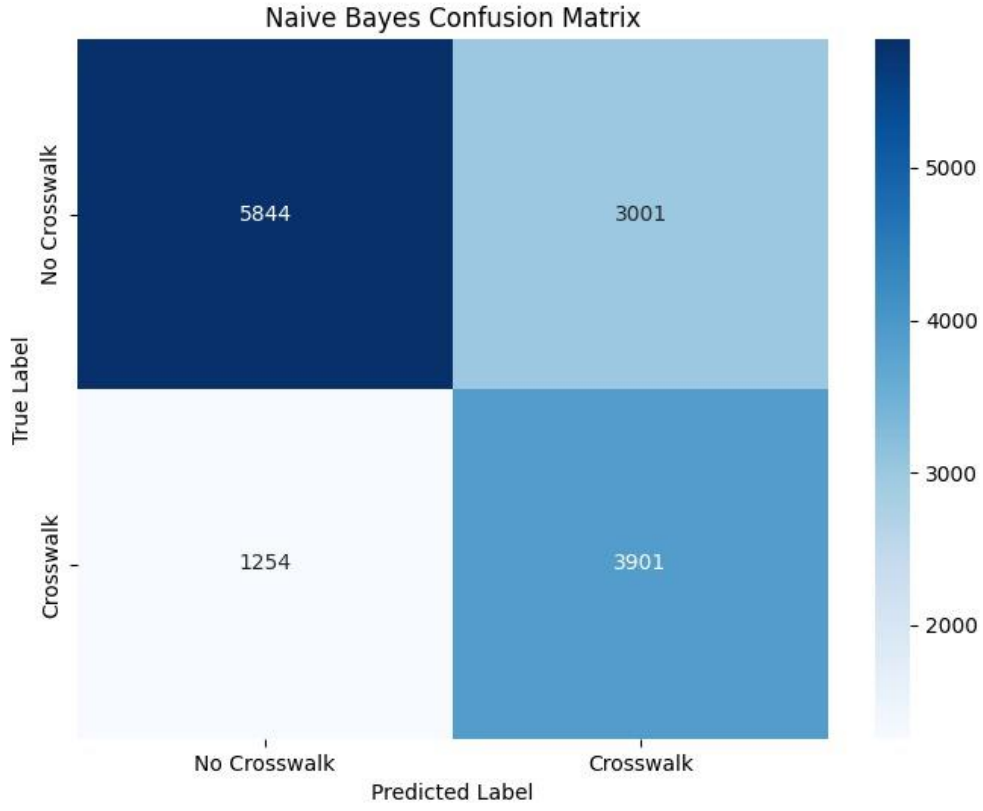

Figure 2: Confusion matrix of Naïve Bayes

Table 2 gives the evaluation results of our model with Naïve Bayes classification.

Table 2: Evaluation results of Naïve Bayes

	Precision	**Recall**	**F1-Score**	**Support**
No Crosswalk	0,80	0,66	0,73	8845
Crosswalk	0,57	0,76	0,65	5155
Macro avg.	0,69	0,71	0,69	14000
Weighted avg.	0,73	0,70	0,70	1400

Accuracy of the model with Naïve Bayes classification is 69,61%.

1.3 K-NN

Figure 3 gives the confusion matrix of our model with k-NN classification. In this case, the image is classified as either a crosswalk exists or not.

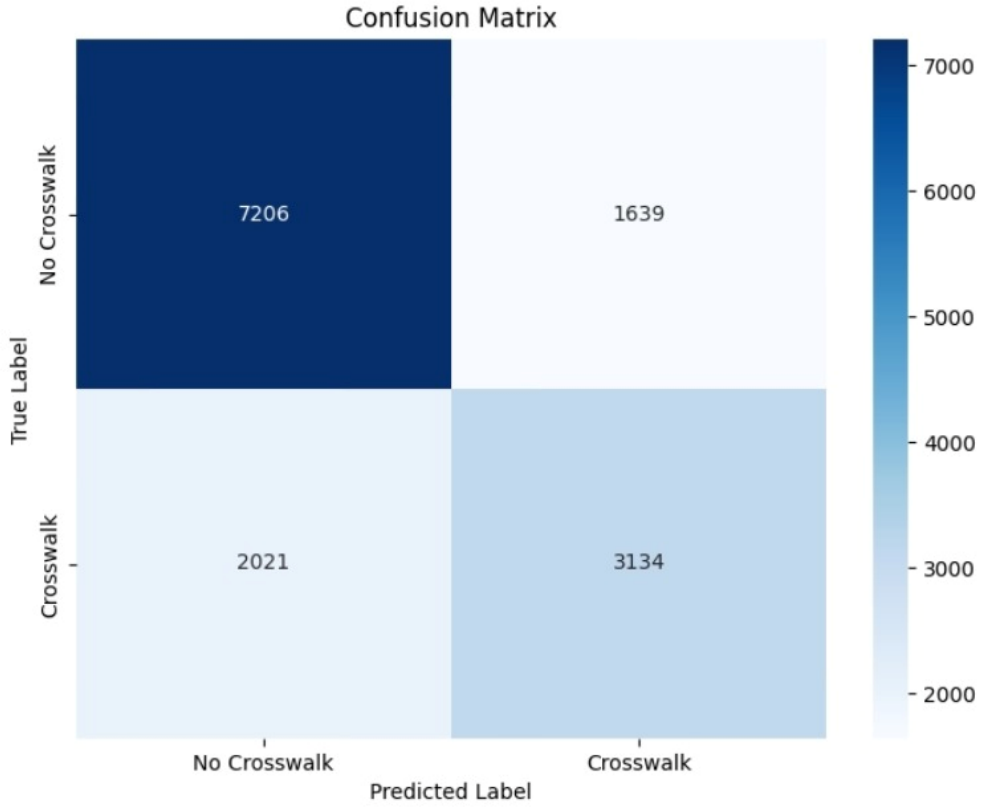

Figure 3: Confusion matrix of k-NN

Table 3 gives the evaluation results of our model with k-NN classification.

Table 3: Evaluation results of k-NN

	Precision	**Recall**	**F1-Score**	**Support**
No Crosswalk	0,78	0,81	0,80	8845
Crosswalk	0,66	0,61	0,63	5155
Macro avg.	0,72	0,71	0,71	14000
Weighted avg.	0,74	0,74	0,74	1400

Accuracy of the model with k-NN classification is 73,86%.

1.4 Logistic Regression

Figure 4 gives the confusion matrix of our model with Logistic Regression classification. In this case, the image is classified as either a crosswalk exists or not.

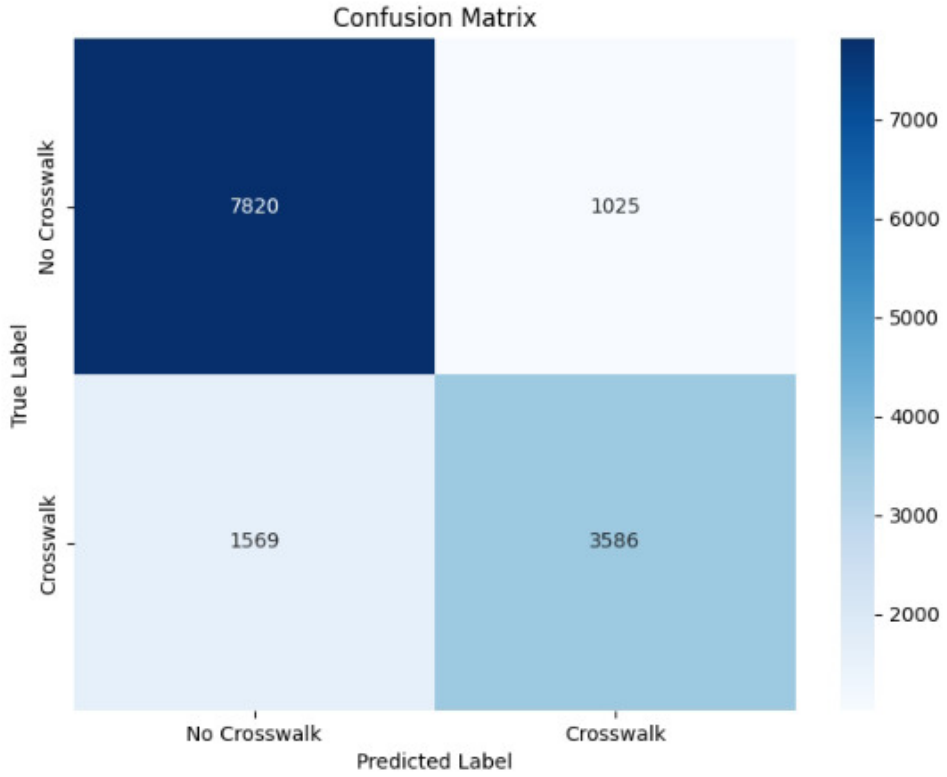

Figure 4: Confusion matrix of Logistic Regression

Table 4 gives the evaluation results of our model with Logistic Regression classification.

Table 4: Evaluation results of Logistic Regression

	Precision	**Recall**	**F1-Score**	**Support**
No Crosswalk	0,83	0,88	0,86	8845
Crosswalk	0,78	0,70	0,73	5155
Macro avg.	0,81	0,79	0,80	14000
Weighted avg.	0,81	0,81	0,81	1400

Accuracy of the model with Logistic Regression classification is 81,45%.

1.5 YOLO

Figure 5 indicates the confusion matrix of our YOLO model. Crosswalk is the area of the image that either has a crosswalk or a crosswalk is detected. Background means either the image has no crosswalk or no crosswalk is detected, meaning the part of the image is detected as background.

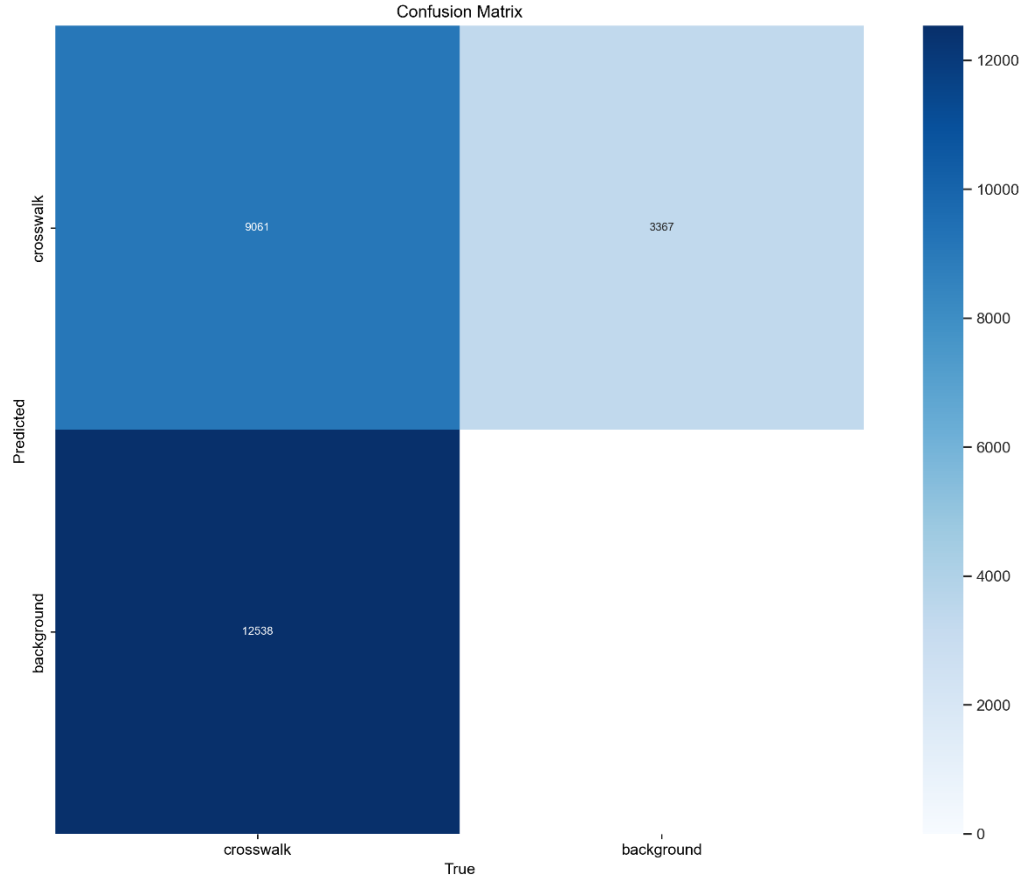

Figure 5: Confusion matrix of YOLO

Figure 6 indicates the confidence curve of the YOLO model. Currently the confidence level is quite low with a value of 0,5. This means that the model needs further optimization. The optimization steps are stated in Section 3.

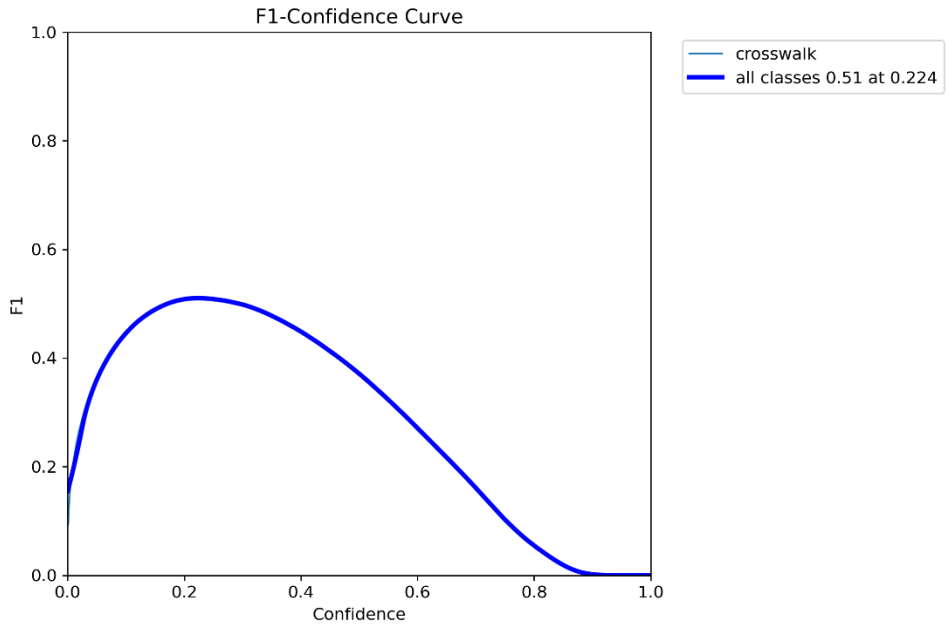

Figure 6: Confidence curve of the YOLO model

Figure 7 indicates the training result of the YOLO model. These results are the bounding boxes of YOLO and how it is applied to our dataset.

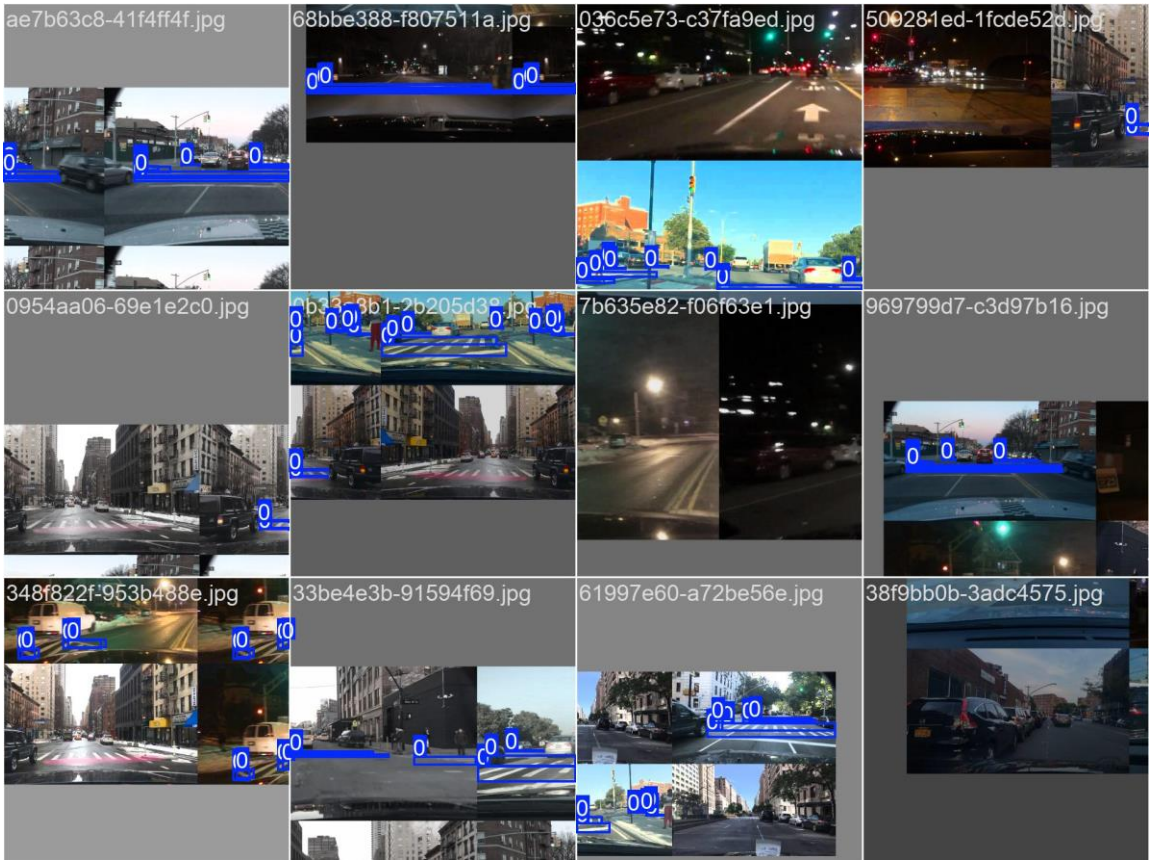

Figure 7: Training result of YOLO model

2. Comparison Between Models

In this section the models are grouped as Decision Tree, k-NN, Naïve Bayes and Logistic Regression as the first group; and YOLO alone as the second group. The first group of models determines whether the image includes a crosswalk or not. However, the YOLO model tries to find where the crosswalk is if there exists any. Therefore, the first group is classification, and the second group is detection. Among the classification models, the Logistic Regression performed better than other models with the highest accuracy of 81,45%. The models can be ordered from best performance to lower performance as stated in Table 5.

Table 5: Classifications model comparison by accuracy

Classification Model	**Accuracy**
Logistic Regression	81,45%
k-NN	73,86%
Naïve Bayes	69,61%
Decision Tree	67,08%

Currently due to several reasons stated in Section 3, although a crosswalk can be identified, the boundaries may not be set properly or the number of crosswalks in an image cannot be calculated perfectly. In Figure 8, a test result of our YOLO model is given.

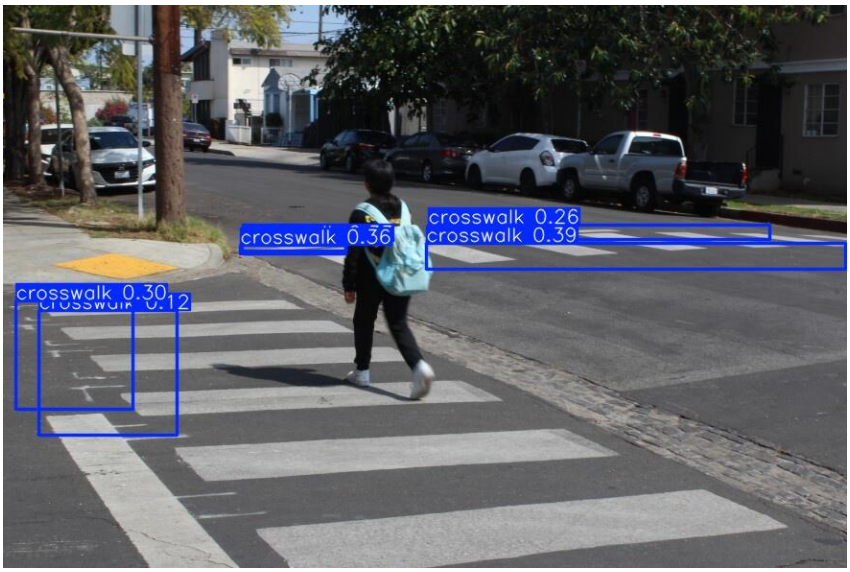

Figure 8: Test result of YOLO model

Although the crosswalks are not identified accurately and the confidence is quite low for setting the boundaries, YOLO model has a great advantage when compared to classification algorithms, since YOLO not only classifies but also detects the object location in an image.

3. Optimization Steps Taken

For the optimization, there were not much thing to do about the implementation. The main challenge was the variety of the dataset that we are using. Although we have a great dataset with many photos, some images can cause the model to act poorly. Some characteristics of the poor images are:

- They can be quite blurry.
- The crosswalks may be on corners, quite far away from the camera. They can be barely seen even with human eye.
- The images can be taken on rainy days, resulting in waterdrops on camera.

Despite believing that these images are the key to having a comprehensive model that detects the crosswalks nearly perfect, they can also be the reason of dataset uncleanliness.

The fact that these poor images cannot be detected in software level, we decided to clean the data by ourselves, determining the images by hand and deleting them from our dataset of crosswalk images.

Another optimization step was to upgrade the YOLO model from 8 to 11. YOLOv11 performs better in terms of time efficiency. Since we had problems with long training time, it was better to upgrade our model to a higher version.