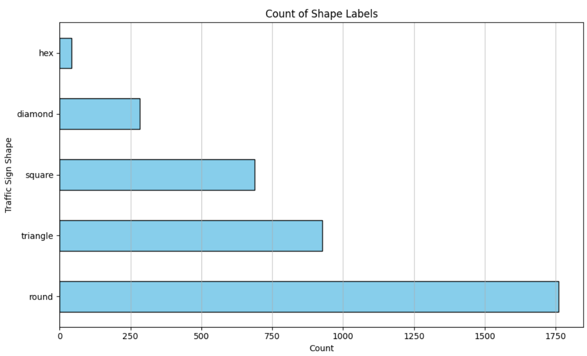
Assignment 2 - Machine Learning Project

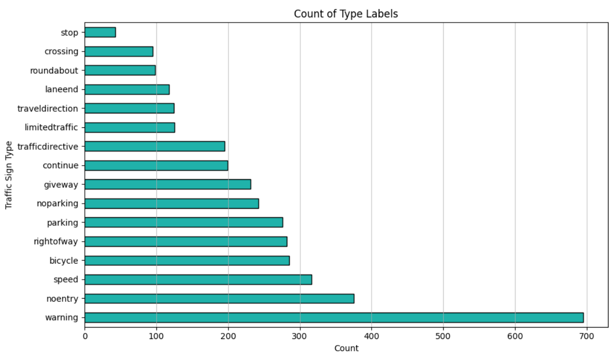
Student ID	Contribution
s3934581	**50%**
s3954428	**50%**

Literature Review:
The problem involves classifying traffic signs from the Belgium Traffic Sign dataset, which includes five unique shapes and fifteen distinct signs. After reviewing several research papers, "Classification of Traffic Signs: The European Dataset, 2018" was identified as the most relevant study. This research successfully classified 164 different traffic signs using a Convolutional Neural Network (CNN), achieving an accuracy of 98.52%. Given the similarity to the current classification task and the high accuracy reported, a CNN-based approach has been selected for the model.

Target:
Based on what we have learned from previous studies, our goal for this project is to train the model to achieve an accuracy of over 98%.

EDA:
The dataset comprises a relatively small sample size of 3699 entries, and the images are of low quality, being grayscale with a resolution of 28x28 pixels. A thorough exploratory data analysis revealed a significant class imbalance, with vast differences between the majority and minority classes in both sign and shape labels.





Multi-output: Given the computational demands associated with training and classification tasks, a unified approach is decided. Rather than developing individual models for each output label, we have chosen to employ a single model capable of concurrently predicting both output labels. This strategy enhances efficiency and optimizes resource allocation throughout the training and classification processes, particularly advantageous as the dataset remains consistent for predicting both output labels.

Data Preparation:

Split: The dataset is divided into an 80-20 ratio for balanced model training and validation. This split ensures ample data for learning patterns while reserving 20% for unbiased model evaluation, preventing overfitting and assessing generalization ability.

Custom Data-loader: To fulfil our multi-output model goal, a custom data loader is developed tailored to the model's requirements. This loader accommodates two input labels alongside image inputs, aligning with the model's architecture.

Data-preprocessing: Labels are assigned numerical values for processing to facilitate efficient handling by the neural network. To address the small image size, data augmentation techniques such as rotation, cropping, and padding are applied during the loading of training data. The batch size is set at 24 to balance computational efficiency and effective learning. Data shuffling is implemented during loading to guarantee batch diversity, ensuring that no single class label predominates within a batch. Printing the labels and images serves to verify the integrity of the pre-processing steps and ensure that the dataset is correctly loaded for training the Model.

Evaluation Metrics:
For the classification task on the dataset, the f1-Macro has been chosen as the primary evaluation metric due to the dataset's imbalance. This metric offers several advantages in this scenario: it provides a balanced assessment of model performance across all classes, incorporating both precision and recall. Precision and accuracy are utilized as secondary metrics for continual assessment. These metrics complement the f1-Macro by offering insights into specific aspects of the model's performance.

Building CNN Model:

Model: The baseline model comprises convolutional and pooling layers, followed by a flattening layer and dense layers for feature extraction and classification. Convolutional layers (Conv2D) with ReLU activation, employed to introduce non-linearity, capture spatial features from input images, while max pooling layers (MaxPooling2D) reduce feature map dimensions, retaining essential features. The flattening layer (Flatten) converts multidimensional feature maps into a one-dimensional vector format, facilitating input to dense layers. Dense layers (Dense) process features for shape and type classification, utilizing SoftMax activation for probabilistic outputs.

Compiler: The Adam optimizer is selected for its efficiency in training deep neural networks, utilizing adaptive learning rates and momentum-based updates. A learning rate of 0.003 is chosen to balance speed and stability, aiding effective convergence. Categorical cross-entropy is used for its effectiveness in multi-class classification, penalizing model predictions based on disparities between predicted and true distributions.

Class weights: To deal with the data imbalance, we have calculated class weights for each class for both label types. This will help penalize the labels with high entries.

Training the Model:
The baseline model provides a fundamental architecture against which we can benchmark subsequent, more complex models. By establishing a straightforward Convolutional Neural Network (CNN) as our baseline, we can better understand the strengths and limitations of our initial approach and identify areas for potential enhancement.

The baseline model training over 32 epochs showed initial improvement in loss and F1 scores, with training and validation losses decreasing significantly in the early epochs. By the 10th epoch, the model exhibited strong performance with training F1 scores for shape and type outputs around 0.98 and validation F1 scores near 0.96.
However, from the 18th epoch onwards, there were signs of overfitting, as evidenced by increasing validation losses despite continued improvement in training metrics. By the 32nd epoch, the model achieved high F1 scores for both training and validation sets but the growing gap between the training and validation losses indicated slight overfitting. This overfitting indicated that while the model was learning well on the training data, it was not generalizing as effectively to the validation data.
To address this, Model V2 introduced dropout layers with a 10% dropout rate after each convolutional block and the final dense layer. Dropout is a regularization technique that randomly drops units and their connections during training, which helps to prevent the network from becoming too finely tuned to the training data and thus improves its ability to generalize to new data.
Adding the dropout layer and training the model has improved the stability of the loss curve (Appendix 2) and mitigated some overfitting, but some remains.

Comparison with EfficientNetBo:
Evaluating the model with a detailed classification report provides better insights into its performance. For Shape classification, the model achieved an accuracy of 99%, demonstrating high precision, recall, and F1-score across all classes. Notably, classes like rightofway, stop, and noentry achieved perfect precision, recall, and F1-score, indicating robust performance. Similarly, for Type Codes, the model achieved a perfect accuracy score of 100%, with consistent high precision, recall, and F1-score across all classes. These impressive results validate the model's efficacy in accurately classifying traffic signs based on their shape and type characteristics. Given the high-performance metrics and perfect accuracy scores observed in the classification report, the model highlights its reliability and suitability for real-world deployment in traffic management and safety systems. With its capability to identify a diverse range of traffic signs with precision and confidence, the model has the potential to significantly enhance traffic safety and efficiency on roadways.

Classification Report of test data:
Evaluating the model with a detailed classification report provides better insights into its performance. For Shape classification, the model achieved an accuracy of 99%, demonstrating high precision, recall, and F1-score across all classes. Notably, classes like rightofway, stop, and noentry achieved perfect precision, recall, and F1-score, indicating robust performance. Similarly, for Type Codes, the model achieved a perfect accuracy score of 100%, with consistent high precision, recall, and F1-score across all classes. These impressive results validate the model's efficacy in accurately classifying traffic signs based on

their shape and type characteristics. Given the high-performance metrics and perfect accuracy scores observed in the classification report, the model showcases its reliability and suitability for real-world deployment in traffic management and safety systems. With its capability to identify a diverse range of traffic signs with precision and confidence, the model has the potential to significantly enhance traffic safety and efficiency on roadways.

Confusion Matrix of test data:
The confusion matrix complements the classification report, reaffirming the high performance of our model. It reveals minimal to no misclassifications for both shape and sign class labels. This alignment between the confusion matrix and the classification report underscores the model's accuracy and reliability in correctly identifying traffic signs based on their shape and type characteristics. The absence of significant misclassifications further validates the effectiveness of our model in accurately classifying a diverse range of traffic signs, contributing to improved traffic management and safety systems. (Appendix Figure 3)

Evaluation with Independent Test data

Source:
We have made the independent through the help of few kind peers in Europe who clicked few traffic images and sent, so that we are doing the independent testing on the same type of data we trained on. We have included 12 such images. The rest images were obtained from Classification of Traffic Signs: The European Dataset. We manually assigned the images to their respective classes to build the test data.

Classification Report:
The classification report of the independent data demonstrates a prominent level of performance by our model, with notable F1 scores. For the sign types, the weighted average F1 score is 0.82, indicating strong overall performance. Despite challenges in certain classes such as "bicycle" and "trafficdirective," the model achieves high precision and recall for classes like "noparking" and "stop." Similarly, for sign shapes, the model achieves an impressive weighted average F1 score of 0.98, demonstrating excellent performance across all shape categories. Notably, the F1 score for the minority class "hex" shape is 1.00, indicating perfect precision and recall. Overall, with F1 scores of 0.82 for sign types and 0.98 for sign shapes, our model demonstrates impressive performance on unseen data. This reinforces the model's ability to generalize well to new instances, with an elevated level of accuracy and reliability in classifying traffic signs based on their shape and type characteristics.

Confusion Matrix
The confusion matrix confirms the model's strong performance, with very few misclassifications. The most significant one is only 9 notable misclassifications where "traveldirection" was misclassified as "bicycle." Apart from this, there are minimal misclassifications in shape or type labels, demonstrating the model's precision and reliability in traffic sign classification. (Appendix Figure 4)

Ultimate Judgement:

The conducted study has culminated in the development of a prominent traffic sign classification model. Through meticulous analysis, training, and evaluation, the model has achieved the desired level of accuracy, demonstrating proficiency in accurately identifying traffic signs of several types and shapes. its performance on unseen real-world data underscores its reliability and applicability in practical scenarios. With results comparable to professional models, the proposed model stands as a viable solution for traffic sign classification tasks.

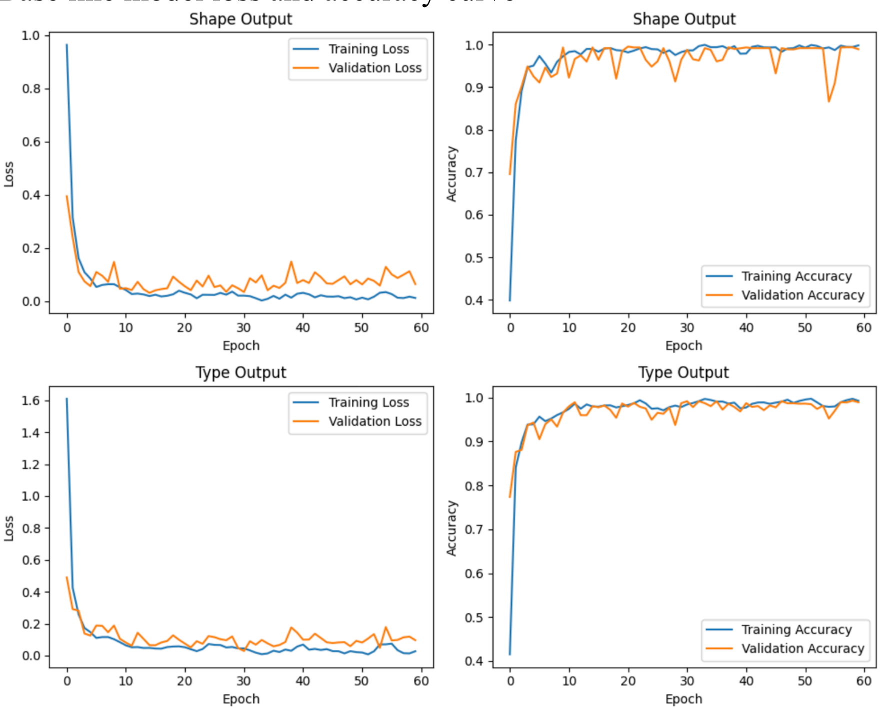
Future Improvements:
To enhance the model's accuracy, obtaining higher-quality images could significantly improve feature extraction. Additionally, acquiring more training data for minority classes would provide the model with a more balanced dataset, leading to better performance.

References:

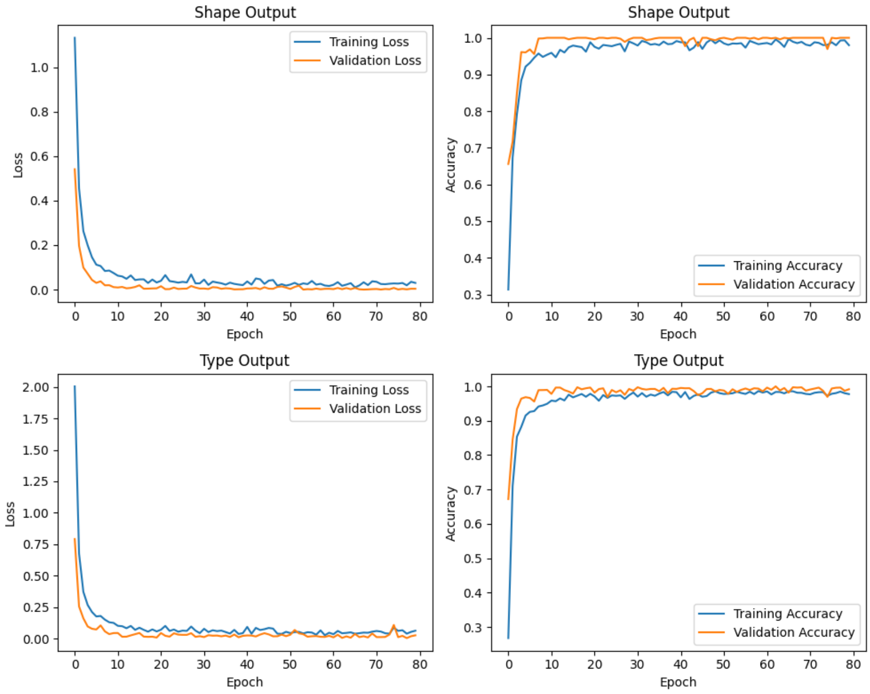
1. C. Gámez Serna and Y. Ruichek, "Classification of Traffic Signs: The European Dataset," in *IEEE Access*, vol. 6, pp. 78136-78148, 2018, doi: 10.1109/ACCESS.2018.2884826. keywords: {Europe;Roads;Shape;Autonomous vehicles;Feature extraction;Training;Benchmark testing;Convolutional neural networks;dataset;traffic signs;traffic sign classification},

Appendix

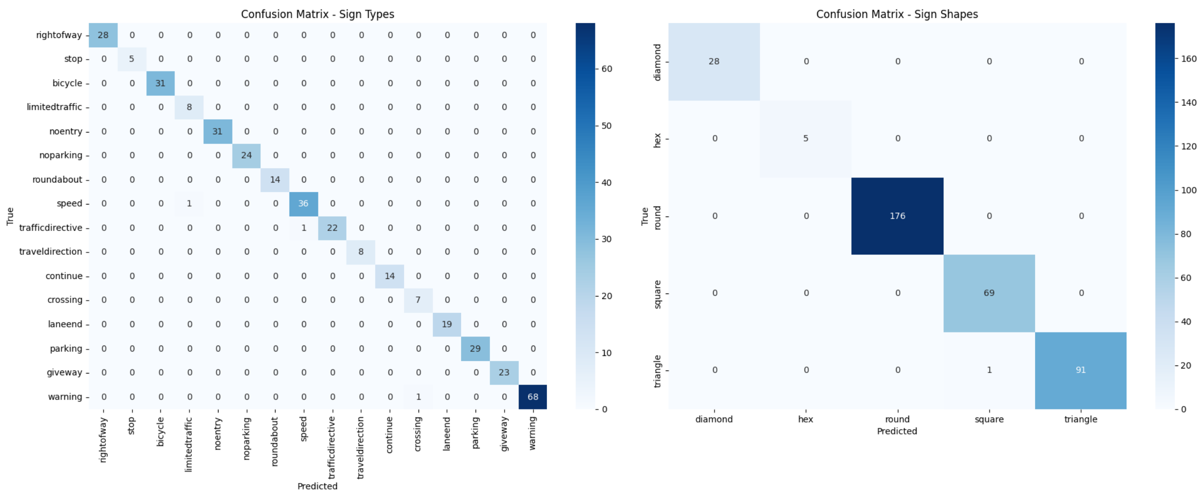
1. Base line model loss and accuracy curve



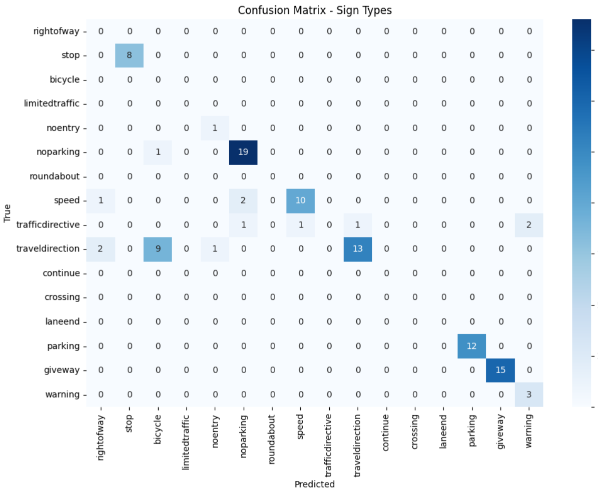
2. Final model loss and accuracy curve

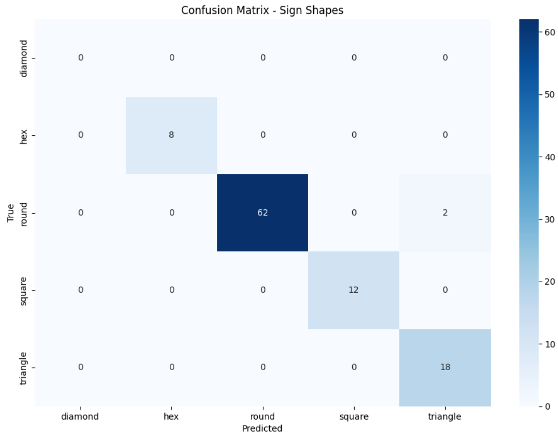


3. Test Data Confusion Matrix.



4. Independent Test Data Confusion Matrix


Confusion Matrix - Sign Types
True
Predicted
rightofway
stop
bicycle
limitedtraffic
noentry
noparking
roundabout
speed
trafficdirective
traveldirection
continue
crossing
laneend
parking
giveway
warning


Confusion Matrix - Sign Shapes
True
Predicted
diamond
hex
round
square
triangle