Road Signs Classification for Extreme Driving Conditions

Team 7: Abhijith Tammanagari, Athena Liu, Tianyun Hou, Weiliang Hu

April 13th, 2022

I. Abstract

Driving in adverse weather conditions such as heavy rain, heavy fog, and nighttime is difficult for drivers to see road signs. It is even more dangerous if drivers have impaired vision which poses a great risk to the drivers and those around them. Our goal is to develop a model which can accurately identify traffic signs in harsh weather conditions and warn drivers of those signs. Typical datasets tend to only include normal weather conditions and exclude situations such as heavy rain or fog. In the project, we utilize a Convolutional Neural Network(CNN) model which was built off the LeNet architecture to accurately classify signs and compare the performances across multiple datasets which include simulated adverse weather conditions. The model utilizing a training dataset that contains only dark images yields the highest accuracy for normal and adverse conditions. Therefore using this model we can accurately identify and warn visually impaired drivers of road signs up ahead and prevent potential accidents.

II. Introduction

Street signs play a pivotal role in driving and are one of the many means that keep us safe on the road and prevent unwanted accidents. Being able to properly see and identify street signs is crucial but this may be difficult when the driver is visually impaired or driving through adverse weather conditions. We hope to develop a model that can be deployed in a camera to identify and warn drivers of upcoming relevant signage in normal conditions as well as adverse weather conditions. Doing so will give drivers more confidence and an extra safety net to ensure they are confident in identifying signs that might otherwise be difficult to discern and will result in fewer accidents. Typically such models are trained on the dataset available but we want to take that one step further and focus on reliability in varying weather and thus propose to use a data set consisting of simulated weather conditions and see how that holds up to the normal dataset.

III. Background

In the field of traffic sign recognition (TSR), a number of notable works have been published over the years. Most of them include the module of detecting, tracking, and classifying images in a single pipeline while others specialize in some of them. Beginning with a multi-layer perceptron classification method (De La Escalera., 1997), followed by a system consisting of AdaBoost, Haar wavelet features and Gaussian probability density model (Bahlmann., 2005).

Two years later, a three-way neural network pipeline of color segmentation, shape detection and classification was proposed for Italian road signs recognition (Broggi., 2007). That same year, American and European traffic signs were classified by segmenting potential digits inside traffic sign candidates and then applying multi-layer neural networks(Moutarde, 2007). Following the upsurge of machine learning techniques in the 2010s, many significant works had been done for road signs classification. For example, a multi-class classification challenge held at the International Joint Conference on Neural Networks 2011 (Stallkamp., 2012) paved the way for future scientists. During the competition, Ciresan.(2012) topped the final phase with 97.46% accuracy by combining five deep neural networks over each of the five preprocessing inputs and averaging individual predictions of the DNN which resulted in beating the best individual human performance (Stalkamp., 2012). In Jin et all. (2014), a local response normalization as discussed in Krizhevsky.(2012) and an SVM-like hinge loss cost is used to create an ensemble as described in Ciresan.(2012) that resulted in 98.65% accuracy.

In recent years, convolutional neural networks(CNN) dominated the realm of traffic sign image classification. Islam and Raj (2017), Huang (2017), and Wasif (2021) built their own artificial neural network for classifying various types of traffic signs using CNN and they all received pretty impressive prediction accuracy, ranging from 97% to 99%. Meanwhile, pre-trained models using CNN, such as ResNet and Lenet, were used in some academic journals on road sign classification (Zaibi., 2021 and Huo., 2020) or Road Network Traffic Prediction(Jeon., 2021), obtaining similarly fantastic results as well.

While studies using pre-trained and customized models of CNN to classify real-world images (including but not limited to road signs) bloomed in the last 10 years, few studies considered classifying road sign images under different weather conditions, which is, as a matter of fact, an important task in reality. So far, there's no existing public dataset for different weather conditions. Therefore, researchers would need to simulate their own images under different weather conditions. Previous traffic sign research either generated synthetic data unrelated to weather conditions (Lunge., 2014 and Chen., 2021) or simply generated noises simulating problems in detection systems (Glorot., 2011 and Zhou., 2017).

Overall, the larger body of work tries to address the challenges of recognizing and classifying images using various architectures. In our work, we aim to design an adaptive classification tool for visually impaired individuals under extreme weather and lighting conditions. We begin to tackle this problem by devising an NN-based classifier trained on our data selected from reliable sources, starting from pre-trained models that were proven to perform well on traffic sign classification tasks in previous research.

IV. Data

The data used in this analysis is obtained from two Kaggle datasets: German Traffic Sign Recognition Benchmark (GTSRB) which originally consists of more than 50,000 pictures of traffic signs in 43 classes, and Road Sign Detection dataset which contains 877 images of 4 distinct classes of traffic signs. The two datasets are aggregated together and 4 classes of traffic signs are selected for our classification purpose: speed limit sign, stop sign, crosswalk sign, and traffic light sign. Examples of images from each class are displayed below (Figure 1).

Figure 1. Examples of Images of Different Classes

After subsetting the GTSRB data to the desired four classes and aggregating with the Road Sign Detection dataset, the final dataset contains 10,297 images in total. Each image is represented by a 48✕48✕3 array of pixel intensities after resizing. As shown in Figure 2 below, there is a significant imbalance across the 4 classes: the speed limit sign and stop sign has a large number of samples while the crosswalk and traffic light sign contain much fewer images. The imbalanced class size poses the challenge that the trained model could have a biased performance towards over-representative classes. The final dataset is split into training, validation, and testing dataset with percentages of 55%, 25%, 20%, respectively. To mitigate the discrepancy issue across classes, data augmentation is implemented to create additional images through rotation, shifting, and flipping to extend the dataset for under-representative classes. Images are normalized before fitting into the model to ensure the input pixel has a similar distribution which can lead to faster convergence while training.

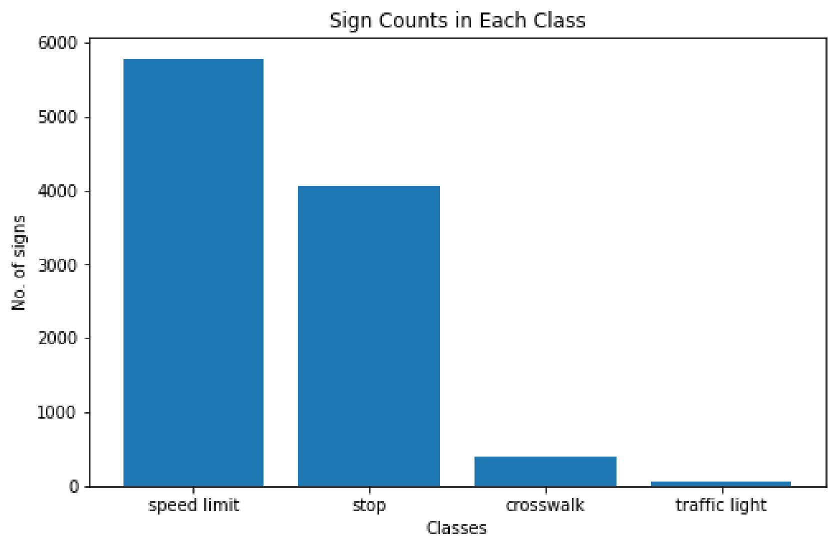

Figure 2. Data Size of Each Class

V. Methods and Experiments.

A. ResNet

ResNet is one of the popular deep convolutional neural networks that is developed in recent years for boosting performance on image classification tasks. As the training process of a deep neural network usually requires large-scale datasets, transfer learning provides an alternative to develop well-performing deep neural networks efficiently by leveraging pre-trained models in the related domain (Weiss, et al., 2016). Therefore, we chose pre-trained ResNet models of ResNet18, ResNet50, and ResNet101 as the starting point and fine-tuned the model by re-training the last layer on our original dataset to identify which model performs the best. The ROC curve from the ResNet18 model (Figure 3) shows the model seemed to perform decently but we might be able to do better.

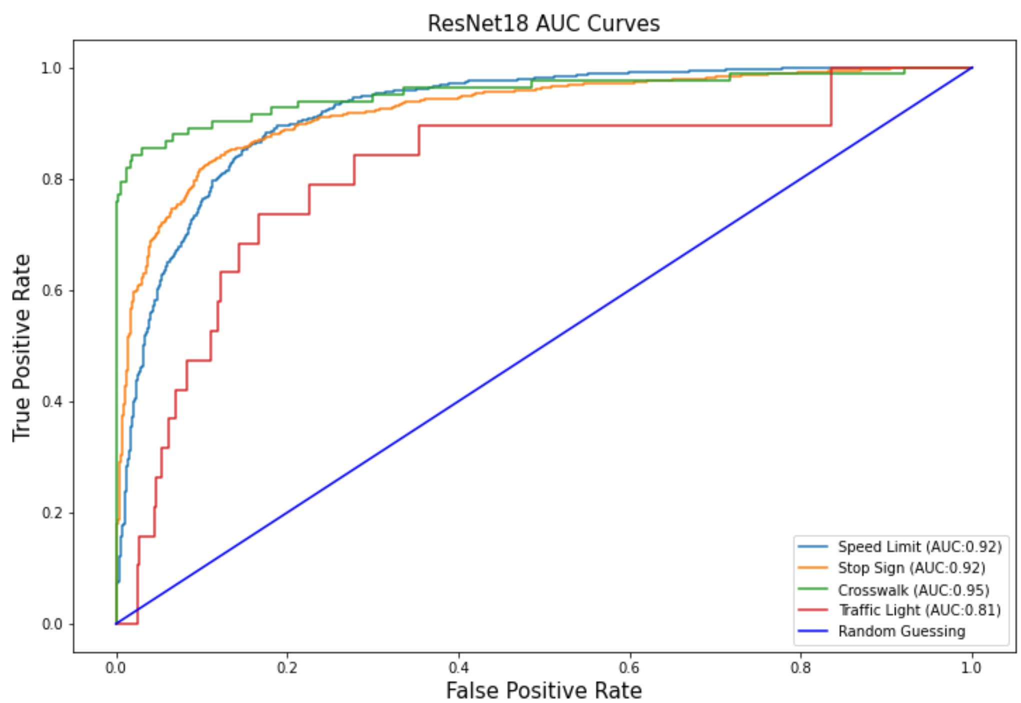

Figure 3. ResNet18 ROC Curve

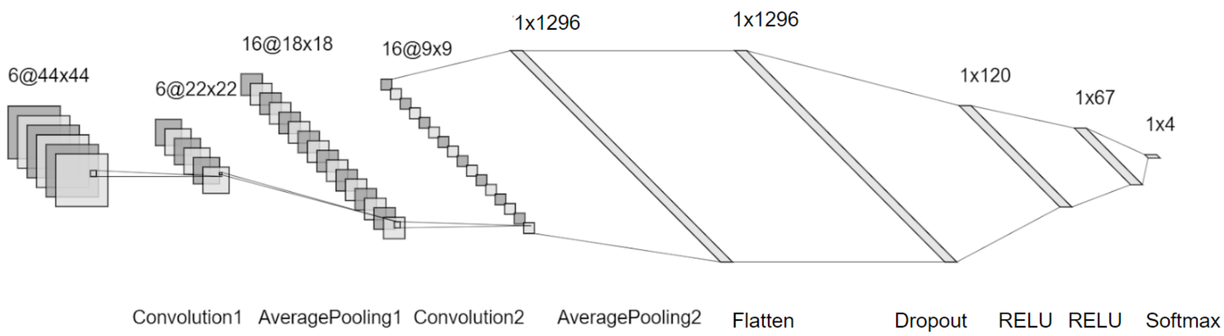

Figure 4. Customized LeNet Architecture

B. Customized LeNet-5

LeNet-5 is another simple but effective CNN architecture in the realm of character recognition and image classification. Based on the traditional structure of LeNet-5, we further customized the CNN architecture to better fit our data. As shown in Figure 4, the customized CNN consists of a convolution layer with 6 filters and a kernel size of 5x5, a 2x2 pooling layer, another convolutional layer with 16 filters and a kernel size of 5x5, a 2x2 pooling layer, a flattening layer, a dropout layer and followed by 3 dense layers consisting of 120, 67 and 4 units each. The model yielded a validation accuracy of .9719, which vastly outperformed the ResNet18 model. To further investigate the performance of the customized LeNet model, we utilized the confusion matrix below (Figure 5) and observed that it did a really good job of classifying all our varying cases and also performed moderately acceptable in correctly classifying our low sample size class 3 (crosswalk).

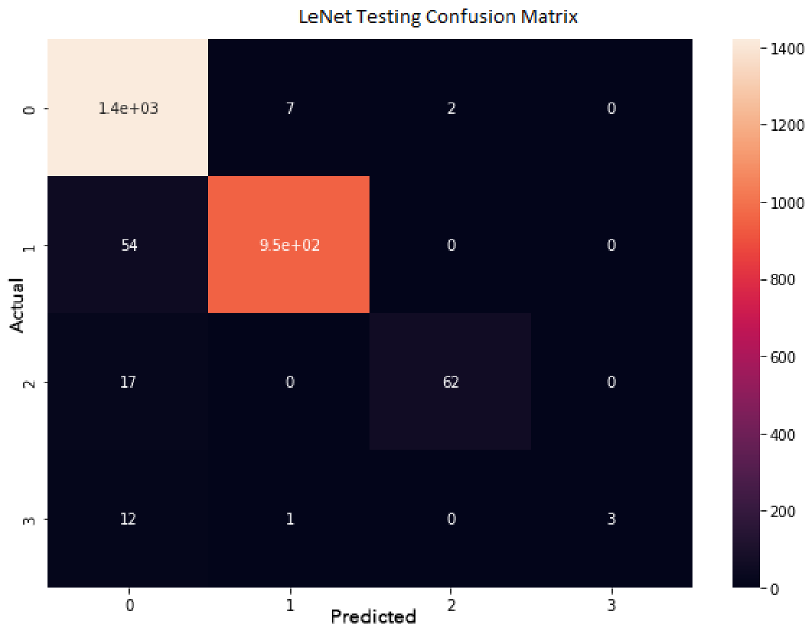

Figure 5. Confusion Matrix for the LeNet Model Trained and Tested on Original Dataset

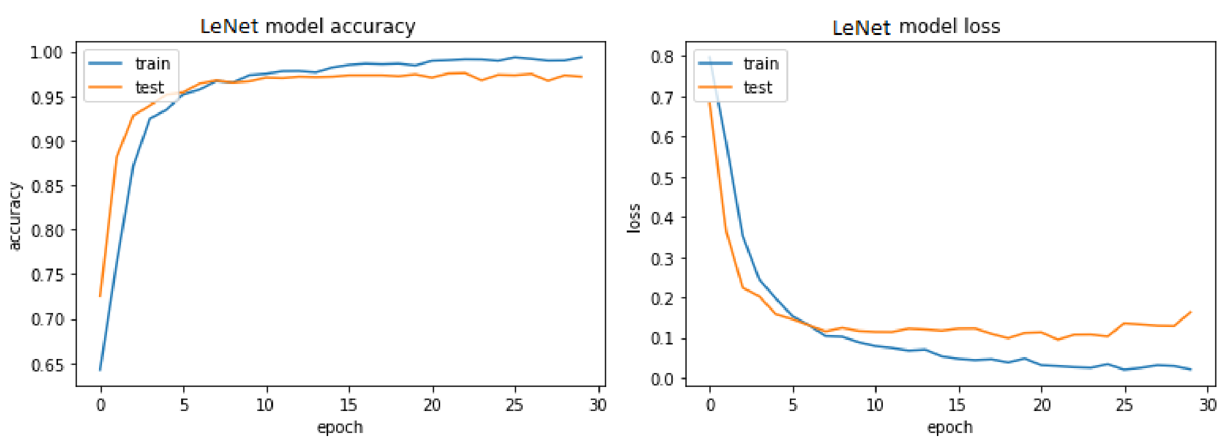

Figure 6. Plots of LeNet Model Accuracy and Loss Over Time

In addition, according to the accuracy and loss plot shown above (Figure 6), this model seemed to keep learning without overfitting till epoch 20 but we will investigate this further below through hyperparameter tuning.

The below confusion matrix is the result of our test accuracy on a LeNet model with Data Augmentation (Figure 7). It achieved a validation accuracy of .91 but resulted in a large amount of misclassification of classes 0 and 1, which performed worse than without data augmentation.

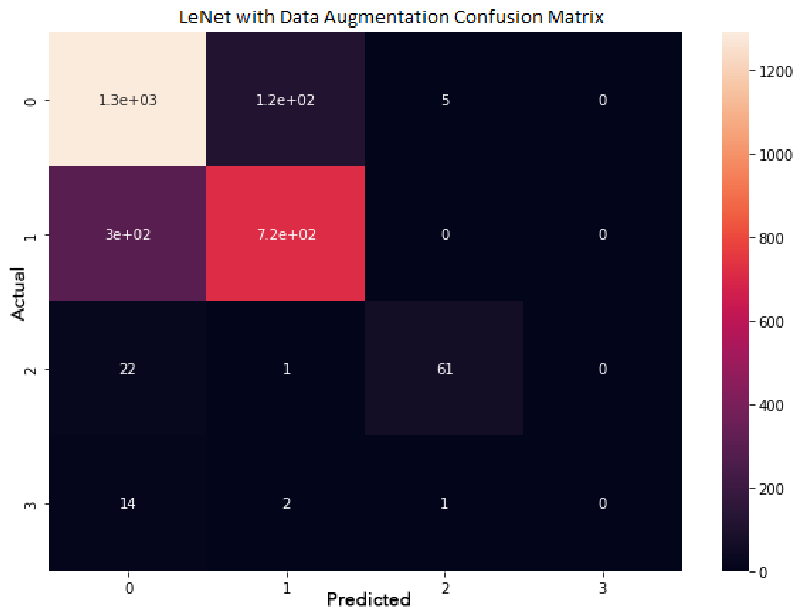

Figure 7. Confusion Matrix for LeNet Model Trained on Data Augmented Dataset

Through this entire process of testing the various models above we have determined that the best classifier is the LeNet algorithm with training on the original dataset and will be used throughout the rest of our study as the sole baseline model.

To optimize the model performance, we performed hyperparameter tuning on the LeNet model. We are interested in finding the optimal combinations of batch size, epochs, and optimizer learning rate for the model by performing Grid Search Cross-Validation with 3 folds. Among the combinations of hyperparameters, batch size of 200, epochs of 20, and learning rate of 0.001 yields the best model performance.

C. Image Distortion and Experimental Design

Thus far, the model was trained and validated only on the original dataset, which contains mostly clean and clearly visible images. However, these original images are not representative of bad driving conditions in real life, and we only have limited sign images captured in bad conditions. Therefore, we generated datasets that are simulations of rainy weather, foggy weather, dark environment, and bright environment for our experiments as they are common challenging driving conditions for individuals with impaired vision. The image distortions are simulated by blurring the pixels, altering brightness and situations, and adding randomized noises to the images (Figure 8).

8a. Original *8b. Rain* *8c. Fog* *8d. Dark* *8e. Bright*

Figure 8. Examples of Image Distortions

We designed a 6-components experimental process that allows us to compare how a model trained by different images would respond to adverse driving conditions (Figure 9). For each component of the experiment, one or multiple types of distortion effect(s) are applied to the original training dataset. Component 1 contains the original training images and serves as the baseline for the experiment. Models trained in other components are compared to the baseline model. Component 2 comprises images with mixed distortion effects -- 20% original images, 20% rainy images, 20% foggy images, 20% dark images, and 20% bright images. Components 3 to 6 comprise images with only one distortion effect, where rainy, foggy, dark, and bright effects were applied separately to the original training images. Then, all training images are normalized prior to training. We then trained a LeNet model for each component and tested them on six testing datasets -- original, rain, fog, dark, bright, and mixed (same effects breakdown as the mixed training dataset). Each experiment component would output six AUC scores corresponding to each testing dataset, and the model performance is evaluated by comparing average AUC scores across different components.

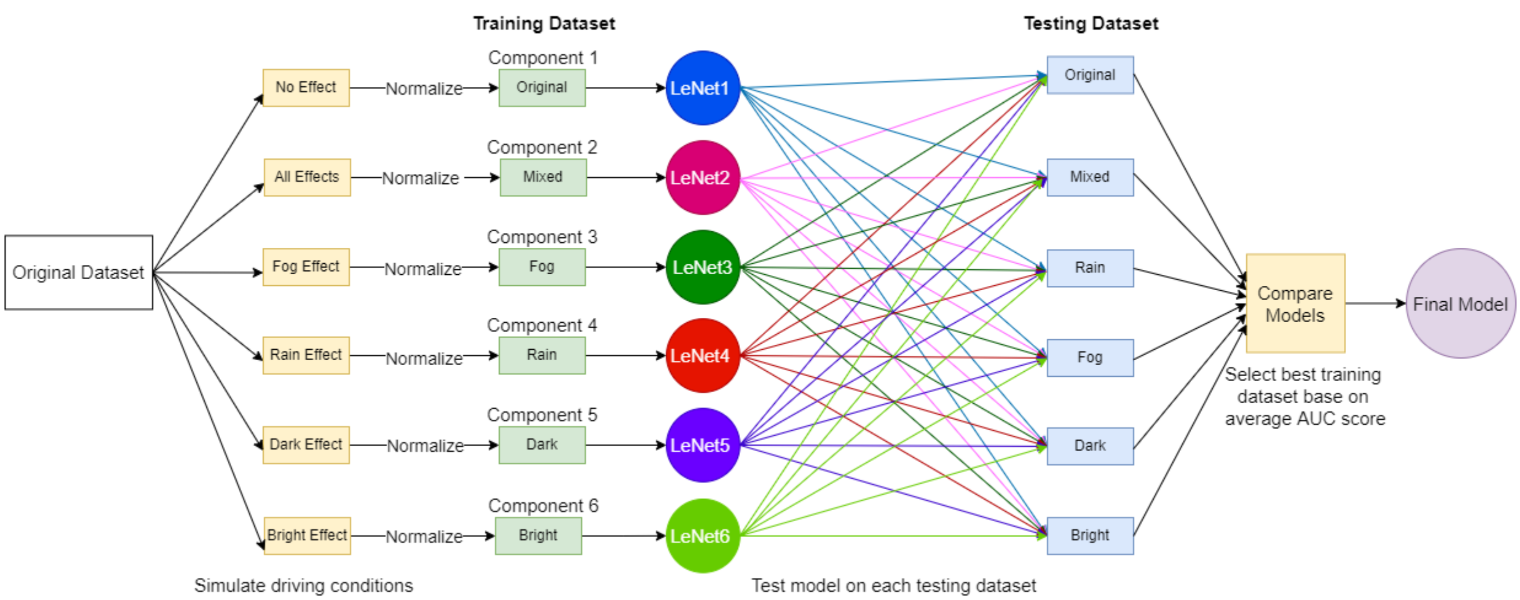

Figure 9. Experimental Design Flow Chart

VI. Results

Among the four models we tested -- ResNet18, ResNet50, ResNet101, and LeNet -- LeNet achieved the best classification accuracy on the original image dataset (Table 1). When LeNet was trained with the original dataset, its classification accuracy of the original testing dataset could reach 0.98.

Model	ResNet18	ResNet50	ResNet101	LeNet
Test Accuracy	.86	.83	.87	.97

Table1. Model Accuracy Trained and Tested on Original Dataset

The 6-components experimental process provided us with better insights into what kinds of training datasets could help the model to achieve the best classification performance in adverse driving conditions. By following the experimental process described in the Methods section, we obtained the average AUC score table displayed below (Table 2).

	original	mixed	rain	fog	dark	bright
Testing Dataset						
original	0.970	0.961	0.967	0.960	0.972	0.974
rain	0.635	0.641	0.684	0.590	0.725	0.585
fog	0.652	0.633	0.683	0.683	0.757	0.616
dark	0.643	0.627	0.659	0.574	0.717	0.581
bright	0.717	0.685	0.757	0.722	0.803	0.686
mixed	0.682	0.637	0.673	0.653	0.758	0.626

Table 2. LeNet Average Testing AUC

The columns represent the training dataset used, while the rows represent the testing datasets used. Regardless of the training dataset, almost all models achieve the AUC score of 0.97 when classifying road signs from the original testing dataset. However, different training datasets could have varying impacts on classifying road signs in adverse driving conditions. When the model was trained using the original training dataset (baseline model), the average AUC score for different driving conditions could range between 0.635 to 0.717. The worst-performing model was trained by only using bright images, in which the mean average AUC is 0.626. The best-performing model was trained by using only dark images. The average AUC scores for different driving conditions range from 0.717 to 0.803, with a mean average AUC score of 0.758. This indicates that the model has been introduced to all types of poor driving conditions and noisy data performs the best. We further compared the ROC curves of the baseline model and the best-performing model.

The ROC curves below displayed how the baseline model performed in classifying road signs in different driving conditions (Figure 10). For most scenarios, the model demonstrated some abilities in accurately classifying the four classes. The model, in almost all scenarios, lacks the ability to classify traffic lights, in which its AUC score is near equivalent to random guessing (AUC = 0.50). This could be caused by the limited number of traffic light images in the training dataset.

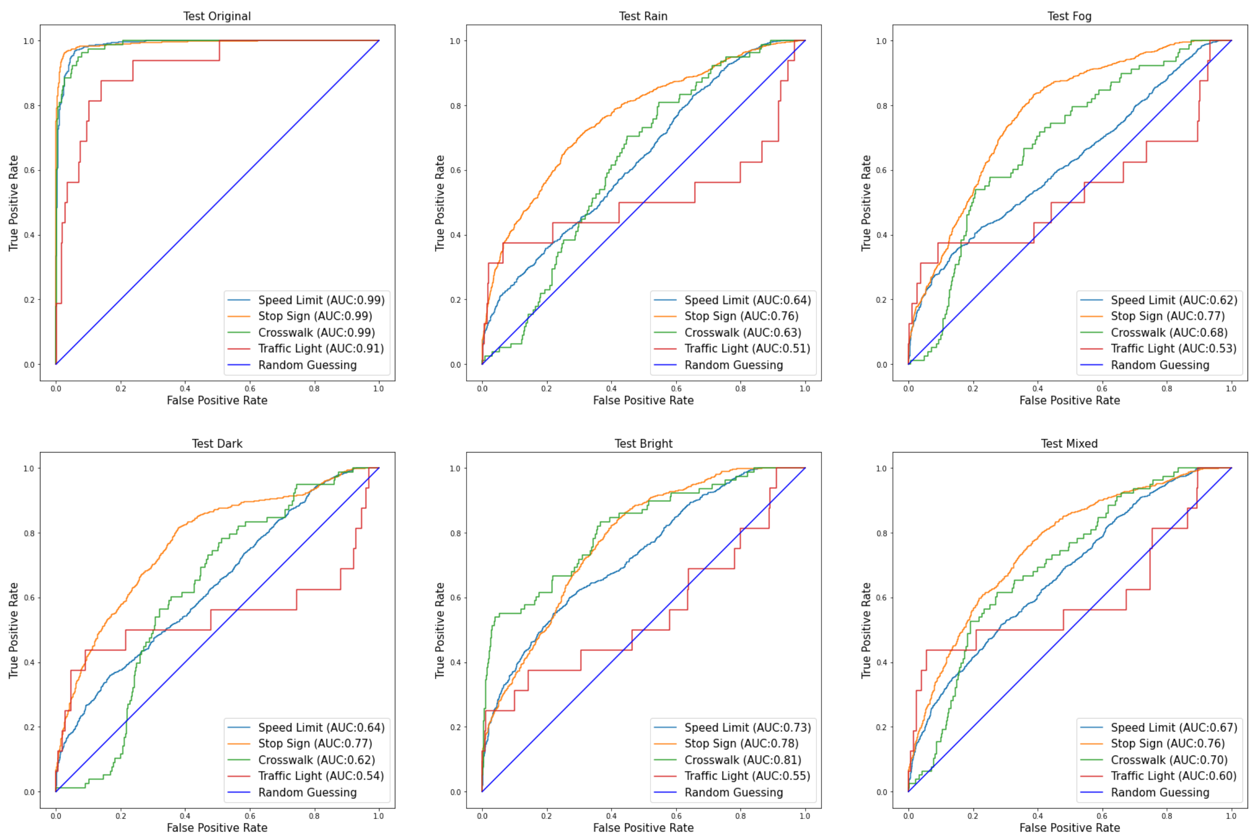

Figure 10. Original Images Training Dataset LeNet Model ROC Curve Breakdown (Baseline)

The ROC curves below display the performance of the best-performing model (Figure 11). Different from the baseline model, the model trained with dark images showed the ability to classify stop signs, speed limit signs, and crosswalk signs with AUC above 0.70 in most scenarios. Performance for traffic lights classification remained roughly the same. The improvement in the performance of this model from the baseline could be contributed by the random noises added to the training images. This experimental process confirmed that having a relatively noisy training image dataset could help to boost the model's performance in classifying road signs in adverse driving conditions. However, while the model has better accuracy, it is likely biased. The confusion matrix below showed that the model is incapable of classifying crosswalks and traffic lights and is only relatively more trustworthy when classifying speed limits and stop signs (Figure 12). Since we cannot resolve the biases with our current dataset, we decided to proceed with the highest accuracy model. The final model for our project is a LeNet CNN model trained with dark images, with the hyperparameter setting of 200 for batch size, 20 for epochs, and 0.001 for optimizer learning rate.

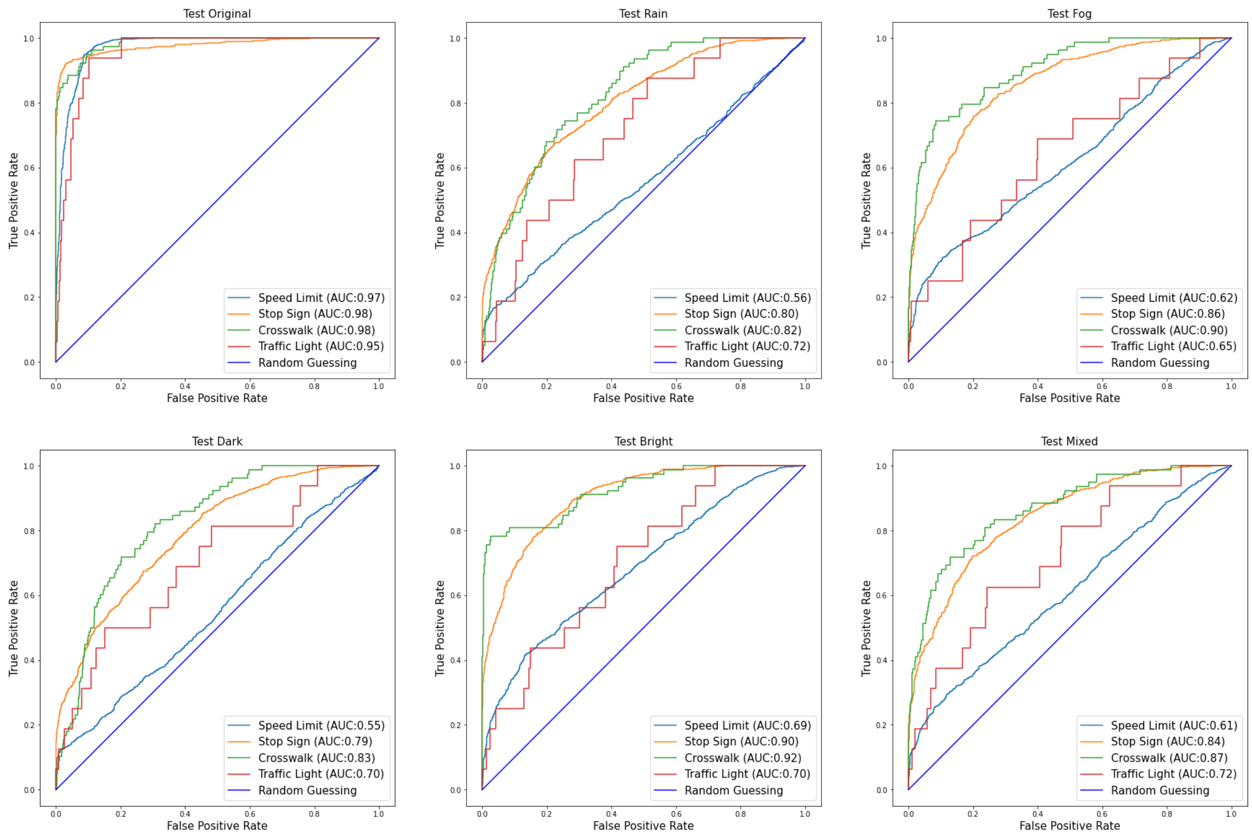

Figure 11. Images Training Dataset LeNet Model ROC Curve Breakdown (Best)

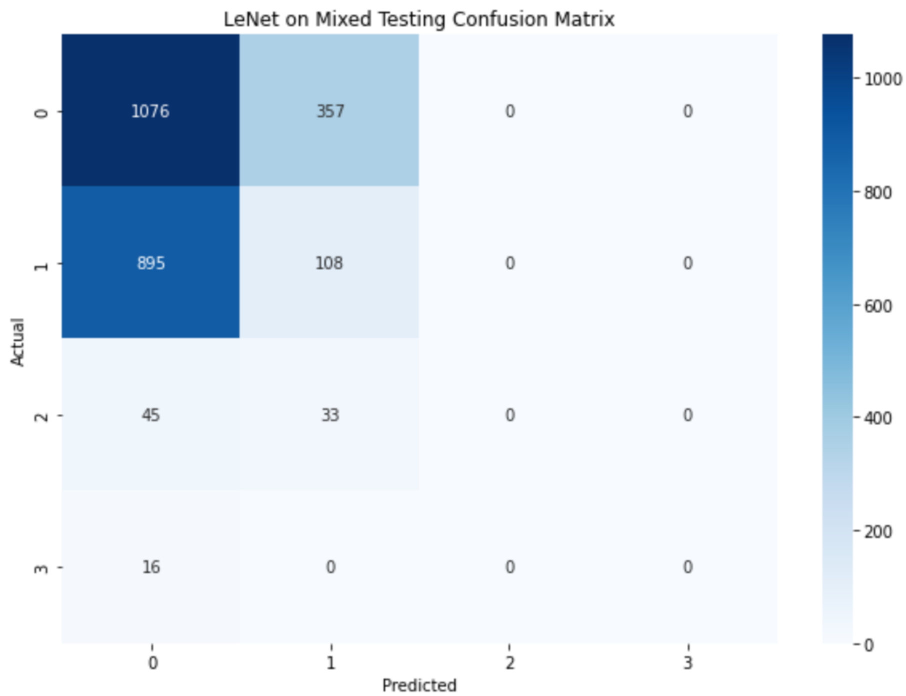

Figure 12. LeNet on Mixed Testing Confusion Matrix

VII. Conclusions

A road sign classifier that could reliably recognize and warn drivers of upcoming road signs in adverse driving conditions could be helpful and potentially life-saving. After extensive model testing in different low visibility scenarios, we found that LeNet (batch size 200, epochs 20, learning rate 0.001) is the best performing model and is trained on the dark dataset which consistently classifies road signs with an average AUC score of 0.758. Therefore, the takeaway of our project is that to maximize the performance of street sign classification across all conditions it is most beneficial to train on a nighttime dataset.

It is important to acknowledge the limitations of our project. The dataset we used is imbalanced as 55% of the dataset consists of speed limit signs and only 0.5% of the dataset consist of traffic light images. As a result, our model performed poorly in classifying under-representative classes. Moreover, our simulated images may not accurately reflect the real-world circumstances perceived through the camera and only cover some of the adverse driving conditions that drivers can possibly encounter. How close the simulated images can imitate the real world will significantly affect the performance of our neural network if applied to real settings.

For future work, we would like to collect more images of stop signs, crosswalk signs, and traffic lights. We wish to train the model with a more balanced dataset, so that the classifier could recognize underrepresented classes more accurately. If possible, we should repeat our experimental procedure on real-life images instead of simulated images to investigate what kinds of training data would be most helpful in terms of boosting classifier accuracy on low visibility road sign images.

VIII. Roles

Abhijith Tammanagari	Took the role of taking the initial andrew sign data set and processing those images and xml metadata into arrays to be used for modeling as well as cropping images from both datasets into an even size. Implemented the data augmentation step as well as the ResNet Models.
Athena Liu	Image preprocessing (generate rain, fog, dark, and bright), experimental design, worked on report methods, results, and conclusion, note keeping/meeting management.
Tianyun Hou	Performed literature review on related topic and researches; Contributed to exploratory data analysis and model building; Worked on hyperparameter tuning the models
Weiliang Hu	Took the role of doing literature review and reading up similar work. Trained and tuned Lenet-5 model and customized the final Lenet model on the original dataset. Implemented evaluation metrics for performances and examined the results.

IX. Reference

Escalera, A. et al. (1997). Road traffic sign detection and classification. *IEEE Trans. Ind. Electron.* 44 (1997): 848-859.

Bahlmann, C. & Zhu, Y. & Ramesh, V. & Pellkofer, M. & Köhler, T.. (2005). A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. 255 - 260. 10.1109/IVS.2005.1505111.

Broggi, A. et al. (2007). Real Time Road Signs Recognition. 2007 IEEE Intelligent Vehicles Symposium: 981-986.

Moutarde, F. et al. (2007). Robust on-vehicle real-time visual detection of American and European speed limit signs, with a modular Traffic Signs Recognition system. 2007 IEEE Intelligent Vehicles Symposium: 1122-1126.

Stallkamp, J. et al. (2011). The German Traffic Sign Recognition Benchmark: A multi-class classification competition. The 2011 International Joint Conference on Neural Networks: 1453-1460.

Ciresan, Dan C. et al. (2012). Multi-column deep neural networks for image classification. 2012 IEEE Conference on Computer Vision and Pattern Recognition: 3642-3649.

J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel (2012).
Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Neural Networks, ISSN 0893-6080,https://doi.org/10.1016/j.neunet.2012.02.016.

Jin, J. et al. (2014). Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks. *IEEE Transactions on Intelligent Transportation Systems* 15: 1991-2000.

Krizhevsky, A. et al. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM 60: 84 - 90.

Huang, S. & Lin, H. & Chang, C. (2017). An in-car camera system for traffic sign detection and recognition. 1-6. 10.1109/IFSA-SCIS.2017.8023239.

Haque, W., Arefin, S., Shihavuddin, A.S.M., Hasan, M. (2021).
DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements. https://doi.org/10.1016/j.eswa.2020.114481.

Zaibi, A., Ladgham, A., Sakly, A. (2021). A Lightweight Model for Traffic Sign Classification Based on Enhanced LeNet-5 Network, *Journal of Sensors*, vol. 2021.
https://doi.org/10.1155/2021/8870529

Huo, T. et al. (2020). Traffic Sign Recognition Based on ResNet-20 and Deep Mutual Learning. 2020 Chinese Automation Congress (CAC): 4770-4774.

Jeon, S. et al. (2021). Road Network Traffic Data Prediction Using ResNet. *Journal of Korean Society for Geospatial Information Science*.

Islam, KT., Raj, RG. (2017). Real-Time (Vision-Based) Road Sign Recognition Using an Artificial Neural Network. Sensors (Basel). 2017 Apr 13;17(4):853. doi: 10.3390/s17040853.

Ms. Rani V. Lunge, Prof. B. V. Pawar. (2014). Detection and Recognition of Road Traffic Signs in Various Illumination and Weather Conditions, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)* Volume 03, Issue 07 (July 2014).

Zhou, Y., Song, S., Cheung, N. (2017). On classification of distorted images with deep convolutional neural networks. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1213-1217, doi: 10.1109/ICASSP.2017.7952349.

Rifai, S., Glorot, X., Bengio, Y., & Vincent, P. (2011). Adding noise to the input of a model trained with a regularized objective. ArXiv, abs/1104.3250.

Dewi, C., Chen, RC., Liu, YT. et al. (2021). Synthetic Data generation using DCGAN for improved traffic sign recognition. *Neural Comput & Applic*. https://doi.org/10.1007/s00521-021-05982-z

Weiss, K., Khoshgoftaar, T.M. & Wang, D. (2016). A survey of transfer learning. *J Big Data* 3, 9. https://doi.org/10.1186/s40537-016-0043-6