Traffic signs classification using deep learning

Ameer Talha
Computer Science And Engineering
AUST
Dhaka, Bangladesh
180104036@aust.edu

Sheikh MD Rezone Ullah
Computer Science And Engineering
AUST
Dhaka, Bangladesh
180104037@aust.edu

Samin Ul Alam
Computer Science And Engineering
AUST
Dhaka, Bangladesh
180104043@aust.edu

I. INTRODUCTION

Traffic sign classification is one of the most important systems which can recognize traffic signs by using some machine learning models. Traffic sign classification is a technology that can automatically read traffic signs and display them to vehicles. Image processing techniques and computer vision are used in this technology. Image datasets are used in this classification. Deep learning models are worked so well in this classification. This is very vital for reducing accidents, and speed limits. There are used CNN based on some pre-train models such as AlexNet, VGGNet, LeNet, GoogleNet, ResNet, etc which gives good accuracy. Vehicles detected traffic signs by using automated system technology. The detection method can be divided into three methods these are the colour method, the shape-based method, and the learning-based method. There are many traffic sign shapes. These are hexagons, circles, and rectangles. There are use some activation functions like ReLU, ELU, FELU, etc.

Most of the machine learning models use data from several camera sensors mounted on the vehicle roof at different angles. Traffic Sign Recognition (TSR) system ensures the safety of drivers, and automated vehicles as well as increase the comfort of driving. Automate systems are used in automobiles, kitchen tools, power backup devices, medical, entertainment, industrial machinery, sports, etc.

II. RELATED WORKS

1) In paper[1] the author introduced a real-life dataset of traffic signs and classified them using different convolutional neural network architecture for autonomous driving systems.They have chosen GTSRB dataset that consists of 50,000 images divided in 43 classes and European dataset which consists of 80,000 images divided in 164 classes. In the paper we observe that the author has used CNN models to achieve high accuracy and low loss scores for the experiment.So, the author have used five different CNN models that can achieve the best performances in the state of the art regarding traffic sign classification.From the 5 CNN model CNN 8-layers performed best in GTSRB dataset which accuracy is 98.52% and in European dataset CNN asymmetricK performed best among 5 models that accuracy is 98.48%.

2) In paper [2], the author has applied a total of three models are MobileNet, inception, and ResNet. The result of mobileNet is 81.30% Precision, 71.10% Recall, and 75.90% F-measure. The result of inception is 31.90% Precision, 78.40% Recall, and 81.10% F-measure. And the result of ResNet is 76.30% Precision, 85.20% Recall, and 80.50% F-measure. There are used two datasets, one is mobileNet and another is MS COCO dataset.

3) In paper[3] author introduced deep inception based neural network which is used in classify different traffic signs. The famous traffic sign dataset GTSRB is used here. In these datasets there are more than 50,000 post samples. By using the CNN model they get a 98.6% accuracy and which is better than other deep learning models

4) In paper [4], the author has implemented four CNN pre-train models: AlexNet, VGGNet, Inception, and ResNet. The accuracy of these models are AlexNet(85.21%), VGGNet(91.74%), Inception(95.54%), and Resnet(95.57%). The accuracy of these models is pretty good. The dataset used in this paper is GTSRB. This dataset contains more than 50,000 images and more than 40 categories.

5) In the paper [5], the author has applied four CNN pre-train models: Zheng et al, AlexNet, GoogLeNet, Zhou et al. The accuracy of these models is Zheng et al(94.62%), AlexNet(95.90%), GoogLeNet(96.50%), Zhou et al(98.20%). The methods used in these models are CNN+PCA+LVQ, CNN, CNN+ Inception, and ReduceV2Net. There is also used GTSRB (German

Traffic Sign Recognition Benchmark) dataset.

6) In paper [6], the author has used only two CNN pre-train models: VGG16, and ResNet. There are a total of three activation functions used: ReLU, ELU, and FELU. The accuracy rate of VGG16 is 93.66% (ReLU), 93.72%(ELU), and 94.44%(FELU). The accuracy rate of ResNet50 is 92.08% (ReLU), 94.24%(ELU), and 95.57%(FELU).GTSRB dataset has been used in this paper.

III. BACKGROUND STUDY

CNN (Convolutional Neural Network)

In CNN it takes input as images, assigning weight of a images and biasing various aspects of images and finally differentiates among them in different classes of images. Compared to other classification algorithm CNN required lower preprocessing and also successfully it can filtering through the application from different dependencies. Compared to other neural network model CNN architecture perform a better fitting in image dataset.Number of parameter is also less than other neural network model.

VGGNet

VGG is another CNN-based pre-train model. It is used in medical imaging, speech recognition, computer vision, image recognition, and machine translation. Deep CNNs, in particular, is composed of several layers of processing. The VGGNet architecture has 16-19 layers. Image recognition architecture is very popular architecture in VGGNet.

The input size of VGGNet is 224-by-224. And use 1500 object categories and 16 layers. VGGNet architecture consists of input, convolutional layers, fully-connected Layers, and hidden layers.

LeNet

LeNet is a convolutional neural network. LeNet is CNN based pre-train model which detects handwritten documents. LeNet is a convolutional network designed for handwritten and character recognition convolutional neural networks. LeNet was used primarily for OCR and character recognition in documents. LeNet (CNN) architecture is built up of 7 layers.

This LeNet architecture has five weighted layers to solve this classification problem.

IV. METHODOLOGY

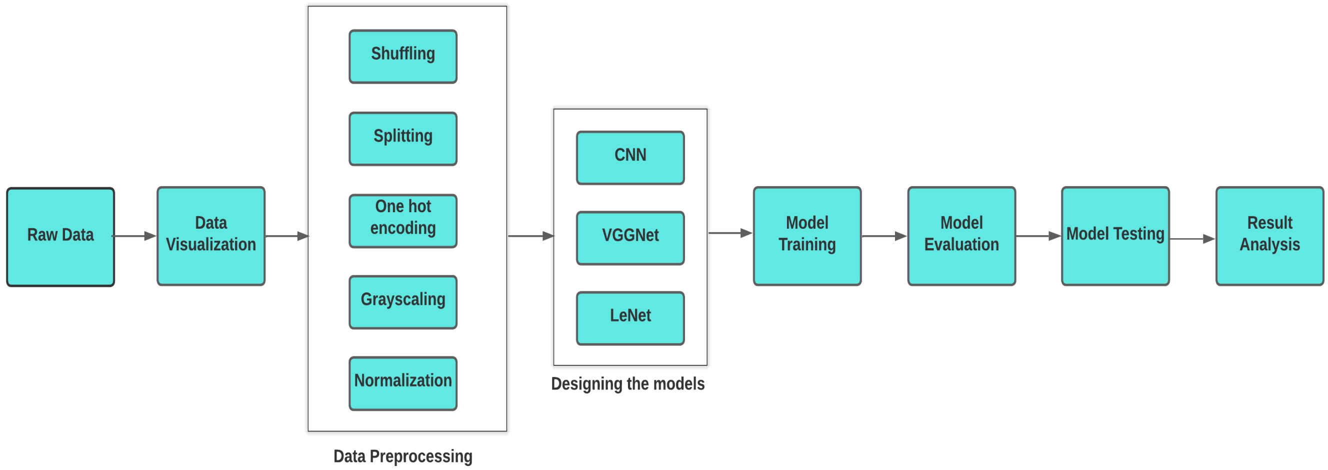

Fig : Flowchart of our project

Step 1: Data Preprocessing

Shuffling:
In our project after collecting the training data, we shuffled the dataset. Then we split the the data into train and validation set

Splitting:
Dataset splitting is a technique to divide the data into two or more subsets. Generally,in two part splitting one part is used to train the data and the other part is to test. Here,we are splitting the dataset into train,test and validation. We have splitted 34799 data into train, 12630 data in test and the rest of them in validation set to perform the experiment.

One hot encoding:
One hot encoding is a system of converting data for machine learning models and to acquires a good prediction. It converts the categorical value into a binary value of 0 and 1. Machine learning models assume the data to be numeric. Data must be pre-processed because machine learning models can not work with any kind of non-numeric data. That's why we need one hot encoding for machine learning models.

Grayscaling:
Grayscale is a range of shades of gray without apparent color. Grayscling is very important for image color. It is a collection of gray shades. Grayscling only contains brightness information and no color information..

Normalization:
The main purpose of normalization is to minimize duplicate data. It also converts the source data.

Step 2: Design the Model Architecture

This is an important step that is fundamental in our project. We have used CNN to classify the images because ConvNets is designed to recognize visual patterns directly from pixel images with minimal pre-processing.It automatically learns hierarchies of invariant features at every level from data. Our goal is to reach higher accuracy in train,validation and test set so that our project can be succesful. We will start explaining each network architecture one by one below:

Setup - 1 :
In CNN three types of a layer are used convolutional layer, pooling layer, and fully connected layer. Different types of a parameter are used in making the model. So it uses 3 Convolution and Pooling layers. Here we use 30 epochs while using the CNN model.

Setup - 2 :
VGGNet is a CNN pre-trained model which is renowned to be one of the excellent computer vision model architecture. In VGGNet architecture there are input, convolutional,hidden,fully connected layer and output phases.Special thing about VGGNet is that without using huge number of hyper-parameters it focused on having convolution layers of 3x3 filter and used same padding and maxpool layer of 2x2 filter. It also have used the pattern of convolution and max pool layers throughout the whole architecture.Lastly,it has 2 fully connected layers followed by a softmax for output.However, it has total of 16 layers that have weights overall.

Setup - 3 :
The LeNet-5 model is another computer vision model that we have used in our project.In first two convolution layers, it uses 5x5 filter,average pooling and relu activation function.Secondly,it flattens the network. Thirdly, it has two fully connected layer that utilizes relu activation function. Lastly, softmax activation function is used in another fully connected layer.

Step 3: Model Training

Model training is a process in which a deep learning algorithm is fed with enough training data to learn from.
At first, we are building the models and train them with our German traffic sign recognition benchmark dataset. Then, we have implemented the models and compiled the different layers of models.However, training a model sometimes arise different types of problem such as overfitting.To eradicate this problem, we have randomly dropped units to prevent it from adapting too much. This method significantly reduced overfitting problem and gave us major improvements over other regularization methods.

Step 4: Model Testing

Model testing is the process where the performance of a fully trained model is evaluated on a testing set. The testing set consisting of a set of testing samples which has similar probability distribution as the training set. In our project, we have used our trained model from model training and fit the model into test set for testing its accuracy and loss.In testing phase, we have to define epochs,batch size,iterations,total dataset to achieve the required output from our models.

Different types speed limits and shapes make our model misclassify between signs which is a challenge for our models. We will overcome these problems using hierarchical CNN architectures. TensorFlow

V. DATASET

Fig. 1. Sample of images from the dataset

Dataset Collection
The dataset used for this project is provided by German Traffic Sign Recognition Benchmark(GTSRB) and is available on Kaggle.
Dataset Link- Click here

Statistic of Dataset
The dataset consists of more than 50,000 images in total and more than 43 classes. We used 39,209 images used in training and 12691 images selected for testing.

VI. RESULT ANALYSIS

Performance of our model
The performance of different models are reported with accuracy, precision, recall and F1 score.The performance of different architectures are summarized in the table below.

Performance Model Table

Model	Accuracy	Precision	Recall	F1-score
CNN	98.7	98	98	98
VGGNet	97.4	97	97	97
LeNet	88.1	89	88	88

We have observed from the table that CNN and VGGNet models have better accuracy,precision and recall performance score against LeNet model.We can see that these models have achieved accuracy over 95% which is very good.CNN have scored 98.7% and VGGNet got 97.5% and there precision,recall,f1 score are also good.However, LeNet model did not perform very well and scored 88.1% which is not impressive at all. Though LeNet model didn't perform like the other two models still it did get a pretty good performance matrix scores. So, from our result analysis we can say that CNN and VGGNet worked the best in this experiment.

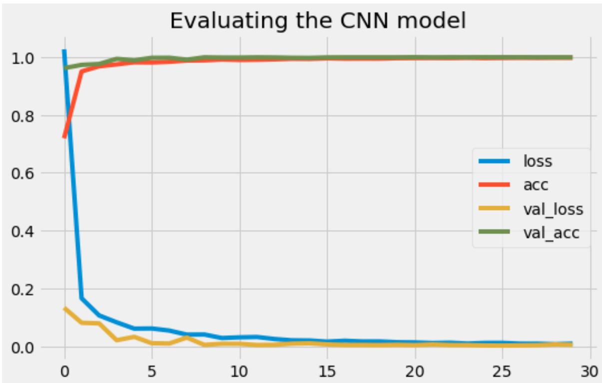

Fig. 2. CNN Model Curve

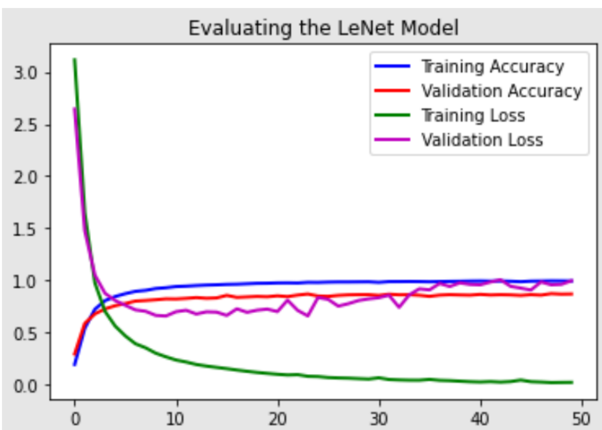

Fig. 3. LeNet Model Curve

VII. CONCLUSION

Traffic sign classification mainly classifies the traffic sign features to read the traffic sign by using the GTSRB dataset. CNN-based pre-train models are used in traffic sign classification. The pre-train models are such that CNN, LeNet, VGGNet etc. There are also use some activation functions such that ReLU, ELU, and FELU. The pre-train models capture images and predict them accurately.

The images of the dataset are pre-processed and enhanced the image by histogram equalization. The accuracy rate is almost 98% by using a convolutional neural network. There are some drawbacks also. These are it could be times covers or not visible clearly. Sometimes it might not able to keep check the speed of vehicles which could be the reason for the accident. We will have to research on this topic to improve it further in order to get a better result.

REFERENCES

[1] Citlalli Gamez Serna and Yassine Ruichek, "Classification of Traffic Signs: The European Dataset", 2018, IEEE.

[2] Dong Li, Dongbin Zhaoy, Yaran Cheny, Qichao Zhangy, "DeepSign: Deep Learning based Traffic Sign Recognition", 2018, IEEE.

[3] Mrinal Haloi, "Traffic Sign Classification Using Deep Inception Based Convolutional Networks", 2016.

[4] Xiucai, Guo, Changxuan, Zhao, Yaodong, Wang, "Traffic Sign Recognition Based on Joint Convolutional Neural Network Model", August 2019, ACM.

[5] Zhilong, He, Zhongjun, Xiao, Zhiguo, Yan, "Traffic Sign Recognition Based on Convolutional Neural Network Model", 2020, IEEE.

[6] Zheng Qiumei, Tan Dan, and Wang Fenghua, "Improved Convolutional Neural Network Based on Fast Exponentially Linear Unit Activation Function", 2017, IEEE.