Traffic Sign Classification Comparison Between Various Convolution Neural Network Models

[1], Jonah Sokipriala ,[2], Sunny Orike

Abstract: Fast detection and accurate classification of traffic signs is one of the major aspects of advance driver assistance system (ADAS) and intelligent transport systems (ITS), this paper presents a comparison between an 8-Layer convolutional neural network (CNN), and some state of the Arts model such as VGG16 and Resnet50, for traffic sign classification on The GTSRB. using a GPU to increase processing time, the design showed that with various augmentation applied to the CNN, our 8-layer Model was able to outperform the State of the Arts models with a higher test Accuracy, 50 times lesser training parameters, and faster training time our 8 -layer model was able to achieve 96% test accuracy.

Index terms: AlexNet, Convolutional Neural Network, ResNet50, Traffic Signs, VGG16

1. INTRODUCTION

Computer vision has helped solved various complex image classification task, because of the huge dataset which has been made available from the ImageNet classification challenge. Various neural network architectures have helped to push the boundary of research in this field, leveraging on the various new implementations of state of the arts models, which allows for fast experimentation and developments. New implementations of deep learning frameworks have gained large amounts of popularity, with their implementations leveraging the state of the arts and wrapping it in an easy to use package, allowing the fast experimentation and developments seen today. These frameworks also make use of GPU acceleration, leading to significant improvements of training times. Over the years the number of road vehicles have continued to increase, this has resulted in an increase in road accidents, one of the leading cause of these accidents is ignorance of traffic sign [1] .Efficient traffic sign classification has been one of the major areas of research in intelligent transport and advanced driver assistance systems (ADAS) [2]. Advances in computer vision have achieved major milestones, through the development of state of arts models. The deep neural network revolution emerged from the breakthrough of LeNet Model [3]. LeNet model gives a very good performance in image classification, after which several other more sophisticated networks emerged, which were entered into the image Net challenge. From the Alex Net model in 2012, VGG in 2014 [4] etc., to various other models. This work focus the performance of two of the most widely used models the VGG16 and the ResNet50 model for traffic sign classification, however most of this models showed efficient results even better that human performance they require expensive hardware and are computationally expensive. Using up high GPU and millions of parameters making them unsuitable for embedded devices and real time performance [5], this work we develop an 8 layers CNN network. The network was compared with other state of the arts model, such as VGG16 and Resnet50, winners of 2013 and 2015 ImageNet classification challenge respectively [6]. We used the German Traffic Sign Recognition Benchmark (GTSRB), to test the performance of our model. Other parts of this work consist of reviews of other related research, methods used, results and conclusion

Jonah Sokipriala is a lecturer at rivers state university, has interest in computer vision application email: Sokipriala.jonah@ust.edu.ng

2. LITERATURE REVIEW

Image classification is an important aspect of self-driving cars to enable them to distinguish between different traffic signs, different objects on the road and classify them correctly. Convolutional neural networks have been found to give good performance in image classification task. [7] Various research have been extensively conducted using various CNN models on various dataset. These models have shown to vary in performance based on the dataset, such as LeNet model which outperformed other state of the arts models on the Fashion MNIST classification, with a 98% accuracy [8] .One of such CNN model is the Vgg16 Network which consist of more layers, 16 learning layers,13 conv layers and 3 fully connected layers this won the 2014 ImageNet localization challenge and the runner up in the classification challenge developed by [4]. VGG is computationally expensive with a total of 138M parameters , various modifications have been performed on VGGNet, to reduce the number of neurons in the fully connected layers and added more filters to the fifth Conv layer which shows a reduction in the network parameters being used from 138 M to 32M [9],giving a lesser training time and less computationally expensive. One of the major challenges faced by CNN's was that their performance reduces as the network layers increase. Residual network which was developed by [10] the winners of the image Net (ILSVRC) 2015 challenge solves this challenge by introduction of a short cut model, where the input from the previous layer is added with the features of the current layer, before been feed into the activation layer. Various research has been carried on traffic signs classification, some approach involves the application of color-based models with Neural Network approaches such as in [11], [12], this approach suffers from problems such as fading of traffic signs color. Other papers apply machine learning such as HOG, SIFT for feature extraction such as in [13], [14] where hand crafted features are used, this approach, proves to be invariant to illumination and rotation making it better than color-based approach however the method tends to be very difficult and complex [15]. the dimension of the extracted features been too high making it unsuitable for real time performance [16]. Others include feature descriptors and Artificial Neural Network [1], artificial neural networks perform poorly on large complex images [17].Some researchers have approach the challenge using transfer learning, where the network weights from one of the state of arts models which have been trained on large data set is applied to a similar task with lesser data set as in [18], [19].Transfer learning have seen wide spread use following publications of various dataset such as CIFAR-10,COCO,PASCAL,ImageNet, [20]. transfer learning has been proven to be only very effective with problems with limited dataset [21]. various convnet have also been adopted such as [22],[23], most of the research using CNN's attempted to tackle the problem using LeNet 5-layer architecture.

3. METHODOLOGY

CNN usually takes input of fix sizes hence various prepossessing steps are taken to make the training images suitable for our network to train effectively.

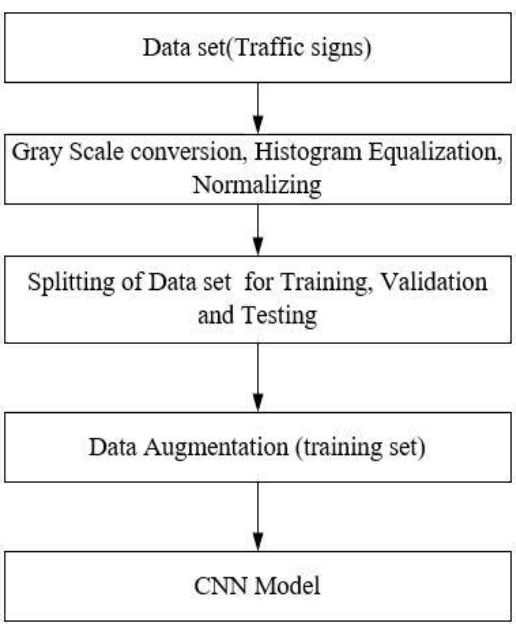

Fig. 1. flow of proposed model

3.1 Description of Dataset

For evaluation of our proposed model we used GTSRB, the GTSRB was extracted from bit bucket repository. The data set contains 43 different classes of labelled traffic signs images of 32 pixels. The data set is split into training (34799, 32, 32, 3), validation (4410, 32, 32, 3) and testing (12630, 32, 32, 3), the data-set in the various classes are non-uniformly distributed fig 2. show some training data set and fig 3. shows the distribution of the data set

Fig. 2. Sample data set

3.2 Pre-processing of Dataset

Data-set is pre-processed to improve the feature extraction from the GTSRB data-set, the training data-set are converted to grey scale to reduce the intensity and reduce the computational cost required ,histogram equalization is then applied to the training set for contrast stretching to ensure uniform distribution of the pixel intensities, this improves the feature extraction, the pixel values are normalized to range between (0 and 1) by dividing each pixel intensity value by 255.this is done to increase convergence speed during training.

3.3 Data Augmentation

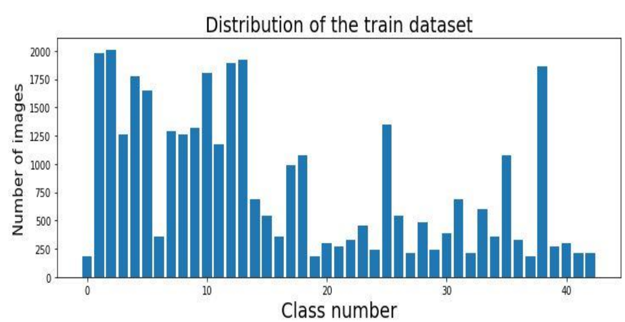

Fig. 3. Distribution of Dataset

3.3 Data Augmentation

one of the major challenges in deep learning models is insufficient training data to enable them to learn relevant features, and an imbalance in the training sets. The insufficiency in the training data usually results in over fitting due to

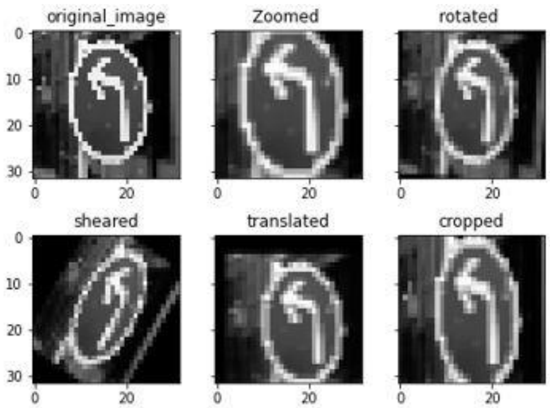

Fig. 4. Augmentation on a sample image

biased results, while the imbalance results in poor classification [24]. Various data augmentations where performed to add variations to the training set, these would allow for increase diversity in the training set without having to add more images to the actual training set. The augmentation where carefully selected so the traffics signs are not wrongly classified as a complete rotation of a left direction sign by 180 might mean a right direction arrow sign.

Table 1 : Augmentation applied

Variations	Value
Translation (x, y)	10%
Zoom	20%
Rotation range	10°
Shear range	10%
Crop	15%

4. MODEL CONFIGURATION

The CNN used was based of some modification of Alex Net, the filter sizes was reduced from that of Alex Net and smaller stride sizes where applied due to the size (32 pixel) and low-resolution image, AlexNet one of the least complex models, which won the image Net challenge and has low computational cost [25] The model we applied modified AlexNet using smaller filter sizes and stride size of 1, we used 5 conv layers, Maxpool with stride 2 for down sampling of feature maps and flatten layer followed by 3 fully connected layer. The final fully connected layers consist of 43 neurons for the 43 classes of the traffic signs and a SoftMax activation function for final classification as shown in fig 5

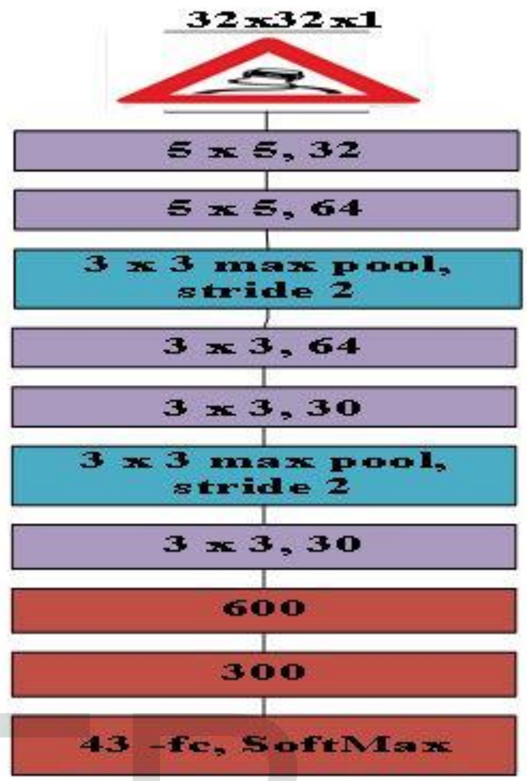

Fig. 5. 8-layer CNN

4.1 Convolution Layer

the conv layer is the major building block of the entire CNN model this layer extracts the features in the input image using convolutional operation, with various learn-able filters of varying filter sizes to extract relevant features. The features extracted are the feature maps which are passed as input to other layers as the network goes deeper.

4.2 Max Pool Layer

Pooling layers provide an approach to down sampling maps by summarizing the presence of features in patches of the feature map, a 3 x 3 max pooling with stride 2 is used to downscale the feature map after the second and fourth convolution layers the 3x3 max pool, moves using 3x3 grid to extract the most relevant feature in that window, as shown in fig 6

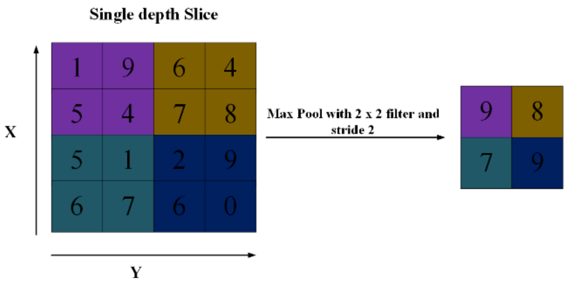

fig 6: Max Pool Operation

4.3 Activation Layer

Relu activation function is mainly applied after each convolutional layer, to prevent the image pixels obtained after the convolution from averaging to zero, therefore all the negative values of the pixels after convolution are converted to zero using the RELU activation function as shown in (1)

$$f(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x \geq 0 \end{cases} \qquad 1$$

4.4 fully connected

After the fifth conv layer block we used a flatten layer, to flatten the feature maps, which are connected to a fully connected layer which updates it weight and bias values based on the gradient decent algorithm during back propagation, finally the final fully connected layer with SoftMax activation, is used to calculate the probability of an input belonging to one of the 43 classes for our traffic signs using (2)

$$\sigma(\vec{r})_i = \frac{e^{r_i}}{\sum_{j=1}^{n} e^{r_j}} \qquad 2$$

Where

$\sigma = softmax$

$\vec{r} = input\ vector$

$e^{r_i} = standard\ exponential\ function\ for\ input\ vector$

$e^{r_j} = standard\ exponential\ function\ for\ ouput\ vector$
$n = number\ of\ classes$

4.5 Model Training

The loss function of the model is computed using categorical cross entropy(CE) loss, as the model is a multi-class data set(43 -different classes of traffic signs) and multiplied with our SoftMax activation function as shown in (3) to output the probability of a given traffic sign to belong to one of the 43 classes, as the 43 class labels are one hot encoded only one element of the target vector t which is not zero ti = tp so all elements which summation are zero can be discarded and (3) can be written as (4).Adam optimizer is used for the classification of our training data due to its ability in handling of sparse gradients, with a learning rate of 0.0001 for slower learning, 10 complete epochs where made in the data sets in mini-batches of 40 for effective generalization of the features to avoid over fitting

$$CE = -\sum_{i}^{C} t_i \, log(\sigma(s)_i) \qquad 3$$

$$CE = -log\left(\frac{e^{s_p}}{\sum_{j}^{C} e^{s_j}}\right) \qquad 4$$

c = number of classes

s_p= CNN score

t_i= ground truth

s_j = CNN score for each class.

5. RESULTS AND ANALYSIS

we compared the validation accuracy of three different Models, Resnet-50 with 50 trainable layers, Vgg-16, with 16 trainable layers, and our 8 layer as shown in fig 7, the 8 layer model showed the best performance. It had a validation accuracy of 99 % after 10 epochs, performing better than the state of arts models.

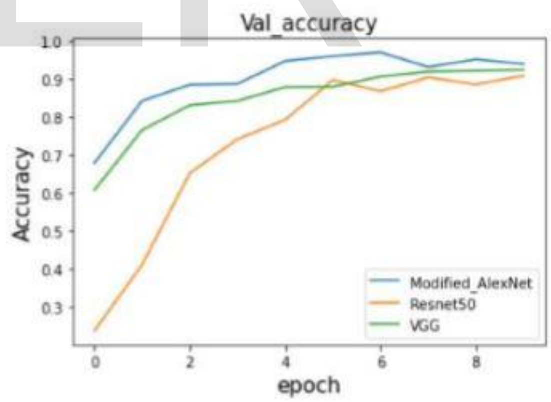

fig 7:accuracy of Resnet-50, vgg-16, 8-layer CNN

The three models were compared for just 10 epochs which showed the list epoch for their convergence, the validation loss for the 8-layer CNN was also the least percent, for the GTRSB used, it was observed that the deeper the layers went the poorer the network performed which can be seen from the comparison of the test accuracy between where VGG 16 performed better than ResNet50. The 8-layer CNN also used the least training parameters as shown in fig 8 is a Resnet model and the vgg used rgb channels and overfitted the training data as shown in fig 9 and fig 10 where the training accuracy overlaps that of the validation, fig 11 shows that the 8 layer CNN did not experience overfitting as the validation accuracy was higher than that of the training accuracy. After the augmentation of the dataset, the modified Alexnet(8-layer CNN) had a higher validation accuracy of 99% as compared to the other state of arts models which shows that it can generalize better in classifying various unlabeled images the model was plot before and after augmentation it showed that the model validation increased after augmentation. this shows that augmentation increase the performance of the model as shown in fig 12.

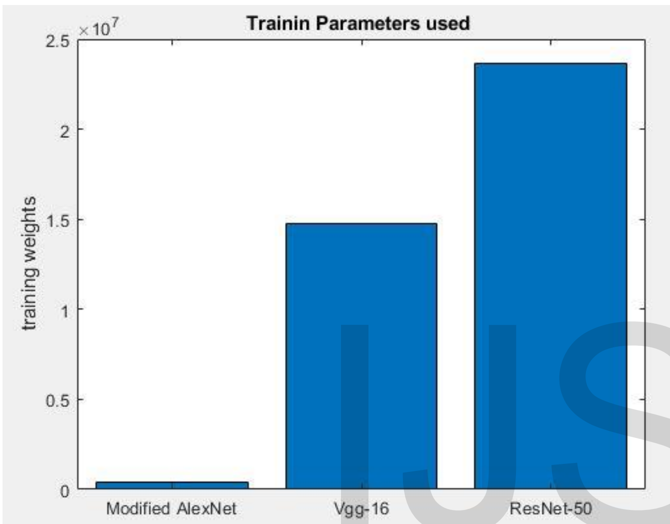

fig 8: Training parameters used by CNN Models

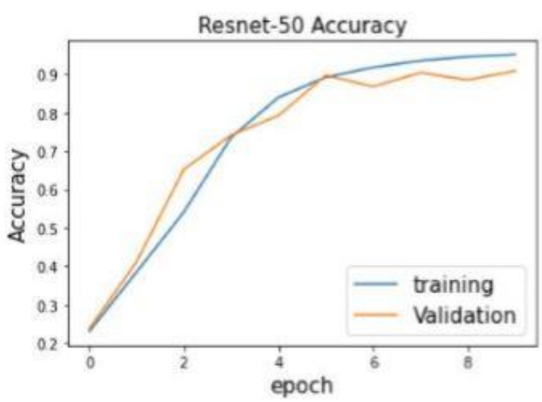

fig 9: Resnet-50 Validation vs Training accuracy

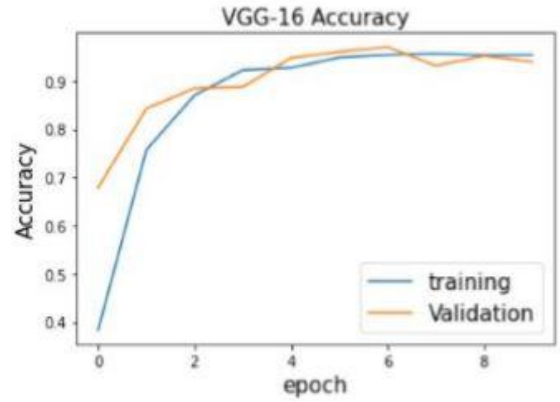

fig 10:validation and training accuracy of 8-layer CNN

After 10 epochs the validation and the training accuracy seems to strongly overlap which shows that the model is overfitting to the training set, hence would not be suitable for use in classifying unlabelled traffic sign

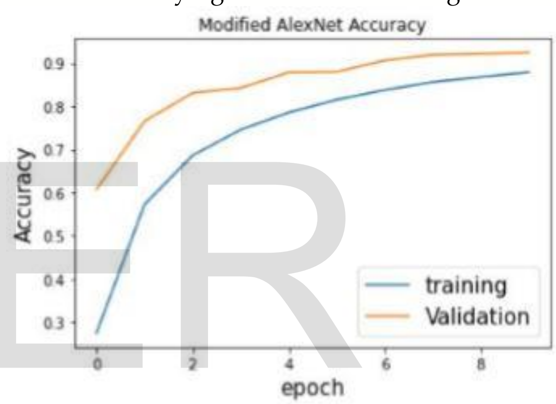

fig 11:validation and training accuracy of 8-layer CNN

From fig 10 the variation brought by the various augmentation helped improve the performance of the model from its normal state.

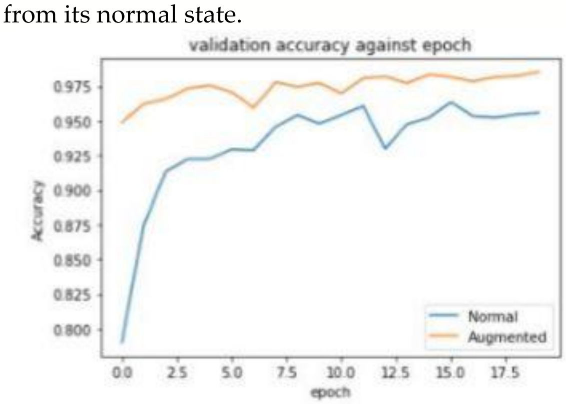

fig 12:Augmented vs Normal for 8-layer CNN

Model Evaluation parameter for test set

$$Precision = \frac{True\ Positive}{True\ Positive + false\ positive}$$

$$Recall = \frac{True\ positive}{True\ Positive + False\ Negative}$$

$$f1\ score = \frac{2\ x\ precision\ x\ recall}{Precision + Recall}$$

The f1 score was used to measure the performance of the model on the test set as the f1 gives a good measure of the accuracy of the model since it combines the recall and precision together in which case errors with extreme values would be well punished making the f1 score a better metric in analyzing the accuracy of the various models as shown from table 2 our 8 layer CNN performed better than VGG-16 and Resnet-50 with 96.0% accuracy on test set

Table 2: of Test Evaluation

SN	Model	Test Accuracy
1	Vgg-16	95.5
2	Resnet-50	95.4
3	8 -Layer CNN Model	96.0

6. CONCLUSION

The State of the arts Model such as VGG16 and ResNet50 did performe poorly on the GTSRB, while they have performed better than on a larger data set such as the ImageNet, this would have been result of the pixel sizes been of the images 32x32 while those for the image net where 224x 224,addition of more augmentation would have resulted in a more training image could improve the performance of the state of arts Models, further finetuning of simpler model would be required for embedded devices. Resizing of the image to the actual sizes used in the ImageNet or use of higher resolution images would have giving better performance for the State of the Arts models data augmentation played a very important role in reducing over-fitting by addition of variety of images more augmentation technique can be applied to improve the model performance.

REFERENCES

[1] M. Z. Abedin, P. Dhar, and K. Deb, “Traffic sign recognition using hybrid features descriptor and artificial neural network classifier,” in 2016 19th International Conference on Computer and Information Technology (ICCIT). IEEE, 2016, pp. 457–462.

[2] W. Li, Q. Chen, T. Dong, L. Wei, and Q. Zhang, “Traffic signs classification based on local characteristics and elm,” in 2017 10th International Symposium on Computational Intelligence and Design (ISCID), vol. 1. IEEE, 2017, pp. 127–130.

[3]Y.LeCun,L.Bottou,Y.Bengio,andP.Haffner,“Gradient-basedlearning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014

[5] A. Mhalla, T. Chateau, S. Gazzah, and N. E. B. Amara, “An embedded computer-vision system for multi-object detection in traffic surveillance,”IEEE TransactionsonIntelligentTransportationSystems,vol.20, no. 11, pp. 4006–4018, 2018.

[6] M. A. E. Muhammed, A. A. Ahmed, and T. A. Khalid, “Benchmark analysis of popular imagenet classification deep cnn architectures,” in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE, 2017, pp. 902–907.

[7] E. Chen, X. Wu, C. Wang, and Y. Du, “Application of improved convolutional neural network in image classification,” in 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). IEEE, 2019, pp. 109–113.

[8] M. Kayed, A. Anter, and H. Mohamed, “Classification of garments from fashion mnist dataset using cnn lenet-5 architecture,” in 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE). IEEE, 2020, pp. 238–243.

[9] Y. Zhiqi, “Gesture recognition based on improved vggnet convolutional neural network,” in 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2020, pp. 1736–1739.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[11] M. A. A. Sheikh, A. Kole, and T. Maity, “Traffic sign detection and classification using colour feature and neural network,” in 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI). IEEE, 2016, pp. 307–311.

[12] C. Rahmad, I. F. Rahmah, R. A. Asmara, and S. Adhisuwignjo, “Indonesian traffic sign detection and recognition using color and texture feature extraction and svm classifier,” in 2018 International Conference on Information and Communications Technology (ICOIACT). IEEE, 2018, pp. 50–55.

[13] P. H. Kassani, J. Hyun, and E. Kim, “Application of soft histogram of oriented gradient on traffic sign detection,” in 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). IEEE, 2016, pp. 388–392.

[14] S. Vashisth, S. Saurav et al., “Histogram of oriented gradients based reduced feature for traffic sign recognition,” in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2018, pp. 2206–2212.

[15] D. Co¸tovanu, C. Zet, C. Fos¸al˘au, and M. Skoczylas, “Detection of traffic signs based on support vector machine classification using hog features,” in 2018 International Conference and Exposition on Electrical And Power Engineering (EPE). IEEE, 2018, pp. 0518–0522.

[16] M. D. RADU, I. M. COSTEA, and V. A. STAN, “Automatic traffic sign recognition artificial inteligence-deep learning algorithm,” in 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, 2020, pp. 1–4.

[17] X. He and B. Dai, “A new traffic signs classification approach based on local and global features extraction,” in 2016 6th International Conference on Information Communication and Management (ICICM). IEEE, 2016, pp. 121–125.

[18] H. Fleyeh, R. Biswas, and E. Davami, “Traffic sign detection based on adaboost color segmentation and svm classification,” in Eurocon 2013. IEEE, 2013, pp. 2005–2010.

[19] L. Wei, L. Runge, and L. Xiaolei, “Traffic sign detection and recognition via transfer learning,” in 2018 Chinese Control And Decision Conference (CCDC). IEEE, 2018, pp. 5884–5887.

[20] L. Studer, M. Alberti, V. Pondenkandath, P. Goktepe, T. Kolonko, A. Fischer, M. Liwicki, and R. Ingold, “A comprehensive study of imagenet pre-training for historical document image analysis,” in 2019

International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019, pp. 720–725.

[21] Z. Nadeem, A. Samad, Z. Abbas, and J. Massod, "A transfer learning-based approach for pakistani traffic-sign recognition; using convnets," in 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). IEEE, 2018, pp. 1–6.

[22] I. Amin, S. Hassan, and J. Jaafar, "Semi-supervised learning for limited medical data using generative adversarial network and transfer learning," in 2020 International Conference on Computational Intelligence (ICCI). IEEE, 2020, pp. 5–10.

[23] S. Mehta, C. Paunwala, and B. Vaidya, "Cnn based traffic sign classification using adam optimizer," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE, 2019, pp. 1293–1298.

[24] F. Shahidi, S. M. Daud, H. Abas, N. A. Ahmad, and N. Maarop, "Breast cancer classification using deep learning approaches and histopathology image: A comparison study," IEEE Access, vol. 8, pp. 187531–187552, 2020.

[25] F. R. Mashrur, A. D. Roy, and D. K. Saha, "Automatic identification of arrhythmia from ecg using alexnet convolutional neural network," in 2019 4th International Conference on Electrical Information and Communication Technology