Traffic sign recognition using CNN and Res-Net

J Cruz Antony[1], G. M. Karpura Dheepan[1], Veena K [1*], Vellanki Vikas[1] and Vuppala Satyamitra[1]

[1]Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India

Abstract

In the realm of contemporary applications and everyday life, the significance of object recognition and classification cannot be overstated. A multitude of valuable domains, including G-lens technology, cancer prediction, Optical Character Recognition (OCR), Face Recognition, and more, heavily rely on the efficacy of image identification algorithms. Among these, Convolutional Neural Networks (CNN) have emerged as a cutting-edge technique that excels in its aptitude for feature extraction, offering pragmatic solutions to a diverse array of object recognition challenges. CNN's notable strength is underscored by its swifter execution, rendering it particularly advantageous for real-time processing. The domain of traffic sign recognition holds profound importance, especially in the development of practical applications like autonomous driving for vehicles such as Tesla, as well as in the realm of traffic surveillance. In this research endeavour, the focus was directed towards the Belgium Traffic Signs Dataset (BTS), an encompassing repository comprising a total of 62 distinct traffic signs. By employing a CNN model, a meticulously methodical approach was obtained commencing with a rigorous phase of data pre-processing. This preparatory stage was complemented by the strategic incorporation of residual blocks during model training, thereby enhancing the network's ability to glean intricate features from traffic sign images. Notably, our proposed methodology yielded a commendable accuracy rate of 94.25%, demonstrating the system's robust and proficient recognition capabilities. The distinctive prowess of our methodology shines through its substantial improvements in specific parameters compared to pre-existing techniques. Our approach thrives in terms of accuracy, capitalizing on CNN's rapid execution speed, and offering an efficient means of feature extraction. By effectively training on a diverse dataset encompassing 62 varied traffic signs, our model showcases a promising potential for real-world applications. The overarching analysis highlights the efficacy of our proposed technique, reaffirming its potency in achieving precise traffic sign recognition and positioning it as a viable solution for real-time scenarios and autonomous systems.

*Corresponding author. Email: veenakanagaraj07@gmail.com

1. Introduction

In recent years, Computer Vision techniques have proven immensely effective in addressing diverse image classification tasks. Particularly, Deep Learning has emerged as a powerful paradigm within the realm of Computer Vision, demonstrating exceptional accuracy across a wide spectrum of image-related tasks. Neural networks, notably CNN, have emerged as a cornerstone for image classification tasks, surpassing traditional algorithms such as SVM, KNN, and Decision Trees, particularly in scenarios involving extensive databases. The architectural design of neural networks often involves a human-guided exploration of experimental outcomes. CNN, with its layered structure encompassing convolution, max-pooling, and activation layers, has gained remarkable popularity for tasks like object classification, handwriting recognition, digit identification, speech recognition, and face detection, attributing its success to its precision and rapid computational capabilities.

As the volume of vehicles on roads continues to rise, the issue of road safety has become more pressing due to

an increasing number of accidents. A primary contributing factor to these accidents is a lack of awareness regarding traffic signs. Consequently, the classification of traffic signs has emerged as a pivotal research area in the development of Advanced Driver Assistance Systems - ADAS, aimed at furnishing drivers with safety measures and guidance based on their surrounding environment. Vital features of ADAS encompass understanding the environment, detecting various objects, and classifying them into categories like vehicles, roads, signs, and pedestrians. An effective ADAS must accurately recognize traffic signs under diverse conditions such as adverse weather (rain, dust), low-light scenarios (nighttime), or instances where parts of the image are obscured. This recognition process encompasses two steps: first, accurately locating and detecting traffic signs, followed by their classification. This paper centers on constructing an efficient CNN architecture augmented with Residual blocks for traffic sign recognition, specifically focusing on the Belgium traffic signs dataset. The incorporation of Residual networks is advantageous due to their mitigation of the Vanishing Gradient Problem through skip connections, rendering them more efficient than other deep neural networks.

Localization and classification are the two fundamental processes in traffic sign classification. The primary colors on most traffic sign boards are red, white, and black. The authors used image-filtering and template matching to locate the signs. The authors have introduced a traffic signal classification methodology that heavily relies on color and shape characteristics. This approach encompasses three main stages:

Segmentation of the image, carried out within the HSV color space. Identification of traffic signs based on their distinctive geometric shapes, such as circles, triangles, and other relevant forms. Feature extraction and subsequent classification utilizing Gabor filters and Support Vector Machines (SVM) [1].

2. Literature review

The authors proposed a traffic signal classification method using Support Vector Machines and image segmentation. They outlined it in three main steps. First, image pre-processing was performed using Canny edge detection to eliminate noise by applying a Gaussian filter. In the second stage, they separated the traffic signal from the background using the Hough transform algorithm, which identifies shapes like circles and triangles. SVM algorithms with various kernels were used for classification [2]. In this approach, the authors designed a hierarchical architecture to enhance the recognition accuracy of traffic signs [3]. The authors have introduced a technique for traffic sign recognition that combines deep convolutional features with an extreme learning classifier [4]. The authors introduced a traffic sign recognition approach that employs a CNN and leverages the positions

determined by max-pooling [5]. The authors' primary focus was on the development of a real-time CNN for traffic sign recognition. Additionally, their paper incorporates performance metrics like accuracy, precision, recall, and F1-score to showcase the efficacy of their CNN design for real-time traffic sign recognition [6]. The authors explored the use of various CNN architectures for detecting emotional states. The paper also aimed to employ deep learning techniques to automatically detect these emotional states from visual data, such as images or video frames [7].

The main advantage of color-based filtering is its shorter computational time. However, when color-based filtering is applied, it may yield lower accuracy when some parts of the image are occluded in real-time. The Hough transform algorithm is a widely used approach, but it can be challenging to implement in real-time, especially when dealing with high-definition video frames. The authors proposed an algorithm to address these issues by utilizing color shape Regular Expressions, which are implemented on DFA and N-DFA [8].

The authors suggested a customized AlexNet configuration for the task of traffic sign recognition using the German Traffic Sign Recognition Benchmark (GTSRB) dataset. They conducted a comparative analysis against the original ResNet-50 and VGG-16 architectures to evaluate the performance differences. During the preprocessing stage, they converted the colored image into a grayscale image to reduce intensity and computational costs. They then applied histogram equalization for the uniform distribution of pixel intensities. A modified AlexNet architecture was used for classification, employing smaller window sizes due to the very small image size (32x32). The model they proposed achieved a 96% accuracy rate, which is greater when compared to the original ResNet-50 and VGG-16 architectures. Many traffic sign classification algorithms primarily utilize Convolutional Neural Networks for feature extraction and categorization. After completing all the convolution and max-pooling operations, the layers are flattened, resulting in what are called fully connected layers. The fully connected layers, when formed into Classical Neural Networks and trained using the conventional Gradient Descent technique, have limited generalization potential. To overcome this constraint, they eliminated the fully connected layers from the CNN architecture. Instead, they used the feature extractors as input for the Extreme Learning Machine (ELM) classifier, which is a Single Hidden Layer Feedforward Neural Network. In this ELM classifier, they introduced random initialization of biases and weights for the input layer, and they applied a variety of activation functions to each neuron in the hidden layer. Their proposed model achieved an impressive accuracy of 99.40% with a hidden layer containing 12,000 nodes in the ELM classifier [9].

In the traditional multi-class CNN classification approach, the output layer calculates the probability for each class, and the one with the highest probability determines the input object's class. However, using this

N-way classification method, some categories are more frequently misclassified than others. To address this issue, they clustered the 43 classes of traffic signs into 6 subsets. They used a similarity matrix for clustering and CNN for partitioning into 6 subsets. Then, they designed separate CNN models for each subset. This method offers several advantages. Firstly, the hierarchical structure enables more optimized resource allocation, allowing the network size to be customized for each category problem. Secondly, a smaller network with fewer classes to recognize is easier to train than a larger one. The model they proposed achieved an accuracy of 99.67%. As per their model, following the completion of all convolution operations, 250 feature maps comprising 4x4 neurons each are obtained. Subsequently, they employed a max-pooling operation with a window size of 2x2, encoding the positions of the maximum values into 4-bit binary representations. Finally, the entire Max Pooling Positions (MPPs) sequence is obtained by concatenating all the binary values. They proposed a classification approach in 5 stages: Data Collection, Data Processing, Activation Selection, Classifier Initialization, and Classifier Fine Tuning. Furthermore, they conducted a classification similarity analysis on Max-Pooled Positions (MPPs) belonging to the same class. The model they introduced yielded an accuracy rate of 96.45%.

Their work represents significant advancements in this field. Their study introduces an inventive approach that harnesses CNN technology to create a resilient Traffic and Road Sign recognition system. The research evaluates the efficacy of this architecture through an original dataset, specifically the Tunisian traffic signs dataset. Notably, the authors have streamlined the LeNet network by minimizing its layer count, thereby reducing network parameters to enhance computational efficiency. The architecture's adaptability to diverse parameters is emphasized, aimed at optimizing recognition rates in challenging real-world scenarios encompassing variables such as adverse weather, intricate backgrounds, fluctuating lighting, and fading sign colors. Remarkably, the experimental outcomes showcase a substantial enhancement in accuracy, surpassing previous benchmarks in similar studies [10].

In their recent research, the authors emphasized the crucial role of traffic signs in maintaining road order, ensuring driver adherence to regulations, and preventing accidents, injuries, and fatalities. Their paper introduced a groundbreaking autonomous framework grounded in deep learning to facilitate the efficient recognition of traffic signs, particularly within the context of India [11]. The authors published a paper in "Expert Systems with Applications" in 2022, wherein they conducted a study involving the application of deep learning for the detection and classification of Mexican traffic signs. In their research, they incorporated performance metrics like accuracy, precision, recall, F1-score, and possibly IoU (Intersection over Union) to evaluate the effectiveness of both detection and classification outcomes [12]. The authors recognized a significant shortcoming in current

methods, namely their inefficiency in extracting essential features. This limitation resulted in decreased detection accuracy and a higher frequency of misclassification errors. In response, they introduced a pioneering automated number plate recognition method. This innovative approach sought to attain exceptionally precise number plate identification while mitigating the occurrence of errors [13].
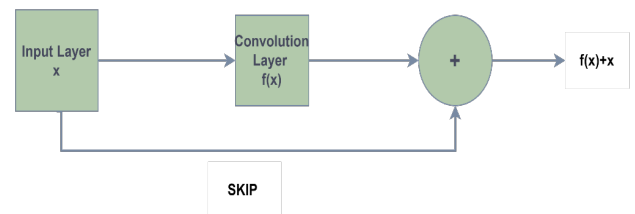

Figure 1. Representation of Skip Connection

2.1. Model proposed

Numerous traditional machine learning algorithms, such as Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Random Forest, have demonstrated impressive accuracy across various image classification tasks. However, their effectiveness tends to diminish when confronted with exceedingly large databases. To address this challenge, Neural Networks, particularly CNNs, come into play. One of the prominent drawbacks of conventional Neural Networks is their requirement for a substantial number of parameters when dealing with high-quality images in classification tasks. This issue is mitigated by the use of CNNs in numerous classification tasks, where they excel in Feature Extraction, leading to Dimensionality Reduction. In the CNN technique, a sequence of operations, including convolution, max-pooling, and activation, are applied to the data.

Convolution Layer
The convolutional layer is the fundamental building block of the whole CNN model; this convolutional layer does the feature extraction from its input using convolution operations (dot product of the input image with a kernel) by using various filters of different window sizes. In the proposed model six Convolution layers are used. After the convolution operation RelU activation function is performed. This output will be the input to the next layers of the model.

Activation Layer
The activation function plays a pivotal role in determining whether a neuron should be activated or not, achieved by calculating a weighted sum and adding bias to it. Its primary function is to introduce non-linearity into a

neuron's output. In our version ReLU activation function is applied after convolution operation, to save you the image pixels acquired after the convolution from averaging to 0, consequently all of the terrible values of the pixels after convolution are transformed to 0 using the ReLU activation feature:

$$F(x) = x \; if \; x \geq 0 = 0 \tag{1}$$

Max Pool Layer
Max Pool layer extracts the most relevant features from the window.

Fully Connected Layer
After going through all the convolutional layers, at last the output will be flattened and this layer is known as Fully Connected layer. Some dense layers were added after a fully connected layer. Output layer is generated by applying a soft-max activation function which gives probability. The CNN-Architecture used in this report is a combination of several convolutional layers, max pool layers along with Residual block in ResNet Architecture.

Residual Block
Res Net is known as Residual Network. It was the first architecture which introduced the concept of skip connection. Residual network is more efficient when compared to other deep neural networks because with resnet there won't be any Vanishing Gradient Problem due to skip connections.

Vanishing Gradient Problem:
The Vanishing Gradient Problem arises when training deep neural networks using back-propagation and the chain rule of derivatives. In cases where the model has a large number of layers, gradients can become extremely small as they are multiplied together during the backward pass. This can lead to some layers not effectively contributing to the training process, making it challenging to train the model effectively.

To address the Vanishing Gradient Problem, skip connections was implemented. Skip connections involve taking the original input and adding it to the output of a convolutional block, ensuring that no features or data from the input are lost during the process. This helps gradients flow more efficiently during training and enables better learning in deep neural networks.

Let X be the input layer, F(X) be the output after Convolution, Y be the output after including the input image to the output of the convoluted layer. Y = F(X) + X. Here F(X) is made as zero so that there won't be loss of any features. Since there is no loss of information of the input, there will be an increase in overall accuracy.
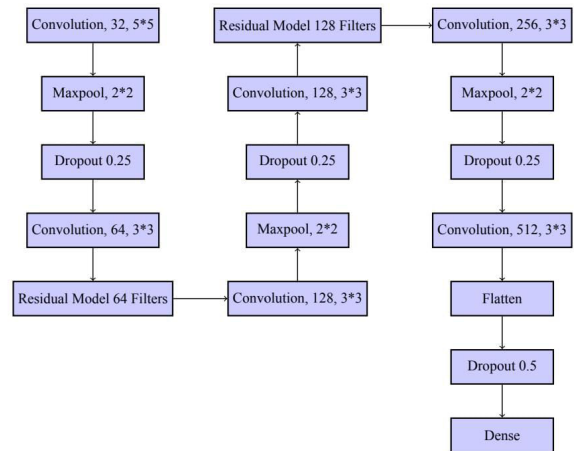
Architecture Layout

Figure 2. Architecture diagram

Experiment

Dataset Description
For the evaluation of our model Belgium Traffic Signs BTS dataset was used which contains traffic signs of 62 different classes of various sizes. The road sign images in these databases extracted from the environment do not take into account the interference caused by motion blur and rainy weather. The dataset is divided into two classes: Train (4575) and Test (2520). Some of the images in the dataset are of low quality and low brightness. If this dataset is directly trained with the proposed model, there will be less accuracy. So before training our model pre-processing of the data is to be done.
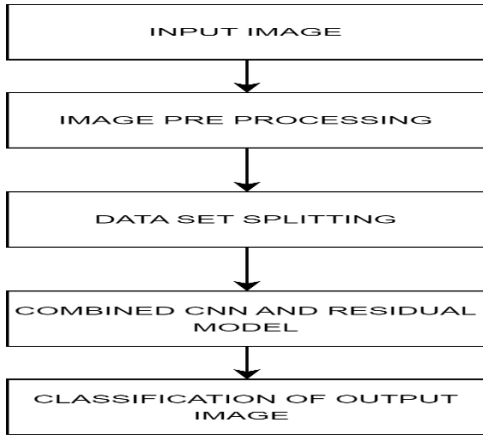

Figure 3. Different Classes of Belgium Traffic Signs

Figure 4. Flow chart Traffic sign recognition using CNN and Res-Net

Pre-Processing

There won't be any ideal images hence some changes are required before further processing. Pre-processing is done to improve feature extraction. In the BTS dataset there are many images with low light and brightness which affects the performance of our model. As CNN is done on images of same size every image is resized to 100*100 size. Since some of the images in the dataset are of low brightness indicates that pixel intensities in the histogram of the image are concentrated at one place only, i.e; there is no uniform distribution of pixel intensities. Histogram Equalization is then implemented to the schooling set for comparison stretching to make certain unique distribution of pixel intensities. Normalizing is done through dividing the pixel values with 255.Normalizing is done for better performance of the model in terms of both speed and accuracy. In BTS there are two sets Train and Test. The train set is divided into train and validation sets with different split percentages.

Test Setup

The proposed system is implemented with the Keras library. Experiments are carried out by implementing our proposed model (CNN + Residual) on the Belgian Traffic Sign dataset. A human optimizer is employed for network training. The ReLU activation function is applied to each neuron in the CNN. The loss function utilized is categorical entropy.

Table 1. Details of Architecture.

Layer Type	Output shape	Parameter
Sequential	(100,100,3)	0
Conv2d_2	(50,50,32)	2432
max_pooling_2d_1	(25,25,32)	0
dropout_1	(25,25,32)	0
Conv2d_3	(25,25,64)	18496
Conv2d_4	(25,25,64)	36928
Conv2d_5	(25,25,64)	36928
Conv2d_6	(25,25,64)	36928
add	(25,25,64)	0
Conv2d_7	(25,25,64)	36928
add_1	(25,25,64)	0
Conv2d_8	(25,25,128)	73856
max_pooling_2d_2	(12,12,128)	0
dropout_2	(12,12,128)	0
Conv2d_9	(12,12,128)	147584
Conv2d_10	(12,12,128)	147584
Conv2d_11	(12,12,128)	147584
Conv2d_12	(12,12,128)	147584
add_2	(12,12,128)	0
Conv2d_13	(12,12,128)	147584
add_3	(12,12,128)	0
Conv2d_14	(12,12,256)	295168
max_pooling_2d_3	(6,6,256)	0
dropout_3	(6,6,256)	0
Conv2d_15	(6,6,512)	1180160
flatten	18432	0
dropout	18432	0

Results

The proposed architecture is applied to the Belgium traffic sign dataset (BTSD) for efficient classification of Belgium traffic signs. The detailed information about our proposed model is presented in Table 1 below, which includes the output shape of the tensor after each layer and the number of trainable parameters. The data collected for Belgian traffic signs is used to evaluate the suggested architecture (BTSD). This document effectively categorizes traffic signs in Belgium and provides a comprehensive view of our proposed model in Table 1, showcasing the number of trainable parameters and the output shape of the tensor after each layer. In order to determine the best split, our model was experimented with different split ratios, such as 9:1, 8:2, and 7:3. The corresponding results are shown in Table 2.

Table 2. Accuracy comparison of different Validation split

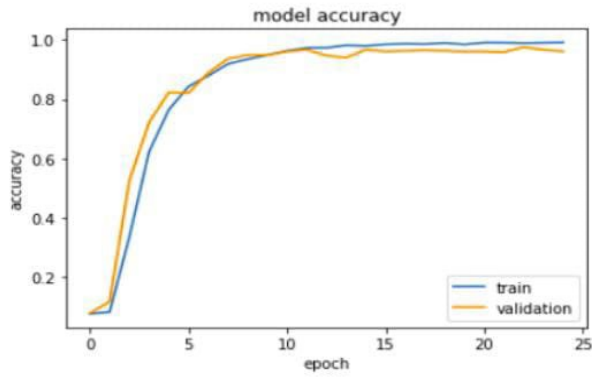| Train: Validation Split Ratio	Train : Accuracy	Validation: Accuracy	Test : Accuracy
9:1	98.6%	96.70%	93.17%
8:2	99.0%	96.9%	91.1%
7:3	99.19%	95.44%	92.66%

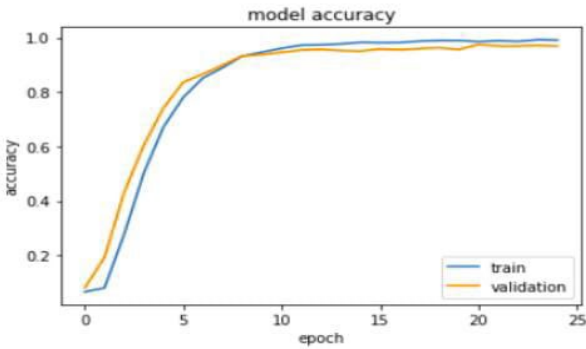**Figure 5.** Accuracy vs Epoch for 10% Split

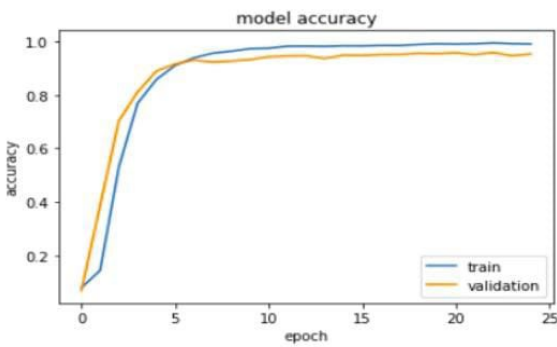**Figure 6.** Accuracy vs Epoch for 20% Split

Figure 7. Accuracy vs Epoch for 30% Split

From the results obtained, it is evident that in each case, the training accuracy surpasses the validation accuracy, indicating the presence of overfitting. This overfitting issue is attributed to the small size of the dataset. To mitigate this, data augmentation techniques were applied, including Horizontal Flip, Rotation, and Zoom. Data augmentation increases the dataset size by introducing slightly modified copies of existing data or generating synthetic data from the existing dataset. The results when Data Augmentation is applied are presented in Table 3.

Table 3. Accuracy comparison of different Validation splits with Data Augmentation

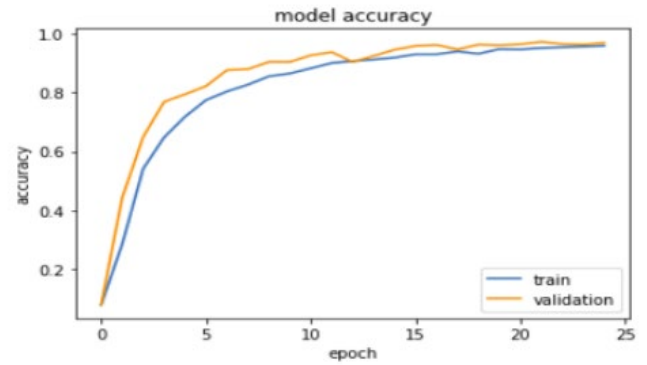| Train: Validation Split Ratio	Train : Accuracy	Validation: Accuracy	Test : Accuracy
8:2	95.76%	96.81%	94.25%
7:3	95.97%	96.22%	93.53%

Figure 8. Accuracy vs Epochs for 20% Split with data augmentation

The plot shows how the model's accuracy improves over the course of training epochs when using a 20% data split for training and employing data augmentation techniques to enhance the model's ability to generalize and recognize traffic signs.
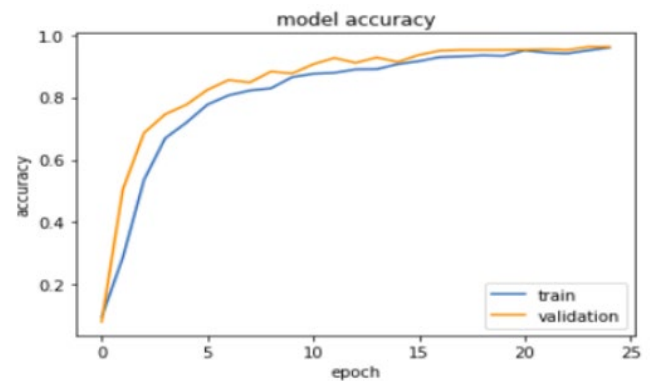

Figure 9. Accuracy vs Epochs for 30% Split with data augmentation

The plot illustrates how the model's accuracy evolves over training epochs when trained on a 30% data split, and data augmentation techniques are applied to improve the model's generalization and its ability to recognize traffic signs. By using Data Augmentation, the problem of overfitting was resolved, and test accuracy also increased.

Table 4. Metrics v/s Score representation or 30%.

Metrics	Score
Precision	95.2%
Recall	94.2%
F1 Score	94.4%

Conclusion

In this study, a combination of CNN and residual models for traffic sign identification was presented, demonstrating their performance using the BTS dataset. The Residual block offers the advantage of mitigating the Vanishing Gradient Problem in this context. Additionally, tested the model proposed using the GTSRB dataset and Indian Traffic Signs dataset, achieving accuracies of 95.76% and 95%, respectively. The model was trained using various alternative architectures, including LeNet and Inception Net, as part of the ongoing research. These models have shown exceptional performance in accurately identifying and classifying traffic signs. The precision score of 95.2% underscores the system's proficiency in correctly identifying a significant portion of positive instances among all those predicted as positive. Moreover, the system exhibits a recall rate of 94.2%, affirming its capability to effectively capture a substantial portion of actual positive instances within the dataset. The F1 Score, at 94.4%, signifies a harmonious equilibrium between precision and recall, showcasing the system's strength in achieving both accurate classification and comprehensive detection of traffic signs. Collectively, these results affirm the effectiveness and reliability of the approach in enhancing road safety through advanced traffic sign recognition technology.

References

[1] Chen, Y, Xie, Y, Wang, Y. Detection and recognition of traffic signs based on hsv vision model and shape features. Journal of Computers. 2013; Vol. 8: pp. 1-8.

[2] Chaiyakhan, K, Hirunyawanakul, A, Chanklan, R, Kerdprasop, K, and Kerdprasop, N. In: Seiichi Serikawa, Traffic sign classification using support vector machine and image segmentation Proceedings of the 3rd International Conference on Industrial Application Engineering 2015; March 2015; pp. 52–58.

[3] Mao, X, Hijazi, S, L, Casas, R, A, Kaul, P, Kumar, R, Rowen, C. Hierarchical CNN for traffic sign recognition. Proceedings of IEEE Intelligent Vehicles Symposium (IV), 19-22 June 2016; Gothenburg, Sweden; 2016. Pp 130-135.

[4] Zeng, Y, Xu, X, Fang, Y, Zhao, K. Traffic sign recognition using extreme learning classifier with deep convolutional features. Zhi-Hua Zhou, Proceedings of the 2015 International Conference on Intelligence Science and Big Data Engineering. June 2015, Suzhou, China, 2015, pp. 272-280.

[5] Qian, R, Yue, Y, Coenen, F, Zhang, B.: Traffic sign recognition with convolutional neural network based on max pooling positions. Proceedings of the 12th International conference on natural computation, fuzzy systems and knowledge discovery. Aug 2016, China, pp. 578–582.

[6] Shustanov, A, Yakimov, P.: CNN design for real-time traffic sign recognition. Procedia engineering, April 2017, Russia, Vol. 201, pp. 718–725.

[7] Theckedath, D, Sedamkar, R. Detecting affect states using vgg16, resnet 50 and se-resnet 50 networks. SN Computer Science. 2020; Vol. 1(2), pp. 1–7.

[8] Nikonorov, A, V, Petrov, M, V, Yakimov, P, Y.: Traffic sign detection on gpu using color shape regular expression neural network models. International Journal of Scientific and Engineering Research. 2016; Vol.12, pp.165–171.

[9] Jonah, S, Orike, S. Traffic sign classification comparison between various convolution neural network models. International Journal of Scientific and Engineering Research. 2021; Vol. 12, pp. 165–171.

[10] Hana, B, F, Amani, C, Jamel, B, Hassen, F, Chokri, S. An efficient implementation of traffic signs recognition system using CNN, Microprocessors and Microsystems. 2023; Vol. 98. pp 1-8.

[11] Rajesh, K, M, Kondareddy, T, Sreevatsava, R, M, Hemanth, N, Lokesh, G. Indian traffic sign detection and recognition using deep learning. International Journal of Transportation Science and Technology. 2023 Vol. 12(3), pp. 683-699.

[12] Rúben, C, R, Carlos, M, C, Osslan, O, Vergara, V. Mexican traffic sign detection and classification using deep learning. Expert Systems with Applications. 2022; Vol. 202, pp 1-8.

[13] Sathya, R, Ananthi, S, Vaidehi, K, Prasanth, A. A Hybrid Location-dependent Ultra Convolutional Neural Network-based Vehicle Number Plate Recognition Approach for Intelligent Transportation Systems. Concurrency and Computation: Practice and Experience. 2023; Vol.35, pp.1-25.