International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1504-1511www.ijaem.net ISSN: 2395-5252

Traffic Sign Board Recognition and Alerting System

M.Rama Krishna Murthy [1] P. Rakesh[2] S. Akriti[3] S. Vamsi Kiran[4] L. Vamsi Krishna[5]

[1] *Professor, Department of computer science & engineering, Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, Andhra pradesh, India*

[2345] *Students, Department of computer science & engineering, Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, Andhra pradesh, India.*

Submitted: 15-05-2022 Revised: 20-05-2022 Accepted: 25-05-2022

ABSTRACT: Road traffic accidents are one of the leading causes of deaths in today's generation. The reckless driving and a busy road are the fuel for this developing problem. As streets get denser with various vehicles, it is hard for the person who is driving the vehicle to recognize all the traffic signs on the way. Chances of passing up crucial signs that might give rise to deadly mishaps and wounds can't be neglected. To inform and help drivers to overcome this growing problem, camera-based traffic sign recognition and alerting systems are used as a part of the latest driver assistance system. In the proposed system, a deep learning-based road traffic signs boards acknowledgment technique is created which is exceptionally uplifting in the improvement of Advanced Driver Assistance Systems and also plays a crucial role in semi-autonomous vehicles. The proposed system is trained with the help of a Convolutional Neural Network (CNN) which is useful in traffic sign classification and identification. The output produced by the convolutional layer is connected to the Multilayer Perceptron (MLP). Traffic Signs Recognition by building a deep neural network model can classify live captured images of traffic signs into different categories. After the identification of the sign board by the system, a voice alert is sent through the speaker of the system which alerts the driver about the type of sign captured by the webcam. The region of interest (ROI) in the captured frame is detected by the hue saturation value(HSV) technique, this technique is used for color filtering. There are various kinds of traffic signs boards that delivers instructions to the drivers, like speed limits sign boards, warning signs , traffic signals , children crossing, no passing of heavy vehicles, etc. Traffic signs classification is the procedure of identifying to which class a traffic sign belongs. The training of the model is done using the German Traffic Sign Dataset and accomplishes great outcomes in perceiving traffic signs.

Keywords: Identification, Classification, Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Deep Neural Network, Road signs, Semi-Autonomous vehicles, Region of Interest(ROI), Hue Saturation Value(HSV).

I. INTRODUCTION

In current traffic management systems, there is a high probability that the driver may miss some of the traffic signs on the road because of overcrowding due to neighboring vehicles. To beat this problem our proposed system can be used in semi-autonomous vehicles which may lead to a reduction in death caused by road accidents. Current traffic sign board classification and identification systems are made by using convolutional neural networks. The detection system has to recognize a variety of traffic sign boards of different shapes and not just speed limits. A convolutional neural network can be trained by taking predefined traffic sign boards and learning it using Deep Learning strategies. Image Processing along with Computer Vision techniques are applied to train the network with its likely results. The trained neural network can then be utilized in real-time to identify new traffic sign boards. Advanced computer vision and neural network strategies make this objective efficient and feasible in real-time. The construction of this type of system is done in mainly two phases. In the first phase, the system is trained by a huge number of traffic sign boards images. These traffic sign boards images are fed to the convolutional neural network(CNN) so that the signs can be classified accurately and later the model is saved. In the second phase, the system will be tested. The image which is captured live by

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1504-1511www.ijaem.net ISSN: 2395-5252

the system webcam is converted into a binary image using HSV(Hue Saturation Value) technique. This technique helps in finding the ROI (region of interest). After the ROI is detected, the image will be cropped and will be given to the saved model to predict the correct class for the image. This proposed system not only detects and recognizes the traffic sign but also generates an alerting message about the traffic sign captured.

Dataset Collection:

The dataset used in this project refers to the German Traffic signs DataSet from Kaggle, it contains 32,100 plus images in total.The dataset contains a considerable amount of variation in nature, as some traffic signs have several images while others have fewer images in the records[6].In our separate folders, the images are organised into numerous classes, from which the data is taken into our Python script. To convert raw sample data into a comprehensible format, data is translated into a suitable format. This dataset which is used for training our system consists of fortythree folders each representing a different class. The folders are named from “0” to “42”. Each folder contains images of traffic sign boards. for eg. folder “0” contains 180 images of traffic sign “speed limit 20 km/h” , same like this folder “1” contains images of traffic sign board - “speed limit 30 km/h”.folder 42 -not allowing passing by vehicles over 3.5 metric tons.The images in these folders are of type .png and width =32 pixels,height=32 pixels.

Fig 1 : Some sample images of german traffic sign dataset

Let’s see how many samples we have for each traffic sign class. Below, there is a histogram of sample occurrences in the data set for each label.

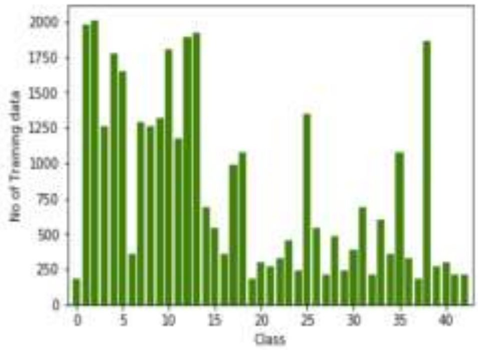

Fig 2 : Traffic sign frequency graph

II. LITERATURE SURVEY:

In this paper, the system mainly works with capturing traffic signs through a web camera and classifying the respective traffic sign and alerting. It mainly consists of extracting the features, detecting the traffic sign and recognizing the traffic signs. Some related work presented in [1] by Djebbara Yasmina, Rebai Karima, Azouaoui Ouahiba is on deep learning based road traffic sign recognition system. Their system uses the modified LeNet-5 for extracting the main features and recognizes the symbol. From the paper[2], study on Traffic sign recognition. The authors Mohamed Yusof Radzak , Mohd Fauzi Alias and Mohamad Rosydi Ahmad used the concept of ROI on Malaysian and Singapore traffic sign dataset to detect the traffic signs and recognize them easily from the black and white pixels. Additional information found in paper[3] for this paper that authors followed a 3 step process to enhance the recognition of traffic signs. One important step is tracking which track the traffic sign in motion. More information related to ROI and HSV concepts are taken from paper [4] , that uses HSV to detect the traffic signs.The paper[6] authors suggest using HSV for object detection and tracking the position of the object, so using this concept to detect the main object i.e. traffic sign in captured image. After going through the research paper, the best method for traffic sign detection is using the deep neural network CNN as a model for ROI using HSV will be the best way for traffic sign recognition.

Existing Methodology:

Traffic sign detection is useful in semi autonomous vehicles in order to support the driver. The existing approach proposes a light weight convolution neural network for recognising traffic sign boards . The existing system uses a LeNet-5 convolutional neural network architecture. The goal of employing a convolutional neural network

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1504-1511www.ijaem.net ISSN: 2395-5252

design is to preserve the spatial relationship between pixels by using kernel filters to get the major properties of the inserted data or picture, which are then injected into a fully connected network to determine the class. In this current application, the LeNet-5 architecture with nine layers is employed, starting five layers are convolution and simplification functions, and the remaining four layers are fully connected layers with minor modifications. That is the output of the first convolution operation is added to the first layer of the fully connected layer. The ReLu activation function is used in each of these convolution layers, but the soft-max activation function is used in the fully connected neural network's final layer for classification. The Kernel/filter in the convolution layer moves across the inserted picture, multiplying the values in the filters with the image's original pixel values. The multiplications are added together to produce a single value for that field. Feature maps are the output of the convolution layer, the number of feature maps created is equal to the number of filters utilised. Max-pooling is performed on the feature maps to downscale the image. In order to achieve an appropriate accuracy value, the training phase of the neural network adjusts its parameters weights and biases. The dropout strategy is used during the training phase to allow the network a flexible margin to react to inputs outside of the training examples data, which helps to minimise overfitting. A dropout of 31% and 53% of neurons is conducted in the second and third layers of the fully connected network, respectively, while a 90% dropout (shutting down 10% of neurons) is performed on the layer before the fully connected network. When testing the system, the image of the traffic sign board has to be uploaded from the computer so that the detection and recognition of the sign can be done. At a time only one sign board can be detected and the output will be displayed on the screen.

Proposed Methodology:

The suggested approach uses deep learning neural network framework - CNN – to construct traffic sign board identification and warning systems. The proposed system is able to recognize the traffic sign board live using the webcam and gives an alerting message about the traffic sign board recognized by the system.

The algorithm for proposed system is as follows:

1. Exploring the dataset : the dataset used is german traffic sign dataset available on kaggle
2. Data set splitting into train, test and validation .
3. Applying image preprocessing.
4. Performing data augmentation to generate more generic data
5. Creating the CNN model
6. Training the model
7. Classifying images using trained CNN models.

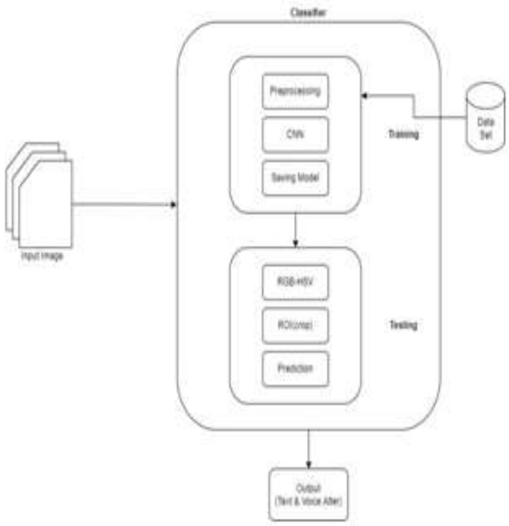

Fig 3: Block Diagram of Proposed approach

Data Preprocessing:

To carry out image processing the images should be turned into NumPy arrays before proceeding (i.e., numeric values). The keras.utils packages are imported to carry out one-hot encoding which as a result enhance the model's prediction and classification accuracy. By dividing the column into many columns, one-hot encoding turns categorical data into numeric data. The data has the shape (22271, 32, 32, 3), indicating that there are 22,271 pictures of 32x32 pixels and that the last three show that the data comprises coloured images (RGB value). The Sklearn package's train_test_split() function is used for splitting.

Epochs: 15

Image Dimensions= 32,32,3

Test Ratio = 0.2 which is 20% of dataset

Validation Ratio =0.2 which is 20% of train dataset.

We have 34799 images in total. Out of which 20% of total images are assigned for test dataset which will be equal to 6960 images falling under test dataset. The shape of the test array will be (6960,32,32,3).

Out of the remaining 80% of the dataset 20% will be assigned for validation and rest will be assigned for train dataset.hence the shape of train array will be(22271,32,32,3) and validation array will be (5568,32,32,3).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1504-1511www.ijaem.net ISSN: 2395-5252

Numerical instabilities are decreased by normalising the visual data by having pixel values range between -1 and +1. Some adjustments may be done on the obtained augmented and finely refined data, such as adjusting the brightness, rotation of the picture, and so on, to improve model performance.

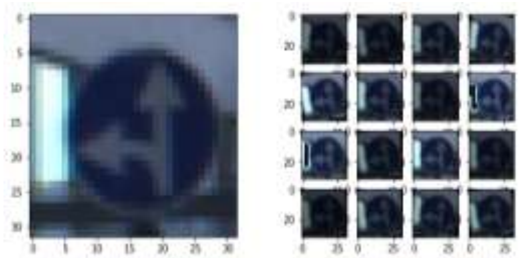

Fig 4: Image generated after transformation

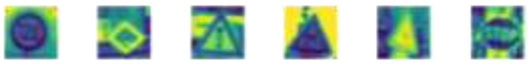

Fig 5: Sample Image after Preprocessing.

Training:

Our proposed model is trained by Convolution neural network.In our proposed architecture after two convolution layers we have placed a max-pooling layer.The Rectified Linear Units (ReLU) activation function is utilised in both convolution and the first layer of a fully connected network to incorporate non linearity in the model. The equation 2 describes the ReLu activation function.

f(x)=max(0,x) --(1)

Where x is input and f(x) is ReLu activation function.The range of ReLu activation function is 0 to infinity. A dense layer is formed when all of the neurons in one layer are linked to neurons in the next layer, resulting in a fully connected layer. To compute the probability distribution of classes, the SoftMax unit is employed. Finally, a Softmax classifier is utilised to divide the images into categories. The softmax activation function is described by equation 2.

$$\sigma(\mathbf{z})_{in} = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}} \quad \text{--(2)}$$

Where σ is softmax, z is the input vector,e^{zi} is a standard exponential function for input vector. K is the total classes in the multi-class classifier . e^{zj} represents standard exponential function for output vector.

In our CNN model we have eight important layers . Convolution layers, pooling layers, and fully-connected (FC) layers are the three types of layers that make up the CNN. To create a CNN architecture, these layers must be layered. The convolution layer is the initial layer, and it extracts numerous information from the input pictures. The basic mathematical multiplication operation between the input picture and a kernel of a specific size is done at this layer. Kernels of sizes 5x5 and 3x3 were utilised in our suggested system. The workings of convolution operation is shown in Figure 6.

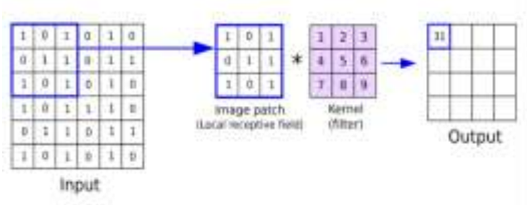

Fig 6: Convolution operation

The output of convolution layer feature maps.The number of feature maps produced is the same as the number of filters in the convolution layer. The dimension of the feature map can be calculated by the equation number 1.

$$\mathbf{n_{out}} = \left[\frac{(\mathbf{n_{in}}+2\mathbf{p}-\mathbf{k})}{\mathbf{s}}\right] + \mathbf{1} \quad \text{---(3)}$$

Where $\mathbf{n_{in}}$=size of input image

n_{out}=size of output image (also known as feature map)

p=padding size

k=dimension of filter/kernel

s=stride size

The designed CNN model consists of four convolution layers, two convolution layers one max pooling layer is repeated. The feature extraction contains convolutional layer (Conv2D) layer, Max pooling 2D layer, Activation layer, dropout layer and for classification part a fully connected layer a dense layer is added. The training is important in building the system to this model. To get high accuracy and performance it is necessary to select the parameters like no of filters , size of filter to be used, type of pooling , activation function is used. The first convolutional layer will be considered as an input layer which takes input images in the form of 32 X 32 gray scale images from the database with default value of stride 1. The Conv 2D layer extracts all possible features from the pictures and provides us with the feature maps and helps us in finding the internal representation. The max-pooling operation is performed after convolution operation. The max-pooling technique is used to minimise the size of the feature maps produced by the convolution layer's output. The fig 7 describes the working of max-pooling. For our proposed system the size of max-pooling is taken as 2x2.

.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1504-1511www.ijaem.net ISSN: 2395-5252

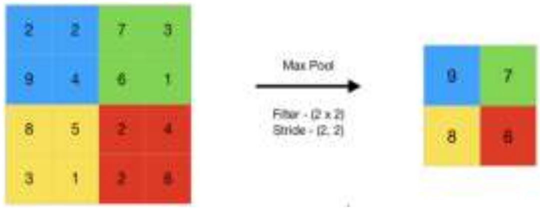

Fig 7: max-pooling operation

After the feature extraction is done the 2D feature maps are flattened into 1D and are fed to the fully connected layer. For the system we have two layers in the fully connected layer.The first layer consists of 500 nodes and the second layer consists of 43 nodes. 50% nodes are dropped out in the first layer of a fully connected network. Dropout is performed to avoid overfitting of data.

The following are the parameters that will be used for our system.The input image given to the CNN model is of size 32 x 32 x 1, The no.of filter used is 40, size of filter/kernel: 5x5 ,size of filter-2 / kernel2=3x3, size of pool=2*2, no.of nodes= 500, batch size=32.

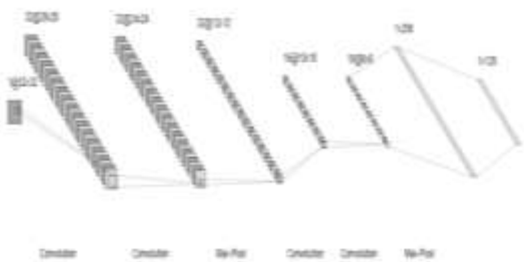

Fig 8: CNN model related to the project

Neural Network uses an optimizing approach to reduce the error in the algorithm. We have used Adam optimizer in our proposed system. An optimizer is used to alter the weights and learning rate of the neural network that results in reduction of overall loss and improves the accuracy. The error is calculated by using a Loss Function. It is used to quantify how good or bad the model is performing. The loss function used in training the model is Categorical cross entropy, this type of loss functions are very much useful in multi-class classification tasks.

Testing :

In this phase, we monitor the validation accuracy and loss of the model and we will be saving it in its best possible state which is having higher test accuracy on the data set used , based on final network weights trained the model gives the determined traffic sign board. We are perforing live detection of the traffic sign boards hence after capturing the image hence ,the output displayed is the probability value and the class name matched to the image captured by the webcam . The model is trained with the training dataset and in testing phase, images of dataset which consists of all 43 different types of traffic signs are used for testing , real values are compared with the predicted values for accuracy of the model .The proposed system architecture has a total 412,003 parameters and all are trainable parameters.The test accuracy is about 98%. Later in order to make predictions from the live webcam we have used HSV algorithm which helps in filtering the colors. The RGB image is converted to the binary image by using HSV algorithm so that region of interest(ROI) can be detected in the frame captured by the webcam. The ROI detection is a very crucial step as the background may be noisy sometimes which may create hurdles in detecting the correct sign. The ROI is cropped and preprocessed. In the preprocessing of the image the RGB(Red Green Blue) image is changed into grayscale image and then Histogram equalization and then normalization is performed on the grayscale image. The preprocessed image is given to the trained model to predict the respective class. A voice alert is generated through the speaker by the help of gTTs (Google Text-to-Speech), which is a python library used to convert the entered text into audio.

Experimental Results and analysis :

A traditional CNN classifier trained by applying all the necessary augmentation and preprocessing techniques. This model has achieved a testing accuracy of 98% which is more than 3% of existing model architecture

This model predicted fine with better accuracy and loss values. The following graph shows the accuracy and loss

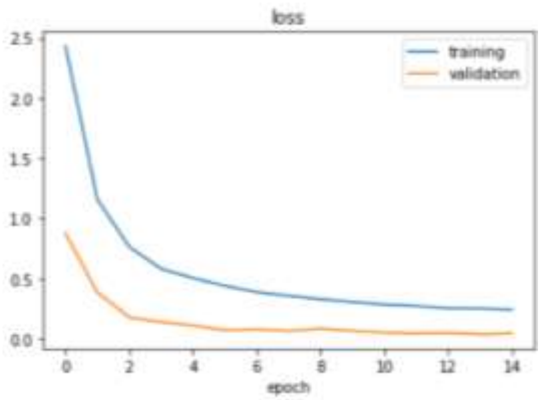

Fig 9 :Loss curve

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1504-1511www.ijaem.net ISSN: 2395-5252

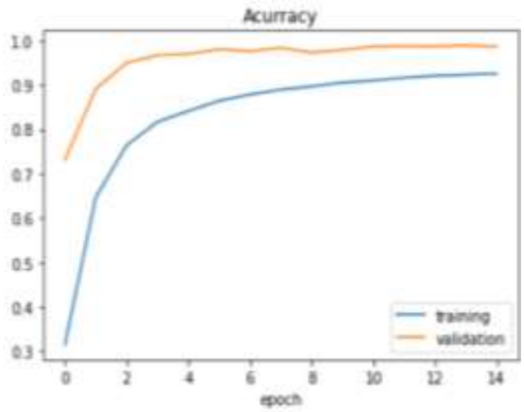

Fig 10: Accuracy curve

From fig 9 and fig 10 we can see that loss , accuracy curves were constant after 12 epochs. From this we can see that training the model till 13 epochs to get good accuracy. The reason for the great accuracy is due to the vast dataset and the use of data augmentation. We utilised the Data Augmentation approach to expand the amount of data by inserting slightly changed copies of already existing data. To see how effectively our system can predict or categorise, we may use one of two ways. The confusion matrix is one of the categorization metrics. The important factors in the classification metrics are precision ,recall, f1-score which are very much useful in performance evaluation. The better the model, the higher the score. Table number 1 shows the classification metrics.

	Precision	Recall	F1-score	Support
0	1.00	0.94		
1	0.99	0.99	0.97	35
2	0.99	0.99	0.99	368
3	0.96	0.98	0.99	413
4	0.99	0.99	0.97	249
5	0.93	0.95	0.99	352
6	1.00	1.00	0.94	315
7	1.00	0.89	1.00	74
8	0.99	0.99	0.94	259
9	1.00	1.00	0.99	252
10	1.00	1.00	1.00	268
11	1.00	0.96	1.00	356
12	1.00	1.00	0.98	223
13	1.00	1.00	1.00	379
14	1.00	1.00	1.00	374
15	0.98	0.99	1.00	157
16	1.00	1.00	0.99	107
17	1.00	1.00	1.00	83
18	0.97	1.00	1.00	186
19	1.00	1.00	0.99	232
20	1.00	1.00	1.00	39
21	1.00	0.95	1.00	69
22	1.00	1.00	0.97	59
23	1.00	0.99	1.00	57
24	0.97	1.00	0.99	95
25	1.00	1.00	0.99	37
26	0.99	0.96	1.00	276
27	1.00	0.98	0.97	101
28	1.00	1.00	0.99	49
29	0.98	0.96	1.00	88
30	0.96	0.96	0.97	46
31	0.97	1.00	0.96	84
32	1.00	1.00	0.99	142
33	1.00	1.00	1.00	47
34	1.00	1.00	1.00	119
35	1.00	1.00	1.00	71
36	0.99	0.99	1.00	218
37	1.00	1.00	0.99	74
38	1.00	0.99	1.00	33
39	0.98	1.00	0.99	370
40	1.00	1.00	0.99	54
41	1.00	1.00	1.00	63
42	1.00	1.00	1.00	48
			1.00	39
Micro avg	0.99	0.99	0.99	6960
Macro avg	0.99	0.99	0.99	6960

Weighted avg	0.99	0.99	0.99	6960
Samples avg	0.99	0.99	0.99	6960

Table 1: Classification metrics

Precision is the percentage of accurately detected positives among all positives projected.

Precision=$\frac{\text{True Positives}}{(\text{True Positives} + \text{False Positives})}$ --(4)

Recall is the percentage of positives accurately detected by the system out of all positives.

Recall = $\frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})}$ ---(5)

The harmonic mean of the model's accuracy and recall is used to calculate the F1 Score.

F1 Score= $2\times\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ ---(6)

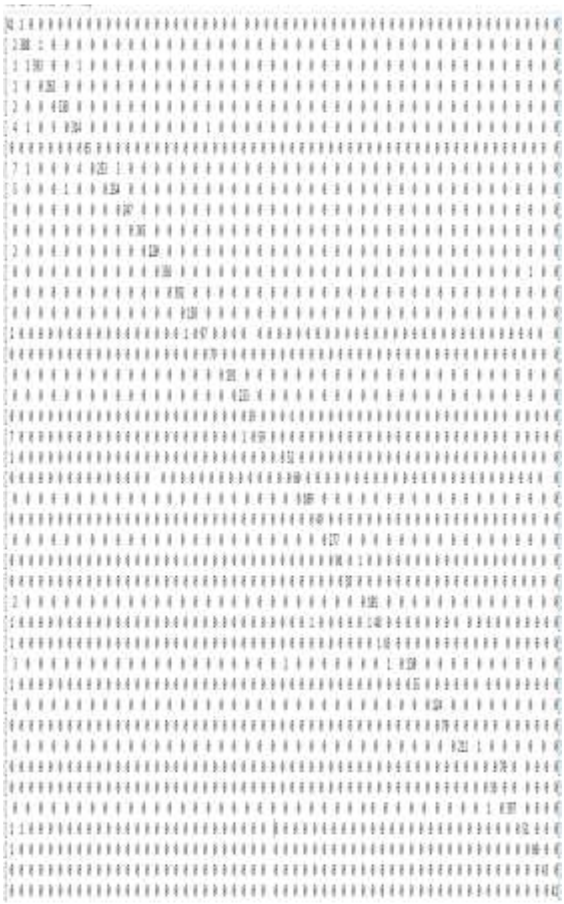

Fig 11: Confusion matrix

Figure 10 shows the confusion matrix for our classifier. It is a tabular representation of the model prediction versus the actual labels. This confusion matrix is particularly useful for analysing performance and measuring classification accuracy, which is equal to the proportion of right guesses out of total predictions. The examples in an actual class are represented by each row of the confusion matrix, whereas the occurrences in a predicted class are represented by each column.

III. CONCLUSION:

In this study, we expressed our idea to automatically detect and recognize traffic sign boards live. To decrease human effort when driving, the newest deep learning neural network frameworks are being used. We investigated the neural networks' ability to accurately detect traffic sign boards. The findings emphasise the value of deep learning neural networks in image categorization. As a result, these network frameworks can be used to reduce human effort while also assisting in the reduction of road accidents by alerting the driver about the captured traffic sign.

REFERENCES

[1]. Djebbara Yasmina; Rebai Karima; Azouaoui Ouahiba ‘Traffic signs recognition with deep learning’DOI:10.1109/ICASS.2018.8652024 Publisher: IEEE

[2]. Mohamed Yusof Radzak, Mohd Fauzi Alias, Mohamad ROSYIDI Ahmad ‘Study on Traffic Sign Recognition’ Conference: International Journal of Research Studies in Computer Science and Engineering (IJRSCSE).

[3]. B. Wali ,Majid A. Abdullah , Mahammad A. Hannan ,Aini Hussain ,Salina A. Samad ,Pin J. Ker and Muhamad Bin Mansor ‘Vision-Based Traffic Sign Detection and Recognition Systems: Current Trends and Challenges’ Safat

[4]. Siti Sarah Md Sallah , Fawnizu Azmadi Hussin, Mohd Zuki Yusoff ‘Road sign detection and recognition system for real-time embedded applications’ July 2011 DOI:10.1109/INECCE.2011.5953878 SourceIEEE Xplore

[5]. Priyanto Hidayatullah, Miftahudin Zuhdi , “Color-Texture Based Object Tracking Using HSV Color Space and Local Binary Pattern ” ,DOI:10.15676/ijeei.2015.7.2.1 , 2015.

[6]. Ying Sun; Pingshu Ge; Dequan Liu ‘Traffic Sign Detection and Recognition Based on Convolutional Neural Network’

DOI:10.1109/CAC48633.2019.8997240Publisher: IEEE
[7]. Prashengit Dhar; Md. Zainal Abedin; Tonoy Biswas; Anish Datta 'Traffic sign detection — A new approach and recognition using convolutional neural network' DOI:10.1109/R10-HTC.2017.8288988 Publisher: IEEE
[8]. Md Tarequl Islam ' Traffic sign detection and recognition based on convolutional neural networks' DOI: 10.1109/ICAC347590.2019.9036784 Publisher: IEEE
[9]. Nazmul Hasan, Tanvir Anzun , Nusrat Jahan , "Traffic Sign Recognition System (TSRS): SVM and convolutional Neural Network.",DOI:10.1007/978-981-15-7345-3_6, 2021.
[10]. Zhong Liu, Weihai Chen , Yuhua Zou , Cun Hue , "Regions of interest extraction based on HSV color space "DOI: 10.1109/INDIN.2012.6301214, 2012.
[11]. Alexander Shuastnov, Pavel Yakimov , "CNN Design for Real-Time Traffic Sign Recognition" , Volume 201, 2017, Pages 718-725.
[12]. Enoch ArulPraksh, Martin Aruldoss . " A study on generic object detection with emphasis on future research directions.
[13]. Djebbara, Yasmina ,Rebai Karima, Azouaoui Ouahiba.'Traffic signs recognition with deep learning.'2018 International Conference on Applied Smart Systems (ICASS'2018)