Vehicle Detection in Bhutan Using Convolutional Neural Network

Karma Wangchuk[1], Tenzin Phuntsho[2], Penjor Tshering[3], Tshering Pem[4], Thinley Phuntsho[5]
Information and Technology Department, College of Science and Technology, Royal University of Bhutan

E-mail: karma.cst@rub.edu.bt[1], 02180426.cst@rub.edu.bt[2], 0217520.cst@rub.edu.bt[3], 02180432.cst@rub.edu.bt[4], 02180429.cst@rub.edu.bt[5]

Abstract

Manual vehicle entry at different checkpoints in Bhutan by police personnel creates traffic congestion and delay. Drivers wait in a queue to register vehicles by providing details such as vehicle type and number. There is no automatic system to detect vehicles. The purpose of the study was to create a machine-learning model to detect different types of vehicles in Bhutan. For this study, a total of 20 popular light vehicle classes were identified. The images and videos were captured. The number of frames was extracted from the videos and different types of data augmentation approaches were then adopted to create variations in the curated dataset for greater model generalization. Then different algorithms were evaluated on this dataset. However, the Convolutional Neural Network outperformed all other algorithms. The training and testing accuracy obtained was 99.85% and 99.62% respectively. The model was then deployed using the Flask web framework.

Key Words: *CNN, classification, detection, dataset, Augmentation, Feature extraction*

1. INTRODUCTION

Bhutan is a small and landlocked mountainous country wedged between China in the north and India in the south. Bhutan remained in isolation for decades (Devkota et al., 2022). However, in the 1960s development of industrial infrastructure and transportation marked the beginning of modernization (Yoshikawa et al., 2019). The trajectory of modernization has been slow but steady. Within four decades, Bhutan has been seeing tremendous progress in socio-economic activities. Furthermore, Bhutan is embracing emerging technologies to be on par with the developed countries. Nevertheless, people are reluctant to accept and adopt new technologies. His Majesty the King of Bhutan, Jigme Khesar Namgyal Wangchuck emphasizes the importance of making use of emerging and trending technologies such as Artificial Intelligence, Machine Learning, Deep Learning, Blockchain, Internet of Things, Big Data, Quantum Computing, Virtual Reality, Augmented Reality, and Digital Identification (BBS, 2019). His Majesty the King in his royal speech stated, “Being a small nation makes us a smart nation, this is not out of choice but out of necessity. Technology is an indispensable tool that will be necessary to realize this aspiration”. Bhutan has a small population and the King in his speeches accentuates and reminds people about having to match skills with manpower and human resources.

In recent times, having a car has become a necessity for everyone in Bhutan. Without a car, one must spend a significant amount of money on traveling. As the socio-economic of the country is improving and enhancing, there have been a number of challenges, one of which is traffic congestion. Vehicles have to register at various checkpoints in Bhutan by the police personnel manually. The manual entry is slow. Drivers have to wait in a queue to get registered their vehicles such as vehicle type and number. This creates traffic congestion and delay. However, these issues can be resolved using technologies such as image processing and machine learning techniques.

Instead of using human laborers to monitor the traffic, the data can be analyzed by integrating machine learning and computer vision techniques. With the help of these new technologies, human activities can be recognized and monitored (Khan & Al-Habsi, 2020). Even most of the developed Western countries have shown interest in these fields. Automatic traffic surveillance systems have been widely developed in recent years. A system that can detect and classify vehicle types is one of the innovative ideas used by many other developed countries to aid guidance in intelligent traffic

systems (Butt et al., 2021). Such a system is required for effective real-time traffic management systems that can detect changes in traffic characteristics promptly allowing concerned regulatory agencies and authorities to quickly respond to traffic situations (Kanakala & Reddy, 2023).

Researchers have shown great interest in the domains of machine learning and computer vision (Bayoudh et al., 2021). A large number of traffic videos are analyzed and informed decisions are made. Not only researchers, but the implementing agencies are also benefitting from technologies (Khan & Al-Habsi, 2020). Today, road traffic video monitoring is at the center of several issues. It provides a useful method for analyzing highway traffic. Road traffic video monitoring can aid in the resolution of a variety of issues that can compromise road safety. In Morocco, a study done by Moutakki et al. (2018) described a real-time management and control system that uses a stationary camera to assess road traffic. Based on three modules: segmentation, classification, and vehicle counting, the proposed system can measure the number and characteristics of traffic in real time. Their contribution entails the creation of a feature-based counting system for vehicle identification and recognition under settings that have proven difficult in existing systems, such as occlusions and lighting conditions. By removing the influence of various factors on system efficiency, their technology could detect and classify vehicles. The collected findings suggest that the system proposed in their paper has a greater counting rate than certain other methods.

Chandrika et al. (2019) studied vehicle detection and classification using Image Processing methods. They have shown how efficient traffic management can reduce the valuable time spent in traffic congestion and how traffic issues can be reduced by studying the traffic flow pattern. Moreover, traffic congestion can also be prevented with a good traffic plan. So, based on those limitations, this paper came up with a solution to implement vehicle detection and classification, including counting the vehicles. The classification of vehicles was based on four classes: bicycles and motorcycles, motor cars, mini-buses, pickup vans, and trucks. The researchers studied different types of vehicle detection algorithms. They implemented six phases: image acquisition, image analysis, object detection, counting, classification, and displaying the result. The dataset was prepared using CCTV videos. The extracted frames from the videos were fed as input to the model. They detected vehicles successfully from the video frames and counted the vehicle during the daytime. However, their biggest shortcoming was that the accuracy of the detection and classification decreased in bad lighting conditions (Chandrika et al., 2019).

Several machine-learning models have been proposed to classify vehicles. A comprehensive overview of vehicle detection and classification strategies and alternative methodologies for detecting automobiles in inclement weather was done (Keerthi Kiran et al., 2020). The study provides a general review of all the existing strategies other researchers have done till the year 2021. The researchers generalized the techniques of detecting and classifying vehicles into appearance-based characteristics and motion-based and aided researchers in the identification, categorization, and accessibility of vehicle data sets.

The development of a vehicle detection system in Bhutan would be more efficient and resolve issues of traffic congestion at different checkpoints and delays. To date, no work has been done in vehicle detection using computer vision and machine learning methods. Therefore, there is no benchmark to compare our results with. The main objective of this study was to detect vehicles in Bhutan. In the following sections, methodology, data acquisition, data pre-processing, results, and conclusion are discussed.

2. METHODOLOGY

There are four stages of the study namely data acquisition, image pre-processing, model training, and deployment as shown in Figure 1. The details of each stage are given in the following sections.

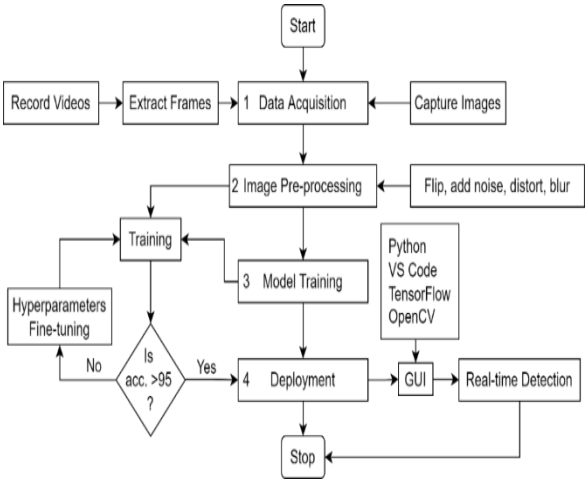

Fig. 1 Pipeline of the proposed system

2.1 Data Acquisition

The dataset is the most valuable asset in any machine learning and deep learning project (Algorithmia, 2020). The accuracy and robustness of the model are determined by the size and quality of the dataset. We have considered 20 classes of vehicles that are popular in Bhutan namely *alto, alto800, astar, Baleno, Bolero, Brezza, Celero, Creta, Eon, i10, i20, Maruti car, X-cross, Santa fe, Santro, Seltos, Swift, Tucson, and Wagonr*. Both videos and images were captured as shown in Figure 2. At the time of video recording various angles, distances, lighting conditions, and colors of the cars were taken into consideration. This creates variation in the dataset and the camera can be placed at different angles and distances at the time of real-time detection. The dataset was collected with permission from the car owners with the help of a consent letter. A video of each vehicle was captured to reduce the workload of collecting data. From the videos, 10 frames per second were extracted as shown in Figure 3.

Fig. 2 Recorded videos of various types of vehicles.

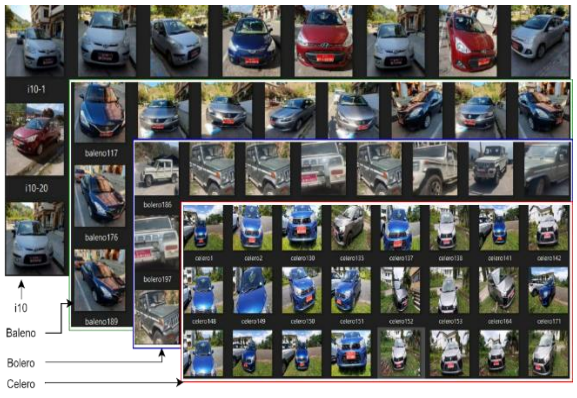

Fig. 3 Extracted frames sample from four classes of videos

2.2 Image Pre-processing

In the image pre-processing stage, image augmentation and size reduction were performed. The image augmentation facilitates the generation of more images from the existing images. Furthermore, augmentation helps to generalize the model well. The images are flipped, blurred, added noise, contrast, and distorted as shown in Figure 4. Next, the images were resized to 128 x 128. The image size was reduced due to the unavailability of high-end systems to train the model.

After the reduction of image size and augmentation, randomly 1200 images per class were selected. The dataset consisted of 24000 images. As suggested by Wangchuk, Riyamongkol, & Warranust (2021), the dataset was further converted into byte streams using the Python pickle module. This conversion reduces the time taken to train to model. The images were shuffled and divided into training and testing sets. The training and testing sets consisted of 20400 and 3600 images respectively.

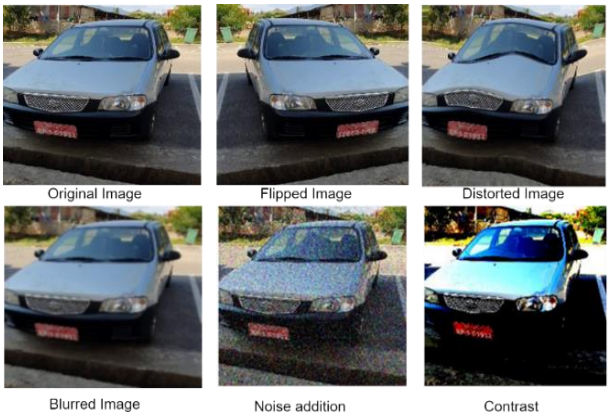

Fig. 4 Augmented images

2.3 Model Training

Different models were evaluated on the dataset. However, Convolutional Neural Network (CNN) outperformed all other algorithms. CNN is one of the variants of artificial neural networks. It is commonly applied to process both spatial and temporal data (Sun et al., 2020). CNN has three fundamental layers: convolutional layer, pooling layer, and fully connected layers. These layers are repeated in the deep CNN models as shown in figure 5. Furthermore, the dropout layer and activation function are used.

The convolutional layer consists of a number of varying-size filters that are applied to images to extract features. The pooling layer reduces the parameters and

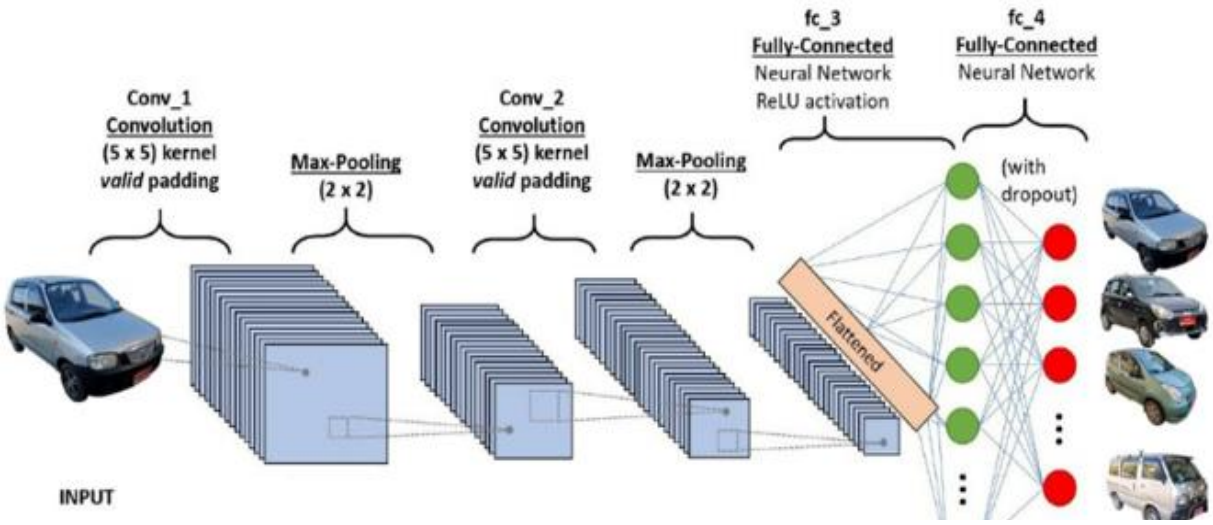

Fig. 5 Convolutional neural network architecture.

size of the images. This allows the model to train faster and required less memory. There are two types of pooling namely average pooling and max pooling. However, only one of this two pooling would be used. The fully connected layer consists of the number of neurons. It makes predictions based on the features learned from the previous layers. The dropout layer is used to address the overfitting issues by randomly dropping a number of neurons from the layers. The activation function learns complex relationships between the variables of the network.

Table 1 shows the model summary of vehicle detection. Four convolutional layers were used for features extraction and two dense layers for the classification. The images of size 128 x 128 x 3 pixels were used as the input for the model. In the first convolutional layer, 32 filters of size 5 x 5 were used and as the result, image size were reduced to 124 x 124 as shown in the Table 1. Similarly, in the second convolutional layer 64 filters of size 5 x 5 were used. However, in the following two convolutional layers, filters size used was 3 x 3 but the number of filters used was 128 and 256 respectively. The dropout of 20% and 50% were implemented in the third and fourth convolutional layers respectively. Next, two types of activation functions were used. The *ReLU* activation function was used in each convolutional layer and the first dense layer. However, in the second dense layer *Softmax* activation function was used.

Table 1 Bhutanese vehicle detection model summary

Layer (Type)	Output shape	Parameters
conv2d_1	(None, 124, 124, 32)	2432
max_pool2d_1	(None, 62, 62, 32)	0
conv2d_2	(None, 58, 58, 64)	51264
max_pool2d_2	(None, 29, 29, 64)	0
conv2d_3	(None, 27, 27, 128)	73856
dropout	(None, 27, 27, 128)	0
max_pool2d_3	(None, 13, 13, 128)	0
conv2d_4	(None, 11, 11, 256)	295168
dropout_1	(None, 11, 11, 256)	0
max_pool2d_4	(None, 5, 5, 256)	0
flatten	(None, 6400)	0
dense_1	(None, 512)	3277312
dense_2	(None, 512)	10260

The model was trained with 128 batch size and different hyper-parameters were fine-tuned. In the forward propagation, features are extracted and weights are learned with the *ReLU* activation function. Using two dense layers in the classification phase, the loss is calculated by predicted class value $\hat{y}_i$ and actual class value y_i as illustrated in Eq. (1). The goal of the model

training is to obtain higher accuracy with the minimum loss. The loss is minimized using an optimizer in the backpropagation. The softmax activation function classifies the vehicles by providing probabilistic values. The *sparse_categorical_crossentropy*, *ReLU*, and *softmax* are defined by Eq. (1), Eq. (2), and Eq. (3) respectively.

$$loss = \sum_{i=1}^{M} y_i \,.\, log\, \hat{y}_i \quad (1)$$

$$f(x) = max(0, x) \quad (2)$$

$$\sigma(Z)_j = \frac{e^{Z_j}}{\sum_{k=1}^{k} e^{Zk}} \quad (3)$$

The model was trained for 21 epochs. However, an early stopping mechanism was used with the patience of 3 to stop training. Early stopping resolves the issue of overfitting by stopping the training when there is no learning taking place.

4. RESULTS AND DISCUSSION

The experiments were conducted using Google Collaboratory called Colab. It provides Collaboratory notebook to run arbitrary Python program codes directly from their browser. Google created it to provide researcher unfettered access to GPUs and TPUs integrated with PyTorch, Tensor Flow, and OpenCV. It offers 12 hours of free usage with a steady connection.

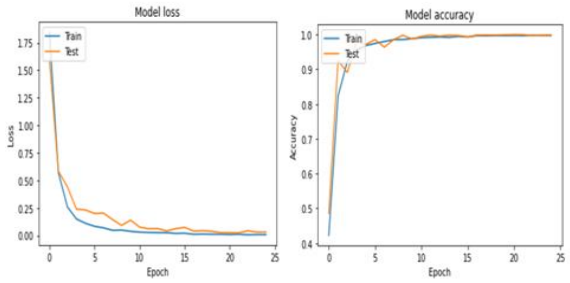

Fig. 6 Train and test accuracy: (left) accuracy vs. epoch (right) Loss vs. epoch

After partitioning the dataset into 85 % of the training set and 15% of the testing set with a batch size of 128, several deep learning models for vehicle detection and classification were trained with the TensorFlow backend. Table 2 shows the results of different algorithms evaluated on the Bhutanese vehicles' dataset.

The data that was collected previously was then fed to different models based on different machine learning algorithms to check which model will be more suitable for the dataset. CNN and CNN with ResNEt50, VGG6, and AlexNet architecture were used to train the different models. It was observed that CNN with six convolutional layers achieved the highest testing accuracy of 99.62% and train accuracy of 99.85%. VGG6 also scored the second-highest but the time taken for the model to train was maximum as compared to CNN. However, the minimum and maximum training times were observed with Alexnet and VGG6. Upon observing the result, the traditional architecture of CNN known as vanilla CNN was selected to train our model with the dataset curated.

We observed 99.85 % train accuracy with our trained model, indicating that the model is learning well with the data we feed it. Finally, we identified a test accuracy of 99.62%, which seemed adequate for us to move forward with development. The loss and accuracy are shown in figure 6. The model learned well in the first 21 epochs, but after that, learning stalled and training stopped. The early stopping with patience value 3 was used to avoid overfitting.

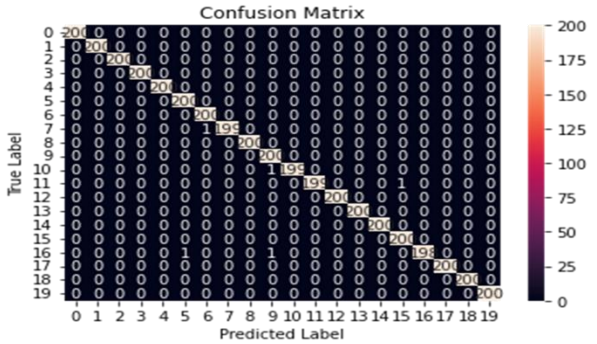

Fig. 7 Confusion matrix

Table 2 Shows the results of different algorithms evaluated on the Bhutanese dataset

Model	Batch size	Epoch	Time (s)	Train Accuracy (%)	Test Accuracy (%)
CNN	128	21	8798	99.85	99.62
AlexNet	128	7	2929	81.71	84.82
VGG6	128	23	9538	90.90	95.82
ResNet50	128	17	6858	39.9	43.15

Table 3 Illustrates precision, recall, and f1-score of 20 classes of vehicles

Class	Precision	Recall	F1-score	Class	Precision	Recall	F1-score
Alto	1.00	0.96	0.98	i20	1.00	0.97	0.99
Alto800	1.00	1.00	1.00	Maruti	1.00	1.00	1.00
Astar	1.00	1.00	1.00	s-Xcross	1.00	1.00	1.00
Baleno	0.97	0.99	0.98	Santafe	1.00	1.00	1.00
Bolerao	1.00	1.00	1.00	Santro	1.00	0.99	1.00
Brezza	1.00	1.00	1.00	Seltos	0.99	1.00	1.00
Celero	0.99	1.00	0.99	Swift	1.00	1.00	1.00
Creta	1.00	1.00	1.00	Tucson	1.00	1.00	1.00
Eon	1.00	1.00	1.00	Van	0.99	1.00	1.00
i10	1.00	1.00	1.00	Wangonr	1.00	1.00	1.00
Accuracy							1.00
Macro avg					1.00	1.00	1.00
Weighted avg					1.00	1.00	1.00

Figure 7 shows the confusion matrix. The test dataset contains a variety of photos that were not used to train the model. These matrices also assist in visualizing the model's true and erroneous assumptions. Following that, the trained model's confusion matrix revealed that it properly predicfor 17 classes, except for three classes, such as *Tucson*, where just two data items were misidentified, one as *brezza* and the other as *i10*. Similarly, just one data from the *s-XCross* class was misidentified as *seltos*, and only one data item from the *i10* class was misidentified as *celero*.

Since the vehicles in the study were all light vehicles with similar designs and characteristics, several test data items were incorrectly classified into different categories. As shown in Figure 7, the trained model was used to predict the Bhutanese car in real-time using a webcam, as well as on a simple website built using Flask, where a picture of the vehicle from the test dataset was given and then predicted. To provide diversity in the dataset, multiple data augmentation approaches were used. Also, several films from which the frame was retrieved at various angles

According to Table 3, the model can accurately forecast all vehicle classes with an f-score of 100 percent, except for four vehicle classes: *alto* and *Baleno*, which have an *f-score* of 98%, and *celero* and *Maruti*, which have an *f-score* of 99%.

Fig. 8 Real-time prediction for the vehicle

5. CONCLUSION

The goal of this study was to train machine learning model to detect and classify vehicle in Bhutan. The data was evaluated using different CNN-based models. However, the evaluations demonstrated that the vanilla CNN was an effective classification approach with training and testing accuracy, 99.85% and 99.62%, respectively. We observed that training with a minimal dataset and a smaller image size is insufficient to get the greatest accuracy. Overfitting was revealed and these drawbacks can be solved by utilizing a larger volume of dataset and higher resolution images.

In the future, researchers can deploy a trained model to create a system. Furthermore, a new dataset comprising all vehicle types can be curated and trained with state-of-the-art models such as Vision Transformer.

6. REFERENCE

Algorithmia. (2020). *2020 State of Enterprise Machine Learning*.

Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2021). A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer, 1-32.

BBS. (2019, August 10). *Use Technology to Solve Problems: His Majesty's Address to RIM Graduates - BBSCL*. http://www.bbs.bt/news/?p=119143

Devkota, D., Miller, D. C., Wang, S. W., & Brooks, J. S. (2022). Biodiversity conservation funding in Bhutan: Thematic, temporal, and spatial trends over four decades. Conservation Science and Practice, e12757.

Butt, M. A., Khattak, A. J., Shafique, S., Hayat, B., Abid, S., Kim, K., Ayub, M., Sajid, A., & Adnan, A. (2021). Convolutional Neural Network Based Vehicle Classification in Adverse Illuminous Conditions for Intelligent Transportation Systems. Complexity, 2021, 1–11. https://doi.org/10.1155/2021/6644861

Chandrika, R., Ganesh, N., Mummoorthy, A., & Raghunath, K. M. K. (2019). Vehicle Detection and Classification using Image processing. https://doi.org/10.1109/icese46178.2019.9194678

Kanakala, R., & Reddy, K. (2023). Modelling a deep network using CNN and RNN for accident classification. Measurement: Sensors, 100794.

Keerthi Kiran, V., Parida, P., & Dash, S. (2021). Vehicle detection and classification: A review. In *Advances in Intelligent Systems and Computing: Vol. 1180 AISC* (Issue January). Springer International Publishing. https://doi.org/10.1007/978-3-030-49339-4_6

Khan, A. I., & Al-Habsi, S. (2020). Machine Learning in Computer Vision. *Procedia Computer Science*, *167*(2019), 1444–1451. https://doi.org/10.1016/j.procs.2020.03.355

Moutakki, Z., Ouloul, I. M., Afdel, K., & Amghar, A. (2018). Real-Time System Based on Feature Extraction for Vehicle Detection and Classification. *Transport and Telecommunication*, *19*(2), 93–102. https://doi.org/10.2478/ttj-2018-000

Sun, P., Liu, P., Li, Q., Liu, C., Lu, X., Hao, R., & Chen, J. (2020). DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system. Security and communication networks, 2020, 1-11.

Wangchuk, K., Riyamongkol, P., & Waranusast, R. (2021). Real-time Bhutanese Sign Language digits recognition system using Convolutional Neural Network. ICT Express, 7(2), 215–220. https://doi.org/10.1016/j.icte.2020.08.002

Yoshikawa, M., Yoshikawa, S., & Yang, J. (2019, September 5). Urbanization and Transportation in Bhutan. https://storymaps.arcgis.com/stories/5af3ca1f0e1f4c799bed70d564397107