JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

# Classification of Asphalt Road Damage Based on Images Using the Convolutional Neural Network (CNN) Method

M. Rivan Padila[1], Arie Qur'ania[2*], Mulyati[3]
[1,2,3]Study ProgramComputer Science, Universitas Pakuan Bogor, Indonesia
[1]rivan065118311@unpak.ac.id, [2*]qurania@unpak.ac.id, [3]mulyati@unpak.ac.id

**Abstract**

Damage to roads can cause inconvenience in driving and can even lead to accidents. Some of the damages that are often found on the road network are such as fine cracks, alligator skin cracks, potholes, asphalt grain release and others. The damage needs preventive handling because it is the main infrastructure in land transportation that is used every day plus areas with very high rainfall such as Indonesia, Damage to the road surface can occur more quickly. One method in artificial intelligence that can be used in identifying damaged roads is Convolutional Neural Networks (CNN). This method is capable of self-learning for object recognition, object extraction and classification and can be applied to high image resolution. The Citra data is taken from the results of google street view mapping with the application of the CNN model using YOLOv5, which is expected to be able to classify images specifically more effectively, objectively and safely in road maintenance efforts later. This research aims to classify image-based asphalt road damage using the Convolution Neural Network (CNN) method. The stages of this research consist of Data Selection, Preprocessing, Data Transformation, Data Mining and Pattern Evaluation using confusion matrix. The results obtained F1 score model of 73.5%, the value of mean Average Precision (mAP) of 75%, this shows that this model is able to classify fairly against all categories of data used.

**Keywords:** Artificial Intelligence; Convolution Neural Network (CNN); Google Street View Map; Road Damage; YOLOv5;

## 1. INTRODUCING

Roads are one of the main infrastructures to support land transportation. Damage to roads can cause discomfort while driving and even lead to accidents. Some common problems found on road networks include fine cracks, crocodile cracks, potholes, asphalt aggregate detachment, and so on. In 2021, there were approximately 3,848.15 km of national roads that were damaged, and 2,901 km of marginal roads that required preventive maintenance[1]. Based on this, in order for the road network to maintain its condition, it is necessary to carry out maintenance continuously due to its daily use as a primary infrastructure in land transportation. Additionally, in areas with very high rainfall like Indonesia, road surface damage can occur more quickly. The initial step usually taken for road maintenance is by conducting an identification. Currently, identification is still done manually, which involves inspecting the roads, monitoring road conditions with cameras, organizing damaged areas, determining the level of damage according to its type, and then calculating and documenting it in the form of a report. Using this method is very time-consuming, labor-intensive, and costly. Another major problem in making decisions to improve road conditions using this method is that it is prone to subjectivity and can result in low accuracy in determining the condition of the road[2]. Manual monitoring of road

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

conditions is currently considered ineffective and inefficient. With the advancement of technology, several artificial intelligence-based models have been widely developed to produce more accurate predictions, one of which is the Convolutional Neural Networks (CNN) model. CNN is a deep learning method capable of performing self-learning processes for object recognition, object extraction, and classification, and can be applied to high-resolution images[3]. The identification of road damage through smartphone cameras was successfully carried out using CNN[4] YOLOv5 model. YOLO is a detection and classification system capable of detecting objects quickly and with good accuracy. Image data was taken from Google Street View mapping results. Using Google Street View maps can make data collection easier, and combined with the use of 360-degree camera technology, it produces image data with a wide field of view.

Research on CNN has been widely used for image identification, including by Fan et al., (2018)[5] who conducted image identification of cracked and normal roads using 3 types of datasets and several types of training data, resulting in the highest precision rate of 96%. Subsequent research by Triardhana et al., (2020)[6] identified road damage using the Yolov4 Tiny model, using as many as 13,376 image data. The highest accuracy achieved was 85.34%. In addition, Dais et al., (2021) [7] also applied deep learning methods on a different topic using 8 different convolutional neural network (CNN) architectures, namely VGG-16, ResNet34, ResNet50, DenseNet121, DenseNet169, InceptionV3, MobileNet, and MobileNetV2, to classify cracks on the surface of brick masonry in a building using a dataset of 351 photos, resulting in the highest accuracy of 95.3% for the MobileNet architecture and 88.0% for VGG-16.

In the introduction, researchers are expected to be able to explain the existing phenomena or background information such as prior work, hypotheses, problems to be discussed. This is followed by a statement of the purpose of the research issue or problem and/or set of questions you attempt to answer in your research.

## 2. METHODOLOGY

The method used in this research is KDD (Knowledge Discovery and Data Mining). The KDD process in the context of computer science and Big Data is a system or technique for determining knowledge when mining data. The KDD method process can be seen in Figure 1.

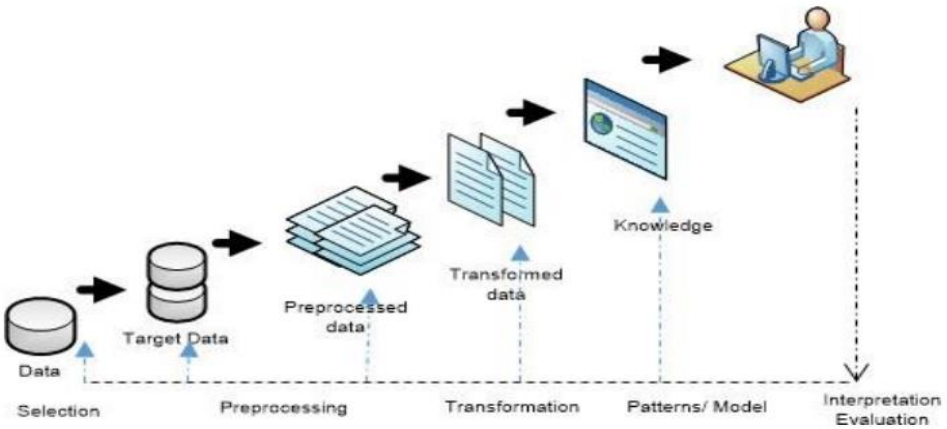


**Figure 1.** KDD Method Process

The explanation of the KDD stages is as follows:

1. Selection, the selection process carried out was choosing the dataset to be used, which was taken from the IEEE Road Damage Dataset, mapping results obtained using Google Street View, and photos taken directly of damaged road conditions using a cellphone camera.

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

2. Preprocessing/Cleaning, at this stage, the existing data undergoes a process (image enhancement), which is an image processing procedure carried out to improve poor-quality images, such as those with noise, overly dark or bright images, or images that are not sharp, to prevent data bias. Quality improvement operations at this stage include brightness adjustment, image smoothing, noise filtering, and so on.
3. Transformation, the image transformation process carried out at this stage aims to produce a richer image dataset. The image transformations performed include adjusting the lighting, cropping the images to focus on existing road damage. Other processes include rotation and changing the contrast of the images.
4. Data Mining, the mining process is the process of finding interesting patterns or information in selected data using certain techniques or methods. This mining process uses a Convolutional Neural Network (CNN) algorithm approach. At this stage, the process of creating a CNN model with the Yolov5 architecture is carried out.
5. Evaluation, the results from the model that has been created are then evaluated using the confusion matrix method, which is a table used to measure the performance of a classification model in machine learning. This method displays and compares actual values or true values with the model's predicted results, which can be used to generate evaluation metrics such as accuracy, precision, recall, and F1-score.

### 2.1 Convolutional Neural Network (CNN)

CNN is a model specifically designed to process data that has high dimensions, such as data consisting of three two-dimensional matrices representing color intensity for three color channels in the context of image processing[9]. The CNN architecture is divided into four main parts[10], namely:

a. Input layer: The layer that will store the pixel values of the input image.
b. Convolution layer: This is the core block of the CNN. This layer functions to simplify the input features into a smaller matrix, and this is where the computations are performed. If a convolutional layer has a sheet of neurons containing 28x28 pixels and each sheet is connected to a small area in the input image sized 5x5 pixels, which is the receptive field, it can be said that the example has one filter with a size of 5x5 pixels [11]. The image convolution operation can be written as follows:

$$s(t) = (x * t)(t) = \sum\infty\, x(\alpha) * w(t - \alpha) \quad (1)$$

Where: S(t) = The result function of the convolution operation
x = Input
w = Weight (kernel) In the function S(t), it can produce a single output in the form of a feature map.
First, the input is x, and second, w acts as the kernel or filter. If the input is viewed as a two-dimensional image, then t can be considered as a pixel and replaced with i and j. Therefore, the convolution operation on an input with more than one dimension can be written as follows:

$$s(i, j) = (K * I)(i, j) = \sum\infty\, \sum\infty\, I(i - m, j - n)K(m, n) \quad (2)$$

$$s(i, j) = (K * I)(i, j) = \sum\infty\, \sum\infty\, I(i + m, j + n)K(m, n) \quad (3)$$

Equations (2) and (3) are the basic calculations in a convolution operation, with i and j being a pixel from an image. These calculations are cumulative and appear when K is used as a kernel, and then I as the input and the kernel can be reversed relative to the input.
c. c. Pooling layer, functions to further simplify by performing pooling or searching for certain elements in the convolution results, for example, max pooling which only takes certain elements from the convolution output, or average pooling which takes the average of the elements from the convolution output.
d. d. Fully-connected Layer, in this layer performs the same task as an artificial neural network and tries to produce class values from the activations, which are used for classification. This layer aligns with the neurons arranged in an ANN.

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

### 2.2 You Only Look Once (YOLO)

YOLO is an object detection algorithm that is part of a convolutional neural network algorithm designed to detect objects in real time. The detection system used employs a reuse classifier or locator to perform detection. This model is applied to images at various positions and scales. The regions of the image that obtain the highest scores are considered detections [12]. YOLO uses an architecture similar to CNN, but YOLO only uses convolutional and pooling layers. The last convolutional layer is scaled according to the number of classes and the number of bounding boxes. The way YOLO detects objects is by dividing the image into an SxS grid. The grid size is determined by the YOLO architecture used. The grid forms various types of grid cells, each tasked with predicting objects within that cell. Each bounding box contains five predictions and a confidence score. YOLO has a framework that predicts the total width and height of the input image, and consists of:

1. Backbone: A convolutional neural network that combines and forms image features at various levels of image detail.
2. Neck: A series of network layers that mix and combine image features and pass them on to the prediction layers.
3. Head: This component can predict image features, generate bounding boxes, and predict categories. Confidence represents the accuracy of classification under certain conditions.

After an object is detected, it will produce a confidence score in the form of a probability value for the object within the bounding box. The next process in YOLO is to predict the object into a class, called the class probability map. To achieve a high probability value, only those exceeding the threshold will be used.

## 3. RESULT AND DISCUSSIONS

### 3.1. Data Selection

The collected data is divided into three types: training data, validation data, and test data. The secondary data (data for model training) was taken from the IEEE Road Damage Dataset, which contains images of road damage in three different countries: Japan, India, and the Czech Republic. This dataset has 7 different classes. The labels and classes of the dataset can be seen in Table 1.

**Table 1.** Road Damage Class

| Type of Damage | Details | Code |
|---|---|---|
| Longitudinal Crack | Longitudinal cracks, vehicle wheel marks | D00 |
| Lateral Crack | Transverse cracks | D10 |
| Alligator Crack | Irregular cracks | D20 |
| Other damages | Bumps & potholes | D40 |
| | Storm drain covers | D50 |
| | Faded zebra crossing | D43 |
| | Faded road markings | D44 |

Next, the data was collected by utilizing Google Street View map technology as a tool to map road conditions as a dataset that will later be tested. The data collection process was carried out by using the road tracking feature on Google Street View map, and then the resulting photos were captured via screen capture to produce road mapping image data. The following is an example of road damage mapping using Google Street View map shown in Figure 2.

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

**Figure 2.** Road Damage Mapping Results Using Google Street View Map

Primary data (data for model validation) was collected directly using a smartphone camera. This process was carried out by pointing the smartphone camera at the road area that was damaged and then photographing the condition of the damaged road. Images were taken in 5 different districts, namely Cigombong, Caringin, Ciawi, Ranca Bungur, and Cibadak. The collected data was then subjected to quality improvement operations, brightness adjustment, image smoothing, noise filtering, and so on. During the training, testing, and validation data split, proportions of 70%, 15%, and 15% were used. The data was separated into three folders, which would later be used as a source for creating a TensorFlow dataset.

### 3.2. Model Evaluation

The model evaluation is conducted using test data and the interpretation of the model's performance is done using confusion matrix, accuracy, precision, recall, and F1 score methods to measure the accuracy and performance metrics of the model. Additionally, changes in the model's loss and accuracy at each iteration will also be used as indicators of whether the created model experiences overfitting or underfitting. The model performance results in this study can be seen through the F1 Confidence Curve and Precision Curve charts shown in Figure 3.

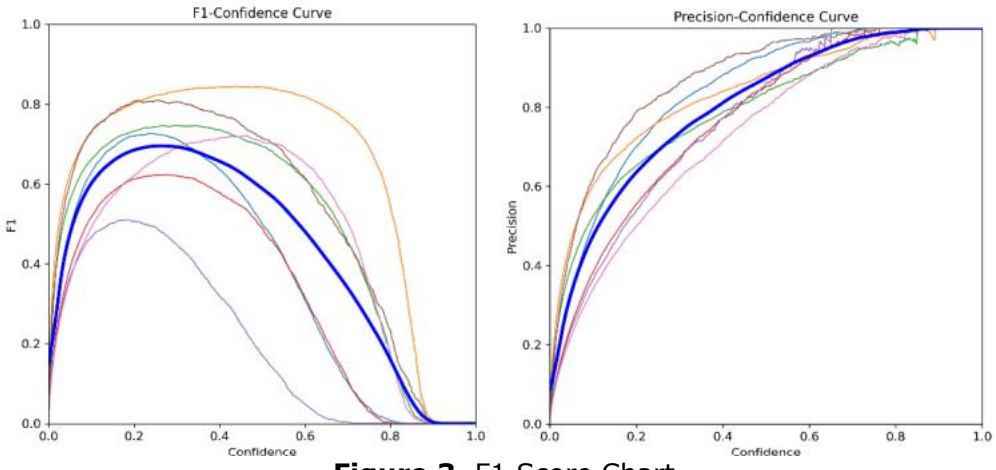


**Figure 3.** F1 Score Chart

Each line on the curves and graphs represents the class of road damage tested in this study. As can be seen in Figure 3, the F1 Confidence Curve graph for each tested class shows good accuracy and confidence results, with a confidence level above 0.6%. The model's precision in classifying and detecting road damage produces good performance, and the performance results can be seen in the Precision Confidence graph, which shows the curve for each class rising close to 1.0. In this study, the angle of image capture and

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

road conditions greatly influence the determination of the accuracy results and the model's effectiveness in detecting road damage according to its class. The images were taken during the daytime, with the capture angle from the side of the road with damaged conditions (images were taken directly using a mobile phone).

### 3.3. Discussion

Based on the results of this study, the model that has been successfully built using YOLOv5 is capable of detecting road damage based on previously trained classes. The advantage of this successfully built model is that it can detect road damage quickly compared to a standard CNN model. This can be seen from the mean average precision (mAP) score of 75%, with the highest AP value achieved by class D44 at 90.1%. However, the drawback of this model is that it requires a relatively long model-building process and high device specifications to build the model.

Model performance is evaluated using four metrics: accuracy, F1 score, training time, and inference time. The best model has the highest accuracy and F1 score and the lowest modeling time and model size.

**Table 2.** Model Accuracy

| Model | Accuracy | F1 Score | Training Time | Inference Time |
|---|---|---|---|---|
| CNN-V | 21,0% | 33,5% | 61 Minutes | 2.1 Seconds |
| CNN-M | 54,3% | 50,3% | 50 Minutes | 2.1 Seconds |
| CNN-Y | 75,3% | 73,5% | 1 Hour 28 Minutes | 1 Minute 40 Seconds |

Based on the metrics in Table 2, it can be concluded that the CNN-Y model, which is the YoloV5 transfer learning model, is the best model. The accuracy achieved after 20 epochs is 75.3%. Although the CNN-Y model has the highest accuracy, the model testing process takes quite a long time compared to other models with the same dataset and the same number of epochs. In addition, models that do not use transfer learning tend to have more fluctuations in accuracy. This is not only due to the image augmentation process but also because modeling without transfer learning has the potential not to reach convergence due to classic problems in image processing using deep learning, such as dataset size and computational capability for modeling [14].

Based on the best CNN-Y model, the next step is to perform validation using a confusion matrix to see how many classifications are correct and incorrect. In addition, precision, recall, and F1 score metrics are also calculated to measure the model's performance in detail.

**Table 3.** Classification Summary

| Code | Precision | Recall | mAP %0,5 |
|---|---|---|---|
| D20 | 78,4% | 66,9% | 75% |
| D50 | 77,1% | 88,5% | 77,5% |
| D44 | 70,4% | 79,2% | 90,1% |
| D00 | 63,8% | 60,9% | 80% |
| D10 | 63,2% | 60% | 65,5% |
| D43 | 83,6% | 86,6% | 86,6% |
| D40 | 58,2% | 76,3% | 76,3% |
| Mean Average Precision (mAP %0,5) | | | 75% |

Based on Table 3 above, the overall mAP@0.5 value obtained is 75%, so the model that has been created is considered sufficient to conduct testing. This model also has a fairly high F1 score, indicating that it is capable of performing fair classification across all data categories used in the modeling process. The following Figure 4 shows the loss curve

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

for the training and validation data, demonstrating that as the number of epochs increases, it converges closer to zero.

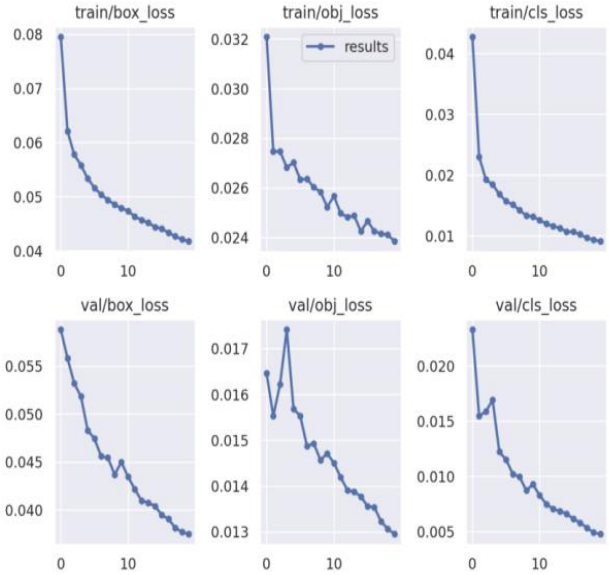


**Figure 4.** Epoch Value

The graph shows that during the training process, the train loss values for the box loss, objectness loss, and classification loss components consistently decreased from the beginning to the end of the epochs, indicating that the model is increasingly able to learn data patterns effectively. A similar trend is observed in the validation loss, where val/box_loss, val/obj_loss, and val/cls_loss tend to decrease, although there are slight fluctuations in the first few epochs, particularly in val/obj_loss. Overall, the stable downward trend in training and validation loss indicates that the training process is effective, the model does not experience significant overfitting, and it has good generalization capability for previously unseen data.

# 4. CONCLUSION

Based on the research results for performing classification and detection on road damage consisting of seven categories using a convolutional neural network (CNN) model with the YOLOv5 method, a model has been successfully built with an F1 score of 73.5% and a mean Average Precision (mAP) value of 75%. Based on these results, it can be concluded that the YOLOv5 method can be used to perform accurate and innovative classification and detection of asphalt road damage images by leveraging Google Street View Maps technology as a tool for mapping damaged roads. Some suggestions for further development of this research include retraining the model with a new dataset with varied image capture angles, as well as using other base models such as the latest YOLO versions, YOLOv6 and YOLOv7.

# 5. REFERENCES

[1] Dirjen Bina Marga, "Kondisi Permukaan Jalan Nasional." [Online]. Available: https://data.pu.go.id/dataset/kondisi-permukaan-jalan-nasional/resource/caebcef7-3273-41b0-b042-a453569aefb6#%7B%7D

[2] R. H. Pramestya, "Deteksi dan klasifikasi kerusakan jalan aspal menggunakan metode Yolo berbasis citra digital," SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY, 2018.

[3] C. Zhang, I. Sargent, X. Pan, A. Gardiner, J. Hare, and P. M. Atkinson, "VPRS-Based regional decision fusion of CNN and MRF classifications for very fine resolution

Arie Qur'ania: *Corresponding Author

JURNAL TEKNOINFO
Volume 20, Number 1, January 2026, Page 40-47
E-ISSN 2615-224X P-ISSN 1693-0010
https://publikasi.teknokrat.ac.id/index.php/teknoinfo/index
DOI: https://doi.org/10.33365/teknoinfo.v20i1.771

remotely sensed images," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 8, pp. 4507–4521, 2018, doi: 10.1109/TGRS.2018.2822783.
[4] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images," Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 12, pp. 1127–1141, 2018, doi: 10.1111/mice.12387.
[5] Z. Fan, Y. Wu, J. Lu, and W. Li, "Automatic Pavement Crack Detection Based on Structured Prediction with the Convolutional Neural Network," no. February 2018, 2018.
[6] F. H. Yoga Triardhana, Bandi Sasmito, "Identifikasi Kerusakan Jalan Menggunakan Metode Deep Learning (Dl) Model Convolutional Neural Networks (Cnn)," Jurnal Geodesi Undip, no. Dl, pp. 1–8, 2020.
[7] D. Dais, İ. E. Bal, E. Smyrou, and V. Sarhosis, "Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning," Automation in Construction, vol. 125, no. January, 2021, doi: 10.1016/j.autcon.2021.103606.
[8] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.
[9] S. Suyanto, Ramadhani, Kurniawan Nur, Mandala, Deep Learning Modernisasi Machine Learning untuk Big Data. Bandung: Informatika, 2019.
A. Saxena, "An Introduction to Convolutional Neural Networks," International Journal for Research in Applied Science and Engineering Technology, vol. 10, no. 12, pp. 943–947, 2022, doi: 10.22214/ijraset.2022.47789.
[10] X. Li, C. Ratti, and I. Seiferling, "Mapping urban landscapes along streets using google street view," Lecture Notes in Geoinformation and Cartography, no. May, pp. 341–356, 2017, doi: 10.1007/978-3-319-57336-6_24.
[11] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018, pp. 2503–2510, 2018, doi: 10.1109/BigData.2018.8621865.
[12] D. Arya et al., "Transfer Learning-based Road Damage Detection for Multiple Countries," pp. 1–16, 2020.
[13] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

Arie Qur'ania: *Corresponding Author