JMM: Journal of Mechanical and Manufacture
Vol. 3 No 1, October 2023 , pp. 12-21
DOI: 10.31949/jmm.v3i1.6371

e-ISSN: 2829-4424

CNN-BASED ARTIFICIAL INTELLIGENCE (AI) IMPLEMENTATION TO IDENTIFY BASMATI RICE IN SUBANG DISTRICT

Ari Ajibekti Masriwilaga [1], Ari Wibowo [2], Dian Susanto [3]

[1] Electrical Engineering-University of Subang, [3]Architecture-University of Subang, [3]Mechanical Engineering-University of Subang,
[1]ariajibekti@unsub.ac.id, [2]ariwibowo@unsub.ac.id, [3]diansus4154@unsub.ac.id

Abstract: The existence of Basmati rice among the upper middle class in Indonesia is increasingly popular. Unfortunately, this rice is only grown in northern India and Pakistan. Fulfillment of rice must be imported and the price in Indonesia is relatively expensive. Responding to this phenomenon, the Center for Rice Research (BB Padi), the Agricultural Research and Development Agency succeeded in assembling a special rice variety Basmati. And given the name Baroma, an abbreviation of type Basmati Aromatic rice. And Baroma rice was launched in Subang in 2019 until now it has been recorded that several agricultural lands in Subang have planted this type. The more types of rice varieties, the more types of rice will be found. So that it will make consumers difficult to distinguish between types of rice with one another. Therefore, we need a solution to overcome this problem. And one solution that can be used is to use AI technology, as in the research we did. Using the CNN algorithm produces very good accuracy for detecting types of rice such as the type of data used for training data and test data. From the results of the model training carried out, it produces an accuracy rate of 98,52% while model testing to see how well the model predicts the label correctly is 97,80%.

Keywords: *Basmati Rice, CNN, Rice Type Detection*

Submitted: 2023-06-15	Revised: 2023-10-01	Accepted: 2023-10-10

Introduction

Rice is one of the staple crops which is an important commodity in Asian societies, including in Indonesia. Rice produced by rice plants is the main ingredient to meet the food needs of Indonesian people in all walks of life. So that it can be said that the agricultural sector, especially rice, is a very important factor related to food security (Abdullah, F. 2022). Based on data from BPS (central statistics agency) in 2022 rice production has increased by 54.7 million rice or 0.6135% compared to the harvested area in 2021 of 54.4 million rice. It was recorded that the results of this harvest produced various types of rice for consumption by people in Indonesia. The various types of rice can be based on the shape, characteristics of the rice, the quality of the rice and the nutritional content of the rice. However, the many types of rice can actually produce problems in choosing the type of rice that will be consumed by the community.

Accuracy in choosing the type of rice is not only good for consumption, or fulfilling market demands, but also on the quality of the rice harvest produced. Such as relatively large harvests, intense harvest periods, resistance to rice pests, resistance to weather and climate and maintenance of rice plants that are easily tolerant of environmental stress, according to the conditions of the planting area (Khamid, M.B., et al., 2019) . This is the background of the Rice Research Center (BB Padi) in collaboration with the Agricultural Research and Development Agency to continue to strive to create new, quality rice varieties that are expected to meet the nutritional and health needs of the Indonesian people. It was from this form of cooperation that they succeeded in assembling a special rice variety for the Basmati type which was named Baroma, short for Aromatic Basmati rice. As the name implies, Basmati itself in Sanskrit means Fragrant or fragrant. In general, Basmati rice has a slender and elongated shape and when cooked it gives off a very fragrant aroma with a savory taste (Afifah, N., et al., 2020).

The large number of rice varieties such as Baroma rice makes it necessary to identify rice varieties. Identification of rice varieties can be done by various methods. One way is to describe the morphology of rice. Such as length, width, ratio of length and width of rice and endosperm color which is genetically inherited (Budiwati, et al., 2019). Differences in the size and shape of different rice can be affected by the length of the rice grains, and the width of the rice grains. The length of a grain of rice is measured between the two ends of a grain of whole rice (National Standardization Agency, 2023). It is the length, width and texture that are quite significant differentiating factors (Adnan, et al., 2013). So that the characteristics of each rice can be used as parameters in classifying and identifying types of rice, using image recognition technology (Cinar, I., 2022).

Techniques for identification of varieties that are cheap, fast, and do not use specific equipment need to be developed to help identify rice reliably. One alternative technique in differentiating rice varieties is using technology that is currently growing rapidly, namely AI. By utilizing the data contained in two-dimensional images. The smallest unit of a digital image in the form of a pixel contains digital information that can be extracted using a digital image application or commonly termed feature extraction (Endang, et al., 2023). The purpose of this feature extraction is to get information about the characteristics or parameters needed to recognize an image. The process of selecting the best features takes a long time and will probably result in quite high errors, especially when the number of datasets is small (Wibowo, A., 2021). The use of Machine Learning, especially Deep Learning, is generally used for cases with complex feature selection and can be done in a large number of datasets, so that it can improve performance in selecting features with a fairly fast time (Cogan, T., et al., 2019). The deep learning method that is often used is the convolutional neural network (CNN), where this method can extract features automatically and more efficiently so that it can produce accurate classifications that are accurate (Zhang, et al., 2018).

Based on the background above, this research was conducted with the aim of obtaining class or type information of some rice, which was used as input test data from the program that had been made. And the information in question is whether the input is Basmati rice, or other types of rice. It is hoped that this research can become one of the innovative steps in realizing food adequacy and security in Indonesia. Apart from that, this research can be a form of collaboration in various fields. Both in the fields of agriculture and computer science in utilizing computer vision systems that can be used to identify types of rice.

Research Methodology

In this study the method used is the Convolutional Neural Network method. The design and implementation process consists of several stages. The research stages are presented in Figure 1 below.

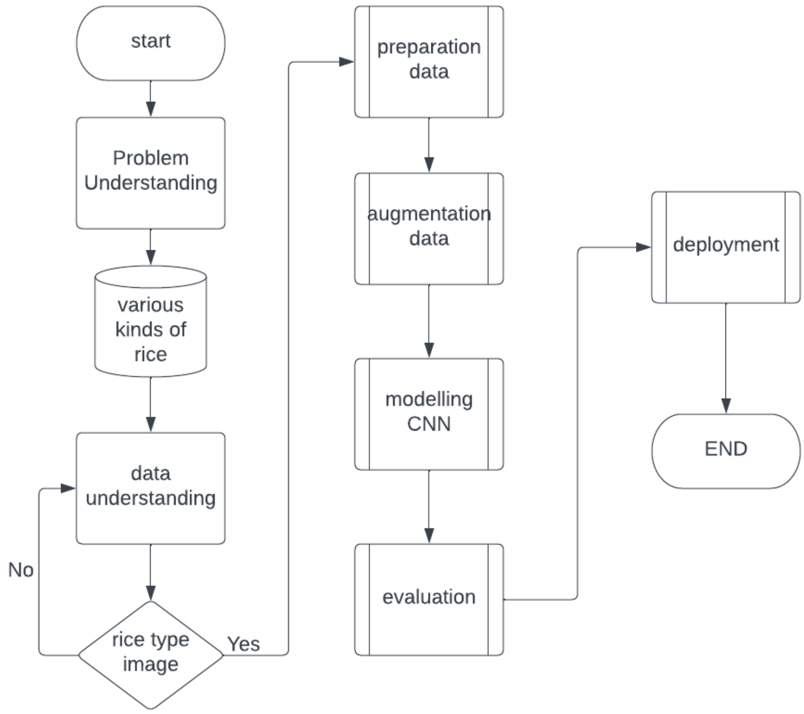

Figure 1. Flow Chart

The flow of this research work began with an understanding of the importance of knowing and detecting the types of rice in Subang district, understanding the differences in morphological types of rice, selecting sample data to be used as input for data training, validation and testing. Then design a network with the CNN method to perform classification and detection of images of rice types. The CNN network design is applied with training data, so that the created network can learn to learn and recognize objects. If network learning yields good results in distinguishing types of rice, then the network is then tested on validation data. If the validation data also shows good results, the network can be used for classification and detection of test data.

1. Data collection

   The data collection on the types of rice used was in the form of types and classes, which included 5 types of rice, including Basmati, Karcadag, Ipsala, Jasmine, and Arborio types. We take this data from the Kaggle catalog. Where the amount of each data that we use in making this system is 46,500 images. Where the images consist of 9300 images of Arborio rice, 9300 images of Basmati rice, 9300 images of Ipsala rice, 9300 images of Jasmine rice, and 9300 images of Karacadag rice. The sample images in each image can be seen in Figure 2 below.

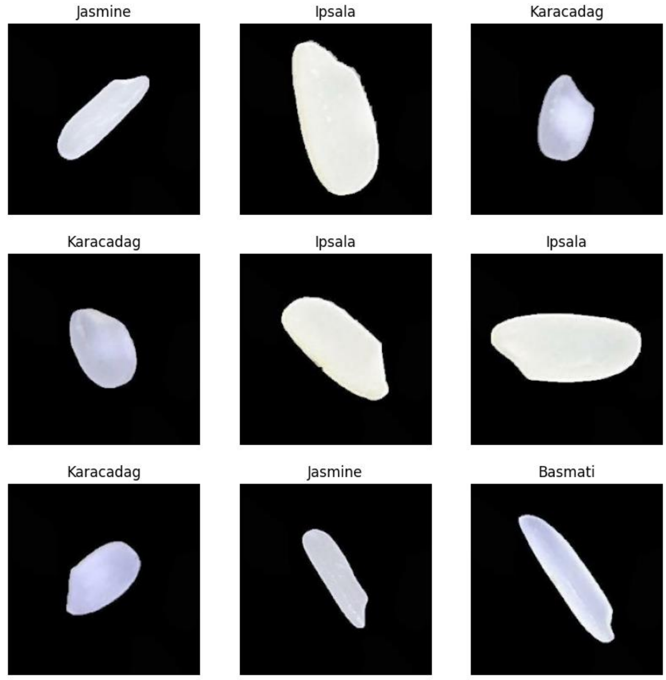

Figure 2. is an example of an image for training data

Overall from the data above, we divide it into training data and test data. Which training data is used to train the model, while test data is used to test how well the model works on data that has not been seen before. The amount of each training data and test data is 37,500 images for training data and 9,000 for test data. The details can be seen in table 1 below.

Table 1. Distribution of Research Data

Rice type	Total training data	Number of validation data	Number of test data
Arborio	7500	1500	1800
Basmati	7500	1500	1800
Ipsala	7500	1500	1800
Jasmine	7500	1500	1800
Karacadag	7500	1500	1800
total	37500	7500	9000

The training data is used to carry out the network learning process, then it is evaluated. If the accuracy in the network model training process has not increased, it is necessary to modify the CNN layer, network parameters and the data samples. If the accuracy results are good, then the next process is carried out, namely testing with validation data. Validation data is data that is not used in the training process. The network validation process uses 7500 data to test the network or model created with the same number of images in each class, namely 1500 images. If the accuracy of the validation data is not good, there is a possibility of overfitting, therefore the network needs to be modified again. So that the results are good or there is an increase, this network can be used to process test data. The test data contains a set of data samples whose classification type you want to know.

2. Preprocessing data

Image preprocessing is done by changing the size of the rice image from a length and width of 250 pixels to a size of 200 pixels in each image for the width and length of the image. The next step is to divide the data that has been collected into two, namely training data and testing data. The labeling process is continued for each image data according to the class of each type of rice. Where the results of this labeling are done using numbers 0 to 4. Number 0 is for the Arborio image label, number 1 is for the Basmati image label, number 2 is for the Ipsala image label, number 3 is for the Jasmine image label, and number 3 is for the Karacadag image label.

3. CNN design

This study proposes a method for detecting rice types based on the shape and size of rice consisting of 5 classes as described in the previous chapter. In the process of detecting the convolution layer, it is carried out on the server until it is finished. Images or datasets will be retrained to get a good CNN model and that will be used to generalize and make good predictions on new data, such as kernel size, filters and layers.

Table 2. Structure of the CNN Model

Lapisan	Ukuran Piksel	Node
rescaling_1 (Rescaling)	(None, 200, 200, 3)	0
conv2d_16 (Conv2D)	(None, 200, 200, 16)	448
max_pooling2d_5 (Max Pooling 2D)	(None, 100, 100, 16)	0
conv2d_17 (Conv2D)	(None, 100, 100, 32)	4640
max_pooling2d_6 (Max Pooling 2D)	(None, 50, 50, 32)	0
conv2d_18 (Conv2D)	(None, 50, 50, 64)	18496
max_pooling2d_7 (Max Pooling 2D)	(None, 25, 25, 64)	0
flatten_1 (Flatten)	(None, 40000)	0
dense_4 (Dense)	(None, 128)	5120128
dense_5 (Dense)	(None, 4)	516

Total params: 5,144,228
Trainable params: 5,144,228
Non-trainable params: 0

In the input layer, the data used is training data. Then the input data is processed in the first convolution layer using maxpooling and the ReLU activation function. The output in the first convolution layer is used as input in the second convolution process. Then the results of the convolution process are collected in the fully connected layer. In this layer, features are determined that have a correlation with certain classes so that the end result of this process is features that are classified into four classes.

Research Results and Discussion

1. Implementation of CNN

There are three stages in implementing CNN, namely training, validation and testing. The training stage is the main stage for training the model or network to study the input data. This model will learn from the training dataset or training data to adjust the parameters or patterns in the data so that it can make accurate predictions. Then the performance of the learning model or algorithm is measured and evaluated using validation data to get how well the model can generalize to data that has never been seen before. The hope is that the model that has been made is accurate and reliable. During the training and validation process, adjustments to model hyperparameters such as learning rate, number of layers and size of the convolution filter are made to obtain better model performance. After the model is considered good enough, based on the evaluation of the validation dataset, the next step is to evaluate the model using the test dataset to test the final performance of the model that has been made. This evaluation will provide an overview of how well the model can generalize to new data that has never been seen.

2. Network or Model training

In the model training process is done using training data. The percentage of training data used is 80% of the total data. In this data, there is also data used for the validation process and this data is separate from the training data. From the training data 80% for the training process and 20% for the validation process. In detail, 30,000 images are used for the training process, and 7,500 are used for the validation process. Because there are 5 classes of rice types, the data for each class is training data of 6000 and 1500 for the validation process. While the computational process is carried out using single GPU mode. The training process uses the following parameters:

Learning rate	: 0,0001
Mini-batch size	: 16
Epoch	: 10

The training results are presented in table 3. Network training provides good accuracy. The graph of the accuracy and error of the training process is presented in Figure 3.

Table 3. CNN network training results

```
accuracy
      training          (min:    0.924, max:    0.992, cur:    0.992)
      validation        (min:    0.952, max:    0.969, cur:    0.967)
Loss
      training          (min:    0.026, max:    0.211, cur:    0.026)
      validation        (min:    0.135, max:    0.229, cur:    0.221)
450/450  [==============================]  -  8s  19ms/step  -  loss:
0.0264 - accuracy: 0.9915 - val_loss: 0.2210 - val_accuracy: 0.9672
```

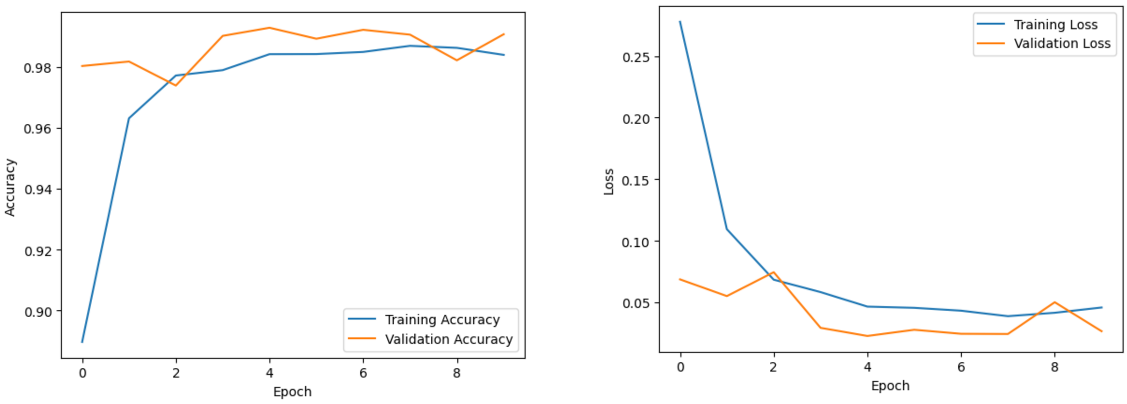

Figure 3. Graph of Accuracy and Training Error

3. Model testing

The interpreter can enter data samples that you want to know the type of classification on the network, then the network will issue a plant type label based on the data entered. The results of the classification labels that come out of the network can be taken into consideration by the interpreter in determining the types of plant objects that are difficult to distinguish visually. In this study the test data scenario uses 9000 data with the number of each class according to the information in table 1. This test produces a good accuracy of 95.74% with a correct number of 8910 data. While the error in predicting the image into the highest rice type is 46 images as Basmati rice type, and 6 as Ipsala rice type, even though the 46 images should belong to Jasmine rice type. This is because jasmine rice has almost the same size characteristics as basmati rice, both in length and width. The results of the classification accuracy of the complete test data can be seen in the confusion matrix in Figure 4 and the results of the semantic classification on the test we also tested as a detection process on an image of a certain type of rice, and the results of the detection are as shown in Figure 5.

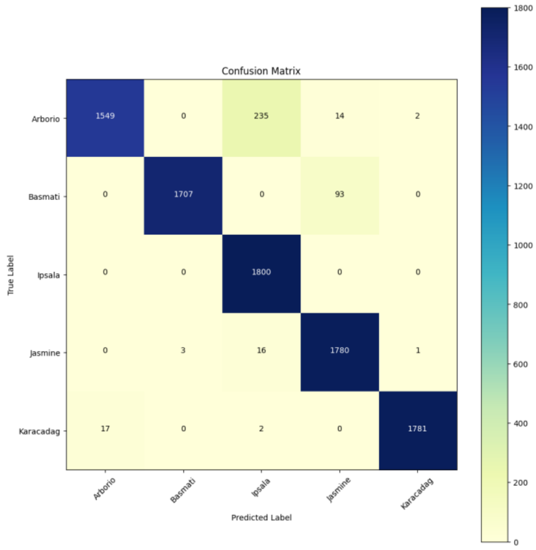

Figure 4. Confusion Matrix From Test Data

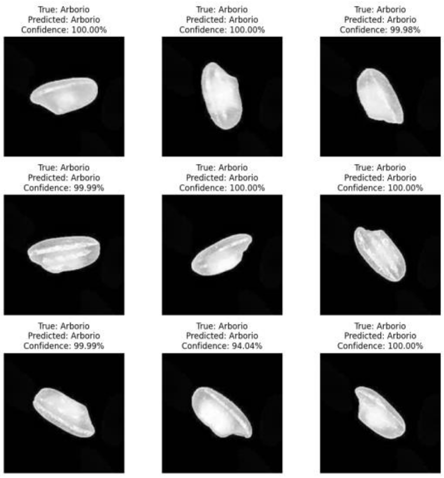

Figure 5. Semantic classification results on test data and test data samples on some data.

Classification Report:

	precision	recall	f1-score	support
Arborio	0.99	0.99	0.99	1800
Basmati	0.98	1.00	0.99	1800
Ipsala	0.99	1.00	1.00	1800
Jasmine	0.99	0.97	0.98	1800
Karacadag	1.00	0.99	1.00	1800
accuracy			0.99	9000
macro avg	0.99	0.99	0.99	9000
weighted avg	0.99	0.99	0.99	9000

Figure 6. Classification report

In the classification model report that has been made as shown in Figure 6, it can be concluded that the model has good performance with an accuracy rate of around 99% for the test dataset consisting of 9000 samples. This is reinforced by the high f1-score values for each class which can indicate that the model can recognize various classes very well.

Conclusion

The results of the implementation of the CNN method to identify types of basmati rice in Subang district show that the CNN architecture can identify 5 types of varieties very well. Work evaluation of the CNN network architecture on the test data produces an accuracy of 99%. For the test data scenario with a total of 5 classes tested as many as 1800 samples. And from the model or network that has been made, it is able to recognize each rice image input very well too. Although there is an error in identifying the type of rice, the error in predicting is not too significant. The most prediction errors were for Jasmine rice, with 55 images with details of 46 identified as Basmati rice, 6 images identified as Ipsala rice, 2 as Arborio rice, and 1 image as Karacadag rice. There was confusion in identifying it because the type of Jasmine rice does have characteristics and morphology that are almost the same as the type of Basmati rice or has a visual resemblance to Basmati rice. Where morphologically, jasmine rice has a long grain size, with a slightly rounded tip shape and a translucent color. And this characteristic is also shared by the Basmati rice type, although in general Basmati rice has a longer grain size, it can even be said that the Basmati rice type has the longest grain size compared to the type of rice used in this study.

To overcome this, it is better to try using a different model, it needs further handling in the sense that approaches are carried out such as strengthening the data by multiplying the images used both for needs during the training and validation processes, carrying out the augmentation process to create variations the more in the training dataset, the model changes.

Basically the research was conducted to determine the semantic classification in helping the interpreter determine the type of variety of a Rice. The interpreter in question is to enter a data sample of the object whose classification you want to know, then the network gives answers about the type of rice and the accuracy of the object. In this way, of course, it can help interpreters whose varieties were previously unknown and can be used as material for consideration in digitizing rice varieties for progress in the field of computer vision and agriculture.

In a computational fluid dynamics simulation study on gas combustion flares resulting from biomass gasification, it can be concluded that; a) the combustion distribution occurs inside the flare with a temperature of 1106°C in the upper area close to the outlet boundary, b) the wall temperature comparison results show that the CFD simulation has a tendency similar to the test results. This shows that CFD simulation can be used to predict fluid flow rates and combustion reactions.

Reference

Abdullah, F., et al. et al. 2022. Analisis ketersediaan beras di kabupaten Gorontalo selang tahun 2021-2030. Agrinesia. Vol. 6, No. 3, 187-197.

Adnan, et al. 2013. Identifikasi jenis dan mutu beras (Oryza sativa) berdasarkan analisis warna dan bentuk menggunakan teknologi pengolahan citra dan jaringan saraf tiruan (JST). Penelitian pertanian tanaman pangan, Vol. 32, No. 2, 91-97.

Afifah, N., Zakiyah, N. 2020. Review artikel:indeks glikemik pada berbagai varietas beras. Farmaka, Vol. 18, No. 2, 42-49, DOI: https://doi.org/10.24198/farmaka.v18i2.27781.

Çınar, İ. & Koklu, M. 2022. Identification of Rice Varieties Using Machine Learning Algorithms . Journal of Agricultural Sciences , 28 (2) , 307-325 . DOI: 10.15832/ankutbd.862482.

Cogan T, Cogan M, Tamil L. MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. Comput Biol Med. 2019 Aug;111:103351. doi: 10.1016/j.compbiomed.2019.103351. Epub 2019 Jul 10. PMID: 31325742.

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 Remotely Sensed Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 8, pp. 4507-4521, Aug. 2018, doi: 10.1109/TGRS.2018.2822783.

Endang, S. & Wibowo, A. 2023. Klasifikasi dan deteksi keretakan pada trotoar menggunakan metode convolutional neural network. JTSC, Vol. 4, No. 1, 412-427, DOI: https://doi.org/10.51988/jtsc.v4i1.116.

Hanas, D.F., et al. 2017. Karakter morfologi beras sebagai pembeda varietas padi. Indonesian journal of legal and forensic sciences, Vol. 7, 23-28, DOI: https://doi.org/10.24843/IJLFS.2017.v07.i01.p04.

Khamid, M.B.R., et al. 2019. Respon pertumbuhan dan hasil padi (Oryza Sativa L.) terhadap cekaman suhu tinggi. Indonesian Journal of Agronomy, Vol. 47, No. 2, 119-125. DOI: https://doi.org/10.24831/jai.v47i2.23854.

Statistik SNI 2023. Badan Standar Nasional. 17 Juni 2023. 28 Juli 2023. < https://bsn.go.id/main/bsn/isi_bsn/20423/statistik-sni-2023>

Wibowo, A. & Putra, M. 2021. Pemanfaatan kamera DSLR Canon 1200D untuk pengamatan fotometri bintang variabel. Seminar Panorama Antariksa 2021, 978, 78-81, 2022.

Wibowo, A. & Lusiana. 2022. Budidaya Tanaman Aren Sebagai Langkah Strategis Mewujudkan Hutan Lestari Di Subang. Sadeli: Jurnal Pengabdian kepada Masyarakat, Vol. 2, No. 216-24, 2022.

Yulianto, Y. & Wibowo, A. 2023. Deteksi keretakan perkerasan jalan aspal menggunakan metode convolutional neural network. JTSC, Vol. 4, No. 2, 581-593, DOI: https://doi.org/10.51988/jtsc.v4i2.