Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 6, No.1, January 2024, pp. 02401014-01 ~ 02401014-09
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v6i1.17821

The Effect of LAB Color Space with NASNetMobile Fine-tuning on Model Performance for Crowd Detection

Muhammad Rafid*, Ardytha Luthfiarta, Muhammad Naufal, Muhammad Daffa Al Fahreza, Michael Indrawan

Faculty of Computer Science, Universitas Dian Nuswantoro Semarang, Jl. Imam Bonjol No 207 Semarang 50131, Central Java, Indonesia

*111202012803@mhs.dinus.ac.id

Abstract. In the COVID-19 pandemic, computer vision plays a crucial role in crowd detection, supporting crowd restriction policies to mitigate virus spread. This research focuses on analyzing the impact of using the RGB LAB color space in advance to perform NASNetMobile better. The fine-tuning process, involving freezing layers in various NASNetMobile base model variations, is considered. Results reveal that the model with LAB color space outperforms the model with RGB color space, with an average accuracy of 94.68% compared to 94.15%. From all the test iterations, it was found that the highest performance for the NASNetMobile model occurred when freezing 10% of the layers from the back for both model LAB and RGB color spaces, with the LAB color space achieving an accuracy of 95.4% and the RGB color space achieving an accuracy of 95.1%.

Keywords: Convolutional Neural Network (CNN), Fine-tuning, LAB, NASNetMobile, RGB

(Received 2023-12-15, Accepted 2023-12-21, Available Online by 2024-01-02)

1. Introduction

Computer vision, as one of the main branches of artificial intelligence, has had a significant impact on various aspects of human life [1]. With its ability to understand and process visual information, computer vision plays an important role in various social activities. One important context in computer vision applications, especially during the COVID-19 pandemic, is crowd detection. One of the most effective methods in processing visual data is Convolutional Neural Networks (CNN), which has shown its superiority in various image processing applications. In this context, CNNs are used to detect crowds of people to support crowd restriction policies so as to prevent the spread of disease outbreaks. In addition, this technology is also used in other activities, such as detecting the use of masks [2] and automatic human counts [3].

Although there has been significant progress in the application of CNNs to various social activities, research continues to be developed in search of models with better accuracy. Several CNN models, including ResNet, DenseNet, MobileNet, and NASNetMobile, have been proven to excel in solving visual detection problems. NASNetMobile has stood out as an efficient and robust model in image recognition tasks of the various existing CNN architectures.

Several studies focusing on studying popular deep learning architectures have been conducted over the past few years [4]–[6]. Some literature has discussed the CNN model, especially regarding the performance of NASNetMobile.

For example, Anwar Fuadi [4] researched the comparison of MobileNet and NASNetMobile architectures for disease classification in potato leaf images. This research aims to compare MobileNet and NASNetMobile architectures in performing disease detection on potato plant leaves. The data used in this study are divided into images of healthy potato leaves, images of potato leaves infected with Early Blight, and images of potato leaves infected with Late Blight. The research concludes that models using NASNetMobile architecture produce better model evaluation results than MobileNet with certain schemes.

Ahsan M [5] conducted research on patient symptom detection with COVID-19 using eight different models including VGG16, InceptionResNetV2, ResNet50, DenseNet201, VGG19, Mo-bilenetV2, NasNetMobile, and ResNet15V2, using two datasets namely 400 CT scans and 400 chest X-ray images. The results of this study show that the NASNetMobile model outperforms all other models.

Enkvetchakul P [6] presents a plant leaf disease recognition system using two CNN Architectures namely MobileNetV2 and NASNetMobile. The results of the study show that the architecture that has the highest accuracy for plant leaf disease recognition is the NASNetMobile architecture using transfer learning. These results are obtained when combining offline training techniques with data augmentation techniques.

In addition, the accuracy results of several CNN models, including NASNetMobile, are also affected by the selection of color features. Therefore, it is important to explore the potential of some CNN models further and understand how the use of color space can affect their performance. Several color spaces, such as RGB, LCH, LAB, and others, are often used for image processing research. Several studies have conducted discussions on the differences in these color spaces.

Thevarasa N [7] researched the classification of mosquito breeding locations using various approaches such as CNN, SVM, and FSL. The study also compared the use of datasets with RGB and LAB color spaces in training CNN. The results showed that CNN training using LAB color space produced better performance with an accuracy rate of 90%, while CNN training with RGB color space only achieved an accuracy of 84.29%. Meanwhile, SVM and FSL achieved an accuracy rate of 79% and 80%, respectively.

Gowda S [8] conducted a study that explored the effect of various color spaces on the level of accuracy in training deep learning CNN. The study tested several color spaces, such as RGB, HSV, YUV, LAB, YIQ, XYZ, YPbPr, YCbCr, HED, and LCH, using the Cifar-10 dataset. The results showed that the LAB color space achieved the highest accuracy rate than other color spaces, reaching 80.43%, with a training time of 26 seconds.

This research will focus on the application of the NASNetMobile CNN model by applying LAB color space features and fine-tuning for crowd-detection purposes. The selection of NASNetMobile as the main model is based on its proven advantages in previous studies. In contrast, the selection of LAB color space is motivated by previous studies showing its potential in improving visual detection accuracy.

This research aims to investigate the effect of LAB color space by fine-tuning NASNetMobile on the performance of crowd detection. By considering this aspect, this research is expected to provide more insight into how NASNetMobile can be optimized by considering the use of specific color spaces. This research will collect test iteration results of various freezing layer variations in two RGB and LAB color spaces on the NASNetMobile model for crowd detection. We hypothesize that the use of RGB LAB color space can affect the model's performance, and adjusting the freezing layer range will play an important role in strengthening the experimental results.

With these experiments, we expect to obtain more precise information about how color space variables might be used to optimize deep learning models, especially NASNetMobile, for crowd-detection tasks. The findings of this research significantly improve image recognition technology and enhance model performance in various application scenarios.

2. Research Methods

2.1. Research Flowchart

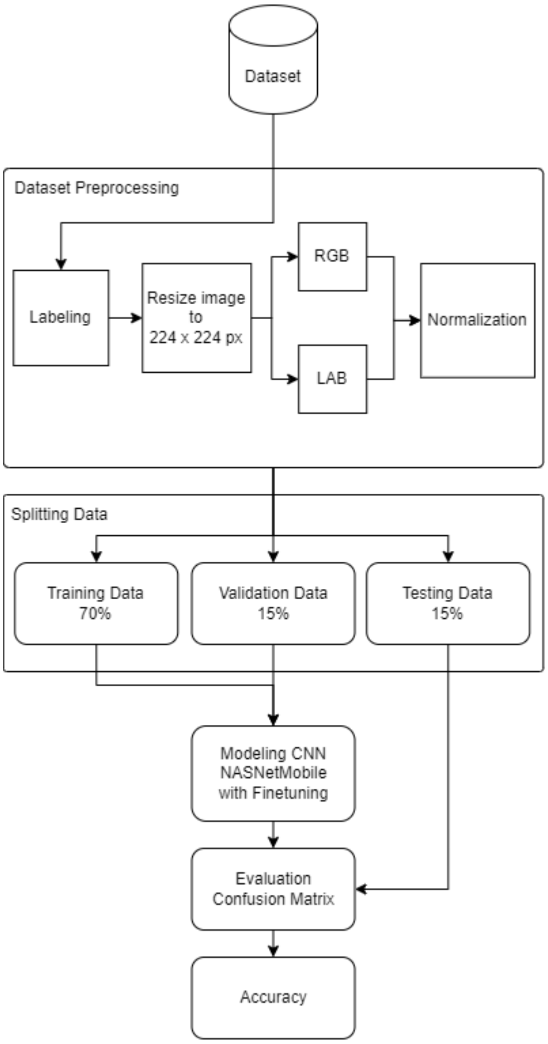

Figure 1. Research Flowchart

Figure 1 illustrates how the research begins with the initial stage of dataset preparation. The dataset will then undergo pre-processing to prepare it for model training. After that, the dataset processed in the earlier stage will be split into three sets: testing, validation, and training.

Next, the NASNetMobile CNN modeling process with fine-tuning using freezing layer. There are several freezing layer ranges used starting from overall freezing, 10% from the front, 20% from the front, 25% from the front, 30% from the front, 40% from the front, 50% from the front, 10% from the back, 20% from the back, 25% from the back, 30% from the back, 40% from the back, and 50% from the back as can be seen in Figure 3.

The model will be trained using the training data during the training phase. The confusion matrix will be used as an evaluation metric to evaluate the model's performance. This metric measure offers insight into the model's functionality, which will be utilized as evaluation material. The accuracy value serves as a standard for a number of the behaviors used in this research.

2.2. Dataset

The Crowd Human dataset, which is an open-source collection of images in a range of sizes, positions, and backdrops, is used in this study [9], [10]. There are 15,000 images in this dataset overall. This dataset has been processed by previous research [11] to highlight the important areas of each image.

2.3. Dataset Preprocessing

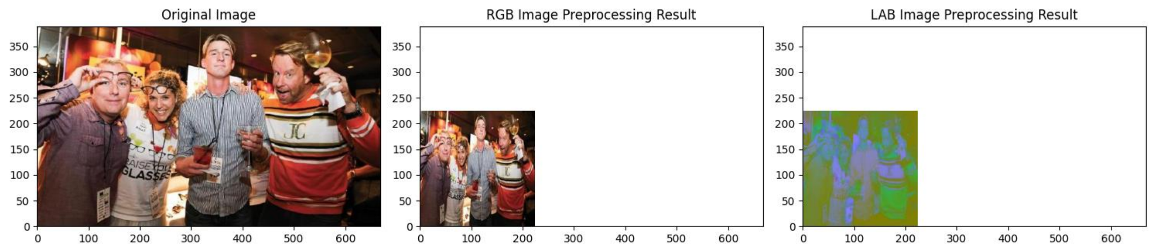

Figure 2. Dataset Preprocessing Results

Of the 15000 images in the Crowd Human collection, only 30% are used in this research, which is 5500 images. These images are going to be set up for model training. Initially, the images are divided into two classes: the crowd class and the non-crowd class. The image is then resized to 224 by 224 pixels. Since the NASNetMobile model requires a certain image size to be learned, this step is carried out to prepare the images or data for training on the model.

After that, this is where the color space distinction will be used. An additional process for the LAB color space is needed, which is converting the color space from RGB to LAB. In contrast, the RGB color space does not need to do an additional process because the original color of the dataset is RGB. An explanation of RGB and LAB color spaces can be seen in section 2.4.

Last, each image's pixel intensity is normalized, applying its value to a range from 0 to 1. This normalizing step is essential to guarantee consistency and improved performance during the model training phase. The results of pre-processing the RGB and LAB versions of the images can be seen in Figure 2.

2.4. Convert RGB color to LAB color

RGB (Red, Green, Blue) color space is one of the additive color space models in images that produce colors at each pixel. RGB color space is very common in electronic devices, such as monitor screens, televisions, and other display devices [12]. This model is relatively easy to understand and use in image processing due to its intuitive nature, but it is not fully capable of representing all colors that the human eye can see.

LAB color space is one of the color space models specifically designed to represent all colors that can be seen by the human eye [12]. Another advantage of the LAB color space is that it segmented accurate colors. By separating the color and brightness components, LAB color space can be used to identify objects based on their color difference from the background or other objects. LAB color space is invariant to lighting changes, meaning that colors in this color space are more stable to changing lighting conditions. This makes LAB color space commonly used in applications that require color accuracy, such as medical image processing and color quality in photography.

Converting RGB color to LAB color is done through two conversion stages, i.e., RGB color conversion to XYZ color and XYZ color conversion to LAB color. RGB color conversion to XYZ color can be done through equation (1).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.00 & 0.066 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

The XYZ color is converted to LAB color using the following equation (2-5).

$$L^* = 116\sqrt[3]{\frac{Y}{Y_n}} - 16, \text{ untuk } \frac{Y}{Y_n} > 0.008856 \quad (2)$$

$$L^* = 903.3\frac{Y}{Y_n}, \text{ untuk selain } \frac{Y}{Y_n} > 0.00885 \quad (3)$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \quad (4)$$

$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \quad (5)$$

Therefore, LAB color space is often used for image processing that requires high precision in color reproduction, color mapping, and color adjustment.

2.5. Splitting Data

The 5500 images will be separated into two classes: non-crowd (2250 images) and crowd (2250 images). Moreover, the dataset will be split into three sets: 70% of the images are training data (3850 images), 15% are validation data (825 images), and 15% are testing data (825 images).

2.6. Modeling CNN NASNetMobile with Fine-tuning

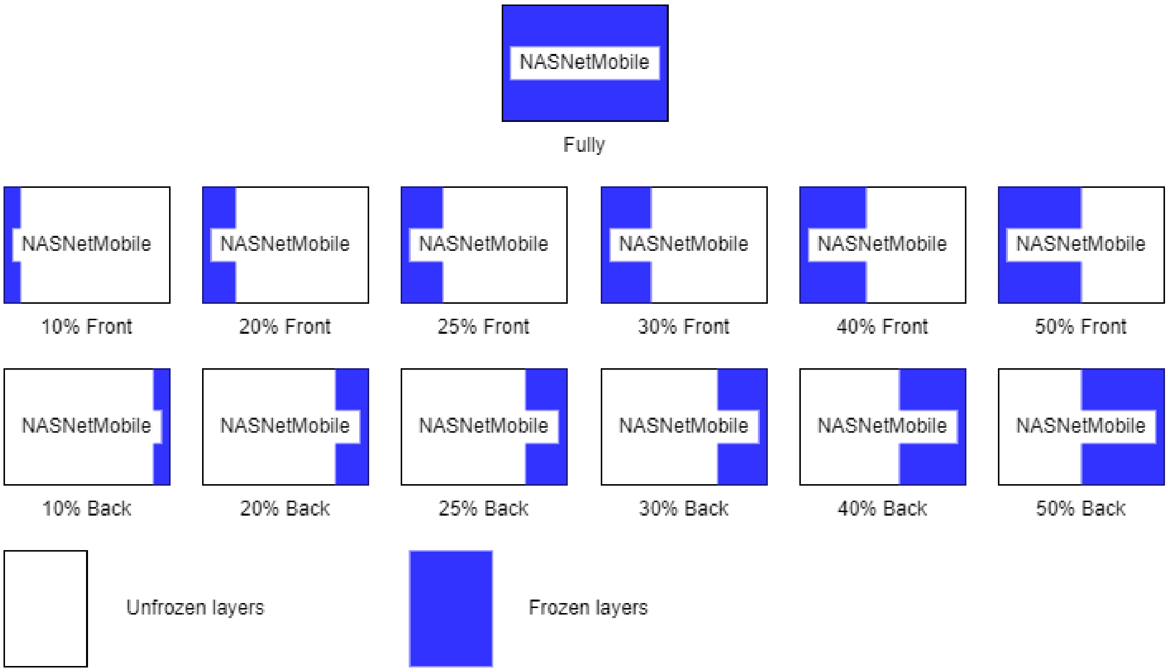

Figure 3. Freezing Layer Variations

In this research, the convolutional neural network model used is NASNetMobile. NASNetMobile has a mobile-friendly design, making it a practical choice for real-world applications requiring image classification on mobile devices. The architecture of NASNetMobile includes multiple convolutional layers designed to extract features from images, making it suitable for investigating the effect of different color spaces on image classification performance.

To sample the performance of the NASNetMobile model, the transfer learning method is carried out and then processed by fine-tuning by freezing the layer with various variations, i.e., fully freezing, 10% from the front, 20% from the front, 25% from the front, 30% from the front, 40% from the front, 50% from the front, 10% from the back, 20% from the back, 25% from the back, 30% from the back, 40% from the back, and 50% from the back as can be seen in Figure 3. The results of these various variations will be used as comparison samples to find out the effect of RGB and LAB color spaces on NASNetMobile performance.

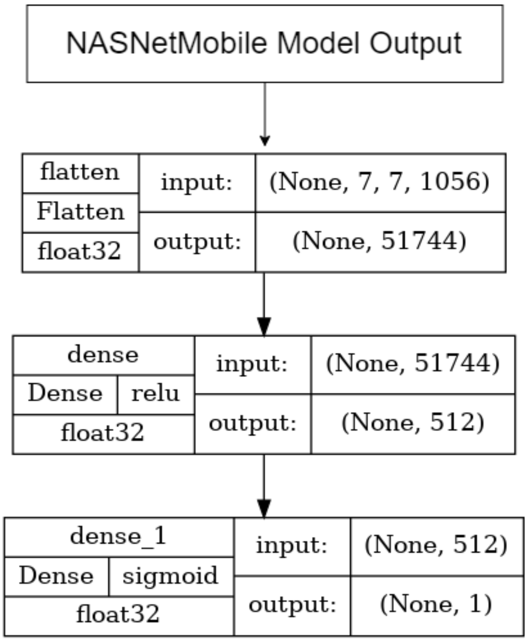

Figure 4. Modelling NASNetMobile

Additionally, three training layers are added, as shown in Figure 4: a flattened, dense, and final dense layer. From the previous output layer, which results from the transfer learning of the NASNetMobile CNN architecture, the flattened layer is utilized to create a one-dimensional vector [13], [14]. The final dense layer contains 1 neuron, while the dense layer contains 512 neurons. In the dense layer, the Rectified Linear Unit (Rectified) activation function is also utilized to resolve the gradient loss issue and enhance training effectiveness [15]. L2 kernel regulation is also used in the dense layer to prevent overfitting. The final dense layer is used to process data and analyze more complex patterns than the previous layers [16], [17]. In the last dense layer, sigmoid is also utilized. Sigmoid is an activation function that is utilized for binary classification activities. The sigmoid function's main objective is to translate the model's output into a probability value between 0 and 1 [18].

2.7. Evaluation

In this research, two color spaces, RGB and LAB, were evaluated with several samples obtained through various freezing layers. The evaluation is done by comparing the confusion matrix and classification report. The confusion matrix evaluates how well the model performs in classification. A classification report is an instrument that provides an overview of the precision and recall for each class. The comparison value used is the accuracy value in the classification report.

3. Results and Discussion

3.1. Implementation

Implementation is done by applying the NASNetMobile model with three-layer modifications. The three layers are the flattened layer, the dense layer, and the final dense layer. Implementation is done with the Crowd Human dataset. The training model employs binary cross-entropy as the loss function, learning rate 1e-4, batch size 32, and SGD optimizer. The early stop and patience techniques were applied for 10 epochs during the 100 epochs of the training phase. Each test run takes about 35 minutes. All test results also show that the learning process runs smoothly until the 100th epoch without an early stop.

3.2. Evaluation of Results

Table 1. Accuracy Test Iteration Results (%)

Model Color Space	**Freezing Layer**													**Avg**
	Fully	**Front**						**Back**						
		10%	**20%**	**25%**	**30%**	**40%**	**50%**	**10%**	**20%**	**25%**	**30%**	**40%**	**50%**	
RGB NASNet Mobile	93,8	93,7	93,5	94,2	94,3	93,9	94,2	95,1	94,3	94,2	94,8	93,7	94,3	94,15
LAB NASNet Mobile	94	94	94	95,2	94,9	94,8	94,4	95,4	95,1	94,8	95,2	94,5	94,6	94,68

From Table 1, we can see the results of the experimental tests that have been carried out, starting from freezing the layer entirely up to freezing the layer 50% from the back for both RGB and LAB color spaces. The results of experimental tests that have been carried out on the NASNetMobile model are to compare the performance differences between the RGB color space and the LAB color space. From 13 freezing layer experiments, all experiments show that the model's accuracy with LAB color space is higher than that of the model with RGB color space.

From all the experimental results, it is also known that the highest accuracy results for the NASNetMobile model occur when freezing the layer 10% from the back. In this case, the LAB color space still shows higher accuracy than the RGB color space, with the highest values of 95.4% and 95.1%, respectively. LAB color space clearly stands out in better pattern recognition capabilities.

The average accuracy calculation results show that the RGB color space gets an average value of 94.25% while the LAB color space gets an average value of 94.58%. This proves that the LAB color space performs better than the RGB color space, with a difference of 0.53%.

3.3. Analysis and Discussion

In the NASNetMobile model, three additional layers are added as the basic model of this research. Furthermore, the model will be trained using the Human Crowd dataset, which has been divided into two classes, such as the crowd class and the non-crowd class, which has passed the pre-processing stage according to the color space to be tested (RGB or LAB). Then the model is iterated with various freezing layer variations, including a fully freezing layer, 10% from front, 20% from front, 25% from front, 30% from front, 40% from front, 50% from front, 10% from back, 20% from back, 25% from back, 30% from back, 40% from back, and 50% from back.

All test accuracy results were stored, recorded, and compared in one table, as shown in Table 1. The table of experimental test results shows that the model with LAB color space provides consistently higher accuracy results than the model with RGB color space. This is proven by 13 freezing layer experiments; all experiments show that the model's accuracy with LAB color space is higher than the model with RGB color space.

The freezing layer, 10% from the back, has the highest accuracy compared to all freezing layer variations. This can happen because, in crowd detection, special features or patterns that detect crowds more accurately are located in the deeper layers of the network. By freezing the layer 10% of the way back, the model can maintain its ability to capture these important features. The table of experimental test results also shows that the average calculated result of the RGB color space gets an average value of 94.15%. In comparison, the LAB color space gets an average value of 94.68%. The difference between the two average values is 0.53%. All these test results show that the LAB color space clearly stands out in better pattern recognition ability in crowd images.

4. Conclusion

Based on the evaluation results, all iterations of test results show good accuracy for crowd and non-crowd image classification, which is above 90%. The experimental test results table shows that the LAB color space performs better than the RGB color space on the performance of the NASNetMobile model for crowd detection. This is proven by all experimental results showing that the model with LAB color space provides consistently higher accuracy results than the model with RGB color space.

The average value of the two color spaces also shows that the LAB color space is higher than the RGB color space, with an average accuracy value of 94.68% and 94.15%, respectively. The difference from the average obtained is 0.53%. Although the difference in numbers looks small, it is also a good improvement, considering the accuracy results of all test iterations are already very high, above 90%.

The table of experimental test results also shows that the freezing layer 10% from the back in the NASNetMobile model is the highest accuracy result. This is because, in crowd detection, special features or patterns that can detect crowds more accurately are located in the deeper layers of the network. By freezing the back 10% layer, the model can maintain its ability to capture these important features.

NASNetMobile models with LAB color space and fully freezing layers are proven to show improvements in model performance. In addition, implementing the tuning parameter of freezing at 10% of the back improved the accuracy significantly. The findings of this research contribute to further development in image processing applications and understanding of optimal fine-tuning, i.e., freezing layer mechanisms in Deep Learning environments.

References

[1] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, “Deep learning in computer vision: A critical review of emerging techniques and application scenarios,” *Machine Learning with Applications*, vol. 6, p. 100134, Dec. 2021, doi: 10.1016/j.mlwa.2021.100134.

[2] A. I. Pradana, “Deteksi Ketepatan Pengunaan Masker Wajah dengan Algoritma CNN dan Haar Cascade,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 3, pp. 2305–2316, Sep. 2022, doi: 10.35957/jatisi.v9i3.2912.

[3] D. T. Laksono, I. N. Husna, M. Ulum, A. K. Saputro, M. F. Fahmi, and D. N. Purnamasari, “SISTEM DETEKSI DAN PERHITUNGAN JUMLAH MANUSIA DALAM RUANGAN MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK,” *Jurnal Simantec*, vol. 11, no. 1, pp. 131–138, Dec. 2022, doi: 10.21107/simantec.v11i1.19745.

[4] A. Fuadi and A. Suharso, “PERBANDINGAN ARSITEKTUR MOBILENET DAN NASNETMOBILE UNTUK KLASIFIKASI PENYAKIT PADA CITRA DAUN KENTANG,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 3, pp. 701–710, Aug. 2022, doi: 10.29100/jipi.v7i3.3026.

[5] M. M. Ahsan, K. D. Gupta, M. M. Islam, S. Sen, Md. L. Rahman, and M. Shakhawat Hossain, “COVID-19 Symptoms Detection Based on NasNetMobile with Explainable AI Using Various Imaging Modalities,” *Mach Learn Knowl Extr*, vol. 2, no. 4, pp. 490–504, Oct. 2020, doi: 10.3390/make2040027.

[6] P. Enkvetchakul and O. Surinta, “Effective Data Augmentation and Training Techniques for Improving Deep Learning in Plant Leaf Disease Recognition,” *Applied Science and Engineering Progress*, Jan. 2021, doi: 10.14416/j.asep.2021.01.003.

[7] N. Thevarasa, G. Ananthajothy, R. Navaratnam, and M. B. Dissanayake, “Weighted Ensemble Algorithm for Aerial Imaging Based Mosquito Breeding Sites Classification,” in *2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*, IEEE, Aug. 2023, pp. 347–352. doi: 10.1109/ICIIS58898.2023.10253588.

[8] S. N. Gowda and C. Yuan, "ColorNet: Investigating the Importance of Color Spaces for Image Classification," 2019, pp. 581–596. doi: 10.1007/978-3-030-20870-7_36.

[9] X. Shao *et al.*, "Multi-Scale Feature Pyramid Network: A Heavily Occluded Pedestrian Detection Network Based on ResNet," *Sensors*, vol. 21, no. 5, p. 1820, Mar. 2021, doi: 10.3390/s21051820.

[10] S. Shao *et al.*, "CrowdHuman: A Benchmark for Detecting Human in a Crowd," Apr. 2018.

[11] P. Langgeng, W. E. Putra, M. Naufal, and E. Y. Hidayat, "A Comparative Study of MobileNet Architecture Optimizer for Crowd Prediction," *Semarang 123 Jl. Imam Bonjol No*, vol. 8, no. 3, p. 50131, 2023.

[12] D. Alamsyah and D. Pratama, "Segmentasi Warna Citra Bunga Daisy dengan Algoritma K-Means pada Ruang Warna Lab," *Jurnal Buana Informatika*, vol. 10, no. 2, Oct. 2019, doi: 10.24002/jbi.v10i2.2458.

[13] B. Kaddar, H. Fizazi, M. Hernandez-Cabronero, V. Sanchez, and J. Serra-Sagrista, "DivNet: Efficient Convolutional Neural Network via Multilevel Hierarchical Architecture Design," *IEEE Access*, vol. 9, pp. 105892–105901, 2021, doi: 10.1109/ACCESS.2021.3099952.

[14] W. Sun, Q. Wei, L. Ren, J. Dang, and F.-F. Yin, "Adaptive respiratory signal prediction using dual multi-layer perceptron neural networks," *Phys Med Biol*, vol. 65, no. 18, p. 185005, Sep. 2020, doi: 10.1088/1361-6560/abb170.

[15] E. Belcore and V. Di Pietra, "LAYING THE FOUNDATION FOR AN ARTIFICIAL NEURAL NETWORK FOR PHOTOGRAMMETRIC RIVERINE BATHYMETRY," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-4/W1-2022, pp. 51–58, Aug. 2022, doi: 10.5194/isprs-archives-XLVIII-4-W1-2022-51-2022.

[16] A. M. Javid, S. Das, M. Skoglund, and S. Chatterjee, "A ReLU Dense Layer to Improve the Performance of Neural Networks," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Jun. 2021, pp. 2810–2814. doi: 10.1109/ICASSP39728.2021.9414269.

[17] C. Fang, H. He, Q. Long, and W. J. Su, "Exploring deep neural networks via layer-peeled model: Minority collapse in imbalanced training," *Proceedings of the National Academy of Sciences*, vol. 118, no. 43, Oct. 2021, doi: 10.1073/pnas.2103091118.

[18] X. Wang, H. Ren, and A. Wang, "Smish: A Novel Activation Function for Deep Learning Methods," *Electronics (Basel)*, vol. 11, no. 4, p. 540, Feb. 2022, doi: 10.3390/electronics11040540.