Tipburn and Leaf Spot Detection on Strawberry Plants Using Convolutional Neural Network

[1] Erika Lety Istikhomah Puspita Sari, [2]I Ketut Agung Enriko, [3]Muhammad Imam Nashiruddin, , [4]Al Yafi,
[1,2]Department of Telecomunication Engineering, Faculty of Engineering, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia,[3]School of Electrical Engineering, Faculty of Engineering, Telkom University, Bandung, Indonesia, , [4]Department of Electrical and Computer Engineering, Faculty of Engineering, Syiah Kuala University, Banda Aceh, Indonesia
[1]erikalety459@gmail.com, [2]enriko@ittelkom-pwt.ac.id*, [3] 2imamnashir@telkomuniversity.ac.id,
[4] alyafi@mhs.unsyiah.ac.id,

Abstract— *Nowadays, the process of identifying plants is still manual and fraught with difficulties owing to human nature. Human nature contains a flaw that renders the desired outcome ineffectual. Another issue is that strawberry plant diseases such as tipburn and leaf spots can damage growth and crop quality and impact the agricultural economy. So the researchers created a Deep Learning model using the Convolutional Neural Network (CNN) VGG16 Algorithm and a dataset of 2897 photos to classify tipburn, leaf spot, and, the healthy state of strawberry plant leaves. To minimize overfitting in classification training, the training dataset will be included. This is done so that the model can recognize the fundamental variance of the strawberry leaf picture object and achieve training and validation accuracy of 95.05% and 97.4%, respectively. Thus, the training loss value is 19.68%, whereas the validation loss value is just 7.54%. The finding accuracy was greater than 90% for both training and validation parameters. This research is expected to be valuable in giving information on the process of data augmentation and disease classification in strawberry plants.*

Index Terms—*srawberry, convolutional neural network, tipburn, leaf spot, healthy, accuracy*.

I. INTRODUCTION

Strawberry output fell significantly between 2014 and 2018, from 58,884 tons in 2014 to 8,831 tons in 2018. Tipburn and leaf spot are the two most serious diseases affecting strawberry fruit output [1]. Tipburn is caused by a local calcium shortage, which is often seen at the shoot tips of quickly developing leaves [2]. In rare situations, tipburn can cause fruit harvest to be delayed, maintenance expenditures to be increased due to the requirement to pluck buds, and in severe cases, growing point death, resulting in production failure [3]. Leaf spot is a fungus-induced illness caused by Mycosphaerella fragiae [4]. Symptoms of leaf spots include the emergence of purple circular spots on the top leaf surface [5].

Given the importance of the two diseases, good monitoring and management are required to control disease transmission. Increased non-stomal transpiration (mass flow) or increased relative humidity in greenhouses at night are excellent ways to decrease tipburn and leaf spots using fungicide treatments [6] [7]. Meanwhile, it has been found manually in the monitoring procedure, which is obviously exhausting and time-consuming for farmers [8]. As a result, one effective way that may be used is deep learning artificial intelligence (AI) technology, which has recently been used to identify plant illnesses [9]. Convolutional Neural Networks (CNN) is one of the most often utilized deep learning algorithms today. pooling, and all linked layers to examine the properties of the training dataset [10].

As a result, the researchers proposed developing a CNN model to identify disease symptoms in strawberry plants, particularly on leaves, as a means of preventing fruit growth, by carrying out the training process and dataset validation using 1982 training data images and 924 validation images. Sort tipburn, leaf spot, and their associated conditions. This article is intended to be useful in providing information about the process of data augmentation and disease classification in strawberry plants, as well as a reference for future research and the development of new, similar theories and research.

II. RESEARCH METHOD

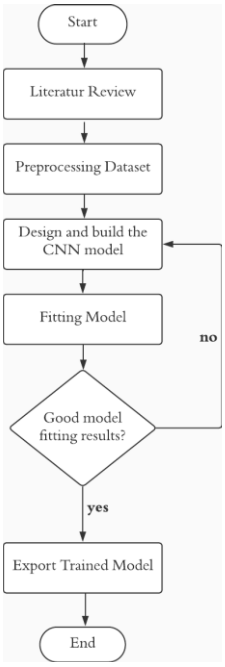
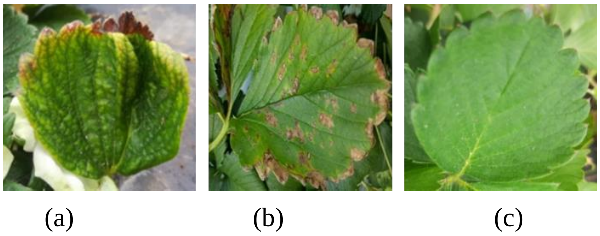


Figure 1. Preprocessing Flowchart

A. *Collecting Dataset*

During the dataset collection process, the main indicators that became the dataset in this study were images of strawberry leaves with symptoms of leaf surface conditions such as tipburn, leaf spot, and healthy leaves. The dataset was obtained from kaggle.com and datadryad.com. Here's a close-up of the strawberry leaves.



Fugure 2. (a) Tipburn, (b) Leaf Spot, (c) Healthy Leaves [14][16][20]

Table 1. Number of Datasets for Each Leaf Classification

Class	Training Data	Validation Data	Amount
Healthy Leaves	626	290	916
Tipburn	805	261	1056
Leaf spot	552	373	925
Total	1983	924	2897

B. *Split Dataset*

The strawberry leaf dataset will then be placed in the categorized dataset folder, which includes Dataset Training and Dataset Validation. The data input process is divided into two parts: 80% of the dataset is input to the training folder and 20% to the validation folder. This data sharing helps to avoid model performance measurement errors.

C. *Preprocessing*

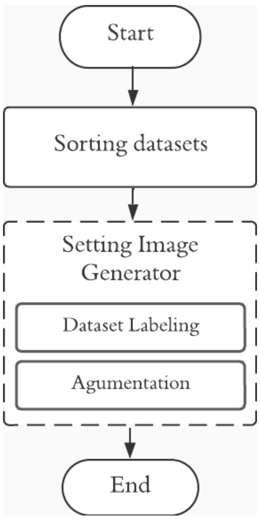


Figure 2. Preprocessing Flowchart

To reduce overfitting in classification training, the training dataset will be augmented. The model is supposed to identify the basic variation in the object image of strawberry leaves. Rotation, zoom, rescale, horizontal reflection, width and height shift, and shear are the transformations used in dataset augmentation. The augmented training data and validated data results are saved in the data storage. The images below show each type of augmentation performed on a leaf with tipburn symptoms.

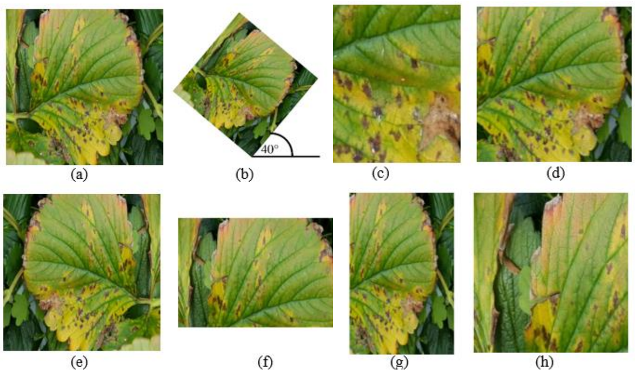


Figure 3. Strawberry Leaf With Criteria; (a) Original, (b) Rotation, (c) Zoom, (d) Rescale, (e) Horizontal Flip, (f) Width Shift, (g) Height Shift, (h) Shear.

D. *Data Classification*

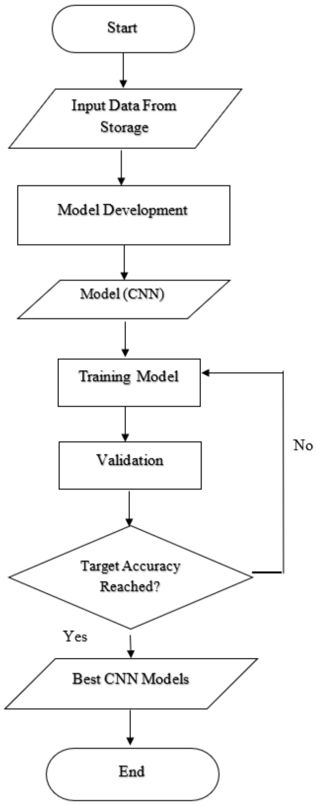


Figure 4. Dataset Classification Process Flowchart

The training dataset will be supplied to reduce overfitting in classification training. The model is supposed to recognize the fundamental variation in the object image of strawberry leaves. Transformations used in dataset augmentation include rotation, zoom, rescale, horizontal reflection, width and height shift, and shear. The validated data findings and additional training data are maintained in the data storage. The photos below depict each augmentation approach being used on a leaf with tipburn symptoms.

In the model-building process, the network has structural features that determine the model's quality. The CNN model utilized is VGG16, which has a convolutional layer with a 3x3 (small size) convolutional filter specification. By raising the size of the convolutional filter, the inclusion of a convolutional layer can enhance the depth of the neural network [35].

The architecture of the CNN model is separated into two key sections: feature extraction and classification. The feature extraction section alters the incoming data so that it may be classified appropriately. Data is transmitted from both linear and non-linear adjustments. The classification section's function is to reduce data such that it may be classified by the softmax classifier.

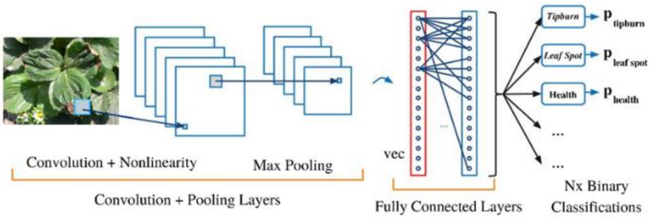


Figure 5. Architecture Layer Of CNN

The layer conducts convolution, max pooling, and dropout functions in feature extraction, aiming to simplify network architecture preparation. At the classification stage, there is a mix of one layer of linear transformation and softmax layers to minimize the number of feature maps.

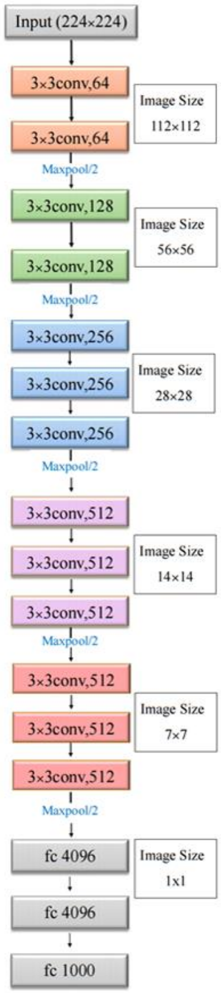


Figure 6. Vgg16 Network Architecture

The process of fitting the CNN model to the Strawberry dataset in order to achieve the optimal model. The procedure is broken down into three stages.

1. Training is the adaption of the model to existing data so that it can 'learn' the data and predict it for comparable data. The training model is also subjected to a hyperparameter tweaking procedure in order to improve its accuracy. The hyperparameter details and values are shown in Table 2.

Table 2. Hyperparameter Model Training

Hyperparameter	Value
Epochs	100
Batch Size	126
Dropout Value	0.5
oss	rmsprop
Optimizer	categorical_crossentropy
Activation Function for Conv. Layer	ReLu
Activation Function for Last Dense Layer	Softmax

2. Testing is a procedure that influences the model's performance measurement determination.

3. Regularization is a procedure performed to a model to avoid overfitting or instances in which the model fits too well with the training data, resulting in suboptimal prediction performance and pattern generalization.

III. Result and Discussion

A. *Research Parameter*

At this stage, the process of detecting Starberry plant objects uses Deep Learning which will be implemented into the system. An object system built using the Python programming language. The classification of plant images is divided into 3 categories, namely "Tipburn", "Leaf Spot" and normal Strawberry leaves.

Figure 7. Strawberry Dataset Sorting Results based on Folders in each Class

B. *Convolutional Neural Network (CNN) Model Result*

Using filter 148-72-34-15, three convolution layers (conv2d, conv2d 1, conv2d 2, and conv2d 3) were used during the creation of the CNN model. Then, there is increased layer dropout to reduce overfitting and some denser layers to classify. There is also maximum unification to help with overfitting by providing data-driven insights into internal representations. The entire data set can be seen in the image below. The total number of parameters is 3,473,475.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 148, 148, 64)      1792

 max_pooling2d (MaxPooling2D  (None, 74, 74, 64)       0
 )

 dropout (Dropout)           (None, 74, 74, 64)        0

 conv2d_1 (Conv2D)           (None, 72, 72, 64)        36928

 max_pooling2d_1 (MaxPooling  (None, 36, 36, 64)       0
 2D)

 dropout_1 (Dropout)         (None, 36, 36, 64)        0

 conv2d_2 (Conv2D)           (None, 34, 34, 128)       73856

 max_pooling2d_2 (MaxPooling  (None, 17, 17, 128)      0
 2D)

 dropout_2 (Dropout)         (None, 17, 17, 128)       0

 conv2d_3 (Conv2D)           (None, 15, 15, 128)       147584

 max_pooling2d_3 (MaxPooling  (None, 7, 7, 128)        0
 2D)

 dropout_3 (Dropout)         (None, 7, 7, 128)         0

 flatten (Flatten)           (None, 6272)              0

 dropout_4 (Dropout)         (None, 6272)              0

 dense (Dense)               (None, 512)               3211776

 dense_1 (Dense)             (None, 3)                 1539

=================================================================
Total params: 3,473,475
Trainable params: 3,473,475
Non-trainable params: 0
_________________________________________________________________
```

Figure 8. Deep Learning Model

C. Fitting Model

The learning rate in a model is used to calculate the corrected value of the weight when training data is carried out. In script code, learning rate is usually written as lr and followed by a value called a constant. In other words, this learning rate shows how many times k changes. These constants will be updated each time the batch is run. In the analysis of learning rate variations, it aims to determine the accuracy and loss values obtained after training, and data validation. In this learning rate variation, a learning rate of 0.01 is used. Learning rate, is one of the training parameters to calculate the weight correction value during the training process. This learning rate value is in the range of zero (0) to (1). The greater the learning rate, the faster the training process will run.

The results of the accuracy in training and validation will be a comparison and prove that the learning rate affects the accuracy value of a model. The following is an experiment of learning rate variations carried out on the image data of strawberry plant leaves to see the resulting accuracy and loss as well as the graphic pattern.

The following is an experiment using a learning rate of 0.001 with an epoch of 100 to produce a good fitting model by looking at its accuracy and loss function.

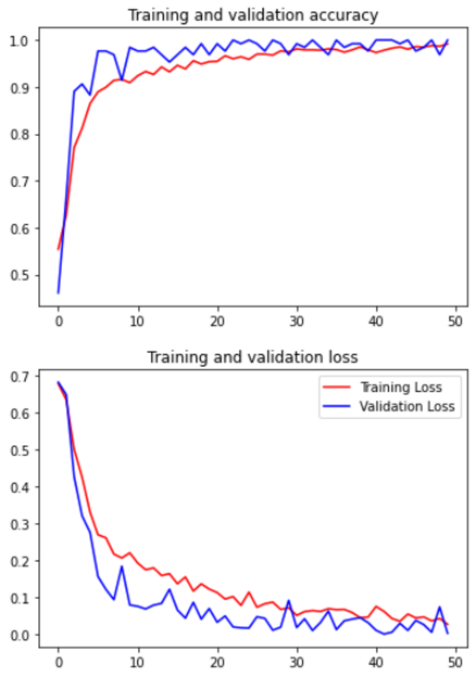


Figure 9. accuracy and validation, loss and loss validation at a learning rate of 0.001

Figure 5.3 shows the results of model fitting using a learning rate of 0.001 and 100 epochs showing high accuracy and low loss. training accuracy is 97% and validation accuracy is 96.99%. Thus, the loss value for training is 8.54%, but the loss value for validation is only 0.26%. The accuracy results meet the accuracy target of >90% for training and validation parameters.

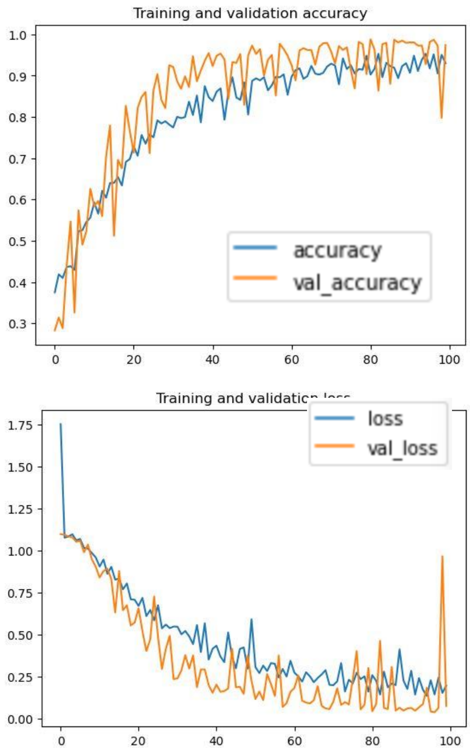


Figure 10. accuracy and validation, loss and loss validation at a learning rate of 0.01

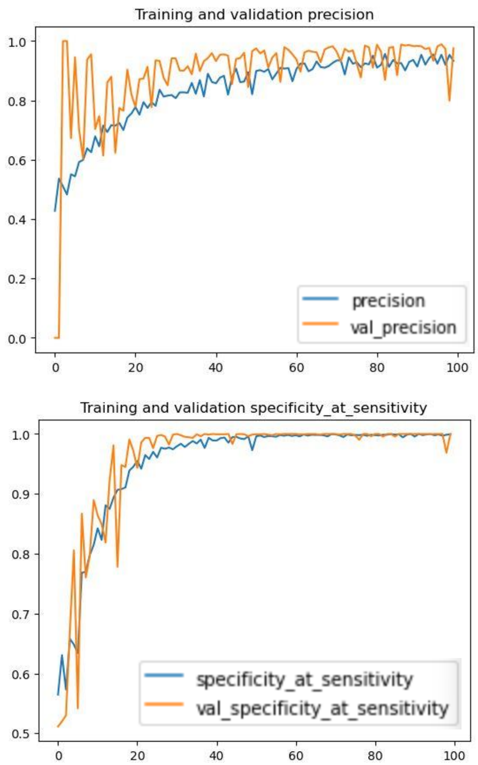


Figure 11. accuracy and validation, loss and loss validation at a learning rate of 0.01

Figure 5.4-5.5 shows the second experiment using a learning rate of 0.01 with an epoch of 100 to produce a model that is appropriate fitting by looking at its accuracy and loss function. The results of a learning rate of 1.01 show high accuracy and low loss. graph showing the accuracy and loss of the model developed to identify healthy, tipburn, or leafspot on strawberry plants during training and validation. Based on Figure 5.4, the training accuracy is 95.05% and the validation accuracy is 97.4%. Thus, the loss value for training is 19.68%, but the loss value for validation is only 7.54%. The accuracy results meet the >90% accuracy target for the training and validation parameters but the model training is slightly unstable.

Training and validation graphs for (a) precision and (b) sensitivity are also shown in Figure 5.5 The accuracy values for training and validation are 93.33% and 97.6%, respectively, as are the sensitivity values of 99.95% and 96.86 %, as shown in Figure 4.6. This demonstrates excellent precision and sensitivity of the training and job validation techniques. but the model training is a bit unstable.

```
Epoch 100/100
15/15 [=====================]
- 582s 39s/step
- train_accuracy     : 0.9505
- val_accuracy       : 0.9740
- train_loss         : 0.1968
- val_loss           : 0.0754
- train_precision    : 0.9333
- val_precision      : 0.9760
- train sensitivity  : 0.9995
- val_sensitivity    : 0.9686
```

Figure 11. Fitting Model Result Summary

CONCLUSION

Optimization of the Deep Learning model for leaf image classification for tipburn, leaf spot, and healthy conditions using the CNN VGG16 architecture was demonstrated to identify and categorize infection in strawberry plants. To improve the performance of the CNN model, the dataset augmentation approach and changing the hyperparameter parameters, especially the Learning Rate parameter, are used in this study. The augmentation and classification procedures included 2,897 photos of strawberry leaves. In the first experiment, the Learning Rate used was 0.001 with an accuracy of 97% and a validation accuracy of 96.99% and a loss value of 0.0054 resulting in a fairly stable and good fitting model. In the second experiment with a Learning Rate of 0.01 it resulted in a training accuracy of 95.05% and a validation accuracy of 97.4%, and the resulting training loss was 19.68% and the validation loss was only 8.54%, %. The accuracy results met the >90% accuracy target for the training and validation parameters but in this second experiment the fitting results were not very stable.

REFERENCES

[1] M. Iqbal, "Menilik Nasib Produksi Stroberi Bandung," detikNews, Bandung, 2021.

[2] M. Kroggel and C. Kubota, "Controlled environment strategies for tipburn management in greenhouse stroberi production," Acta Horticulturae, no. 1156, pp. 529-536, 2017.

[3] K. Kuronuma, M. Saotome, M. Ando and H. Watanabe, "Excessive Calcium Accumulation in the Roots is a Key Factor in Tipburn Incidence under High Ca Supply in Lisianthus (Eustoma grandiflorum) Cultivars," Agronomy 2020, vol. 1, no. 10, pp. 1-10, 2020.

[4] J. S. Baggio, J. C. Mertely and N. A. Peres, "Leaf Spot Diseases of Stroberi," Plant Pathology, pp. 1-6, 2020.

[5] R. Angel, V. S. Hilda, F. Dionico, A. Uriel, M. Merari and E. P. Brenda, "An Emerging Stroberi Fungal Disease Associated with Root Rot, Crown Rot, and Leaf Spot Caused by Neopestalotiopsis in Mexico," Plant Disease, vol. 104, no. 8, pp. 2054-2059, 15 April 2020.

[6] C. Kubota and M. Kroggel, "Soilles stroberi production in semiarid climate: Improving fruit quality by nighttime VPD control," in Poster presented at 2012 Annual Conference, Miami, 2012.

[7] J. Shin, Y. K. Chang, B. Hueng, T. Nguyen-Quang, G. W. Price and A.-M. Ahmad, "A deep learning approac for RGB image-based powdery mildew disease detection on stroberi leaves," Computers and Electronics in Agriculture, no. 183, pp. 1-8, 2021.

[8] J. Xiao, P. Chung, H. Wu, Q. Phan, J. A. Yeh and M. T. Hou, "Detection of Stroberi Diseases Using a Convolutional Neural Network," Plants, vol. 10, no. 31, pp. 1-14, 2021.

[9] R. Jayawardena, J. Huang, B. Jin, J. Yan, X. Li, K. Hyde, A. Bahkali, S. Yin and G. Zhang, "An account of Colletotrichum species associated with stroberi anthracnose in Chine based on morphology and molecular data," Mycosphere, vol. 7, no. 8, pp. 1147-1163, 2016.

[10] S. Lee, C. Chan, P. Wilkin and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in IEEE International Conference on Image Processing (ICIP) 27-30 September, Quebec City, 2015.

[11] Balai Penelitian Tanaman Jeruk dan Buah Subtropika - Kementerian Pertanian, "Teknologi Budidaya Stroberi di

Lahan," Balitjestro, 18 January 2016. [Online]. Available: http://balitjestro.litbang.pertanian.go.id/teknologi-budidaya-stroberi-di-lahan/. [Accessed 13 November 2022].

[12] Lahan, "Potensi Perkebunan Di Indonesia," lahan.co.id, 18 September 2022. [Online]. Available: https://lahan.co.id/perkebunan-di-indonesia/. [Accessed 13 November 2022].

[13] D. O. Oktarina, Armaini and Ardian, "Growth and Production of Strawberries (Fragaria Sp) with the Application of Various Concentration of Liquid Organic Fertilizer (LOF) in Subtrates Hydroponics," JOM FAPERTA UR, vol. 4, no. 1, pp. 1-12, 2017.

[14] M. Hariri and E. Avsar, "Tipburn disorder detection in stroberi leaves using convolutional neural networks and particle swarm optimazation," Multimedia Tools and Applications, vol. 81, no. 8, pp. 11795-11822, 2022.

[15] B. Karki, H. Sitaula, S. Bhandari, G. Pramod and H. Manandhar, "Effect of Different Botanical Extracts and Organic Compound in the Management of Leaf Spot Disease of Stroberi (Fragaria ananassa Duch.) Caused by Pestalotia longisetula under Field Condition," International Journal of Applied Sciences and Biotechnology (IJASBT), vol. 9, no. 3, pp. 193-202, 28 July 2021.

[16] U. Afzaal, B. Bhattarai, Y. Padeya and J. Lee, "An Instance Segmentation Model for Stroberi Diseases Based on Mask R-CNN," Sensors, vol. 21, no. 6565, 2021.

[17] A. Pratiwi, "Ketahui, ini Tanda Tanaman Sehat," Kompas, Jakarta, 2021.

[18] N. Rahmadia, "Kaggle, Komunitas Belajar Data Science yang Bisa Jadi Sumber Uang," Glints, 14 January 2021. [Online]. Available: https://glints.com/id/lowongan/kaggle-adalah/#.Y3F7r8dBw2x. [Accessed 12 November 2022].

[19] The Dryad Team, "New Dryad is Here," Dryad, 24 September 2019. [Online]. Available: https://blog.datadryad.org/2019/09/24/new-dryad-is-here/. [Accessed 12 November 2022].

[20] D.-P. Adrien, S. Dennis, and V. Stavros, "Dryad," 08 February 2018. [Online]. Available: https://doi.org/10.25338/B8V308. [Accessed 05 November 2022].

[21] F. A. Sianturi, "Kompresi File Citra Digital dengan Arithmetic Coding," Jurnal Teknik Informatika Unika St. Thomas (JTIUST), vol. 3, no. 1, pp. 45-51, 3 June 2018.

[22] H. Fahmi, M. Zarlis, H. Mawengkang, and N. Zendrato, "The Using of Thresholding and Region Merging Algorithm for Correcting the Multiple Choice Answer Sheets," Journal of Physics: Conference Series, vol. 1255, no. 1, pp. 1-6, 17 July 2019.

[23] N. Nafi'iyah, "Algoritma Kohonen dalam Mengubah Citra Graylevel Menjadi Citra Biner," Jurnal Ilmiah Teknologi dan Informasi ASIA (JITIKA), vol. 9, no. 2, pp. 49-55, 2015.

[24] A. F. Hastawan, R. Septiana and Y. E. Windarto, "Perbaikan Hasil Segmentasi HSV pada Citra Digital Menggunakan Metode Segmentasi RGB Grayscale," Edu Komputika Journal, vol. 6, no. 1, pp. 32-37, 26 July 2019.

[25] B. M. Sari, "Identifikasi Tingkat Kematangan Buah Stroberi Berdasarkan Warna RGB dengan Menggunakan Metode Regionprops," Terapan Informatika Nusantara, vol. 1, no. 5, pp. 225-230, 2020.

[26] Trivusi, "Pengertian dan Cara Kerja Algoritma Convolutional Neural Network (CNN)," Trivusi, 28 July 2022. [Online]. Available: https://www.trivusi.web.id/2022/04/algoritma-cnn.html. [Accessed 13 November 2022].

[27] C. Bai, L. Huang, X. Pan, J. Zheng, and S. Chen, "Optimization of deep convolutional neural network for large scale image retrieval," Neurocomputing, vol. 303, pp. 60-67, 2018.

[28] dshadid380, "Learn Convolutional Neural Network from basic and its implementation in Keras," Towards Data Science, 25 February 2019. [Online]. Available: https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529. [Accessed 11 November 2022].

[29] D. Mwiti, "Image Classification with Convolutional Neural Networks (CNNs)," KD Nuggets, 4 May 2022. [Online]. Available: https://www.kdnuggets.com/2022/05/image-classification-convolutional-neural-networks-cnns.html. [Accessed 14 November 2022].

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, pp. 1-9, 2012.

[31] M. Mandal, "Introduction to Convolutional Neural Networks (CNN)," Analytics Vidhya,

[32] 1 May 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/. [Accessed 14 November 2022].

[33] R. Dharmadi, "Mengenal Convolutional Layer dan Pooling Layer," Medium, 4 April 2018. [Online]. Available: https://medium.com/nodeflux/mengenal-convolutional-layer-dan-pooling-layer-3c6f5c393ab2. [Accessed 14 November 2022].

[34] A. Kholik, A. Harjoko and Wahyono, "Classification of Traffic Vehicle Density Using Deep Learning," Indonesian Journal of Computing and Cybernetics Systems (IJCCS), vol. 14, no. 1, pp. 69-80, 3 March 2020.

[35] A. Kholik, "KLASIFIKASI MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA TANGKAPAN LAYAR HALAMAN INSTAGRAM," JDMSI, vol. 2, no. 2, pp. 10-20, 24 May 2021.

[36] L. Chen and G. Papandreou, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE, pp. 1-14, 2017.

[37] H. Wu and X. Gu, "Towards dropout training for convolutional neural networks," Neural Networks, vol. 71, pp. 1-10, 2015.

[38] M. Tang, A. Djelouah, F. Peparazzi, Y. Boykov and C. Schroers, "Normalized Cut Loss for Weakly-Supervised CNN Segmentation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City - 2 June, 2018.

[39] W. S. Ahmed and A. A. A. Karim, "The Impact of Filter Size and Number of Filters on Classification Accuracy in CNN," in International Conference on Computer Science and Software Engineering (CSASE), Duhok - Iraq, 2022.

[40] Y. Xu and R. Goodacre, "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning," Journal of Analysis and Testing. Springer Science and Business Media LLC, vol. 2, no. 3, pp. 249-262, 2018.

[41] C. Luo, X. Li, L. Wang, J. He, D. Li, and J. Zhou, "How Does the Data set Affect CNN-based Image Classification Performance?," The 2018 5th International Conference on Systems and Informatics (ICSAI 2018), vol. 18, pp. 361-366, 2018.