International Journal of Plant & Soil Science

Volume 36, Issue 5, Page 688-695, 2024; Article no.IJPSS.115603
ISSN: 2320-7035

Detection of Diseases in Blackgram (*Vigna mungo* L.) Using Machine Learning Models: A Case Study

Venketesa Palanichamy N [a++], Kalpana M [a#*] and Karthiba L [b†]

[a] *Agricultural College and Research Institute, TNAU, Coimbatore, India.*
[b] *Department of Pulses, TNAU, Coimbatore, India.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/IJPSS/2024/v36i54566

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/115603

Original Research Article

Received: 01/02/2024
Accepted: 03/04/2024
Published: 05/04/2024

ABSTRACT

Black gram (*Vigna mungo* L.) is widely used in Indian cuisine and one of the most significant pulses cultivated in India. Identification of plant diseases at their earlier stages is essential to take necessary plant protection measures to reduce yield loss to the farmers. Anthracnose and Powdery Mildew are the major diseases in black gram which causes significant yield losses to the farmers. In this research study, advanced disease detection machine learning models such as Multinomial Logistic Regression, Random Forest Classifier were employed to assist the farmers in detection of plant leaf diseases in blackgram at their early stages of growth. For this present study, Image data sets were collected from Thanjavur block, Thanjavur district, Tamil Nadu. Results of the study showed that accuracy of Random Forest Classifier was higher with train accuracy 99.17% and test accuracy 97.00% when compared to the other machine learning methods for detection of plant leaf diseases in black gram, which aids in promotion of smart agriculture.

[++] *Dean(Agri.);*
[#] *Professor;*
[†] *Assistant Professor;*
**Corresponding author: E-mail: kalpusiva@gmail.com;*

Keywords: Blackgram; machine learning; multinomial logistic regression; random forest classifier; accuracy; smart agriculture.

1. INTRODUCTION

Pulses are one of the major food crops in the world, because of their higher protein content. The world's largest producer of pulses is India. As an essential component of the Indian diet, pulses add much-needed protein to the carbohydrate – rich diet. Blackgram, one of the most profoundly cherished pulses of India and generally utilized in Indian cuisine. In India the black gram is cultivated in both Kharif and Rabi seasons. India producing over 1.5 million tonnes of black gram annually. The yield of the blackgram is reduced significantly because of the most common diseases of black gram such are Anthracnose and Powdery Mildew. The early and timely identification and classification of these diseases are necessary for higher yield. Diseases in Blackgram break out and spread quickly, which reduce production greatly and thus do harm to agricultural economy. The advancement of non-destructive detection and early crop disease identification is crucial to the development of precision and ecological agriculture. Plant diseases cause substantial losses to the farmers. The use of technologies like machine learning and computer vision helps the farmers in detection of plant diseases [1,2,3]. Automatic plant disease detection using machine learning models are blooming in recent years. Convolution Neural Network methods are used to detect the Muti-crops leaf diseases in grapes and tomato. Extraction is used to classify the healthy and unhealthy leaves. CNN based Visual Geometry Group model is used to improve the performance of the model. The accuracy of model was 98.40 per cent for grapes and 95.71 per cent for tomatoes [4,5]. The traditional method of plant disease detection is time consuming and requires skilled labour. Smart and efficient crop disease detection computer vision and machine learning technique are used in five common plants for twenty different diseases with 93 per cent accuracy [6]. Automatic disease detection helps the farmers to diagnose the plant disease accurate and faster. The techniques are used are random forest, ID3, SVM and RBF-SVM [7,8]. Deep learning models helps in automatic identification of crop diseases, which helps the farmers to manage the crops and to get higher yield. Deep convolutional neural networks split the images into training and testing for early disease identification in plants [9]. Convolutional neural network model were used for diagnosis of plant diseases. The pre-trained models are VGG-16, Inception V4, DenseNet-121 and ResNet-50. The image dataset of 54,305 images of different plant disease in 38 cases are taken. The model performance was evaluated with accuracy, specificity, sensitivity and F1 score. DesnseNet-121 is the effective model with higher accuracy compared to other models [10]. Image processing were used with different algorithms to classify the diseases in tomato. KNN algorithm is used for image classification in tomato. The classification based on color, texture and bound were employed to identify the diseases [11]. Rice is the major crop cultivated in India and affected by leaf diseases which affects the crop yield. The major rice diseases are Blast, Brown spot which occurs frequently in plant leaves. The various machine learning and deep learning models are used for detection of leaf diseases. The metrics such as accuracy, precision and recall are used to measure the performance of the model. The deep learning, five layer convolution model has the accuracy of 78.2 per cent and VGG16 has 58.4 per cent accuracy [12]. Lack of awareness among farmers in diagnosis of plant diseases using scientific approaches at their early stages results in yield losses [13]. There are several advanced technologies in disease diagnosis which reduces yield loss in agriculture [14]. CNN and an AlexNet classifier were used for identification of diseases in rice and accuracy was taken as metric to measure the performance. The performance of the model is more than 90 per cent [15]. The classification approach such as ANN, SVM and KNN methods are used to classify the leaf disease images [16]. Plant leaves such as lemon (sun burn disorder), rose/beans (bacterial disorder), beans (fungal) and banana (early scorch) images are captured using digital camera. The genetic algorithms were used to extract the segmented images. The classifier tool such as SVM is used for classification. The accuracy is 97.6 per cent [17]. Leaf diseases were segmented using k-means clustering for classification of disease with SVM. Measures of Central Tendency such as mean, median, mode and standard deviation were used to record the findings [18]. A model was tested with leaf database and the performance of the model was accurate [19].

The traditional methods of identification of plant leaf diseases are based the experiences accumulated by the farmers over the years. Inaccurate identification of plant diseases may lead to severe yield loss. The traditional methods of plant diseases identification are time consuming and laborious task. The availability of skilled persons to identify the plant diseases is low and to overcome these issues, an accurate, intelligent and less time consuming machine learning methods were used to identify plant leaf diseases in black gram [20].

To overcome the challenges discussed above the Machine learning models provide better accuracy to detect the plant diseases at their earlier stages of growth. The present study was carried out with the objective of employing machine learning models in identification of plant leaf diseases in Black gram (*Vigna mungo* L.).

2. MATERIALS AND METHODS

The present study was carried out at Thanjavur district of Tamil Nadu. For this present study, image datasets of blackgram healthy leaves, leaves affected with Anthracnose and Powdery mildew were collected from various fields in Thanjavur block of Thanjavur district, Tamil Nadu. In this study, Machine learning models such as Multinomial Logistic Regression and Random Forest Classifier were employed to detect plant leaf diseases in blackgram. Sampling methods and description of machine learning models employed were discussed here under.

2.1 Sampling

In this study, image dataset (Fig. 1) of black gram were collected from in Thanjavur block, Thanjavur district, Tamil Nadu. They are

- Blackgram healthy leaf
- Blackgram leaves with Anthracnose(*Colletotrichum lindemuthianum*)
- Blackgram leaves with Powdery mildew (*Erysiphe polygoni)*

Machine learning models were employed in Blackgram with total of 3000 numbers of images, including 1000 healthy leaves, 1000 leaves affected with anthracnose and another 1000 leaves affected with powdery mildew were collected from various fields in Thanjavur block were captured by a camera (SONY Alpha ILCE-6100Y APS-C), with image size: 2992 × 2992 pixels and image DPI: 72 pixels/inch.

The dataset was split into training (80 per cent) and test set (20 per cent) to generate machine learning models. The test image set was used to compute the performance of the training model with accuracy as metrics. The entire analysis is carried out using Python 3.8.1.

These selected images of black gram leaves were split into train and test datasets and coded for healthy and unhealthy infected blackgram leaves. Then, the machine learning models such Multinomial Logistic Regression and Random forest classifier were employed to identify and classify the anthracnose, powdery mildew and healthy leaves in blackgram. The images collected were enhanced by improving the picture quality without any loss of information. Blackgram images were converted into 64 x 64 with same dimension sat pixel levels to classify the images. The sequential steps in detection of diseases in blackgram from the image dataset were illustrated in Fig. 2.

Healthy Blackgram leaf **Anthracnose affected leaf** **Powdery mildew affected leaves**

Fig. 1. Healthy Vs disease affected leaf in black gram

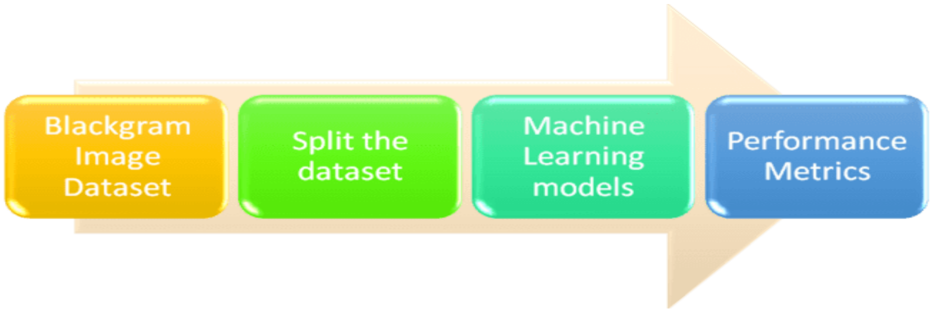

Fig. 2. Architecture of identification leaf diseases in Black gram

List 1. Algorithm for leaf disease identification in Black gram of using Machine learning models

Step 1	Collection of dataset for blackgram leaf disease
Step 2:	Images dataset are split into training (80%) and test(20%)
Step 3:	Machine learning models such as Multinomial Logistic Regression and Random forest classifier were used to create the model for leaf classification
Step 4:	Acquire the Performance metrics for machine learning model.

2.2 Machine Learning Models Employed

In this present study, Machine learning model such as Multinomial Logistic Regression and Random Forest Classifier were employed to detect diseases in black gram. These models were discussed hereunder.

2.3 Multinomial Logistic Regression

Multinomial logistic regression is used to classify the images for more than two target class. The inputs for multinomial logistic regression are the features from the image dataset. The features of the model consist of healthy, anthracnose and powdery mildew disease infected leaves. These features are treated as the inputs for the multinomial logistic regression.

The linear model equation is the same that of linear equation in the image.

Where the X is the set of inputs (healthy leaves, anthracnose and Powdery mildew infected leaves), In the image X is a matrix which has all the feature (numerical values) X = [x1,x2,x3]. W is a matrix for same input with weights W = [w1,w2,w3].

The linear model output = w1*x1, w2*x2, w3*x3. The weights w1, w2, w3, w4 are update in the training phase.

Logits also called as scores, the outputs of the linear model. The Logits change because of the changes in the calculated weights. Softmax function is a probabilistic function which helps to calculate the probabilities for the score obtained. Softmax function gives the high probability for highest score and fewer probabilities for the left-over score. The probabilities ranges from 0 to 1and sum of all the probabilities is equals to 1. The input is the anthracnose image, the target having 3 possible outcomes like healthy, anthracnose and Powdery mildew.

2.4 Random Forest Classifier

A Random Forest Classifier with advanced V3 inception model is used to classify the images of Blackgram. In the research paper, Inception V3 was used for pre-training the models. Inception V3 was an additional design for CNN developed by Google. Inception starts with a sparse structure, increases the network depth and width, and clusters the spare data into a dense structure to enhance the model accuracy [21]. Transfer the learning of plant pathology data to the Inception V3 model pre-trained on ImageNet data to speed up the training process and to improve the model performance; the Inception V3 model has 94 convolution layers, 14 pooling, and dense layers [22].

The Random Forest Classifier uses ensemble learning, which provide solutions to Blackgram images. Many decision trees are generated in Random Forest Classifier. Random Forest Classifier is outcome from the predictions of the decision trees. The final predictions were based on the mean or average of the output from the various trees. The training images dataset contains images such as healthy, anthracnose and Powdery mildew. Image dataset split into subsets for Random Forest Classifier. The Subset are used by every decision tree in the random forest model. Each and every decision tree produces specific output. For example, the prediction for 1 is healthy leaves, 2 is infected with anthracnose and 3 is infected with powdery mildew in blackgram disease identification. The random forest classifier collects the major vote, which provides the final prediction of blackgram leaf diseases as shown in Fig. 3.

Fit the Random forest model to the training set of blackgram images. To fit the model, import the Random Forest Classifier class from the sklearn. Ensemble library with n_estimators= About 50 number of trees are taken in the Random Forest. Model is fitted to the training set for blackgram images, as to predict the test result. For prediction, y_prediction is the new prediction variable. The prediction variable are checked and test set real variable, which helps to determine the incorrect predictions for blackgram leaf disease images done by the classifier.

2.5 Confusion Matrix

The confusion matrix helps to determine the incorrect and correct and predictions as shown in Fig. 4.

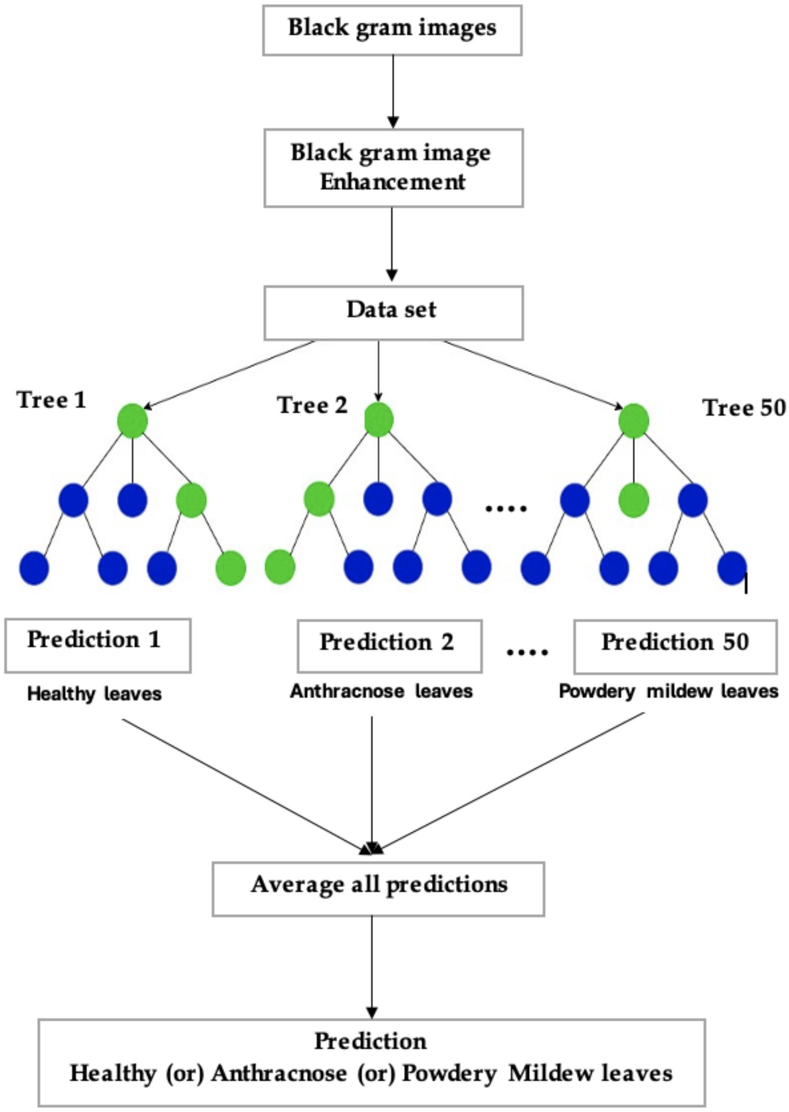

Fig. 3. Random forest classifier for blackgram images

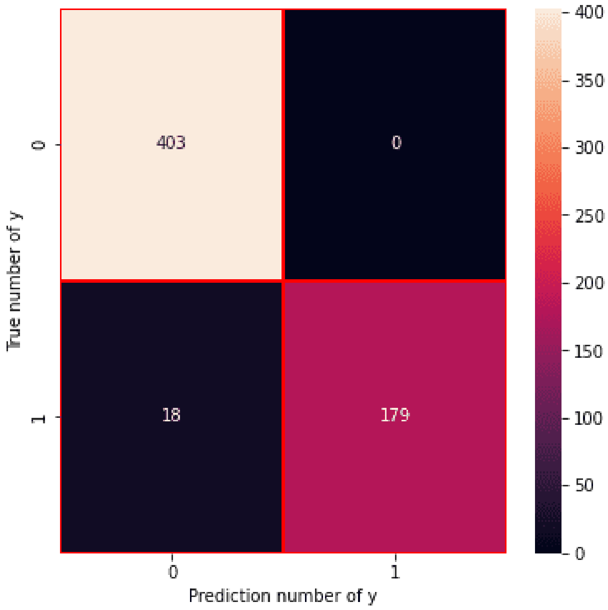

Fig. 4. Confusion matrix for blackgram leaf disease

From Fig. 4 we conclude
True Positive (TP) = 403
True Negative (TN) = 179
False Positive (FP) = 0
False Negative (FN) = 18

3. RESULTS AND DISCUSSION

Results of the study was discussed hereunder. Machine learning model was evaluated using accuracy as metrics. Accuracy is the fraction of total blackgram images that was correctly classified by the model

Accuracy = (TP+TN)/(TP+TN+FP+FN)

True Positive (TP) = number of predictions where the model correctly predicts the positive class as positive.

True Negative (TN) = number of predictions where the model correctly predicts the negative class as negative.

False Positive (FP) = number of predictions where the model incorrectly predicts the negative class as positive.

False Negative (FN) = the number of predictions where the model incorrectly predicts the positive class as negative.

With the training image dataset machine learning model is generated. The test image dataset is used to know the performance of the model using accuracy as metrics. The entire study was analysed using Python 3.8.1. The Table 1 showed the Evaluation metrics for blackgram leaf images.

For this present study, 80 per cent of the black gram images were used to train the data and 20 per cent of the leaf images were used for testing the data. The train and test accuracy was efficient for all the Random Forest Classifier for identification and classification of healthy and unhealthy leaves in black gram.

Results of the study showed that Random Forest Classifier was the best model to diagnosis the healthy leaves, and leaves affected with anthracnose and Powdery mildew when compared to Multinomial Logistic regression.

Table 1. Evaluation metrics for blackgram leaf disease

Methods	Train Accuracy (%)	Test Accuracy (%)
Multinomial Logistic regression	75.5	72.17
Random Forest Classifier	99.17	97.00

Table 2. Performance of the proposed model with other existing models employed

S. No.	Author (s)	Method	Accuracy (%)
1.	Agarwal et al. [23]	CNN network	91.20
2.	Widiyanto et al. [24]	CNN model	96.60
3.	Proposed model	Random Forest Classifier	99.17

4. CONCLUSION

Machine learning models deployed for identification and classification of blackgram leaf images of both healthy, anthracnose and Powdery mildew leaf images. Random Forest Classifier ranked first based on test accuracy followed by Multinomial Logistic regression. The test accuracy would be improved in future by including more images for the subsequent analysis. Use of machine learning model can benefit the farmers by minimizing the yield loss. This study will provide a viable solution to the blackgram farmers, researchers, and extension officer of agriculture to identify the disease accurately, thereby taking proper measure to improve crop production in upcoming years. By identifying diseases at earlier stages in blackgram, farmers can reduce yield loss by following timely management practices for the diseases.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Vu DL, Nguyen TK, Nguyen TV, Nguyen TN, Massacci F, Phung PH. HIT4Mal: Hybrid image transformation for malware classification, Trans. Emerging Telecommun. Technol. 2020;31(11): e3789.
2. Shaikh RP, Dhole SA. citrus leaf unhealthy region detection by using image processing technique, in: IEEE International Conference on Electronics, Communication and Aerospace Technology. 2017;420–423.
3. Yu K, Lin L, Alazab M, Tan L, Gu B. Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system, IEEE Trans. Intell. Transp. Syst. 2020;22(7):4337–4347.
4. Ananda S Paymode, Vandana B Malode. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG, Artificial Intelligence in Agriculture. 2022; 6:23-33
5. Orchi H, Sadik M, Khaldoun M, Sabir E. Automation of crop disease detection through conventional machine learning and deep transfer learning approaches. Agriculture. 2023;13:1-35.
6. Kulkarni Pranesh, Karwande Atharva, Kolhe Tejas, Kamble Soham, Joshi Akshay, Wyawahare Medha. Plant disease detection using image processing and machine learning; 2021.
7. Abu Sarwar Zamani, Anand L, Kantilal Pitambar Rane, Prabhu P, Ahmed Mateen Buttar, Harikumar Pallathadka, Abhishek Raghuvanshi, Betty Nokobi Dugbakie. Performance of machine learning and image processing in plant leaf disease detection. Journal of Food Quality. 2022; 2022:1-7.
Available:https://doi.org/10.1155/2022/1598796
8. Vergni L, Todisco F. A Random forest machine learning approach for the identification and quantification of erosive events. Water. 2023;15(12):2225.
Available:https://doi.org/10.3390/w15122225
9. Tugrul B, Elfatimi E, Eryigit R. Convolutional neural networks in detection of plant leaf diseases: A review. Agriculture. 2022;12(8):1192.
Available:https://doi.org/10.3390/agriculture12081192
10. JA, Eunice J, Popescu DE, Chowdary MK, Hemanth J. Deep learning-based leaf disease detection in crops using images for agricultural applications. Agronomy. 2022;12(10):2395.
Available:https://doi.org/10.3390/agronomy12102395
11. Geetha G, Samundeswari S Saranya, Meenakshi GK, Nithya M. Plant leaf disease classification and detection system using machine learning. Journal of Physics: Conference Series. 2022;1712: 012012.
12. Pallapothala Tejaswini, Priyanshi Singh, Monica Ramchandani, Yogesh Kumar

Rathore, Rekh Ram Janghel. Rice leaf disease classification using CNN, IOP Conf. Series: Earth and Environmental Science. 2022;1032:012017. DOI:10.1088/1755-1315/1032/1/012017

13. Malo M, Hora J. Pulses production in India: Major constraints and way forward. In: Research trends in multidisciplinary research and development. S. Fedorov (Eds.). Weser Books, Germany. pp.35-63.(2020) on Computer Communication and Informatics. 2017;12–16.
14. Priya Rani B, Farheen N, Magda R. Artificial intelligence solutions enabling sustainable agriculture: A bibliometric analysis. PLoS ONE. 2022;17(6):1-19.
15. Krizhevsky A, Sutskever I, Hinton GE. Image net classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe. 2012;60(6):1097-1105.
16. Chaitali G Dhaware, Wanjale KH. A modern approach for plant leaf disease classification which depends on leaf image processing, in: IEEE International Conference; 2017.
17. Vijai Singh, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques, Inf. Process. Agricult. 2017;4(1):41–49.
18. Nema S, Dixit A. Wheat Leaf detection and prevention using support vector machine, in: International Conference on Circuits and Systems in Digital Enterprise Technology. 2018;1–5.
19. Vidyashree Kanabur, Sunil S Harakannanavar, Veena I Puranikmath, Dattaprasad Torse. Detection of leaf disease using hybrid feature extraction techniques and CNN classifier, Springer Comput. 2019;1213–1220.
20. Kalpana M, Karthiba L, Senguttuvan K, Parimalarangan R. Diagnosis of major foliar diseases in black gram (Vigna mungo L.) using Convolution Neural Network (CNN). Legume Research; 2023. DOI:10.18805/LR-5083
21. Al Husaini MAS, Habaebi MH, Gunawan TS, Islam MR, Elsheikh EAA, Suliman FM. Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4. Neural Computing and Applications. 2022;34:333–348. Available:https://doi.org/10.1007/s00521-021-06372-1
22. Chen J, Chen J, Zhang D, Sun Y, Nanehkaran YA. Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture. 2020;173:105393. Available:https://doi.org/10.1016/j.compag.2020.105393
23. Agarwal M, Singh A, Arjaria S, Sinha A, Gupta S. ToLeD: Tomato leaf disease detection using convolution neural network. Procedia Computer Science. 2020;167:293–301. Available:https://doi.org/10.1016/j.procs.2020.03.225
24. Widiyanto S, Fitrianto R, Wardani DT. Implementation of convolutional neural network method for classification of diseases in tomato leaves. In: Proceedings of the 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia. 2019; 1–5. Available:http://doi.org/ 10.1109/ICIC47613.2019.8985909

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/115603