Classification of maize genotype using logistic regression

M. Arief Subchan and N.N. Andayani

Indonesian Cereals Research Institute, Jl. Dr. Ratulangi No. 274 Maros

Email: ningiceri@gmail.com

Abstract. Characterization of maize plants is among the pre-requested document prior to release to the public in Indonesia. Characterization of maize genotype involves various parameters includes agronomic parameters, yield and yield components. Characterization is generally carried out by professionals because it requires special skills in identifying genotypes based on their specific characters. The objective of the study was to classify the genotypes of corn plants based on the characters of the ear and kernel using a logistic regression model. The research was conducted at IP2TP Bajeng in 2020 by planting 4 genotypes, namely DYM-15, N 79, Mal 03 and G102612. A total of 100 plants per genotype were planted for cob characterization. Data analysis was done by using open-source software, Orange Software. The results indicated that the logistic regression model had a very good performance in classifying maize genotypes with an accuracy of > 98%. The values of the five parameters used to access the accuracy of the model are AUC=1.0, CA=0.99, F1=0.99, precision=0.99, recall=0.99. This value indicates that the use of IT-based tools can correctly classifying genotypes with high accuracy and consistency of results. Thus, digital based model can be integrated with manual selection for fast and precise grading of maize genotypes for maintaining seed quality.

1. Introduction

Precision agriculture has a focus on optimizing crop production by considering various aspects including variability and uncertainty in its implementation at field level. Among the aspects that is part of precision agriculture is monitoring the agronomic and yield components aspects of plants which play a role in predicting and monitoring plant performance and the level of yield obtained based on the agronomic character of the plant [1].

Among the agronomic and yields variables that are often used to characterize maize plants are the shape/color of the ear and the shape/color of the maize kernel. This characterization is very important, especially in corn seed multiplication activities to separate types of contamination plants that can pollinate and deteriorate crops purity when planted in the next season. Understanding the ear and kernel character was reported to have a significant effect on the level of yield obtained. Characterization of ear and kernel in the seed production process in addition to taking a long time, must also be carried out by trained agronomists. Limited resources in seed production activities cause important activities such as roguing not to run well, apart from the high cost aspect.

The 4.0 revolution in the agricultural sector is marked by the increasing use of tools such as computer vision that allows for digital-based plant inspections. Computer vision has various advantages including not damaging plants, being able to work on large areas, saving working time and the accuracy of the data produced is very high/consistent.

Utilization of computer vision technology has been applied to maize, including the development of artificial intelligence-based technology for identification of land suitability and soil fertility, detection of pest/disease attacks and drought risk, determination of harvest time and various other applications [2, 3, 4]. The application of computer vision technology can also be carried out to support maize seed
production activities at the level of nucleus seed, parent seed, F1 seed and composites/open pollinated maize seed. This technology is also adaptable to other commodities such as rice, sorghum, wheat and other food crop commodities. As for corn seed, classification of each variety is done by considering that the ears are physiologically mature [5, 6].

This study will explore the application of computer vision technology in increasing the efficiency of ear and kernel roguing activities by utilizing image segmentation-based image technology. Through the application of this technology, identification of off-type plants and dissimilar lines can be done easily. In addition, this research will develop a detection system for purity of maize genotypes based on the character of the shape and color of the cobs and kernel color.

2. Methodology
The research was conducted in January-April 2020 at IP2TP Bajeng, South Sulawesi. The test materials used were including four genotypes, namely DYW-15, G-102612, MAL-03 and N79 lines. Each genotype was planted as many as 100-200 plants with a spacing of 70 cm x 25 cm. The first fertilization was done at 14 days after planting with Urea 150 kg/ha and Ponska 200 kg/ha respectively. The second fertilizer was at 30 days after planting with Urea fertilizer as much as 200 kg/ha. Fertilizer was done manually about 7 cm beside the plant and covered again with soil.

Roguing was carried out twice during plant growth to remove off-type plants, namely at the age of 30 dap and 50-55 dap. Only plants with characteristics fit to the description were kept until harvest. Harvesting was done when the plant enters the physiological maturity phase which was marked by the appearance of a black layer at the bottom of the seed. The photo of the cob was taken using a
3. Results and Discussions

Maize genotype has several parts, including cobs/ear, roots, stems, and leaves. Cobs develop in the segments on the stem. The main cobs are generally found on the sixth to eighth stem segments. The segments usually have 5-7 cobs that develop imperfectly. The content of chemical compounds in corn cobs also affects the color and shape of the maize. Characterization of ear is among important stages in the formation of high-yielding varieties which aims to determine the important characters of economic value and as a characteristic of the variety concerned.

All genotypes data and images was calculated for their feature vector with VGG-16 learning architecture. VGG16 is among popular deep learning architecture used to train two dimensional model. Instead of having a large number of hyper-parameter, VGG model focused on having convolution layers and always used same padding and maxpool layer consistently throughout the whole architecture. The 16 in VGG16 refers to it has 16 layers that have weights. Each prepared image is generated and assigned to each images for further classification. The classification methods used in this study was Logistic regression. Logistic regression is an approach to making predictive models as well as linear regression. The difference is in logistic regression, and researchers predict the probability of high-yielding varieties which aims to determine the important characters of economic value and as a characteristic of the variety concerned.

The process of classification started by incorporating all datasets into the image embedding process, which in addition to generate vector features (Table 1). The output of the process produces data categories, image names, image sizes, and file size. As any as 1000 vector features were calculated with its feature vector.

The process of classification was carried out using open source software Orange. Analysis of the performance model was carried out using 5 measures, namely UAC, CA, F1, precision and recall.
Two confusion matrix tables were created for the whole dataset for exploring the effect of the different ear shapes and color on the classification ability (Table 2 and 3). The table represents rows for predicted classes and columns for actual. Using the confusion matrix table clearly indicated that logistic regression performs the best classification in VGG_16, where only two out of 236 test images of maize ear were misclassified, ensuring high precision over 98 percent. The confusion matrix analysis also showed the most likely misclassification was found in Mal-03 line classification vs N 79 line.

**Table 2. Confusion matrix of tassel classification**

|       | DYW-15 | G-102612 | MAL-03 | N79 | Σ   |
|-------|--------|----------|--------|-----|-----|
| Actual|        |          |        |     |     |
| DYW-15| 100.0% | 0.0%     | 0.0%   | 0.0%| 60  |
| G-102612| 0.0% | 100.0%   | 0.0%   | 0.0%| 59  |
| MAL-03| 0.0%   | 0.0%     | 98.3%  | 1.7%| 59  |
| N79   | 0.0%   | 0.0%     | 1.7%   | 98.3%| 58  |
| Σ     | 60.0   | 0.0      | 59.0   | 58.0| 236 |

**Table 3. Confusion matrix of tassel classification based on instance number**

|       | DYW-15 | G-102612 | MAL-03 | N79 | Σ   |
|-------|--------|----------|--------|-----|-----|
| Actual|        |          |        |     |     |
| DYW-15| 60     | 0        | 0      | 0   | 60  |
| G-102612| 0    | 59       | 0      | 0   | 59  |
| MAL-03| 0      | 0        | 58     | 1   | 59  |
| N79   | 0      | 0        | 1      | 57  | 58  |
| Σ     | 60     | 59       | 59     | 58  | 236 |

Breeding information of the maize ear can be described as follows: G-102612 genotype has average size of the cob is 16.5 cm long and 3.9 cm in diameter. The kernel arranged on the cob straight and regular. Seed type is semi horse tooth (semi dent). The color of the seeds is yellow orange with the number of rows of seeds 12-14 rows of seeds. The genotype has conical cylindrical cob shape. N79 genotype has flint seed type (flint), orange seed color (orange). The color of the seeds is yellow orange with the number of rows of seeds 14-16 rows of seeds Number of rows 14-16 rows per cob. The genotype has cylindrical cob shape. Mal-03 genotype has semi flint seed type (semi-flint), orange seed color (orange). The color of the seeds is yellow orange with the number of rows 2-14 rows per cob. The genotype has conical cylindrical cob shape. DYW-15 genotype has straight and regular kernel arrangement. The color of the seeds is yellow orange with the number of rows 14-16 rows per cob. The genotype has conical cylindrical cob shape. Corn seed with sufficient maturity has a significant effect on seed vigor and seedling rate [7, 8].
Figure 2. Comparison measures to determine the robustness of the models.

The bar diagram of the robustness model using VGG-16 model is shown in Figure 2. Model architecture showed the best performance, based on five measures in terms of accuracy, precision, recall, and F1 score, while AUC measures showed a slightly high score. The values of the five parameters used to access the accuracy of the model are AUC=1.0, CA=0.99, F1=0.99, precision=0.99, recall=0.99. By fine-tuning or change the model version, it is possible to explore a deeper network and increase the model's accuracy.

4. Conclusion
Characterization of maize genotype involves by focusing on ear and kernel characteristics was proved to precisely classify maize genotypes. A total of 100 plants per genotype were tested of their specific characteristics by using modeling image embedding process. The results indicated that the logistic regression model had a very good performance in classifying maize genotypes with an accuracy of >98%. The values of the five parameters used to access the accuracy of the model are AUC=1.0, CA=0.99, F1=0.99, precision=0.99, recall=0.99. This value indicates that the use of IT-based tools can correctly classifying genotypes with high accuracy and consistency of results. Thus, digital based model can be integrated with the field equipment for remote classification of ear type based on physical characteristics.

References
[1] Hayase, H., T. Satake, I. Nishiyama and N. Ito. 1969. Male sterility caused by cooling treatment at the meiotic stage in rice plants. II. The most sensitive stage to cooling and the fertilizing ability of pistils. Proc. Crop Sci. Soc. Jpn. 38: 706-711.
[2] Cires, D.C., U. Meier, J. Masci and L.M. Gambardella. 2003. IJCAI11-210.pdf Proc. Twenty-Second Int. Jt. Conf. Artif. Intell. Flex. 1237–42.
[3] Lu, H., Z. Cao, Y. Xiao, B. Zhuang and C. Shen. 2017. TasselNet: Counting maize tassels in the wild via local counts regression network. Plant Methods 13: 1–17.
[4] DeChant, C., T. Wiesner-Hanks, S. Chen, E.L. Stewart, J. Yosinski, M.A. Gore, R.J. Nelson and H. Lipson. 2017. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning Phytopathology 107: 1426–32.
[5] Abdul Baki, A.A., and J.D. Anderson. 1970. Viability leaching of sugars from germinating barley. Crop. Sci. 10: 31-34.
[6] Delouche, J.C. 1971. Determinants of seed quality. Seed Technology Laboratory, Mississippi State Univ., USA. 51p.
[7] Sadjad, S. 1975. Masalah pengujian benih. Kertas kerad lokakarya benih tanaman reboisasi
dan penghijauan. Direktorat Reboisasi dan Rehabilitasi Ditjen Kehutanan, Deptan.

[8] Sadjad, S. 1975. Pengaruh vigor benih terhadap vigor tanaman di lapang dan daya simpan benih jagung. Magister Sain Tesis. FPS, IPB. 127p.