ITM Web of Conferences **70**, 03025 (2025) https://doi.org/10.1051/itmconf/20257003025
DAI 2024

Dual Convolution Neural Networks of Ensemble Learning with Attention Mechanism for Rice Classification

Linfeng Cheng

Computer Sciences, Chengdu University of Technology, 610059 Chengdu, China

Abstract. Machine vision has been widely applied across fields. Image classification is one of the most classic fields. The aim of this project is to develop a dual convolutional neural network for ensemble learning based on the initial model and the res network model, and apply the ensemble model to the rice classification problem. The ensemble model in this article combines two deep models, InceptionNet and ResNet, and incorporates self-attention block method to construct an attention mechanism that uses multi head attention layers to capture relationships in the input. Attention output is added back to input. At the same time, 10 evaluation indicators were introduced as the results of testing and evaluation. In the result analysis, it can be concluded that the ensemble model has demonstrated excellent training efficiency in these indicators, and the learning rate hyperparameter has been replaced to improve the stability of the model. At the same time, for a more comprehensive comparison, the ensemble model studied in this article was also compared and analyzed reasonably with three pre trained models: VGG-16, ResNet50, and MobileNet. In the future, it is necessary to continuously optimize the structure of integrated models and adjust their hyperparameters to achieve better stability.

1 Introduction

Food is a fundamental necessity in human existence and requires continuous supply and production to meet the escalating needs of an expanding population via sustainable agriculture [1]. However, enhancing agricultural product quality and deciphering methods for selecting high-quality products pose as a significant issue. Rice is an important part of the food in humans' lives. It is also the most important calorie intake for human nutrition. It typically provides 100 calories per 130 grams and contains 1% calcium, iron and 3% magnesium [2]. However, the quality of rice is difficult to define because the quality of rice grown varies from region to region. Quality criteria for mills and growers differ, despite some overlapping qualities sought by all. For instance, milled rice quality measurement relies heavily on post-milling collection rates, while consumer preferences encompass grain appearance (such as length, width, area, and shape). Acceptable aspect ratios lie between 2.5 and 3.0, with a minimum length being 6mm [3]. The rice grain has feed about 67% of the

Corresponding author: cheng.linfeng1@student.zy.cdut.edu.cn

world's population [4]. Classifying rice based on color, presentation, then manual or machine segregation can be limited and prone to errors. Some regions require tactile interaction for classification. Machine vision can detect lower rice layer characteristics and improve classification.

Subsequently, the advent of machine vision has drastically improved efficiency. Machine vision has attracted a lot attention of in various area. The image classification is one of the most important apply area, the main target of it is to categorize images into predefined classes. In today's agricultural development, machine vision has played a lot of convenience. For example, using computer vision and Deep Neural Network (DNN) to classify flowers can greatly improve their accuracy [5]. Deep Convolutional Neural Networks (DCNN) can also be used to classify the four different categories of insects/pests found on soybean crops [6].

For these reasons, this article points its research focus to rice, an important food in agriculture. Computer vision technology presents a feasible alternative method for solving the challenging task of rice classification. A potential solution to this problem is the development of an efficient machine vision system for rice classification. In this study, the utilization of machine vision for the purpose of rice classification will be thoroughly explored, with the utilization of ensemble learning methods playing a key role in the solution's development. This project will also delve into the principles of two of the most popular deep learning models, Inception net and Res-Net, highlighting the key aspects of each as they relate to rice classification. Furthermore, attention block components will be incorporated into these two models, leading to the development of a new ensemble model that incorporates the attention mechanism. This ensemble model will then be utilized to address a five-class classification problem against the Rice Image Data set, providing a valuable and practical approach to developing a robust system for automated rice classification.

2 Method

2.1 Dataset preparation

To construct and evaluate the proposed ensemble net, there are five classes image of Rice varieties are adopted: Arborio, Basmati, Ipsala, Jasmine and Karacadag rice varieties. In 2019, M. Koklu et al. conducted the data set named Rice Image Data Set. These varieties are separated due to their features. Rice has many genetic varieties, and these varieties are separated from each other due to some of their features. These are usually features such as texture, shape, and color [7]. With these features, it is feasible to classify and evaluate the varieties. The five varieties are shown in Fig. 1.

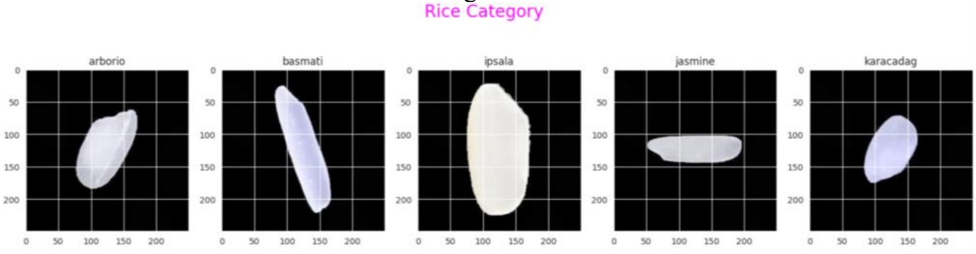

Fig. 1. Different labels of rice (Photo/Picture credit: Original).

A total of 75,000 grain images, 15,000 from each of these varieties, are included in the data set. It contains 106 features including 12 morphological, 4 shape and 90 color features obtained from these images.

In terms of the data preprocessing, the implementation uses image generator method. The rescale is 1.0/255, it scales the pixel values of the images to the range [0, 1]. The rotation_range rotates the image by a specified angle in the range [-20, 20] degrees. And the fill_mode determines the strategy that used for filling in newly created pixels that may result from the above transformations, the 'nearest' fills new pixels with the nearest existing pixel value.

For the data split ratio, the split ratios of 0.7, 0.15, and 0.15 are commonly used in machine learning area. This study chose this ration, arrange the majority of the data to the training set (70%) to allow the ensemble model to learn as many patterns and relationships as possible. And the validation set (15%) is used to tune hyper parameters during the model training process, it helps prevent over-fitting by providing a single set. The remain part is allocated to test set.

2.2 Model

In this section, the proposed model is introduced. This ensemble model combined two different deep learning models: the Inception Net and the Res Net. The primary conceptions of the two nets and the ensemble model are as follows.

2.2.1 The Inception Block

In 2014, C. Szegedy et al. conducted the Inception architecture which improved the utilization of computing resources within the network [8]. The Inception architecture allows for increasing the depth and width of the network, and also constantly 4 keeping the computational budget. It consists of multiple layers, with each layer using a combination of 1x1, 3x3, and 5x5 convolutions to cover clusters of units (Szegedy, et al, 2014). In this study, the Inception block method is defined to achieve an inception net. The block contains three branches. The branches of the inception block apply 1x1, 3x3, and 5x5 convolutions, then the output layer concatenated all these branches.

2.2.2 The Res Net

In 2015, K. He et al. introduces a residual learning framework for training deep neural networks, which allows an easier way of optimization and improved accuracy in image recognition tasks [9]. The Res Net reformulates the layers of the network as learning residual functions with reference to the layer inputs, rather than learning the un-referenced functions. By this way, this team made it becomes easier to optimize the network and achieve better accuracy, especially with the increased depth. In this study, the Res Net block method is defined to achieve a Res Net. The res block first contains a 1x1 filter then followed a batch normalization and ReLU activation. There are also a 3x3 filter and 1x1 filter followed.

2.2.3 Attention block

This study also attempts to construct an attention mechanism by defining self-attention block methods, attempting to integrate attention blocks into Res Net and Inception Net. The self-attention block function implements a self-attention mechanism that uses multi head attention layers to capture relationships in the input. Attention output is added back to input.

2.3 Implementation details

In order to avoid using the in-build loss function and accuracy, this section proposed the self-defined cross-entropy loss method and accuracy method. The self-defined loss function uses the method to predicate the e possibilities.

These two methods are defined to achieve the avoidance of using in-build functions. The cross-entropy method calculates a modified categorical cross entropy loss for classification tasks. And the custom_accuracy method computes classification accuracy by comparing the indices of the true and predicted classes, then the loss is calculated using this formula: Categorical Cross-Entropy Loss = $-\sum$ y_true $\times$ log(y_predicate) [10]. The accuracy 8 method is defined to measure the percentage of correctly predicted instances out of the total instances.

After this, the compile() method will call these self-defined methods to compile the model.

The study shows the training of the model. The compile() method includes the self-defined loss and accuracy, and also contains metrics such as top-k, f1 score, recall etc. For training, the fit() method is adopted, there are 20 epochs to train the first version.

This study introduces 10 evaluation metrics to perform the test and evaluation of over performance. The 10 metrics are Accuracy, Loss, Recall, Precision, Specificity, Sensitive and F1score.

3 Results and discussion

In this section, the current article will introduce and thoroughly explicate the outcomes of the integrated model. For the model under discussion in this piece, a total of ten metrics has been employed to validate its performance. The experimental results pertaining to this model before and after adjusting its parameters are presented here, enabling an impartial examination of its comprehensive capabilities. To ensure objectivity, it is also contrasted with three pre-trained models. This research furnishes ten graphs that illustrate disparity among various indicators for these three pre-trained models.

The present article has undertaken comparative analysis, accentuating the unique advantages of this model. Furthermore, it critically evaluates the limitations of my ensemble model when compared to other models.

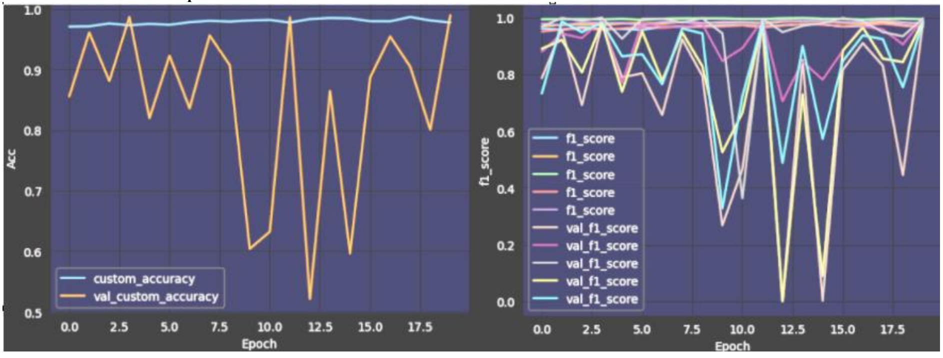

Fig. 2. The results before parameter tuning (Photo/Picture credit: Original).

Regarding the training efficacy, this model demonstrates commendable training efficiency, attaining high accuracy, precision, recall, and F1 score. The training loss diminishes throughout the epochs, verifying that the model is assimilating from the training data. Concerning the validation efficacy, the proposed model exhibits robust generalization, preserving excellent accuracy, precision, recall, and F1 score. The AUC-ROC values are elevated, signifying strong differentiation capacity. Additionally, sensitivity at high specificity (sensitivity_at_specificity_1) and specificity at high sensitivity (specificity_at_sensitivity_1) display commendable performance, demonstrating balanced performance.

However, these metrics lack stability during the training process. As the metrics shown in Fig. 2, these metrics indicated that the model performance had sharply decreased at epoch 10 and epoch 12 since all the 10 metric curves declined steeply. Considering that, this paper tries to tuning parameter to improve this performance flaw. The parameter tuning is shown below.

This paper implemented a dynamic learning rate based on LearningRateScheduler callback that provided by Keras. The formula is:

$$\text{Learning rate} = \text{initial\_learning\_rate} * \text{decay} ** (\text{epoch}//5) \tag{1}$$

The decay factor controls the level that learning rate decreases. This formula means the learning rate decreased each 3 epoch by decay 0.9.

The second tuning is to increase the dense layer of the ensemble model, the size of two dense layers is tuned to 200 and 100 for getting higher overall training performance. Then this report also decreases the number of epochs from 20 to 15. There are 15 epoch for training after tuning.

After parameter tuning, the results display in the Fig. 3:

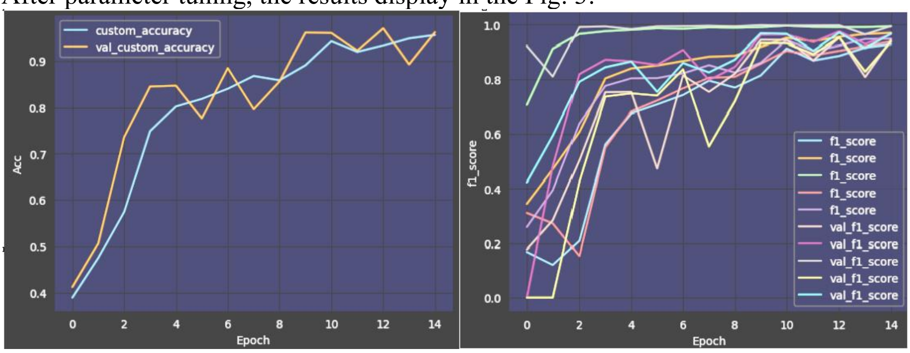

Fig. 3. The results after parameter tuning (Photo/Picture credit: Original).

As shown above, after the implementation of dynamic learning rate, the stability of this ensemble model has improved because the f1_score, the accuracy, the recall, and AUC-ROC curve has stable change during the training process. However, according to precision metric, precision curve has experienced large fluctuation from epoch 0 to epoch 8, then its roes in epoch 9. That means the internal flaw still existed in the model.

This study uses pre-train model to perform the same data set: Rice Image, then fairly comparing the performance of proposed model. The pre-train model, VGG-16, ResNet50, and MobileNet is selected and implemented, and the training results of pre-train models are shown in Table 1:

Table 1. Comparaison of the pre-train models.

Name	Accuracy	Recall	AUC	Precision	Top-3
Proposed model	0.96	0.96	0.93	0.98	0.99
VGG-16	0.92	0.91	0.99	0.93	0.99
ResNet50	0.94	0.96	0.95	0.97	0.98
MobileNet	0.96	0.96	0.99	0.96	0.99

According to Table 1, the proposed ensemble model performance is similar with the pre-train model that provided by Keras. The top-k accuracy meets the average level, the recall is

0.4 higher than the VGG-16 model. The AUC is the lowest indicator among these models. The recall and the precision of the proposed model all perform well.

4 Conclusion

This paper proposed the ensemble model based on the inception net and res net. In the process of fitting image data, this model performs a good overall performance on rice classification, which accuracy reached to 0.96 in the best epoch, and the precision also reached to the range from 0.92 to 0.98. The metrics, including top-k loss, recall, AUC, sensitivity_at_specificity, specificity_at_sensitivity, f1_score, and loss have been displayed.

The advantages of this proposed model are the effective use of residual and inception blocks for feature extraction, and the model performs an acceptable result to avoid over-fitting. However, there are still some flows of this model. The model still exists the risk of over-fitting, and the model need to train on other data set to prove the overall performance. The f1_score stills not stable during the training epoch which performs fluctuations. The precision, loss and AUC performs a sudden decrease from epoch 10 to 12 according to metrics figure in section 3. Although the parameter tuning has improved this problem, it still indicated that the structure exists some internal flaw.The utmost significant future work of this project is to keep optimizing the model structure. And tuning the hyper parameter for stable performance, then trying to reduce the size of this ensemble model. Also, exploring the data argumentation and ensemble method to make better achievements in the future.

References

1. H. S. Abdullahi, R. E. Sheriff, F. Mahieddine, Convolution neural network in precision agriculture for plant image recognition and classification, in 2017 Seventh International Conference on Innovative Computing Technology (INTECH), pp. 1 – 3 (2017).
2. S. F. Ali, H. Jamil, R. Jamil, I. Torij, S. Naz, Low cost solution for rice quality analysis using morphological parameters and its comparison with standard measurements, in 2017 International Multi-Topic Conference (INMIC), pp. 1 – 6 (2017).
3. M. J. Asif, T. Shahbaz, S. Tahir Hussain Rizvi, S. Iqbal, Rice Grain Identification and Quality Analysis using Image Processing based on Principal Component Analysis, in 2018 International Symposium on Recent Advances in Electrical Engineering (RAEE), pp. 1 – 6 (2018).
4. N. M. U. Din, A. Assad, R. A. Dar, M. Rasool, S. U. Sabha, T. Majeed, Z. U. Islam, W. Gulzar, A. Yaseen, RiceNet: A deep convolutional neural network approach for classification of rice varieties, Expert Systems with Applications, vol. 235, art. no. 121214 (2024).
5. P. B. Hai, D. Anh Tuan, H. N. Nam, D. Nguyen Bao, Performance evaluation of the multiclass classification of flowers on diverse datasets, in 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1 – 6 (2023).
6. D. Shah, R. Gupta, K. Patel, D. Jariwala, J. Kanani, Deep Learning based Pest Classification in Soybean crop using Residual Network-50, in 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), pp. 1 – 5 (2022).

7. M. Koklu, I. Cinar, Classification of Rice Varieties Using Artificial Intelligence Methods, International Journal of Information Systems and Applications, vol. 7, no. 3, pp. 188 – 194 (2019).
8. C. Szegedy, C, et al. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9) (2015).
9. K. He et al. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778) (2016).
10. A. Mao, M. Mohri, & Y. Zhong. Cross-entropy loss functions: Theoretical analysis and applications. In International conference on Machine learning (pp. 23803-23828). PMLR (2023).