ELINVO (Electronics, Informatics, and Vocational Education), *Mei 2023; vol 8(1):137-145*
ISSN 2580-6424 (printed), ISSN 2477-2399 (online) DOI: https://doi.org/10.21831/elinvo.v8i1.55763

Classification of Corn Seeds Quality using Residual Network with Transfer Learning Weight

Meidya Koeshardianto[1], Wahyudi Agustiono[2], Wahyudi Setiawan[3,*]
[1]Department of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Jawa Timur, Indonesia
[2,3]Department of Information System, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Jawa Timur, Indonesia
[*]E-mail: wsetiawan@trunojoyo.ac.id

ABSTRACT

Corn is one of the main ingredients in farm animal feed. Currently, corn is preferable because widely available and cheaper in the market than others. However, it needs quality control on corn production. The company that manufactures animal feed has certain quality standards to receive corn material. On the other hand, the quality of corn produced varies greatly. Thus, quality control when receiving corn from suppliers greatly affects the quality of animal feed. The quality of feed ingredients is classified into physical properties and analytical values. Physical properties are determined so that the resulting corn can be accepted or rejected, while the analytical value is used as the basis for formulating the diet. The physical properties of corn are determined by the human senses, such as sight and smell, while the analytical value is by chemical analysis. Physical quality control by relying on human senses is certainly limited and takes time. Based on these problems, it needs to make a classification system of corn seeds automatically. This study uses corn seed images as classification data. The system uses public data from Naagar which consists of four classes: pure, discolored, silk cut, and broken. Image classification uses a Convolutional Neural network (CNN) with ResNet152v2 architecture. The hyperparameters used consist of a learning rate of 0.001, a batch size of 512, and an epoch of 25. Adaptive Moment Estimation (Adam) for the optimizer. Percentage of data training vs validation 80:20. The validation results show an accuracy of 65%, precision of 66%, and recall of 64%

Keywords*:* Corn Seed, Image Classification, Convolutional Neural Network, ResNet152v2, Transfer Learning.

INTRODUCTION

The government shows progress success in corn self-sufficiency, especially for animal feed. In 2018, according to the Ministry of Agriculture, Indonesia was able to export more than 372,000 tons of corn to foreign countries. Domestically, the need for raw materials for livestock is also a problem, because so far, the level of dependence on imported wheat is quite high. The abundance of corn in the country as well as the high price of wheat have finally forced large animal feed processing companies to substitute half of the wheat for corn. As a result, corn is absorbed by large animal feed companies, so small and medium-scale animal or poultry farmers lack corn feedstocks. Until the end of 2018, the government finally re-imported corn on a smaller scale between 50 to 100 thousand tons to secure the stock of domestic needs [1].

Efforts to increase the yield of corn cultivation continue to be carried out. According to data, the need for animal feed, especially poultry, is the largest percentage of corn consumption 60%. Meanwhile, food is only limited to 30%, while the rest is for industry and seeds [2]. In addition, to increase corn production, supervision must be carried out on the quality of corn. A variety of corn has been produced and developed, but pests and diseases attack corn cannot be avoided. The effect is that the quality of corn is not in line with the expectations.

Acceptance of corn material is a key input for quality control programs in feed mills. Corn is the main ingredient used in feed mixtures, and the quality varies greatly. Thus, quality control when receiving corn from suppliers greatly affects the quality of animal feed produced. The quality of feed ingredients is classified into physical properties and analytical values. The physical properties are determined so that the staff receiving the ingredients can determine whether the ingredients are accepted or rejected, while the analytical values are used as the basis for formulating the diet. The physical properties of corn are determined by the senses, such as sight and smell, while the analytical value is determined by chemical analysis.

The sampling process for testing when receiving raw materials at the company is carried out in two stages: sampling one and sampling two. Sampling one was carried out by Quality Control personnel in the parking lot of trucks transporting raw materials. Sampling was carried out from bags that were visible from the outside (about 30% of the total bags). Samples were checked for texture, color, and heat. If the test is declared successful, the moisture content, aflatoxin, percentage of whole seeds, damaged, moldy and foreign seeds will be checked.

After passing the first sampling, the truck is allowed to pass through the weighbridge, then unload and proceed to sample two. In sampling two, the water content, protein, fat, crude fiber, ash, and aflatoxin content were tested. Phase two assays were carried out using near-infrared spectroscopy. The problem of quality control at the time of sampling is the length of the testing process. This is because the determination of the percentage of whole, damaged, and mold seeds is done by manually separating and then weighing them to determine the percentage of each type. [3].

This method is not only time-consuming, but also depends on the analyst's subjective factors, including boredom/fatigue, visibility, bias, pressure, and inappropriate lighting. The time required to determine the physical quality causes long queues of material trucks in the feed mill parking lot. Based on these problems, it is necessary to conduct a technological study to determine the physical condition of corn seeds automatically according to their physical conditions.

One alternative to distinguish the quality of corn kernels is through image classification. Research for the classification of corn quality has been carried out previously, including research on the detection of corn seed quality using a dataset consisting of five classes: corn kernels worm, mold, damages, discoloration, and good. The dataset uses the number of images that varies from 286 to 1,714 per class for training data. The system is pre-processed at the beginning using the grayscale, Otsu, morphology, and watershed methods. The classification uses Google net. The results show an accuracy of up to 95% [4].

Another study used the Convolutional Generative Adversarial Network (CGAN) and Batch Active Learning for fast annotation. The dataset used consists of four classes: broken, discolored, pure, and silk cut. The number of images for each class varies from 1,751 to 7,267. The test results show an accuracy of up to 85.24% [5].

Subsequent research used data on nine types of corn varieties. The steps used are feature extraction and classification. Feature extraction using Gray Level Co-Occurrence Matrix and Linear Binary Pattern. Furthermore, the classification uses a custom Convolutional Neural Network (CNN) with parameters epoch 100, learning rate 0.01, batch 32, and Root Means Square Propagation (RMSProp) optimizer. The data for each class is 1,000 images. The results show an accuracy of 98.1% [6].

Furthermore, research on the detection of corn quality using data from 8,080 images with five classes. The training uses a variety of CNN architectures such as Alexnet, VGG, Google, mobile net, densenet, shufflenet, efficient, and P-ResNet. The parameters used are epoch 30, learning rate 0.001, and Batch Size 32. The results show the best accuracy up to 99.70% by P-ResNet, the architecture created is a modification of ResNet [7].

The differences in the datasets used are difficult to compare. In this study, it is proposed to detect corn quality using data from the research of Nagar et al. Data can be acquired from public data. The data used for detection consists of four classes: pure, discolored, silk cut, and broken [5].

For the method used, many previous studies used deep learning methods such as CGAN or CNN. This research uses a transfer learning Residual Network. Transfer learning is the use of feature extraction results from the CNN pre-trained network.

METHODS

Corn seed quality classification consists of four steps: data exploration, distribution of training and validation data, and classification using ResNet152v2 for training and validation. The block diagram of the system is shown in Figure 1.

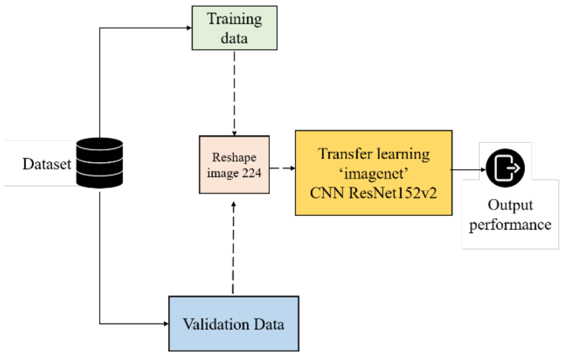


Figure 1. Block diagram of corn grain quality classification

In this study, training data and validation data have been determined. The percentage of training and validation data is around 80:20. The number of training and testing data is 14,322 and 3,479. Classification using Convolutional Neural network with Residual Network architecture 152v2 with weights derived from ImageNet [8]–[12]. Tests were carried out to obtain the best model which was then used to calculate system performance during training and validation.

Dataset

The dataset used is nagar corn seed data which can be obtained free of charge on the web. The dataset address is https://naagar.github.io/cornseedsdataset/. The public dataset consists of four classes. The data per class has different amounts and sizes of the images. The pixel size is between 116 to 184. The comparison of the amount of training data for each class is shown in Figure 2. While the sample images for training and validation data are shown in Figure 3.

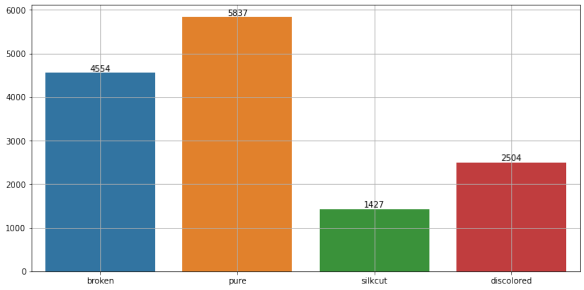
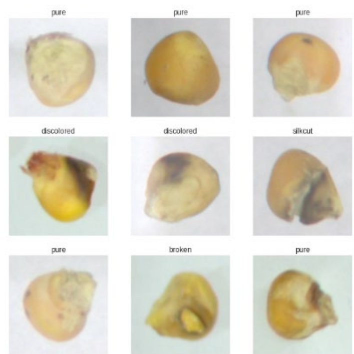
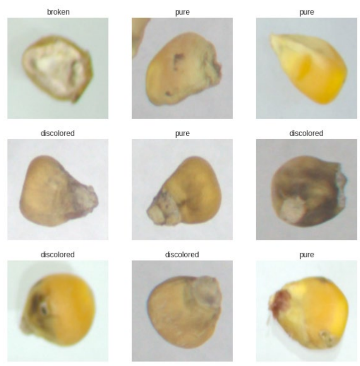


Figure 2. The amount of training data for each class

A comparison of the amount of data shows that the image of each class is not balanced in both the training and validation data.



(a)



(b)

Figure 3 example image (a) training data (b) validation data

Image sizes both length and width are not the same for each image. This is certainly a challenge for the system to be able to recognize validation data better.

Convolutional Neural Network

A Convolutional neural network (CNN) is a method of deep learning to process data with a grid pattern, which is used to learn the special hierarchies of features automatically from low to high-level patterns. CNN has an architecture that is divided into several layers, where these layers are the differentiator in the convolution process with other neural networks. In general, CNN has a layer consisting of a Convolutional layer, a pooling layer, and a fully connected layer [13]–[15].

Residual Network

A variety of architectures have been created since AlexNet won the ImageNet competition in 2012 [16]–[18]. The architectural model that was winning in the following years used more layers to reduce errors. However, there is a problem when adding a deeper number of layers, the vanishing gradient. This can cause the gradient value to be zero or too large. So, when the number of layers increases, errors during training and testing also increase.

The Residual block architecture was introduced as a solution to this problem. Figure 4 shows the ResNet architecture with input-output size. This architecture uses the skip connection technique. This technique connects the activation layer with the next layer by passing several layers in between. This technique forms a residual block. ResNet is created by stacking the residual blocks together. The concept of skip connection is shown in Figure 5 [19]–[21].

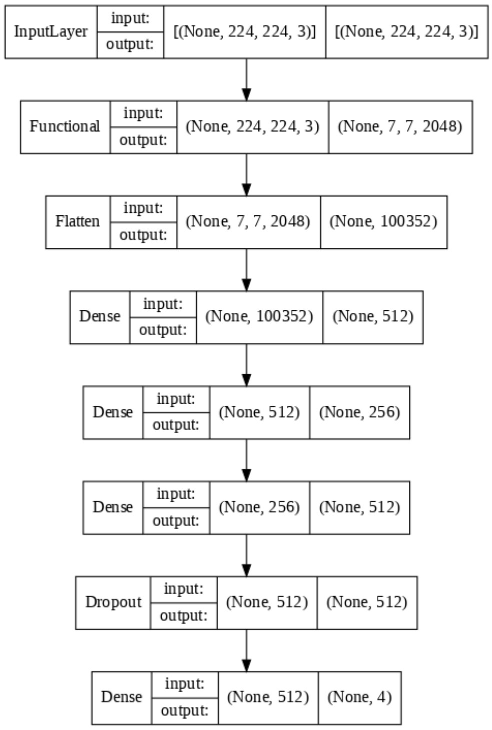


Figure 4. ResNet input and output size

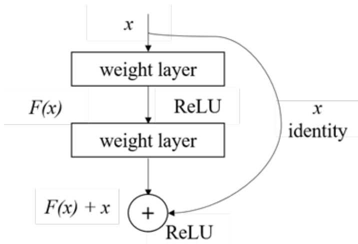


Figure 5. Skip connection

$$F(x) = H(x) - x$$
$$H(x) = F(x) + x \quad (1)$$
$H(x)$ = initial mapping

The advantage of a skip connection is that if there is a layer that causes a performance drop, that layer can be skipped directly with regularization. Residual network architecture inspired by VGG-19 by adding a shortcut connection. Table 1 shows the ResNet152v2 architecture.

Table 1. ResNet152v2 Architecture

Layer name	Output size	152-layer
conv1	112 × 112	7×7,64, stride 2
		3×3 max-pool, stride 2
conv2_x	56 × 56	$\begin{bmatrix} 1\times 1,64 \\ 3\times 3,64 \\ 1\times 1,256 \end{bmatrix} \times 3$
conv3_x	28 × 28	$\begin{bmatrix} 1\times 1,128 \\ 3\times 3,128 \\ 1\times 1,512 \end{bmatrix} \times 8$
conv4_x	14 × 14	$\begin{bmatrix} 1\times 1,256 \\ 3\times 3,256 \\ 1\times 1,1024 \end{bmatrix} \times 36$
conv5_x	7 × 7	$\begin{bmatrix} 1\times 1,512 \\ 3\times 3,512 \\ 1\times 1,2048 \end{bmatrix} \times 3$
	1 × 1	Average pool,1000-fc, SoftMax

Table 1 shows the ResNet152 architecture which consists of the following elements:

1. Convolution_1 contains the kernel size 7×7 and 64 different kernels using stride 2. Next, there is max pooling with stride 2. So, conv_1 consists of 1 layer.
2. Convolution_2 has kernel size 1×1 and 64 different kernels, 3×3 and 64 kernels, 1×1 and 256 different kernels. These three layers are repeated 3 times. So, there are 9 layers in conv2_x.
3. Convolution_3 has kernel size 1×1 and 128 different kernels, 3×3 and 128 kernels, 1×1 and 512 different kernels. It repeated 8 times. There are 24 layers in conv4_x.
4. Convolution_4 has kernel size 1×1 and 256 different kernels, 3×3 and 256 kernels, 1×1 and 1024 different kernels. Those 3 layers have 36 times. It has 108 layers in conv4_x.
5. Convolution_5 has kernel size 1×1 and 512 different kernels, 3×3 and 512 kernels, 1×1 and 2,048 different kernels. These 3 layers are repeated 3 times. There are 9 layers in conv5_x.

At the end of the network, there is an average pool and ends with a fully connected layer 1,000 class, SoftMax. It gives the network a layer.
Total layers are 1 + 9 + 24 + 108 + 9 + 1 = 152 layers Deep Convolutional network
In this study, the ResNet152v2 architecture is used for classified 4 classes. It has a model with more than 109 million parameters as shown in Table 2 [9], [22], [23].

Table 2. ResNet152v2 Parameter

s	Output shape	Param #
resnet152v2 (functional)	(none,7,7,2048)	58331648
flatten	(none, 100352)	0
dense	(none,512)	51380736
dense_1	(none,256)	131328
dense_2	(none,512)	131584
dropout	(none,512)	0
dense_3	(none,4)	2052
total params: 109,977,348		
trainable params: 51, 645,700		
non-trainable params: 58,331,648		

Transfer Learning

The system requires a learning process to perform certain tasks. These tasks include image enhancement, classification, grouping, recognition, and detection. The data is divided into two parts, training data, and validation data. In conventional systems, training data is processed to gain knowledge. Problems arise when the amount of training data is limited, and the learning process does not run well.

An alternative solution to this problem is transfer learning. It is a machine-learning method that works by utilizing existing models. Transfer
learning modifies and updates the parameters of the model. Transfer learning makes the modified model learn with different tasks. The model used for transfer learning has learned from other data, so there is no need to learn from scratch. The model has recognized characteristics such as texture, shape, and color as a result of previous learning.

The benefit of transfer learning is learning well even though the training data is limited. In contrast to traditional machine learning, each learning process always requires a relatively large amount of data. Generally, the type of transfer learning used in deep learning is a pre-trained network. The steps to carry out transfer learning are as follows [10], [24]–[26]:

1. Select a specific model. The previously trained network model is taken from the existing model.
2. Models are reused. Pretrained models can be used as a starting point for carrying out new tasks. New tasks can use all parts of the pre-trained model or part depending on the system requirements.
3. Modify the model. Modifications are made to the Fully connected layer.

In this study, ResNet152v2 was trained with a dataset from ImageNet. The best model of the weights trained from the ImageNet is used to conduct training on the corn seed dataset.

Optimizer

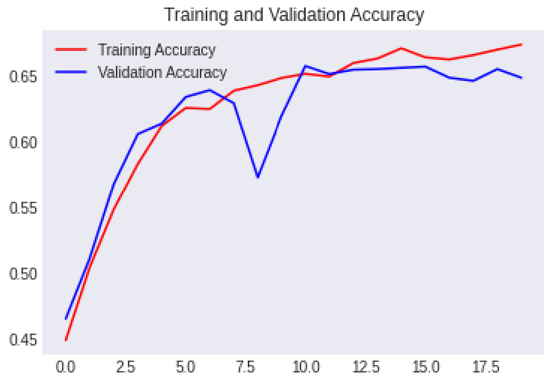
Convolutional neural networks require an optimizer to be able to perform optimal training. Various Gradient Descent optimizers can be used to optimize training. This research uses Adaptive moment Estimation (Adam) [27]–[29]. Adam combines between RMSprop [30], [31] with momentum which is normally used on Stochastic Gradient Descent [32]–[34].

The experiment environment

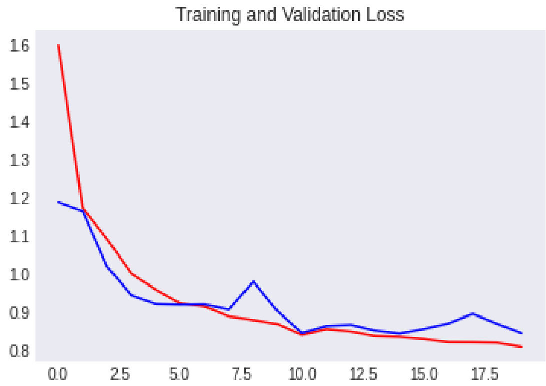
Corn seed quality was classified using Google Collab with specifications hardware of Core i7-7700 processor, GTX 1060 6 Gb D5 amp, 60 Gb Solid State Drive. It also needs a library of phyton such as TensorFlow, hard, imageDataGenerator, os, NumPy, pandas, seaborn, and matplotlib.

RESULT AND DISCUSSION

The experiment was carried out with parameters including epoch 25, learning rate 0.001, batch-size 512, and the optimizer adaptive moment estimation (Adam). The experiment consists of two steps, training, and validation. The results are shown in Figures 6 (a) and (b). It shows the accuracy and loss of training and validation.



(a)



(b)

Figure 6. (a) training and validation accuracy, (b) training and validation loss

The results show the best accuracy during training is 67,59% while for loss is 78,22%. However, in the validation session, the accuracy was only 64.82% and the loss was 84.47%. Furthermore, figure 7 shows the confusion matrix of validation. The validation results showed that each class correctly recognized 51.3% for broken, 73.6% for discolored, 72.8% for pure, and 58.3% for silk cut. The Performa result of corn seed classification shows in Table 3.

Causes of poor performance in the experiment include:

1. Unbalanced number of images in each class
2. The size of the image that is reshaped directly to 224 removes important information from the actual image.

3. Inappropriate classification method

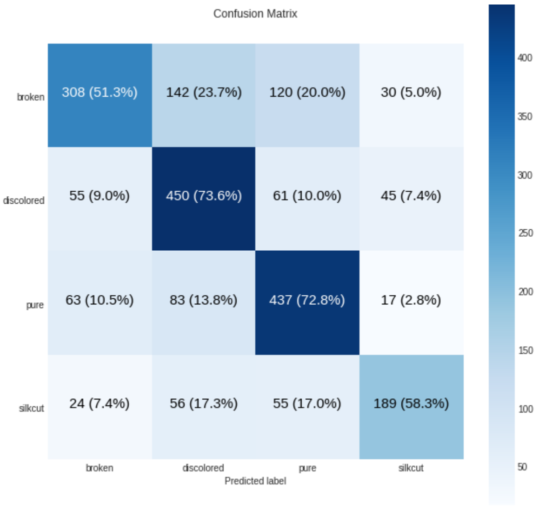


Figure 7. Confusion matrix test results

Table 3. Performa result

classes	Precision	Recall	F1-score
Broken	0.68	0.51	0.59
Discolors	0.62	0.74	0.67
Pure	0.65	0.73	0.69
Silk cut	0.67	0.58	0.62
Accuracy			0.65
Macro avg	0.66	0.64	0.64
Weighted avg	0.65	0.65	0.64

macro avg: all classes contributed to the final averaged metric.
weighted avg: each class contributed to the weighted average by its size.

These three possibilities are hypothesized to be the cause of recognition failure and overfitting. First, an unbalanced amount of data is often the main factor for unsuccessful recognition in the image classification task. Some methods can be used to balance the data such as random under sampling and random oversampling. In addition, we can also take advantage of the Synthetic Minority Over-Sampling Technique (SMOTE). Second, when doing pre-processing that resizes the image directly, it is feared that important information will be lost. Therefore, it can be replaced by doing Region of Interest (ROI) and grayscale or RGB channel pre-processing techniques. Third, it is possible that the model used is not suitable. Furthermore, it can be compared with other ResNet architectures or other CNN architectures such as Alexnet, VGG, or google net. In addition to these three hypotheses, performance improvements can be made by tuning hyperparameters on epochs and learning rates.

CONCLUSION

In this study, the image classification of the quality of corn kernels has been carried out using the Residual network architecture 152v2 which uses the weights from the ImageNet. The validation performance of the system can be improved in the next research process by making the balanced data for each class, the appropriate pre-processing technique, the CNN classification method with the precise architecture, and performing hyperparameter tuning on the epoch and learning rate.

ACKNOWLEDGMENT

This research was funded by Penelitian Mandiri University of Trunojoyo Madura skim Grup Riset No. 159 /UN46.4.1/PT.01.03/2022

REFERENCES

[1] Kementrian Pertanian Republik Indonesia, “Indonesia Ekspor Jagung 372 Ribu Ton dan Menyetop Impor 9,2 Juta Ton,” *https://www.pertanian.go.id/*, 2018. .

[2] S. Panikkai, R. Nurmalina, S. Mulatsih, and H. Purwati, “Analisis Ketersediaan Jagung Nasional Menuju Swasembada Dengan Pendekatan Model Dinamik,” *Inform. Pertan.*, vol. 26, no. 1, p. 41, 2017, doi: 10.21082/ip.v26n1.2017.p41-48.

[3] Adzriral, D. Anggraini, N. Novita, Santosa, and Andasuryani, “Pendugaan Kualitas Fisik Biji jagung untuk Bahan Pakan menggunakan jaringan Syaraf Tiruan berdasarkan Data Citra Digital,” *J. Peternak. Indones.*, vol. 13, no. 3, pp. 183–190, 2011.

[4] S. Huang, X. Fan, L. Sun, Y. Shen, and X. Suo, “Research on Classification Method of Maize Seed Defect Based on Machine Vision,” *J. Sensors*, vol. 2019, no. 1, 2019, doi 10.1155/2019/2716975.

[5] S. Nagar, P. Pani, R. Nair, and G. Varma, “Automated Seed Quality Testing System using GAN & Active Learning,” pp. 1–9, 2021, [Online]. Available:

http://arxiv.org/abs/2110.00777.

[6] S. Javanmardi, S. H. Miraei Ashtiani, F. J. Verbeek, and A. Martynenko, "Computer-vision classification of corn seed varieties using deep convolutional neural network," *J. Stored Prod. Res.*, vol. 92, p. 101800, 2021, doi: 10.1016/j.jspr.2021.101800.

[7] P. Xu, Q. Tan, Y. Zhang, X. Zha, S. Yang, and R. Yang, "Research on Maize Seed Classification and Recognition Based on Machine Vision and Deep Learning," *Agric.*, vol. 12, no. 2, 2022, doi: 10.3390/agriculture12020232.

[8] T. Michalik and O. Polska, "How effective is Transfer Learning method for image classification," in *Conference on Computer Science and Information Systems*, 2017, vol. 12, pp. 3–9, doi: 10.15439/2017F526.

[9] J. L. Mahendra Kumar *et al.*, "The classification of EEG-based wink signals: A CWT-Transfer Learning pipeline," *ICT Express*, vol. 7, no. 4, pp. 421–425, 2021, doi: 10.1016/j.icte.2021.01.004.

[10] H. Salman, A. Ilyas, L. Engstrom, A. Kapoor, and A. Madry, "Do adversarially robust ImageNet models transfer better?," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.

[11] K. Pandya, P. Singhal, K. Pandya, and P. Singhal, "Image Classi fi cation using Transfer Learning," *Int. J. \Control Theory Appl.*, vol. 9, no. 40, pp. 899–905, 2016.

[12] R. Hu, S. Zhang, P. Wang, G. Xu, D. Wang, and Y. Qian, "The identification of corn leaf diseases based on transfer learning and data augmentation," in *Proceedings of the 2020 3rd International Conference on Computer Science and Software Engineering*, 2020, pp. 58–65, doi: https://doi.org/10.1145/3403746.3403905.

[13] S. Albawi, T. A. M. Mohammed, and S. Alzawi, "Layers of a Convolutional Neural Network," in *ICET2017*, 2017, pp. 1–6.

[14] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018, doi: 10.1016/j.patcog.2017.10.013.

[15] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," pp. 1–11, 2015, [Online]. Available: http://arxiv.org/abs/1511.08458.

[16] M. Z. Alom *et al.*, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," 2018, [Online]. Available: http://arxiv.org/abs/1803.01164.

[17] Y. Wu, "Identification of Maize Leaf Diseases based on Convolutional Neural Network," *J. Phys. Conf. Ser.*, vol. 1748, no. 3, 2021, doi: 10.1088/1742-6596/1748/3/032004.

[18] M. Simon, E. Rodner, and J. Denzler, "ImageNet pre-trained models with batch normalization," 2016, [Online]. Available: http://arxiv.org/abs/1612.01452.

[19] T. Gevers and A. Smeulders, "Identity Mappings in Deep Residual Networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9914 LNCS, p. V, 2016, doi: 10.1007/978-3-319-46493-0.

[20] S. Zagoruyko and N. Komodakis, "Wide Residual Networks," *Br. Mach. Vis. Conf. 2016, BMVC 2016*, vol. 2016-Septe, pp. 87.1-87.12, 2016, doi: 10.5244/C.30.87.

[21] T. F. Yu, Fisher, Vladlen Koltun, "Segmentation Dilated Residual Networks," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 472–480, 2017, [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2017/papers/Yu_Dilated_Residual_Networks_CVPR_2017_paper.pdf%0Ahttp://openaccess.thecvf.com/content_cvpr_2017/html/Yu_Dilated_Residual_Networks_CVPR_2017_paper.html.

[22] S. M. Rezaeijo, M. Ghorvei, and B. Mofid, "Predicting breast cancer response to neoadjuvant chemotherapy using ensemble deep transfer learning based on CT images," *J. Xray. Sci. Technol.*, vol. 29, no. 5, pp. 835–850, 2021, doi: 10.3233/XST-210910.

[23] N. M. Elshennawy and D. M. Ibrahim, "Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images," *Diagnostics*, vol. 10, no. 9, pp. 1–16, 2020, doi: 10.3390/diagnostics10090649.

[24] M. Huh, P. Agrawal, and A. A. Efros, "What makes ImageNet good for transfer learning?," pp. 1–10, 2016, [Online]. Available: http://arxiv.org/abs/1608.08614.

[25] R. Marlow *et al.*, "A phase III, open-label, randomized multicentre study to evaluate the immunogenicity and safety of a booster dose of two different reduced antigen diphtheria-tetanus-acellular pertussis-polio vaccines, when co-administered with the measles-mumps-rubella vaccine," *Vaccine*, vol. 36, no. 17, pp. 2300–2306, 2018, doi: 10.1016/j.vaccine.2018.03.021.

[26] S. Kornblith, J. Shlens, and Q. V Le, "Kornblith_Do_Better_ImageNet_Models_Transfer_Better_CVPR_2019_paper," *Proc. IEEE Comput. Soc. Conf. Comput. Vis.*

Pattern Recognit., pp. 2661–2671, 2019.

[27] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *ICLR*, 2015, pp. 1–15, doi: http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503.

[28] S. Y. Sen and N. Ozkurt, “Convolutional Neural Network Hyperparameter Tuning with Adam Optimizer for ECG Classification,” *Proc. - 2020 Innov. Intell. Syst. Appl. Conf. ASYU 2020*, no. 978, 2020, doi: 10.1109/ASYU50717.2020.9259896.

[29] Milan, “Maize Disease using VGG16 and ADAM,” *Kaggle*, 2019. https://www.kaggle.com/milan400/0-00001adam-cornmaizeleaf-vgg16.

[30] G. E. Hinton, N. Srivastava, and K. Swersky, “Lecture 6a- overview of mini-batch gradient descent,” *COURSERA Neural Networks Mach. Learn.*, p. 31, 2012, [Online]. Available: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

[31] S. Ruder, “An overview of gradient descent optimization,” pp. 1–14, 2017.

[32] P. Goyal *et al.*, “Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour,” 2017, [Online]. Available: http://arxiv.org/abs/1706.02677.

[33] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, “Performance of deep learning vs machine learning in plant leaf disease detection,” *Microprocess. Microsyst.*, vol. 80, no. October 2020, p. 103615, 2021, doi 10.1016/j.micpro.2020.103615.

[34] A. Ramezani-Kebrya, A. Khisti, and B. Liang, “On the Generalization of Stochastic Gradient Descent with Momentum,” no. 2015, pp. 1–36, 2021, [Online]. Available: http://arxiv.org/abs/2102.13653.