An Improved Deep Neural Network for Classification of Plant Seedling Images

Catherine R. Alimboyong
Graduate Programs
Technological Institute of the Philippines
Quezon City, Philippines
catherinealimboyong@gmail.com

Alexander A. Hernandez
College of Information Technology Education
Technological Institute of the Philippines
Manila, Philippines
alexander.hernandez@tip.edu.ph

***Abstract*— This scientific pursuit aimed to develop a deep learning architecture tailored to classify plant seedling images. Our architecture encompasses seven learned layers - five convolutions and two fully connected. We performed full training on the network using 4, 234 plant seedling images belonging to twelve plant species from Aarhus University Signal Processing group. The system is fine-tuned for the architecture to have greater processing time and low memory consumption. The architecture was evaluated using different network parameters. Furthermore, we used training loss function, accuracy, sensitivity, and specificity to evaluate the system performance. Experimental results proved that the developed architecture has reached excellent performance with overall accuracy of 90.15%. Results were achieved in 111 minutes and 36 seconds. Future work includes, first, use the model with greater amount of datasets through data augmentation and compare the results to other existing deep learning architectures using same datasets. Second, authors will consider CNN and RNN architectures together using several other plant datasets. Third, create a portable mobile application for plant seedling images classification utilizing the developed model.**

***Keywords*— plant classification, CNN, deep learning**

I. INTRODUCTION

Deep learning comprises numerous layers of nonlinear information processing. This allows learning architectures that implement functions as repeated compositions of simpler tasks, thereby producing learning layers of abstraction with better generalization and representation capacity. A widespread algorithm for deep learning is the CNN. CNNs are characterized as the innovative approach for image classification. Consequently, image recognition tasks have become prominent and essential for which convolutional neural networks have proven very effective[1]. In this milieu, deep CNNs have begun as leading substitutes to analyze immense data, presenting inventive results in image classification[2],[3]. Moreover, deep CNNs are at the core of most of the modern computer vision solutions extensive taxonomic tasks particularly for image classification [4],[1] also known for local connectivity and their weight sharing features.

Recently, we have witnessed tremendous research findings in image classification. The study of Krizhevsky, Sutskever, and Hinton[4], they develop a CNN architecture called AlexNet as a new landscape in computer vision tasks. Esteva et al. [5] showed in their study how the CNNs can outperform a human effort in the task of classifying a huge data sets. Conversely, in countless situations, a sufficient amount of annotated images (ground-truth) is not available. With this, other approaches are needed to improve the accuracy[2],[5]. Frizzi et al. [6] create their own CNN identical to that of the renowned LeNet-5 except that their works has incorporated data augmentation with intensified feature maps in convolution layers. Other researchers utilized actual scenarios to achieve higher precision than the output of conventional machine learning methods.

However, despite the attractive qualities of CNNs, it still has shortcomings. First, when there are too many layers or weights, computational complexity and space is higher [7],[8],[9]. Second, when there are more parameters, there is a bigger peril of overfitting which needs contention using regularization methods such as dropout, L1/L2, augmentation, etc. [10],[4],[11]. Third, when the network is deep, there is a problem of vanishing gradients when the error is back propagated over many layers[12].

Moreover, training a deep neural network is not that easy since it demands numerous input for a superior output [13]. It also requires graphical processing units (GPUs) to process huge data inputs and intensified outputs [9]. Further, the training process requires continual modifications of parameters to guarantee a balanced learning of all layers [10][4]. This observation motivates us to introduce a new approach which implements full training the network despite concerns that neural networks are more difficult to train.

There have been many approaches to optimize the network architecture [14],[15]. Our work is identical to the existing CNN architectures particularly the AlexNet. The main difference is that we build a deep neural network model that consists seven learned layers - five convolutions and two fully connected. We implemented full training and then analyze the ability of the architecture in the task of classifying plant seedling images into twelve species.

The remaining flow of this study is organized in the succeeding manner: In section 2, details of deep learning architectures is discussed. Section 3 underscores the methodology employed in this paper. Section 4 is dedicated to the experimental outcomes and deliberations. Lastly, inferences are drawn in fifth section.

II. Deep Learning Architectures and Transfer Learning

A. Deep Learning

Deep neural networks, usually called Deep Learning (DL), employ complex neural networks to routinely acquire a chain of structures from fundamental injunction regarding component structure. A common variant of the deep neural networks is the Deep CNN. A deep Convolutional Neural Networks (DCNNs) are exceptional varieties of multi-layer neural networks structured explicitly to distinguish visual patterns. Among the conventional methods, the deep NN are drawing remarkable attention owing to its spontaneous characteristics and image learning capability [3].

B. AlexNet

A number of deep learning designs are published, like AlexNet [4], VGG16 [16], GoogLeNet[8], and ResNet [17]. AlexNet done by the group of [4], was a leading pattern in ILSVRC 2012, and it is still fully taken advantage of by countless entities, thus, it is construed as the groundwork for the identification of certain objects. AlexNet is a deep CNN work that was used to win the 2012 ImageNet competition. It was the first of the many Deep CNN architectures to win the challenge, and set the stage for the explosion of research in the area. AlexNet subsuming eight layers – containing five convolutions and three fully allied layers as shown in Figure 1.

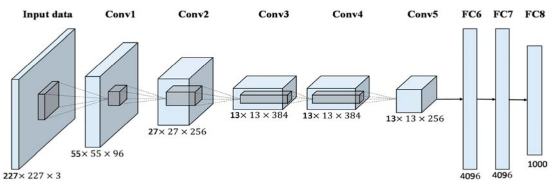

Figure 1. Pre-trained AlexNet Architecture

C. VGG16

Similarly, VGG16 created by Simonyan and Zisserman [18], involves up to 19 layers with pre-trained weights, very small receptive fields of 3 x 3, five max pooling of size 2 x 2, and deep structures of 16 layers and employed about three times more parameters than AlexNet. Just like the other deep learning architectures, VGG utilized dropout regularization in convolutions and a ReLU to all complex layers.

D. GoogLeNet

Another famous architecture for image classification is GoogLeNet. GoogLeNet [8] is another ImageNet winner that improved on AlexNet and VGG by using the inception modules. Inception modules used multiple filter sizes at each layer and concatenate the results together. This allows the network to learn features of different sizes at each layer in the network. Similar to other CNNs, GoogLeNet done by the group of [8], inspired by LeNet. Moreover, it utilizes regularization methods such as batch normalization, image distortions and etc. GoogLeNet architecture involved 22 deep layers.

E. ResNet

ResNet (Residual Network) is another novel CNN architecture utilizing skip connections and features heavy batch normalization [17]. ResNet become the most popular genre of deep learning model. It is deeper than VGG16, but still has a minor intricacy. ResNet won the ILSVRC taxonomy challenge.

F. Transfer Learning

Transfer learning is used to fine-tune the pre-trained models. It is utilized as the jumpstart for another model which will complete the subsequent task. This approach is prevalent in deep learning where pre-trained prototypes are utilized as the starting point for image-based visualization and natural language processing tasks given the massive configuration of resources required to come up with neural network prototypes.

Transfer learning is a good alternative to train a network. In such modality, a network meant to function for a specific task is re-tailored to carry out several related functions. It is understood in two aspects: either as a groundwork or hallmark design[19],[20]. The considerations in an amateur nexus are attuned to the necessary functions utilized as a threshold [20],[21]. Nevertheless, the features are hauled out from the input and then utilized to direct a cutting edge classifier when exploited as a feature generator[22].

III. Methods

A. Datasets

We used a public dataset of images of roughly 4,234 distinctive plants belonging to 12 plant species [23] to assess the performance of the architecture explored in this work. The 12 plant seedling species from Aarhus University Signal Processing group is shown in Table 1. These are the Blackgrass, Charlock, Cleavers, Common Chickweed, Common Wheat, Fat Hen, Loose Silky-bent, Maize, Scentless Mayweed, Shepherds Purse, Small-flowered Cranesbill and Sugar Beet.

TABLE I. Number of Instances and Classes Used For Experiment

Plant	Data
Blackgrass	263
Charlock	390
Cleavers	287
Common Chickweed	611
Common Wheat	211
Fat Hen	475
Loose Silky-bent	654
Maize	221
Scentless Mayweed	516
Shepherds Purse	231
Small-flowered Cranesbill	496
Sugar Beet	385
TOTAL	**4, 234**

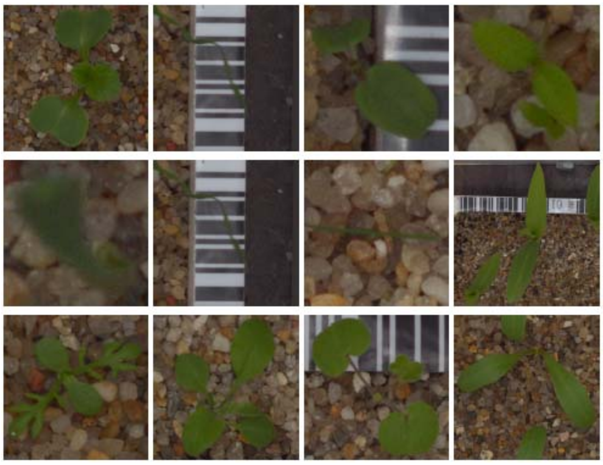

Figure 2: Example of plant seedling images, first row are Charlock, Black-grass, Cleavers, Common Chickweed, second row are Common Wheat, Fat Hen, Loose Silky-bent, Maize, third row are Scentless Mayweed, Shepherds Purse, Small-flowered Cranesbill and Sugar beet plant seedling images.

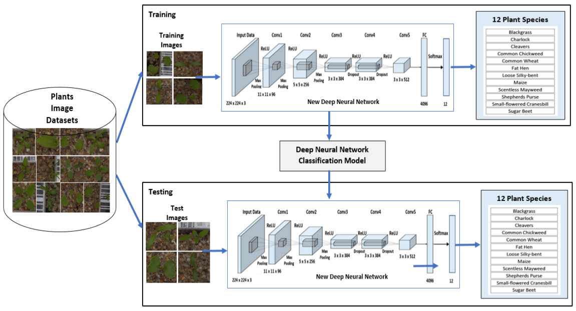

Figure 3. Flowchart of the implementation of Deep Neural Network for predicting plant seedling species.

Figure 3 shows the flow diagram of the deep neural network for predicting plant seedling classifications implementation. A dataset discussed in section 3.1 of pictures of roughly 4,234 distinctive plants at varying development phases [23], this is in collaboration with University of Southern Denmark, is used for training and testing. The dataset is divided randomly into 70% for training, 10% for authentication and 20% for testing.

B. *The Architecture*

The architecture of the network we used for training is depicted in Figure 4. Our model is identical to that of the AlexNet by Krizhevsky et al. [4] (see Figure 1), except that our model consists of only seven learned layers - five convolutional layers and followed by two fully connected (see Figure 4). Conversely, our model consider small datasets and trained on a GPU, compared to over 1.5 million characterized high resolution images in over 22, 000 categories in ILSVRC [4]. In this work, full training in our network containing relatively smaller number of weights, especially with smaller amount of data is employed to overcome severe overfitting. Additionally, data augmentation is not realized in both training and testing phases. Our goal is to experiment full training and fine-tuning making our arcthicture to consume low space requirement and faster processing time.

C. *Implementation*

Figure 4 depicts the design of the setup we employed for training. For coming up with the deeep neural network prototype, images are resized into 224 x 224 x 3 RGB images. It comprises seven learned layers – five convolutions and two fully connected. The prepared images differ in terms of size from a 54 x 54 to as much as 3991 x 2557, M x N sizes of images. Then the layers of the neural network system were defined.

The first layer has a convolutional size of (11×11) with stride of (4×4) pooling of (2×2) and with a depth of 96. To accelerate the process of training, each convolution is succeeded by a cross channel normalization of substance 5, afterwhich is followed by the rectified linear unit (ReLU). ReLU layer executes a basic function where any value less than zero is established at zero succeeded by a pooling window of size three and stride of (2 x 2). The second convolutional network layer has a convolutional size of (5×5) with stride of (1×1) with pooling of (2×2) and with a depth of 256. Another cross channel normalization of size 5 and a ReLU layer followed by a max pooling layer of size three and stride of (2x2).

The third and fourth network layers has a convolutional size of (3×3) with a stride of (1×1) with pooling of (1×1) and with a depth of 384. Followed by a ReLU layer again. The fifth network layer has a convolutional size of (3×3) with a stride of (1×1) with pooling of (1×1) and with a depth of 512. Again, followed by a ReLU layer and a max pooling layer of size of three with stride of (2 x 2).

After the five convolutional layers, the last two fully connected multiplies the input and add the bias vector. A size of 4096 for the first fully connected followed with a ReLU and a droput ratio of 50% to reduce test errors, followed with a softmax. In the last layer is the classification fully-linked layer of size 12.

Finally, for training options we used a stochastic gradient descent with 0.9 momentum including initial acquisition scale information of le-04 with L2 regularizaiton factor of le-04, and 3 mini-batch sizes. A specified positive integer of 40 which represents as the maximum number of epochs is used for training. To minimize the loss function using a mini-batch, iteration is carried out in the gradient descent algorithm. Then classification of the validation set and test set is done. Then assess the network models accuracy after classification.

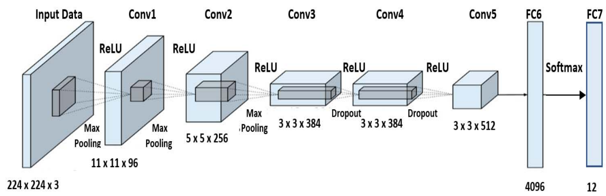

Figure 4. Deep Neural Network for predicting plant seedling species.

D. *Training the Classifier Model*

The training process follows the network architecture described in section 3.2. Before training, the dataset of images are split into 70% training, 20% validation and 10% testing. Then followed by a processing scheme including feature extraction, image resize and classification. The whole training took about 111 minutes and 36 seconds just for one training without augmentation.

To peruse the outcomes of our achitecture, we engaged through an evaluation process on the same data set and compare their the performance as shown in Figure 3. The 10% of images used for validation are done during training, the deep learning neural network classifier model is then used for validation. In testing the classifier model, 20% of the images were used for classification. Moreover, iterations reaches up to maximum of 44480 with 1112 iterations per epoch.

E. Performance Evaluation

To observe the effects of full training using the developed deep neural network that comprised seven learned layers, we conducted an evaluation using the equations described below. Equation (1) is used to compute for the accuracy while Eq. (2) and Eq. (3) is for sensitivity and specificity, respectively. The true positive (TP) is considered as the current plant seedling class category. The (FP) attributed as the false positive signifies the incorrecIty classified plant seedling category.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specifity = \frac{TN}{TN + FP} \quad (3)$$

IV. Experimental results and Discussion

We implemented a deep neural network classifier model from scratch. In this work, we trained on the image dataset of 4,234 plant seedlings belonging to 12 plant species [23]. Then we analyzed the capability performance of our model for the plant seedling classification. The classification accuracy per specie is reported in the confusion matrix in Figure 5. The twelve rows and twelve columns reports the number of correctly classified and misclassified plant seedlings during testing. It detailed the output for analysis of correct classifications.

Output Class \ Target Class	Black-grass	Charlock	Cleavers	Common Chickweed	Common wheat	Fat Hen	Loose Silky-bent	Maize	Scentless Mayweed	Shepherds Purse	Small-flowered Cranesbill	Sugar beet	
Black-grass	14	0	0	0	1	1	3	0	0	0	0	0	73.70%
Charlock	0	38	1	0	0	0	0	0	0	0	0	0	97.40%
Cleavers	0	0	26	0	0	0	0	0	0	0	1	0	96.30%
Common Chickweed	0	0	0	61	0	1	0	0	3	0	0	1	92.40%
Common wheat	0	1	1	0	19	0	2	0	0	0	0	1	79.20%
Fat Hen	0	0	0	0	0	46	0	0	0	1	0	2	93.90%
Loose Silky-bent	12	0	0	0	1	0	57	0	1	0	0	0	80.30%
Maize	0	0	0	0	0	0	0	20	0	0	0	0	100%
Scentless Mayweed	0	0	1	0	0	0	0	1	48	4	0	0	88.90%
Shepherds Purse	0	0	0	0	0	0	0	0	0	18	1	0	94.70%
Small-flowered Cranesbill	0	0	0	0	0	0	0	0	0	0	48	0	100%
Sugar beet	0	0	0	0	1	0	4	1	0	0	0	34	85.00%
	53.8%	97.4%	89.7%	100.0%	86.4%	95.8%	86.4%	90.9%	92.3%	78.3%	96.0%	89.5%	90.1%

Figure 5. Confusion Matrix for the Classification of 12 Plant Species using Deep Neural Network during Testing.

Figure 5 the illustration of the crafted method for the twelve considered classes. It can be observed that Maize and Small-flowered Cranesbill plant species achieve the highest accuracy rates of 100% of which both are agricultural weeds. While the Charlock also yields an excellent performance with an accuracy rate of 97.4%, Cleavers achieves an accuracy rate of 96.3%. It can also be noted that a number of classes achieve more than 90% accuracy except for Scentless Mayweed (a flower weed), Sugar Beet (a crop), Loose Silky-bent (a common wind grass), Common Wheat (a crop) and Black-grass (grass weed) that is below 90%. This result indicates that the classifier exhibits excellent performance of above 90%, hence, the model is capable of determining plant species from the given image dataset.

During the training process, the image classification and validation results achieved by the described architecture is depicted in Figure 6. From the figure, it can be noted that the overfitting problem is successfully resolved using our method, since the validation accuracy moves slowly in the earlier epochs and rises continuously in an ascending way.

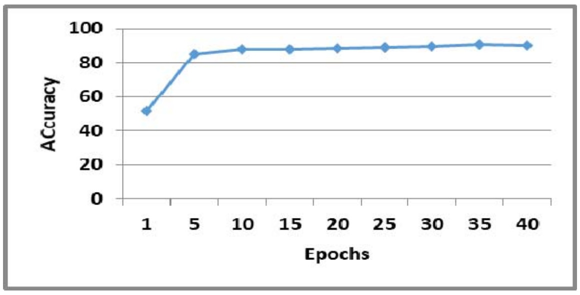

Figure 6. Classification Accuracy during Training using the Deep Neural Network architecture shown in Figure 4.

Furthermore, since accuracy improves significantly as the number of epochs increases, it seems more sufficient to use smaller epochs to train deep networks because our classifier performs better even without augmentation. The model exhibits an excellent classification performance of 90.15% at epoch 40 as shown in Figure 6. This results justifies that our approach is indeed comparable to current image classification methods. This is in part true that deep learning method is a very good prediction classification model for images.

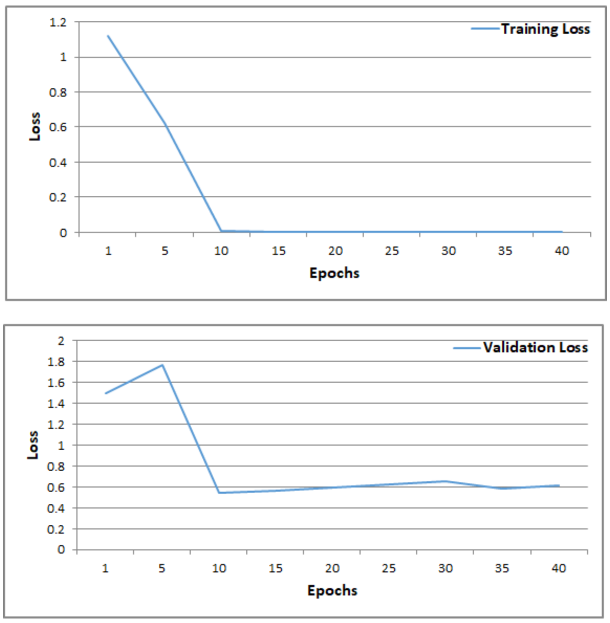

Figure 7. Training loss (top) and validation loss (bottom) during the training of the system.

As shown in Figure 7, the training loss and validation loss are close to each other. The loss value goes down from 1.1187 to 0.00000411. This indicates that the performance and robustness of the model are considerably observed.

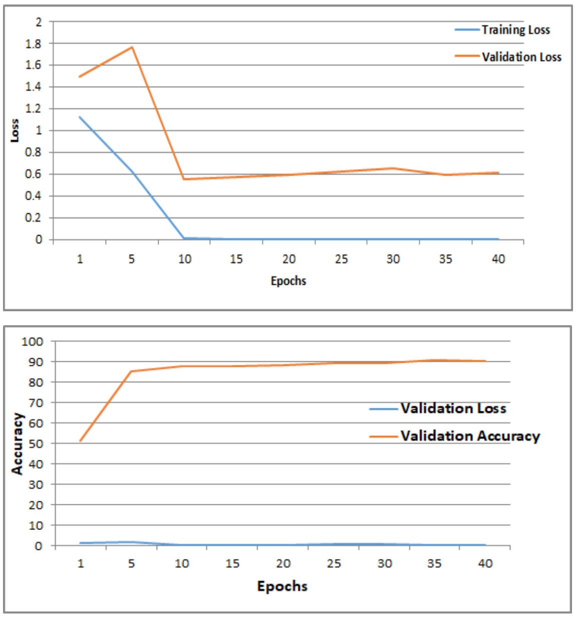

Figure 8. Results our model achieved. Training and validation loss (blue and orange lines) and validation accuracy (orange lines) during training with Deep Neural Network for Plant Seedling Classification.

Figure 8 presents the outcome during the development of the structure. The dipping blue and orange lines (top) correspond to the loss function values for the development and for the authentication sets during preparation. Then the orange line (bottom) represents the validation accuracy. The result shows that when the system acquires suitable development, it values the categories equitably secured and should also be noted that the best model is found.

Moreover, performing full training (training from scratch) in this problem, we see that our work has clearly resulted a high value of validation accuracy which is above 90% with reduced training time and space. To quantify the reliability and capability of our model, we compute the accuracy, sensitivity, and specificity for plant seedling images classification, as the standard measure for system performance, using the equation described in section 3.4. The overall accuracy is 90.15%, with sensitivity of 92.36%, and specificity of 91.67% respectively. Having a high value of sensitivity and specificity of the model leads to the fact that we have been able to reach the avant-garde performance. The overall effect of optimization performed in this work have significantly responded to the problems of time and space complexities.

V. Conclusion

This scientific pursuit aimed to develop a deep complex system capable of classifying plant seedling images. The networks were trained from scratch. Then we analyzed the capability of a fully-trained network in the task of classifying plant seedling images into twelve species. The model achieved an excellent performance with an accuracy rate of 90.15%. We found that removing one fully connected layer resulted by far best results on these datasets described in section 3.1. Additionally, we found out that number of iterations as underscored in this study is one of the factors affecting the system performance. On the other hand, amplifying the batch size also increases the accuracy of the outcome.

We conclude that if deep architecture needs to be fully trained, having a relatively reduced number of weights may perform better, specifically with a relatively small amount of data. In future works, authors will consider first, use the model with greater amount of datasets through data augmentation and compare the results to other existing deep learning architectures using same datasets. Second, authors will consider CNN and RNN architectures together using several other plant datasets. Third, create a portable mobile application for plant seedling images classification using the developed model.

References

[1] Y. Le and X. Yang, "Tiny ImageNet Visual Recognition Challenge," 2016.

[2] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Comput. Electron. Agric.*, no. April, pp. 0–1, 2018.

[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," 2015.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[5] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[6] S. Frizzi, R. Kaabi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional neural network for video fire and smoke detection," *IECON 2016 - 42nd Annu. Conf. IEEE Ind. Electron. Soc.*, pp. 877–882, 2016.

[7] M. Gheisari, G. Wang, and Z. A. Bhuiyan, "A Survey on Deep Learning in Big Data," 2017.

[8] C. Szegedy *et al.*, "Going Deeper with Convolutions," pp. 1–9, 2014.

[9] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[11] Douglas M. Hawkins, "The Poblem of Overfitting," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 1, pp. 1–12, 2004.

[12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015.

[13] B. Microbiana *et al.*, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[14] C. Cortes, X. Gonzalvo, V. Kuznetsov, M. Mohri, and S. Yang, "AdaNet: Adaptive Structural Learning of Artificial Neural Networks," 2016.

[15] B. Zoph and Q. V. Le, "Neural Architecture Search with Reinforcement Learning," pp. 1–16, 2016.

[16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," pp. 1–14, 2014.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Cvpr*, pp. 1–9, 2016.

[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv Prepr.*, pp. 1–10, 2014.

[19] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "{CNN} Features Off-the-Shelf: An Astounding Baseline for Recognition," *Deep. Work.*, 2014.

[20] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection," *Constr. Build. Mater.*, vol. 157, no. September, pp. 322–330, 2017.

[21] O. A. B. Penatti, K. Nogueira, and J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2015–Octob, pp. 44–51, 2015.

[22] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, "From generic to specific deep representations for visual recognition," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2015–Octob, pp. 36–45, 2015.

[23] T. M. Giselsson, R. N. Jørgensen, P. K. Jensen, M. Dyrmann, and H. S. Midtiby, "A Public Image Database for Benchmark of Plant Seedling Classification Algorithms," 2017.