

Classification of Cervical Cytology Overlapping Cell Images with Transfer Learning Architectures

Pallavi V. Mulmule* and Rajendra D. Kanphade

¹Department of E and TC, D. Y. Patil Institute of Technology, Pimpri, Pune, India.

²JSPM's Jayawantrao Sawant College of Engineering, Hadpasar, Pune, India.

*Corresponding Author E-mail: pvmulmule1@gmail.com

<https://dx.doi.org/10.13005/bpj/2364>

(Received: 11 September 2021; accepted: 10 January 2022)

Cervical cell classification is a clinical biomarker in cervical cancer screening at early stages. An accurate and early diagnosis plays a vital role in preventing the cervical cancer. Recently, transfer learning using deep convolutional neural networks; have been deployed in many biomedical applications. The proposed work aims at applying the cutting edge pre-trained networks: AlexNet, ImageNet and Places365, to cervix images to detect the cancer. These pre-trained networks are fine-tuned and retrained for cervical cancer augmented data with benchmark CERVIX93 dataset available publically. The models were evaluated on performance measures viz; accuracy, precision, sensitivity, specificity, F-Score, MCC and kappa score. The results reflect that the AlexNet model is best for cervical cancer prediction with 99.03% accuracy and 0.98 of kappa coefficient showing a perfect agreement. Finally, the significant success rate makes the AlexNet model a useful assistive tool for radiologist and clinicians to detect the cervical cancer from pap-smear cytology images.

Keywords: Alexnet; Cervical cancer; Cytology Images; Convolution Neural Network Models; Deep learning architectures; ImageNet.

In women, the leading cause of mortality is the cervical cancer (CC)¹. It is reported that more cases are found in moderate income countries². The mortality rate of cervical cancer in India is nearly 25%⁴. It is observed that majority of the cases are diagnosed at the severe stage⁵ which is the main cause of increasing death. To reduce the mortality rate it is necessary that the disease must be detected at early stage. In recent years, the artificial intelligence assisted applications using machine learning techniques became very popular in healthcare domain⁶⁻⁸. This helps the doctors in diagnosis as well as prognosis of the disease which will definitely improve the medical aid especially in rural areas where there is lack of expertise.

Recently in biomedical image processing deep convolutional neural network (DCNN) has shown remarkable results. DCNN models are becoming popular because of excellent performance in terms of classification accuracy⁹. With the introduction of deep learning, the classical approach of multi-class prediction or diagnosis with prior segmentation is becoming obsolete. Literature reported more focus of related work on pap smear images for CC detection on single cell with two class classification problem. Srishti et al.¹⁰ has worked on single cell with pap smear data and reported 90% overall patch-based CNN classifier accuracy. Zhang et al.¹¹ worked on single cell pap smear image and 98.3% accuracy for

binary classification with ConvNet. Nirmaljit et al.¹³ have proposed the DCNN architecture (i.e. Deep-Cerv) for binary classification of pap smear images and achieved 99.6% test accuracy. Thus, from the state of the art literature review, it is evident that, transfer learning architecture have not been introduced yet for automated classification of cervical cancer using the pap-smear cytological images. Researchers have either relied on private database or publically available benchmark Harlev database¹².

MATERIALS AND METHODS

The pre-trained transfer learning models are explained which are applied for the classification of the cervical cytology overlapping cell images. Different transfer learning models are examined to evaluate the most appropriate model suitable for cervix cancer detection problem. The method is divided into four sub-sections: database description, data training on Cervix93 database, data classification, and data evaluation. These sub-sections are described in details in below paragraphs.

Database Description

The Cervix93 cervical cytology image database is available publically with annotations¹⁴. The dataset contains 93 image stacks along with their correlated Extended Depth of Field (EDF) images. Each and every image in this database is of size 1280x960 pixels. The cytology cervix images are graded with Bethesda System. There are three grades as Negative, Low-grade Squamous Intraepithelial Lesions (LSIL), and High-grade Squamous Intraepithelial Lesions (HSIL). The details about frames per grade and nuclei per grade are shown in table 1.

Data Augmentation

The traditional practice for data augmentation is to transform color in an image i.e brightness, contrast, sharpening, white balance and blur. Also augmentation is done by modifying similar image attributes i.e rotation, flipping and histogram. Data Augmentation is done here with the mentioned transforms and the augmented images in the dataset are resized to 227 × 227 pixels. Figure 1 shows the augmentation output.

Data Training

Here three different deep convolution

pre-trained models are considered for cervical cytology CERVIX93 dataset image classification. The pre-trained models are trained using the transfer learning approach¹⁵. The pre-training consist of pre-processing, feature extraction, and mapping the existing model into completely new models. The pre-trained model is then fine-tuned by proper adjustment of hyper parameters of the model. Fine tuning is done by replacing last three layers viz; a fully-connected layer, a softmax layer, and a classification output layer. The motive behind using pre-trained transfer learning architectures is that, it is relatively fast and easy to train the network with random initial weights¹⁶. The other motive is that these pre-trained models have low training error than classical ANNs¹⁷. The performance of these transfer learning deep architectures have been evaluated for the cervical cancer detection problem. In the next subsection, these deep transfer learning architectures are described.

AlexNet

The AlexNet is a leading architecture with 8 deep layers, consisting of 5 convolution layers and 3 fully connected layers^{19, 20}. The first five layers are convolutional layers with weights. The output of 5th convolutional layer is fed to next two fully connected layers. The last fully-connected layer feeds to the output softmax classifier distributing into three class labels. Here, the overfitting in fully connected layers is reduced by 'dropout method'²¹. The dropout is process of turning off the hidden neurons having probability of 0.5²² at every iterations. ReLU is used for faster training of the model. The CNN with ReLU is 6 times faster than CNN with *tanh*

GoogleNet: Imagenet and Places365

GoogleNet²³ is 27 layers deep. It is also called as Inception v1 having 9 inception layers, 3 convolutional modules, 4 max-pooling layers, 3 average pooling layers, 5 fully-connected layers, and 3 softmax layers²⁴. The inception layer is combination of 1x1, 3x3, 5x 5 convolution layers and a max pooling layer. The output filter bank is concatenated into single output vector. This vector is input to the next inception module. For detailed explanation about the GoogleNet one can refer the original paper²³. The pre-training the GoogleNet network is done separately using the ImageNet²⁵ dataset as well as Places365 standard²⁶ data set. The network trained on ImageNet dataset contains more

than 14 million of images labeled with more than 5000 different classes. The Places365- standard dataset has around 1.8 million images of various scenes categorized into 365 scenes. Each scene category has at most 5000 images in each category.

CNN Settings

The general architecture of DCNN is shown in figure 1. For fair comparison between the networks, the hyper-parameters in all the experiments are kept same. The setting of the hyper-parameters is done as described in table 2. In DCNN, Stochastic Gradient Descent (SGD) is the widely used optimization algorithm, as it replaces actual gradient with the estimated one³⁰. In SGD, the model hyper-parameters are tuned such as initial learning rate. The learning rate is tuned since, the aim is to find local or global minima of loss function for faster weight adjustments. The momentum helps in accelerating the weight

adjustment in neurons. In all the DCNN’s the over fitting is reduced by using dropout mechanism by employing L2 Regularization by scaling the updated weight by a factor less than one³¹. Every experiment runs 15 epochs and each epoch is a training iteration with batch size of 64.

Classification and Performance Evaluation

Each output layer of the model will have different probability of the corresponding input cytology image. The network will consider the output with highest probability as its predicted class. The higher the value of prediction, the higher will be accuracy of that particular network. In this application there are three output layers as the cytology images are classified in three classes viz; N, LSIL and HSIL.

The performance evaluation of the pre-trained network models under consideration is done using seven performance indices³² namely

Table 1. Database details based on Bethesda System

	Negative	LSIL	HSIL	Total
Number of frame per grade	16	46	31	93
Number of nuclei per grade	238	1536	931	2705

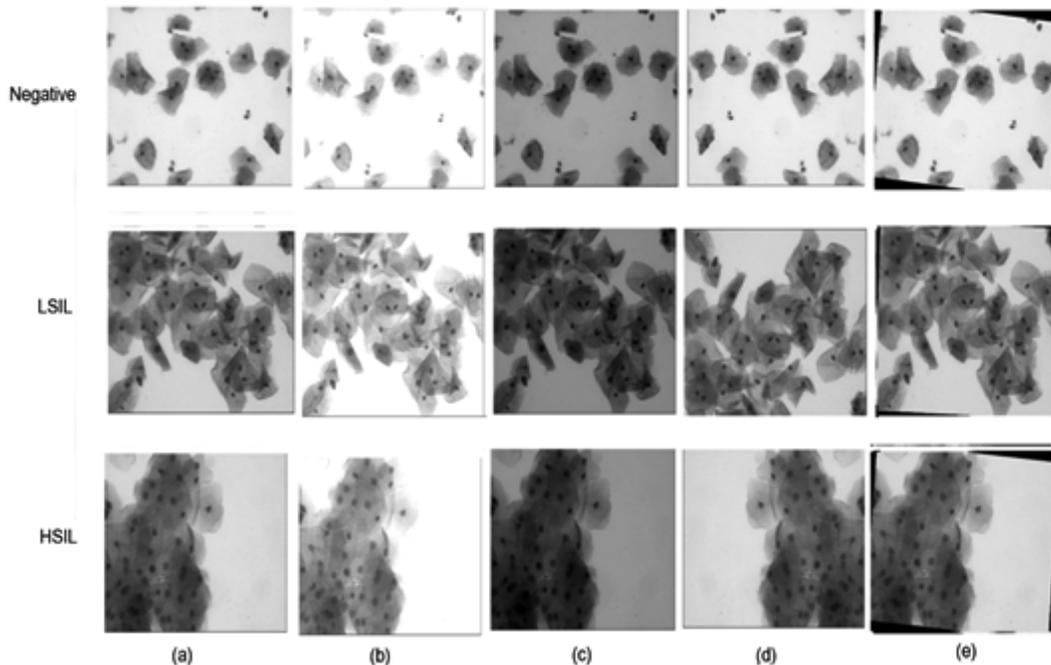


Fig. 1. Data Augmentation Output (a) input image (b)-(e) augmented images

viz; Accuracy (Acc), Sensitivity (S_s), Specificity (S_p), Precision (P_p), F-score, Matthews correlation coefficient (MCC) and Kappa Score.

RESULTS

The main objective of this work is to assess the transfer learning models for the classification

Table 2. Hyper-parameters of the experiments

Hyper-Parameters	Value
Optimization Algorithm	Adam and Sgdm
Momentum	0.9
Initial Learning Rate	0.01
L2 Regularization	0.0001
Epochs	15
Batch Size	64

of cervical cancer images. The assessment is done by comparing the network models based on seven quality measures as listed in previous section. Table 3 shows the results with Adam and Sgdm optimizer.

The performance analysis of all the network models with Adam and Sgdm optimizer is done with reference to Table 3. Starting with the Accuracy, AlexNet outperforms in terms of accuracy with average value of 99.03 for Adam optimizer. The lowest accuracy of 91.21% is reported in Places365 for Sgdm whereas, precision of Alexnet is best having average value of 98.97% followed by ImageNet with 98.15%. The least sensitivity of 88.88% is achieved in Places366 for Sgdm while maximum of 97.78% is achieved with ImageNet with Adam optimizer. In the similar way, the highest specificity of 99.06% is achieved in AlexNet, while lowest 93.57% is

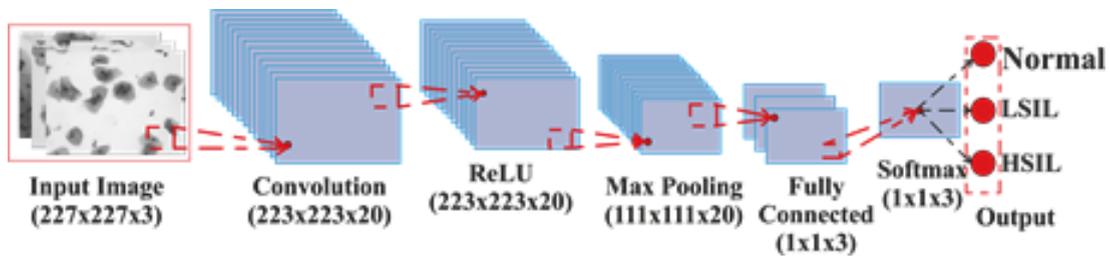


Fig. 2. Representation of deep convolution neural network

Table 3. Performance of Transfer Learning Models with Adam and Sgdm optimizer

Architecture	Optimizer	Class	Sen	Spc	Pre	F1 Score	MCC	Acc.	Avg. Acc.	Kappa Score
Alexnet	SGDM	H	97.16	99.63	99.28	98.21	97.3	98.78	98.78	0.97
		L	100	96.59	96.71	98.33	96.65	98.3		
		N	93.75	100	100	96.77	96.27	99.03		
	ADAM	H	97.87	99.63	99.28	98.57	97.84	99.03	99.03	0.98
		L	100	97.56	97.63	98.8	97.6	98.78		
		N	95.31	100	100	97.6	97.21	99.27		
Imagenet	SGDM	H	99.19	98.79	97.62	98.4	97.6	98.92	98.38	0.97
		L	97.83	97.87	97.83	97.83	95.7	97.85		
		N	95.31	99.68	98.39	96.83	96.2	98.92		
	ADAM	H	99.19	97.98	96.09	97.62	96.43	98.39	98.56	0.97
		L	97.28	98.4	98.35	97.81	95.7	97.85		
		N	96.88	100	100	98.41	98.11	99.46		
Places365	SGDM	H	99.19	85.89	77.85	87.23	81.14	90.32	91.21	0.89
		L	77.17	98.4	97.93	86.32	77.48	87.9		
		N	90.63	96.43	84.06	87.22	84.53	95.43		
	ADAM	H	98.39	96.37	93.13	95.69	93.52	97.04	97.84	0.957
		L	95.11	99.47	99.43	97.22	94.71	97.31		
		N	98.44	99.35	96.92	97.67	97.19	99.19		

achieved with Places365. As far as F-score metric is concern, the highest score of 98.12% is obtained in AlexNet and least score of 86.92% is obtained in places365. Finally, maximum MCC of 97% is obtained in AlexNet and minimum is obtained in 81% in Places365. The highest Kappa coefficient is achieved in Alexnet. Here, it can be concluded that the AlexNet implementation achieved the highest percentage in all the performance indices followed by ImageNet architecture. Places365 gives least results in most the performance metrics.

The image wise classification of each network model for grading as per Bethesda system is shown in table 4. The confusion matrix table is obtained for both Agdm and Sgdm optimizer. The rows in table give the output class as per Bethesda system, while the column indicates the true class with two different optimization. The diagonal cell indicates the correct class, and the off-diagonal cell indicates the mis- classification.

DISCUSSION

Classification of cervical cancer of digital pap smear images using transfer learning approach into three grades based on Bethesda grading system is focused here. For early detection of the cancer, the more robust and speedy network model is essential. The work is based on Cervix93 database consisting of 1536 images of LSIL, 931 images of HSIL, and 238 images of healthy class. The total dataset was divided into 80% training and 20% testing data. The transfer learning pre-trained models were fine-tuned and the performance was evaluated by seven performance indicators. Based on results in Table 3, it can be concluded that Alexnet is the best transfer learning architecture than the other architectures. It can also be concluded that even though GoogleNet: places365 is one of the deepest CNN architecture, it gives low performance.

The computational cost involved in training the network model is around two hours

Table 4. Confusion matrix for all CNN Models with Adam and Sgdm optimizer

Architectures	Bethesda Grade	Sgdm Optimizer			Adam Optimizer		
		Normal	HSIL	LSIL	Normal	HSIL	LSIL
AlexNet	Normal	235	1	2	236	0	1
	HSIL	7	1516	13	4	1521	11
	LSIL	5	7	919	5	4	922
ImageNet	Normal	235	2	1	235	2	1
	HSIL	8	1514	14	7	1514	12
	LSIL	8	5	918	6	7	918
Places365	Normal	217	9	12	233	1	2
	HSIL	51	1401	84	14	1503	17
	LSIL	36	46	849	9	11	911

Table 5. Comparison with the state of the art

Author	Dataset	Methodology	Accuracy
[9]	Private database	DCNN	93.33
[33]	Private and Herlev Dataset	Inception-V3, Xception, VGG-16, Resnet-50	98.6
[34]	Private Dataset	PsiNet-TAP	98.00
[35]	Private and Herlev Dataset	AlexNet, Vgg-16 and 19, ResNet-50 and 101, GoogLeNet	90.00
[36]	Private Dataset	Mask Regional Convolution Neural Network (Mask R-CNN)	91.70
[37]	Cervix 93 Dataset	CNN	89.50
Proposed Model	Cervix 93 Dataset	AlexNet	99.03

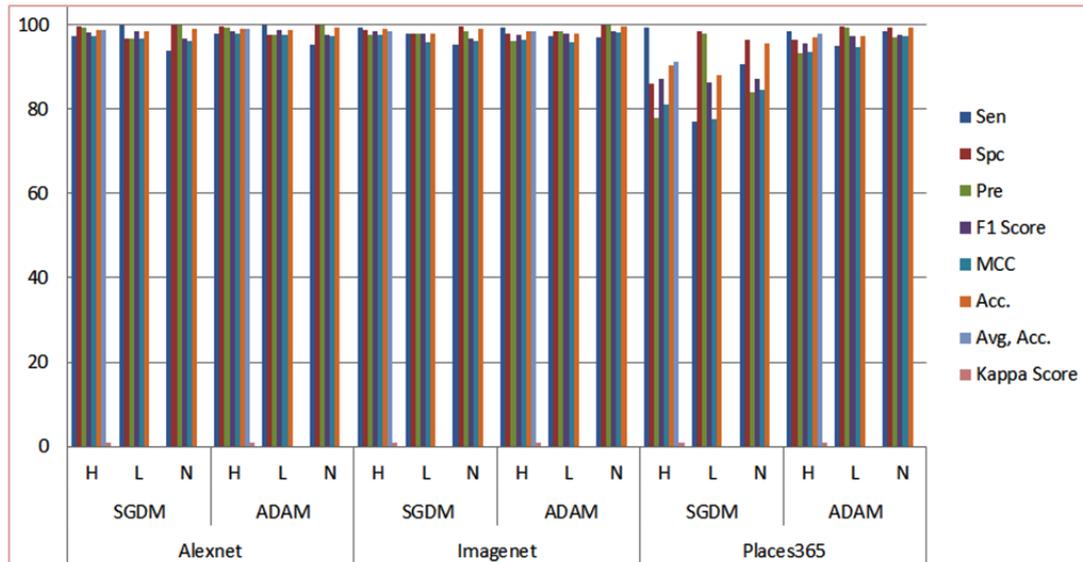


Fig. 3. Performance measure for every pre-trained model for Adam and Sgdm Optimizer

on high performance computer with Central processing unit. Therefore, it is suggested to use Graphics Processing Unit (GPU) for faster training.

The state of the art comparison of existing reported literature is summarized in Table 5. The literature compared here is limited to CNN architecture for multi-cell and overlapped cervical cell cytological images. The proposed method is also compared at the last row in the table for better clarity. It can be concluded from the state of the art comparison that, AlexNet gives best accuracy so far (i.e 99.03%) in overlapping cervical cells. Therefore, it is highly recommended to use AlexNet for cervical cytology image classification.

CONCLUSION

The proposed work employs transfer learning method for comparison and evaluation of various deep transfer learning network architectures for the classification of cervical cytology digital images. The work aimed at applying the pre-trained network models and compares the performances of AlexNet, Imagenet and Places365 with different performance indices. Here, each of the networks has correctly predicted the class label as Normal, HSIL and LSIL cervical cancer from the digital cytological images. It has been found that, AlexNet is able to predict with almost 99.03% accuracy

in all the performance metrics in 15 epochs with batch size of 64. It is also found that, Places365 has least performed as compared to the other networks, representing the poor network inspite of wide and deep layers. It is expected that the proposed AlexNet networks will make an important contribution to the biomedical domain and will be ground work for point of care technology solution.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. W. Small Jr, M. A. Bacon, A. Bajaj, L. T. Chuang, B. J. Fisher, M. M. Harkenrider, A. Jhingran, H. C. Kitchener, L. R. Mileskin, A. N. Viswanathan *et al.*, "Cervical cancer: a global health crisis," *Cancer*, **123**(13): 2404–2412 (2017).
2. G. Jassim, A. Obeid, and H. A. Al Nasheet, "Knowledge, attitudes, and practices regarding cervical cancer and screening among women visiting primary health care centres in bahrain," *BMC Public Health*, **18**(1): 1–6 (2018).
3. Z. Alyafeai and L. Ghouti, "A fully-automated deep learning pipeline for cervical cancer classification," *Expert Systems with Applications*, **141**: 112951 (2020).
4. R. Siegel, J. Ma, Z. Zou, and A. Jemal, "Cancer statistics, 2014." *CA: a cancer journal for clinicians*, **64**(1): 9–29 (2014).

3. I. Mittra, G. A. Mishra, S. Singh, S. Aranke, P. Notani, R. Badwe, A. B. Miller, E. E. Daniel, S. Gupta, P. Uplap *et al.*, "A cluster randomized, controlled trial of breast and cervix cancer screening in mum- bai, india: methodology and interim results after three rounds of screening," *International journal of cancer*, **126**(4): 976–984 (2010).
6. M. Sharma, S. K. Singh, P. Agrawal, and V. Madaan, "Classification of clinical dataset of cervical cancer using knn," *Indian Journal of Science and Technology*, **9**(28): 1–5 (2016).
7. J. Su, X. Xu, Y. He, and J. Song, "Automatic detection of cervical cancer cells by a two-level cascade classification system," *Analytical Cellular Pathology*, **2016** (2016).
8. M. Arya, N. Mittal, and G. Singh, "Clustering techniques on pap smear images for the detection of cervical cancer," *J Biol Today's World*, **7**(1): 30–35 (2018).
9. M. Wu, C. Yan, H. Liu, Q. Liu, and Y. Yin, "Automatic classification of cervical cancer from cytological images by using convolutional neural network," *Bioscience reports*, **38**(6): (2018).
10. S. Gautam, A. Bhavsar, A. K. Sao, and K. Harinarayan, "Cnn based segmentation of nuclei in papsmear images with selective pre-processing," in *Medical Imaging 2018: Digital Pathology*, vol. 10581. International Society for Optics and Photonics, 2018, p. 105810X.
11. L. Zhang, L. Lu, I. Noguez, R. M. Summers, S. Liu, and J. Yao, "Deeppap: deep convolutional net works for cervical cell classification," *IEEE journal of biomedical and health informatics*, **21**(6): 1633–1643 (2017).
12. J. Jantzen and G. Dounias, "Pap smear dtu/herlev databases," <http://mde-lab.aegean.gr/index.php/downloads>, Accessed on 21 December 2019.
13. O. N. Jith, K. Harinarayanan, S. Gautam, A. Bhavsar, and A. K. Sao, "Deepcerv: Deep neural network for segmentation free robust cervical cell classification," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Springer, 2018, pp. 86–94.
14. H. A. Phoulady and P. R. Mouton. Cervix93 cytology dataset. [Online]. Available: <https://github.com/parham-ap/cytology-dataset>
15. S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, **22**(10): 1345–1359 (2009).
16. E. Deniz, A. S. eng u'r, Z. Kadirog'lu, Y. Guo, V. Bajaj, and U' . Budak, "Transfer learning based histopathologic image classification for breast cancer detection," *Health information science and sys- tems*, **6**(1): 1–7 (2018).
17. G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on audio, speech, and language processing*, **20**(1): 30–42 (2011).
18. K. Simonyan and A. Zisserman, "Vgg-16," *arXiv Prepr.*, (2014).
19. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, **25**: 1097–1105 (2012).
20. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of com- puter vision*, **115**(3): 211–252 (2015).
21. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co- adaptation of feature detectors," *arXiv preprint arXiv:1207.0580* (2012).
22. K. Zhang, Q. Wu, A. Liu, and X. Meng, "Can deep learning identify tomato leaf disease?" *Advances in Multimedia*, **2018** (2018).
23. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabi- novich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
24. M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, **235**: 228–235 (2017).
25. P. U. Stanford Vision Lab, Stanford University. Imagenet. [Online]. Available: <https://www.image-net.org/>
26. B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva, "Places: An image database for deep scene understanding," *arXiv preprint arXiv:1610.02055*, 2016.
27. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
28. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
29. E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, **161**: 272–279 (2019).

30. B. Kleinberg, Y. Li, and Y. Yuan, "An alternative view: When does SGD escape local minima?" in *International Conference on Machine Learning*. PMLR, 2698–2707 (2018).
31. T. Laarhoven, "L2 regularization versus batch and weight normalization," *arXiv preprint arXiv:1706.05350*, 2017.
32. N. Japkowicz and M. Shah, *Evaluating learning algorithms: a classification perspective*. Cambridge University Press, 2011.
33. D. Xue, X. Zhou, C. Li, Y. Yao, M. M. Rahaman, J. Zhang, H. Chen, J. Zhang, S. Qi, and H. Sun, "An application of transfer learning and ensemble learning techniques for cervical histopathology image classification," *IEEE Access*, **8**: 104 603–104 618 (2020).
34. P. Wang, J. Wang, Y. Li, L. Li, and H. Zhang, "Adaptive pruning of transfer learned deep convolutional neural network for classification of cervical pap smear images," *IEEE Access*, **8**: 50 674– 50 683 (2020).
35. E. Hussain, L. B. Mahanta, C. R. Das, and R. K. Talukdar, "A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network," *Tissue and Cell*, **65**: 101347 (2020).
36. N. Sompawong, J. Mopan, P. Pooprasert, W. Himakhun, K. Suwannarurk, J. Ngamvirojcharoen T. Vachiramon, and C. Tantibundhit, "Automated pap smear cervical cancer screening using deep learning," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 7044–7048.
37. H. Ahmady Phoulady and P. R. Mouton, "A new cervical cytology dataset for nucleus detection and image classification (cervix93) and methods for cervical nucleus detection," *arXiv e-prints*, pp. arXiv– 1811 (2018).