

Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection

Nour Eldeen M. Khalifa¹,
Mohamed Loey², Mohamed
Hamed N. Taha¹, Hamed Nasr
Eldin T. Mohamed³

¹Information Technology Department,
Faculty of Computers and Artificial
Intelligence Cairo University, Giza, Egypt

²Computer Science Department, Faculty of
Computers and Informatics
Benha University, Qalubia, Egypt

³Ophthalmology Department, Faculty of
Medicine, Al-Azhar University, Cairo, Egypt

Corresponding author: Mohamed Hamed
N. Taha, Lecturer, Information Technology
Department, Faculty of Computers and Artificial
Intelligence, Cairo University, Giza, Egypt.
E-mail: mnasrtaha@cu.edu.eg, ORCID ID: [http://
www.orcid.org/0000-0003-0200-2918](http://www.orcid.org/0000-0003-0200-2918).

doi: 10.5455/aim.2019.27.327-332

ACTA INFORM MED. 2019 DEC 27(5): 327-332

Received: Oct 15, 2019 • Accepted: Dec 22, 2019

© 2019 Nour Eldeen M. Khalifa, Mohamed Loey,
Mohamed Hamed N. Taha, Hamed Nasr Eldin T.
Mohamed

This is an Open Access article distributed under the
terms of the Creative Commons Attribution Non-
Commercial License ([http://creativecommons.org/
licenses/by-nc/4.0/](http://creativecommons.org/licenses/by-nc/4.0/)) which permits unrestricted non-
commercial use, distribution, and reproduction in any
medium, provided the original work is properly cited.

ABSTRACT

Introduction: Diabetic retinopathy (DR) is the most common diabetic eye disease worldwide and a leading cause of blindness. The number of diabetic patients will increase to 552 million by 2034, as per the International Diabetes Federation (IDF). **Aim:** With advances in computer science techniques, such as artificial intelligence (AI) and deep learning (DL), opportunities for the detection of DR at the early stages have increased. This increase means that the chances of recovery will increase and the possibility of vision loss in patients will be reduced in the future. **Methods:** In this paper, deep transfer learning models for medical DR detection were investigated. The DL models were trained and tested over the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 dataset. According to literature surveys, this research is considered one of the first studies to use of the APTOS 2019 dataset, as it was freshly published in the second quarter of 2019. The selected deep transfer models in this research were AlexNet, Res-Net18, SqueezeNet, GoogleNet, VGG16, and VGG19. These models were selected, as they consist of a small number of layers when compared to larger models, such as DenseNet and InceptionResNet. Data augmentation techniques were used to render the models more robust and to overcome the overfitting problem. **Results:** The testing accuracy and performance metrics, such as the precision, recall, and F1 score, were calculated to prove the robustness of the selected models. The AlexNet model achieved the highest testing accuracy at 97.9%. In addition, the achieved performance metrics strengthened our achieved results. Moreover, AlexNet has a minimum number of layers, which decreases the training time and the computational complexity.

Keywords: Diabetic Retinopathy, Deep Transfer Learning, Convolutional Neural Network, Machine Learning.

1. INTRODUCTION

Diabetic retinopathy (DR) is a diabetic disease of the eye in which the retinal blood vessels of people's eyes are damaged because of long-standing Diabetes mellitus (1, 2). As estimated by the International Diabetes Federation (IDF), the number of diabetic patients will increase to 552 million by 2035 (3). In Egypt, more than 6 million people (7.2% of the popu-

lation) suffer from diabetes and are at risk of losing their vision because of DR (4). The early stage of DR is very important and can be controlled or treated if DR is detected and diagnosed on time (5). The best diagnostic options for DR patients differ between stages. Only regular screening is required for patients with no DR or mild non-proliferative DR (NPDR). However, the diagnostic options

vary from laser light scattering diagnostics to vitrectomy for patients with moderate nonproliferative DR or worse (5).

2. AIM

Clinically, the treatment of DR is often made with fundus images, acquired by fundus photographs (6) with roughly 200 million cases worldwide and more than 400,000 deaths per year. Besides biomedical research and political efforts, modern information technology is playing a key role in many attempts at fighting the disease. One of the barriers toward a successful mortality reduction has been inadequate malaria diagnosis in particular. To improve diagnosis, image analysis software and machine learning methods have been used to quantify parasitemia in microscopic blood slides. This article gives an overview of these techniques and discusses the current developments in image analysis and machine learning for microscopic malaria diagnosis. We organize the different approaches published in the literature according to the techniques used for imaging, image preprocessing, parasite detection and cell segmentation, feature computation, and automatic cell classification. Readers will find the different techniques listed in tables, with the relevant articles cited next to them, for both thin and thick blood smear images. We also discussed the latest developments in sections devoted to deep learning and smartphone technology for future malaria diagnosis. (6 Digital image processing (7), computer vision techniques and high computing facilities have been increasing in importance, especially with respect to medical science. A computer-aided diagnosis (CAD) system can reduce the load on ophthalmologists and can contribute more to the field of ophthalmology (8). In papers (9, 10), surveys of DR CAD systems that increase the accuracy and efficiency of diagnosis of DR were described.

3. METHODS

Deep learning

Alternatively, handcraft image recognition techniques have provided reasonable results and accuracy regarding medical DR image disease detection using both uninfected and infected DR images. Deep learning (DL) is a subfield of artificial intelligence (AI) concerned with methods inspired by the neurons of the brain (11) expertly designed constraints and model parameters, and have limited detection performance due to their

maintenance costs. In recent years, deep learning has become a focus in different research fields, because methods based on deep learning are able to directly learn features from training data. In this study, we apply two different convolutional neural network structures to walnut images for automatically segmenting images and detecting different-sized natural foreign objects (e.g., flesh leaf debris, dried leaf debris and gravel dust. Table 1 shows a series of major contributions in the field of DL (12). Today, DL is considered a promising technique in image/video classification and detection. DL depends on algorithms for thinking process simulations and data processing, or for developing abstractions (13) genomics largely utilizes machine learning to capture dependencies in data and derive novel biological hypotheses. However, the ability to extract new insights from the exponentially increasing volume of genomics data requires more expressive machine learning models. By effectively leveraging large data sets, deep learning has transformed fields such as computer vision and natural language processing. Now, it is becoming the method of choice for many genomics modelling tasks, including predicting the impact of genetic variation on gene regulatory mechanisms such as DNA accessibility and splicing. (13. The hidden layers map the inputs to outputs to analyse hidden patterns in complicated data (14) deep artificial neural networks (ANNs). DL improves these CAD systems to realize higher outcomes, widen the scope of diseases, and implement applicable real-time medical image (15, 16) disease detection systems.

Convolution neural networks

Convolutional neural networks (CNNs) have achieved phenomenal success for image/video classification and detection. In 2012, papers in (17, 18) have shown how CNNs based on graphics processing units (GPUs) can improve many vision benchmark records with respect to Chinese characters (19), the NYU object recognition benchmark (NORB) (jittered, cluttered) (20), Modified National Institute of Standards and Technology (MNIST) data (21), Latin letters, traffic signs (22), and large-scale ImageNet (23) benchmarks. In the following years, various advances in deep CNNs further increased the accuracy rate with respect to the image detection/classification competition tasks. CNN pre-trained models have introduced significant improvements in the succeeding annual challenges, including ImageNet Large-scale

Visual Recognition Competition (ILSVRC). Many pre-trained models were introduced, such as VGG-16, VGG-19 (24), GoogleNet (25), ResNet (26), Xception (27), Inception-V3 (28) and DenseNet (29) more accurate, and efficient to train if they contain shorter connections between layers close to the input and those close to the output. In this paper, we embrace this observation and introduce the Dense Convolutional Network (DenseNet).

The remainder of the paper is organized as follows. Section 2 explores related work and determines the scope of this study. Section 3 discusses the dataset used in our research. Section 4 presents the proposed models, and section 5 illustrates the achieved results and presents a discussion. Finally, section 6 provides conclusions and directions for further research.

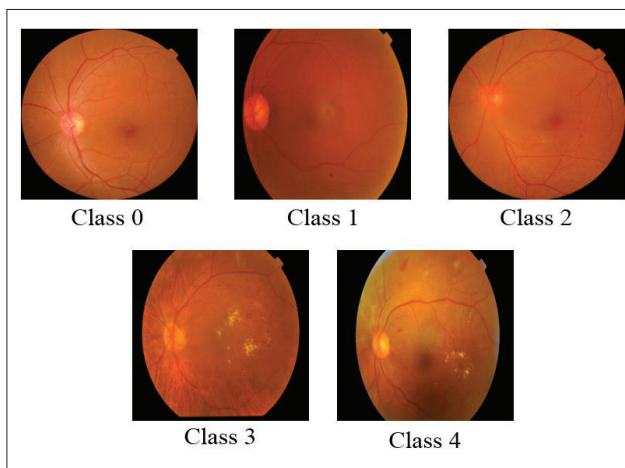


Figure 1. Sample image for each class in the APTOS 2019 dataset

Dataset

The APTOS 2019 blindness detection dataset (30) was selected in this research to be experimented with by the different deep transfer learning models. This dataset was released in the second quarter of the 2019 by the Asia Pacific Tele-Ophthalmology Society. The dataset contains 3662 images and consists of 5 classes as follows:

0 - **No diabetic retinopathy.**

1 - **Mild nonproliferative retinopathy:** Small areas of balloon-like swelling in the retina’s tiny blood vessels, called microaneurysms, occur at this earliest stage of the disease. These microaneurysms may leak fluid into the retina.

2 - **Moderate nonproliferative retinopathy:** As the disease progresses, blood vessels that nourish the retina may swell and distort and may also lose their ability to transport blood. Both conditions cause characteristic changes to the appearance of the retina and may contribute to diabetic macular edema (DME).

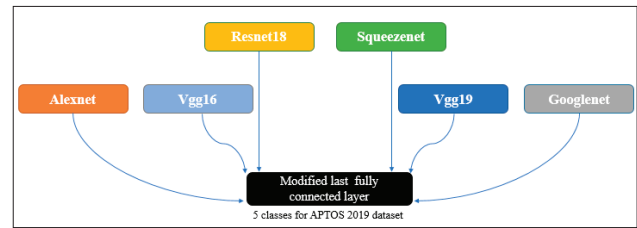


Figure 2. Proposed model’s customization for medical diabetic retinopathy detection

4 - **Severe nonproliferative retinopathy:** Many more blood vessels are blocked, depriving blood supply to areas of the retina. These areas secrete growth factors that signal the retina to grow new blood vessels.

5 - **Proliferative diabetic retinopathy (PDR):** At this advanced stage, growth factors secreted by the retina trigger the proliferation of new blood vessels, which grow along the inside surface of the retina and into the vitreous gel, the fluid that fills the eye. The new blood vessels are fragile, which makes them more likely to leak and bleed. The accompanying scar tissue can contract and cause retinal detachment—the pulling away of the retina from underlying tissue, such as wallpaper peeling away from a wall. Retinal detachment can lead to permanent vision loss.

The number of images for each class is presented in Table 1 and a sample of each class is illustrated in Figure 1.

Class Number	0	1	2	3	4
Number of Images	1805	370	999	193	295

Table 1. Number of images for each class in the APTOS 2019 dataset

Proposed models

The proposed models used in this research relied on the deep transfer learning CNN architectures to transfer the learning weights to reduce the training time, mathematical calculations and the consumption of the available hardware resources. There are a number of studies in (31-33) that have attempted to build their own architecture, but those architectures are problem specific and do not fit the data presented in this paper. The deep transfer learning CNN models investigated in this research are AlexNet (18), ResNet18 (26), SqueezeNet (34) is typically possible to identify multiple DNN architectures that achieve that accuracy level. With equivalent accuracy, smaller DNN architectures offer at least three advantages: (1, GoogleNet (35), VGG16 (36), and VGG19 (36). The mentioned CNN models have only a few layers when compared to large CNN models, such as Xception, DenseNet, and InceptionResNet, which

Accuracy/Model	AlexNet	VGG16	ResNet18	SqueezeNet	VGG19	GoogleNet
Class 0	99.7%	99.8%	99.5%	97.8%	99.6%	99.7%
Class 1	98.0%	96.3%	90.7%	80.0%	98.6%	96.8%
Class 2	96.6%	98.1%	97.3%	87.5%	97.6%	91.4%
Class 3	91.3%	89.1%	91.4%	67.8%	88.8%	92.3%
Class 4	95.8%	92.7%	89.8%	80.9%	88.7%	94.4%
Total Accuracy	97.9%	97.8%	96.8%	90.3%	97.4%	96.3%

Table 3. Classes and total testing accuracy for the different CNN models



Figure 3. (a) AlexNet and (b) VGG16 confusion matrices

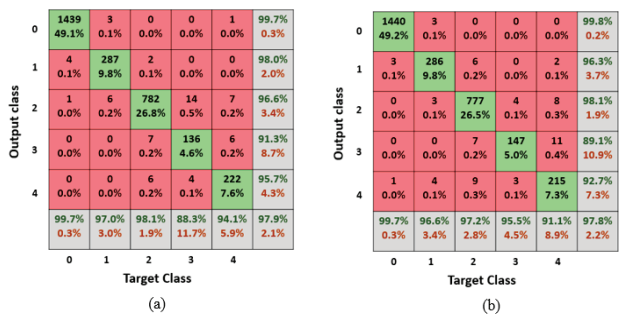


Figure 4. (a) ResNet18 and (b) SqueezeNet confusion matrices

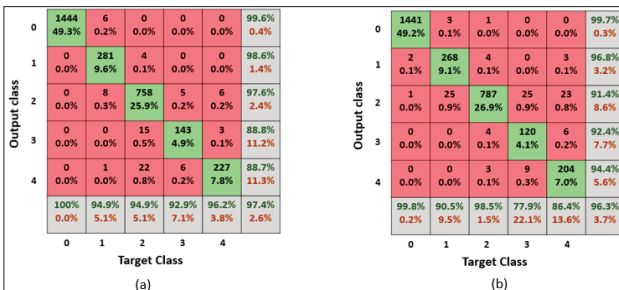


Figure 5. (a) VGG19, and (b) GoogleNet confusion matrices consist of 71, 201 and 164 layers, respectively. The choice of these models reduce the training time and the complexity of the calculations. Table 2 presents the number of layers of the CNN models used in this work.

Model	AlexNet	VGG16	ResNet 18	Squeeze Net	VGG19	Google Net
Number of Layers	8	16	18	18	19	22

Table 2. Number of layers for the different CNN models

The previous CNN models were customized in the last fully connected layer to match the number of classes of the APTOS 2019 dataset, which contains 5 classes, as illustrated in Figure 2.

Data Augmentation Techniques

The most well-known technique to overcome

overfitting is to increase the number of images used for training by applying label-preserving transformations (37). In addition, data augmentation schemes are applied to the training set to render the resulting model more invariant for any kind of transformation and noise. The augmentation techniques used in this research are

- Reflection around the X-axis.
- Reflection around the Y axis.
- Reflection around the X-Y axis.

The adopted augmentation techniques have increased the number of images by a factor of 4 times compared with the original dataset. The dataset increased to 14648 images used in the training and testing phases. This increase will lead to a significant improvement in the CNN testing accuracy, as will be discussed in the following section. Additionally, this approach will render the proposed methods immune to memorizing the data and more robust and accountable for the testing phase.

4. RESULTS

The proposed architecture was developed using a software package (MATLAB). The implementation was central processing unit (CPU) specific. All experiments were performed on a computer server with an Intel Xeon E5-2620 processor (2 GHz), 96 GB of RAM.

Testing accuracy metric

Testing accuracy is an estimation that demonstrates the precision and accuracy of any of the proposed models. Additionally, the confusion matrix is an accurate measurement that provides more insight regarding the achieved testing accuracy. Figures 3, 4, and 5 presents the confusion matrices for the distinctive CNN models used in this research.

Table 3 presents the class and total testing accuracy for different CNN models. The table illustrates that AlexNet, VGG16, and VGG19 achieved a very close total testing accuracy to each other at 97.9%, 97.7%, and 97.4%, respectively. The three models also achieved top-class testing detection accuracy, except for class 3, which achieved the

best accuracy using GoogleNet. According to total testing accuracy, the AlexNet model achieved the highest accuracy at 97.9%. Additionally, AlexNet had the least number of layers with only 8 layers and, as the result, had fewer calculations and less complexity.

Performance evaluation and discussion

To evaluate the performance of the proposed models, more performance matrices need to be investigated through this research. The most common performance measures in the field of DL are precision, recall, and F1 score (38), which are presented from equation (1) to equation (3), respectively.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (1)$$

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

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (3)$$

where TP is the count of true positive samples, TN is the count of true negative samples, FP is the count of false positive samples, and FN is the count of false negative samples from a confusion matrix.

Table 4 presents the performance metrics for the different proposed CNN models. The table illustrates that the AlexNet model achieved the highest percentage for the precision and recall metrics, while VGG16 achieved the highest percentage for the recall metric.

Metric/Model	Alex Net	VGG16	Res Net 18	Squeeze Net	VGG19	Google Net
Precision	96.23%	95.19%	93.75%	82.80%	94.64%	94.92%
Recall	95.42%	96.02%	94.57%	82.16%	95.76%	90.63%
F1 Score	95.82%	95.60%	94.16%	82.48%	95.20%	92.73%

Table 4. Performance metrics for the different CNN models

According to the achieved results for both overall testing accuracy and the performance metrics, AlexNet is the most appropriate CNN model for the APTOS 2019 dataset for medical DR detection with a testing accuracy of 97.9%. Moreover, VGG16 and VGG19 also achieved competitive results, as illustrated in Tables 4 and 5.

5. CONCLUSION

DR is a diabetes complication that affects the eyes. This disease may cause no symptoms or only mild vision problems but eventually, can cause blindness. In Egypt, more than 6 million people (7.2% of the population) suffer from DR. With ad-

vances in computers algorithms, such as AI and DL models, the opportunities for the detection of DR at the early stages increases. Early detection will increase the chances of recovery and reduce the possibility of vision loss in patients. In this paper, deep transfer learning models for medical DR detection were investigated on the APTOS 2019 dataset. According to literature surveys, this research is considered one the first studies that used the APTOS 2019 dataset, as it was released in the second quarter of 2019. Augmentation techniques were used to overcome the overfitting problem and increased the dataset images to be 4 times larger than the original dataset. The deep transfer learning models selected in this paper are AlexNet, ResNet18, SqueezeNet, GoogleNet, VGG16, and VGG19. The overall testing accuracy and performance metrics (precision, recall, and F1 score) showed that the AlexNet model achieved the highest testing accuracy (97.9%), precision, and F1 score percentage. Moreover, this model utilized a minimum number of layers, which decreased the training time and the computational complexity.

- **Acknowledgments:** We gratefully acknowledge the support of NVIDIA Corporation, who donated the Titan X GPU used in this research.
- **Author's contribution:** Each author gave substantial contribution to the conception or design of the work and in the acquisition, analysis and interpretation of data for the work. Each author had role in drafting the work and revising it critically for important intellectual content. Each author gave final approval of the version to be published and they agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
- **Conflicts of interest:** There are no conflicts of interest.
- **Financial support and sponsorship:** None.

REFERENCES

1. Soomro TA, Afifi AJ, Zheng L, Soomro S, Gao J, Hellwich O, et al. Deep Learning Models for Retinal Blood Vessels Segmentation: A Review. *IEEE Access*. 2019; 7: 71696-71717.
2. Sun Y, Zhang D. Diagnosis and Analysis of Diabetic Retinopathy Based on Electronic Health Records. *IEEE Access*. 2019; 7: 86115-86120.
3. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Research and Clinical Practice*. 2014; 103(2): 137-149.
4. Shahin EM, Taha TE, Al-Nuaimy W, Rabaie S El, Zahran OF, El-Samie FEA. Automated detection of diabetic retinopathy in blurred digital fundus images. In: 2012 8th International Computer Engineering Conference (ICENCO). 2012: 20-25.
5. Gao Z, Li J, Guo J, Chen Y, Yi Z, Zhong J. Diagnosis of Diabetic

- Retinopathy Using Deep Neural Networks. *IEEE Access*. 2019; 7: 3360–3370.
6. Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria. *Translational Research*. 2018; 194: 36–55.
 7. Palavalasa KK, Sambaturu B. Automatic Diabetic Retinopathy Detection Using Digital Image Processing. In: 2018 International Conference on Communication and Signal Processing (ICCSPP). 2018: 72–76.
 8. Chandran A, Nisha KK, Vineetha S. Computer aided approach for proliferative diabetic retinopathy detection in color retinal images. In: 2016 International Conference on Next Generation Intelligent Systems (ICNGIS). 2016: 1–6.
 9. Mansour RF. Evolutionary Computing Enriched Computer-Aided Diagnosis System for Diabetic Retinopathy: A Survey. *IEEE Reviews in Biomedical Engineering*. 2017; 10: 334–349.
 10. Agarwal S, Acharjya K, Sharma SK, Pandita S. Automatic computer aided diagnosis for early diabetic retinopathy detection and monitoring: A comprehensive review. In: 2016 Online International Conference on Green Engineering and Technologies (IC-GET). 2016: 1–7.
 11. Rong D, Xie L, Ying Y. Computer vision detection of foreign objects in walnuts using deep learning. *Computers and Electronics in Agriculture*. 2019; 162: 1001–1010.
 12. Shrestha A, Mahmood A. Review of Deep Learning Algorithms and Architectures. *IEEE Access*. 2019; 7: 53040–53065.
 13. Eraslan G, Avsec Ž, Gagneur J, Theis FJ. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*. 2019;
 14. Riordon J, Sovilj D, Sanner S, Sinton D, Young EWK. Deep Learning with Microfluidics for Biotechnology. *Trends in Biotechnology*. 2019; 37(3): 310–324.
 15. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*. 2019; 29(2): 102–127.
 16. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik*. 2019; 29(2): 86–101.
 17. Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012: 3642–3649.
 18. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Curran Associates Inc.; 2012: 1097–1105.
 19. Yin F, Wang Q, Zhang X, Liu C. ICDAR 2013 Chinese Handwriting Recognition Competition. In: 2013 12th International Conference on Document Analysis and Recognition. 2013: 1464–1470.
 20. LeCun Y, Huang FJ, Bottou L. Learning methods for generic object recognition with invariance to pose and lighting. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 CVPR 2004. 2004: II-104 Vol. 2.
 21. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE [Internet]*. 1998 [cited 2017 Oct 22]; 86(11): 2278–2324. Available from: <http://ieeexplore.ieee.org/document/726791/>
 22. Stallkamp J, Schlipsing M, Salmen J, Igel C. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In: The 2011 International Joint Conference on Neural Networks. 2011: 1453–1460.
 23. Deng J, Dong W, Socher R, Li L, Kai L, Li FF. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009: 248–255.
 24. Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). 2015: 730–734.
 25. Szegedy C, Wei L, Yangqing J, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 1–9.
 26. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770–778.
 27. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 1800–1807.
 28. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818–2826.
 29. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2261–2269.
 30. Asia Pacific Tele-Ophthalmology Society. APTOS 2019 Blindness Detection dataset. 2019.
 31. Khalifa N, Taha M, Hassanien A, Mohamed H. Deep Iris: Deep Learning for Gender Classification Through Iris Patterns. *Acta Inform Med*. 2019; 27(2): 96.
 32. Khalifa NE, Hamed Taha M, Hassanien AE, Selim I. Deep Galaxy V2: Robust Deep Convolutional Neural Networks for Galaxy Morphology Classifications. In: 2018 International Conference on Computing Sciences and Engineering (ICCSE). IEEE; 2018: 1–6.
 33. Khalifa NEM, Taha MHN, Hassanien AE. Aquarium Family Fish Species Identification System Using Deep Neural Networks. In: Hassanien AE, Tolba MF, Shaalan K, Azar AT, editors. Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018. Cham: Springer International Publishing; 2019: 347–356.
 34. Iandola FN, Moskewicz MW, Ashraf K, Han S, Dally WJ, Keutzer K. SqueezeNet. arXiv. 2016.
 35. Shah R, Yang Y. GoogLeNet. *Population Health Management*. 2015.
 36. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014 Sep 4 [cited 2017 Aug 4]; Available from: <http://arxiv.org/abs/1409.1556>
 37. Wong SC, Gatt A, Stamatescu V, McDonnell MD. Understanding Data Augmentation for Classification: When to Warp? In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE; 2016: 1–6.
 38. Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. 2010.