Use of convolutional neural networks for the detection of u-serrated patterns in direct immunofluorescence images to facilitate the diagnosis of epidermolysis bullosa acquisita

Shi, Chenyu; Meijer, Joost; Azzopardi, George; Diercks, G.F.H.; Guo, Jiapan; Petkov, Nicolai

Published in:
The American Journal of Pathology

DOI:
10.1016/j.ajpath.2021.05.024

Publication date:
2021

The American Journal of Pathology, Vol. 191, No. 9, September 2021

The American Journal of
PATHOLOGY
ajp.amjpathol.org

SHORT COMMUNICATION

Use of Convolutional Neural Networks for the Detection of u-Serrated Patterns in Direct Immunofluorescence Images to Facilitate the Diagnosis of Epidermolysis Bullosa Acquisita

Chenyu Shi,* Joost M. Meijer,† George Azzopardi,* Gilles F.H. Diercks,†‡ Jiapan Guo,§ and Nicolai Petkov*

From the Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Department of Dermatology, University Medical Center Groningen, Center for Blistering Diseases,† Department of Pathology, University Medical Center Groningen, Center for Blistering Diseases,‡ and the Department of Radiation Oncology,§ University Medical Center Groningen, the University of Groningen, Groningen, the Netherlands*

Accepted for publication
May 18, 2021.

Address correspondence to Chenyu Shi, M.Sc., Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Nijenborgh 9, 9747 AG, Groningen, the Netherlands. E-mail: c.shi@rug.nl.

The u-serrated immunodeposition pattern in direct immunofluorescence (DIF) microscopy is a recognizable feature and confirmative for the diagnosis of epidermolysis bullosa acquisita (EBA). Due to unfamiliarity with serrated patterns, serration pattern recognition is still of limited use in routine DIF microscopy. The objective of this study was to investigate the feasibility of using convolutional neural networks (CNNs) for the recognition of u-serrated patterns that can assist in the diagnosis of EBA. The nine most commonly used CNNs were trained and validated by using 220,800 manually delineated DIF image patches from 106 images of 46 different patients. The data set was split into 10 subsets: nine training subsets from 42 patients to train CNNs and the last subset from the remaining four patients for a validation data set of diagnostic accuracy. This process was repeated 10 times with a different subset used for validation. The best-performing CNN achieved a specificity of 89.3% and a corresponding sensitivity of 89.3% in the classification of u-serrated DIF image patches, an expert level of diagnostic accuracy. Experiments and results show the effectiveness of CNN approaches for u-serrated pattern recognition with a high accuracy. The proposed approach can assist clinicians and pathologists in recognition of u-serrated patterns in DIF images and facilitate the diagnosis of EBA. *(Am J Pathol 2021, 191: 1520–1525; https://doi.org/10.1016/j.ajpath.2021.05.024)*

The u-serrated immunodeposition pattern in direct immunofluorescence (DIF) microscopy on a skin biopsy specimen is a recognizable feature and a reference standard for the diagnosis of the autoimmune bullous disease epidermolysis bullosa acquisita (EBA).[1] EBA is a subtype of pemphigoid disease and characterized by autoantibodies against type VII collagen, located in the sublamina densa zone of the epidermal basement membrane zone (EBMZ). Diagnosis of EBA can be challenging; the disease is likely underdiagnosed because sophisticated laboratory techniques are needed, and circulating disease-specific autoantibodies are only detected in 50% of patients with EBA. Clinical features and histopathology of a skin biopsy specimen are not sufficient to distinguish EBA from other subtypes of pemphigoid diseases.

The disease course of EBA is often chronic and with severe morbidity; the prognosis of EBA differs from other subtypes of pemphigoid diseases, with EBA often being refractory to treatment and with potential scarring of skin and mucosa, strictures, and stenosis. The early identification of EBA is therefore essential for disease management and to prevent disease progression. Current diagnostic techniques for confirmation of EBA encompass direct immune-electron microscopy and fluorescent overlay antigen mapping on a skin biopsy specimen, techniques that are only available in a

C.S. and J.M.M. contributed equally to this work.
Disclosures: None declared.

https://doi.org/10.1016/j.ajpath.2021.05.024

very limited number of referral laboratories and less suitable for routine diagnostics.

Due to differences in the subtypes of pemphigoid disease based on autoantigens and the localization of the targeted structural proteins, different patterns of immunodepositions are formed along the EBMZ. Serration pattern analysis is a detailed examination of the linear pattern of immunodepositions (IgG and/or IgA, complement C3) along the EBMZ by DIF microscopy on a skin biopsy sample. Such analysis involves two different types of patterns: the n-serrated pattern with rounded n-like shapes, and the u-serrated pattern with u- or finger-like ridges. The u-serrated pattern in DIF microscopy corresponds with the localization of type VII collagen, whereas the n-serrated pattern can be observed in other pemphigoid diseases with other autoantigens.[2] DIF microscopy on a skin biopsy specimen is considered the reference standard for diagnosis of autoimmune bullous diseases and is widely performed in medical laboratories.[2–4]

The family of convolutional neural networks (CNN) is the current state-of-the-art for medical image analysis and has seen various applications in dermatology, ophthalmology, radiographic images, and histopathologic slide analysis. CNNs can learn patterns from images by resolving the image content into notable features, followed by selection and aggregation of the most meaningful ones. In dermatology, the use of CNNs has been shown in the diagnosis of skin cancer to distinguish between malignant melanoma, carcinoma, and nevus[5] or between benign lesions and keratinocyte skin cancer lesions.[6] In a previous study of serration pattern analysis, recognition of the serrated pattern in DIF images by observers was achieved with an accuracy of 78.6% in dermatologists, pathologists, and residents in dermatology.[3] Although serration pattern analysis is implemented in diagnostic criteria of EBA and is useable in routine diagnostics, its widespread implementation is still limited because of unfamiliarity with the serrated patterns.[4] A CNN-based approach was therefore investigated to develop an automatic tool for the classification of u-serrated patterns in DIF microscopy images that may facilitate the diagnosis of EBA.

Materials and Methods

DIF Images and Data Set

UserratedPatch2021 data set, consisting of 220,800 DIF image patches (of size 50 × 50 pixels) extracted from 106 images of 46 patients (4800 patches per each patient), was compiled. One-half of the dataset consisted of u-serrated patterns from DIF biopsy specimens of 23 patients with confirmed diagnoses of EBA, and the other half consisted of non−u-serrated patterns from DIF biopsy specimens of the other 23 patients with other pemphigoid diseases. DIF samples with u-serrated patterns were collected from patients with a confirmed diagnosis of EBA. Diagnosis was compliant with the latest consensus on diagnosis of the disease, including the observation of the linear immunodeposition of IgG along the EBMZ with a u-serrated pattern and with serologic analysis performed by using indirect immunofluorescence on salt-split human skin, immunoblotting, and enzyme-linked immunosorbent assay (type VII collagen, or also NC16A BP180 and BP230).[3] The non−u-serrated patterns were extracted from DIF samples of patients with other pemphigoid diseases (DIF samples with a positive linear immunodeposition along the EBMZ). Samples were randomly selected from routine diagnostics and consisted of patients with bullous pemphigoid, nonbullous pemphigoid, mucous membrane pemphigoid, anti−laminin-332 mucous membrane pemphigoid, and one case of porphyria. Patients with linear IgA disease were not included, as only IgG staining was addressed. The original data set of DIF images of serrated patterns was used in a previous study, with the outcome of serrated patterns independently classified by an expert dermatopathologist, two expert dermatologists, and a resident in dermatology.[4]

Different staining intensities and slide thicknesses (4 μm, 6 μm, and 8 μm) were included to generate the data set, representing routine diagnostics. All DIF slides were cut and stained at the Immunodermatology Laboratory of the Center for Blistering Diseases, University Medical Center Groningen (Groningen, the Netherlands). DIF images were taken under standardized settings with the same microscope (Leica DMRA, Leica, Wetzlar, Germany) and 40× magnification dry objective (Leica HCX PL Fluotar 40×/0.75, Leica) and 10× magnification ocular. Figure 1 shows enlarged examples from the DIF images from the data set representing a u-serrated pattern from a patient with EBA and three examples of non−u-serrated patterns from a patient with another pemphigoid disease.

The UserratedPatch2021 data set is publicly available at *https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/ZJTLBB*, and a training website with DIF images is available at *https://rug.eu.qualtrics.com/jfe/form/SV_3adioqW098URBkh?Q_JFE=qdg* (both last accessed June 3, 2021).

The data set was generated by first manually delineating 2300 u-serrated and 2300 non−u-serrated patches, and it was extended by creating rotated versions in 24 orientation intervals of 15 degrees with and without flip in the left-right direction. DIF image patches were of size 50 × 50 pixels each, and the pixel values were extracted from the green channel. The data set was split into 10 subsets: nine subsets that contain one-half u-serrated patterns and one-half non−u-serrated patterns from 42 patients functioned as training sets to train commonly used CNNs. The remaining subset from the other four different patients functioned as a testing data set for evaluation of diagnostic accuracy. For each CNN, this process was repeated 10 times, with a different subset used for validation and the other nine subsets for training (a process called 10-fold cross-validation). All image patches were assigned to the training and testing sets equally and randomly and not with fixed

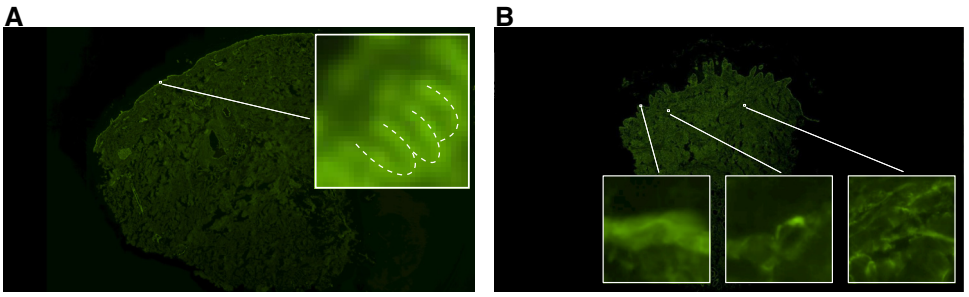

Figure 1 Enlarged examples of a u-serrated pattern (of size 50 × 50 pixels) along the epidermal basement membrane zone in a whole-slide direct immunofluorescence (DIF) image (of size 21,504 × 12,144 pixels) from a patient with epidermolysis bullosa acquisita (**A**) and three examples of a non–u-serrated linear pattern along the epidermal basement membrane zone or background fluorescence in a whole-slide DIF image (of size 23,192 × 13,024 pixels) from a patient with another pemphigoid disease (**B**). One of the main characteristics of u-serrated patterns is the representation of grass or spike shapes in the DIF images, which are marked by **white dashed lines** in **A**.

categories (training or testing). Performance of the CNNs was evaluated based on sensitivity, specificity, receiver-operating characteristic curve, and the corresponding area under the curve (AUC). The performance of CNN using AUC (range, 0 to 1) indicates perfect prediction classification, with an AUC receiver-operating characteristic score of 1 and 0.5 indicating a random classifier.

CNNs and Implementation Details

A set of nine different state-of-the-art CNNs was used: VGG,[7] ResNet,[8] GoogLeNet,[9] DenseNet,[10] MobileNet,[11] DPN,[12] SENet,[13] PreActResNet,[14] and ResNeXt.[15] They were all modified for the binary classification task at hand and are explained in the following text.

VGG-19 contains 16 convolutional layers followed by three fully connected layers. ResNet-18 is an 18-layer, deep plain residual network. GoogLeNet contains 22 layers of inception modules. DenseNet-121 has 121 layers, with each layer containing four dense blocks. MobileNet is based on a streamlined architecture with 28 layers. DPN-92 contains 92 layers that are built by stacking 30 modularized micro-blocks. SENet-18 has 18 layers that have an embedded block termed the "squeeze and-excitation" block. PreActResNet-18 is an 18-layer preactivation ResNet. ResNeXt-29 is an improved version of the 29-layer ResNet.

Images from 42 patients were included in the training set to train the CNN models and the images from the remaining four patients as the validation set to monitor the training process. A maximum number of 100 epochs was used for the training. To avoid overfitting, an early stopping strategy is used if there is no improvement of loss in the validation set after five epochs. Each training image was used to create another seven images by rotation at seven random angles. Thus, the factor of data augmentation was eight. The original architectures of the CNNs used in this study have output layers with many units, one unit per class. The last (classification) layer of each CNN is replaced by one single unit with a graded output between 0 and 1 that gives the likelihood of a pattern being a u-serrated pattern. A pattern is classified as u-serrated if the output is greater than a certain preset threshold value (decision criterion). Using different values of the decision criterion leads to different specificity/sensitivity results. The binary cross-entropy with logits loss was used as the training criterion instead of the cross-entropy loss, which is more suitable for binary classification. The momentum was set to 0.9, weight decay to 0.0005, and the learning rate started as 0.01 and then decreased by a factor of 10 when the accuracy on the validation set stopped improving.

Results

Experiments were performed with the aforementioned CNNs on the UserratedPatch2021 data set to detect

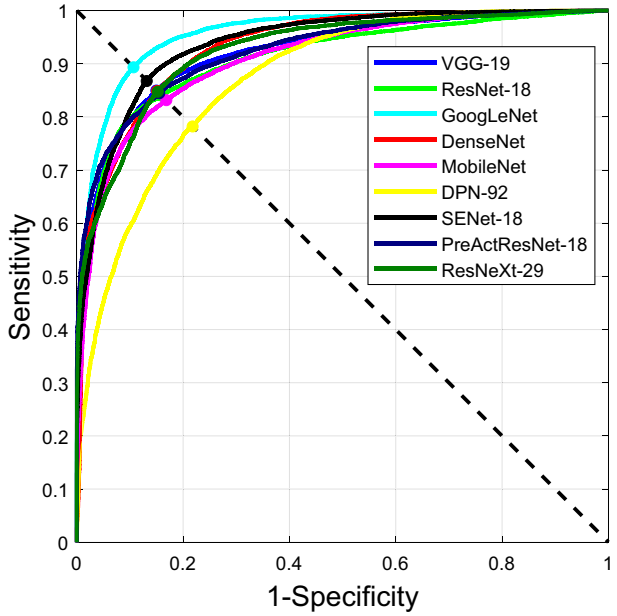

Figure 2 Receiver-operating characteristic curves and area under the curves of nine convolutional neural networks. Each dot marker lies on the **black dashed line,** which indicates the performance for which the sensitivity and specificity are equal. All results are averaged across 10-fold cross-validation.

Table 1 Results (Se, Sp, and AUC with 95% CI) of the Nine CNNs When Se Equals Sp

CNN	Sp = Se = F1-score (%)	AUC, % (95% CI)
GoogLeNet	89.3	95.7 (95.4–96.0)
SENet-18	86.8	93.8 (93.4–94.1)
VGG-19	85.0	92.2 (91.8–92.6)
DenseNet	84.9	93.0 (92.7–93.4)
ResNeXt-29	84.8	92.5 (92.1–92.9)
PreActResNet-18	84.5	92.5 (92.1–92.8)
ResNet-18	84.0	91.2 (90.8–91.6)
MobileNet	83.2	91.1 (90.7–91.5)
DPN-92	78.2	87.3 (86.8–87.7)

All results are averaged across 10-fold cross-validation.

AUC, area under the curve; CNN, convolutional neural networks; Se, sensitivity; Sp, specificity.

u-serrated patterns in DIF images and to compare computer-aided diagnosis versus manual diagnosis based on serration pattern analysis. A total of 220,800 DIF image patches in the data set included 110,400 patches from u-serrated patterns from DIF biopsy specimens of patients with EBA (n = 23) and 110,400 patches from non–u-serrated patterns from DIF biopsy patterns of patients with other pemphigoid diseases. The composition of DIF images in the non–u-serrated group consisted of 98.8% linear n-serrated patterns and 2.2% undetermined linear immunodepositions. Based on the validation data set, for the detection of u-serrated patterns, the respective AUCs of the 9 CNNs ranged from 87.3% to 95.7%. Figure 2 displays the receiver-operating characteristic curves of the nine CNNs used, the results of the AUC of the receiver-operating characteristic curves, and performances of the CNNs when sensitivity equals specificity. The performance of the nine CNNs when sensitivity equals specificity ranged from 78.2% to 89.3%. All results (sensitivity, specificity, and AUC) in Table 1 are averaged across 10-fold cross-validation. The best performance of detection of u-serrated patterns in the DIF images was achieved by the modified GoogLeNet, with an average sensitivity of 89.3%, specificity of 89.3%, and AUC of 95.7% across the 10 experiments in the classification of individual patches along the EBMZ. As a follow-up to this work, one may evaluate the GoogLeNet model on a new set of patients with EBA and with other pemphigoid diseases.

Example

Figure 3 presents an example of the application of the trained GoogLeNet model applied to two studies of DIF images (of size 2088 × 1560 pixels): i) from a skin biopsy specimen from a patient with EBA; and ii) from a skin biopsy specimen from a patient with another pemphigoid disease (non-EBA). The EBMZ region is automatically partitioned into patches by using a sliding window of size 50 × 50 pixels with a stride of 25 pixels. The patches are then automatically classified by the GoogLeNet CNN, using the decision criterion that corresponds to sensitivity and specificity of 89%. The squares in Figure 3 indicate patches that were classified as containing u-patterns. In the DIF image from the patient with EBA, 68 of 165 patches were classified as u-serrated. In the DIF image from the non-EBA patient, only six of 162 patches were mistakenly classified as u-serrated. The shown example could be the output of an online tool for routine immunofluorescence diagnostics of DIF samples with positive linear immunodepositions, suggestive for diagnosis of pemphigoid diseases.

Discussion

In this study, an automated model was developed by using CNNs for the recognition of u-serrated immunodeposition patterns in DIF images to facilitate the diagnosis of EBA. The performance surpassed the manual serration pattern analysis of a previous study and was of expert level.

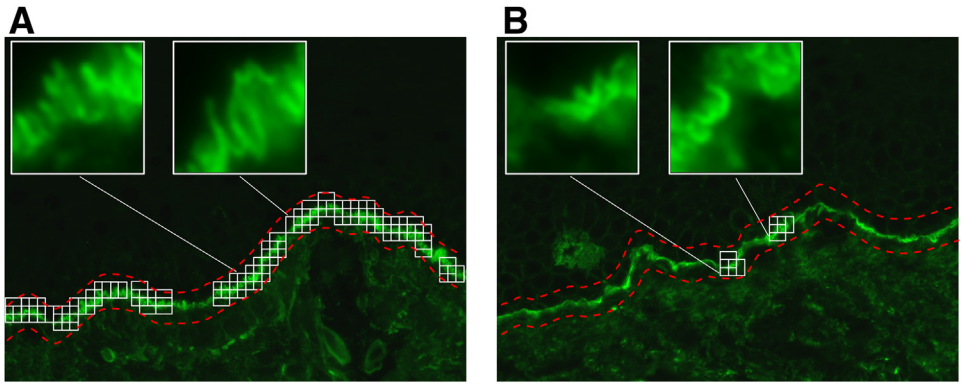

Figure 3 An example of the application of the modified GoogLeNet model to two direct immunofluorescence image crops (of size 2088 × 1560 pixels), one from a skin biopsy specimen of a patient with epidermolysis bullosa acquisita (**A**) and one from a patient with another pemphigoid disease (**B**). Regions between the **red dashed lines** indicate the area of the epidermal basement membrane zone. **Squares** (of size 50 × 50 pixels) indicate patches that were classified according to the convolutional neural networks as u-serrated patterns. In the case of a non–epidermolysis bullosa acquisita patient (**B**), the detected patterns are false-positive findings, and their number is in line with the 89% precision of the method.

Machine learning models have been applied to dermatologic diagnoses based on both clinical and dermoscopic images.[16] With the introduction of whole-slide scanning systems in pathology, CNNs have been shown to improve diagnostic accuracy and aid the pathologist in the examination and quantification of digital histopathologic slide analysis.[17] In a previous study of serration pattern analysis, a mean accuracy of 78.6% of correct manual recognition of serrated patterns was observed in dermatologists, pathologists, and residents in dermatology with various levels of expertise.[3] More recently, a comparative study of DIF serration pattern analysis in two different laboratories showed a very high recognition rate of 97.5% in DIF slides independently classified by four experts: a pathologist, two dermatologists, and a resident in dermatology.[4] With the current introduction of whole-slide immunofluorescence scanning systems, CNNs may have a broader implementation for routine diagnostics and analytical methods. For future implementation of automated slide analysis, an algorithm would require a high specificity for a certain diagnosis and minimal false-positive outcomes.[2]

EBA is a difficult disease to diagnose and has a serious impact on the prognosis of the patient, but the required diagnostic techniques are only available in a few specialized laboratories. The automated pattern recognition shown here may assist in identifying patients with EBA in one simple step, for a confirmed diagnosis based on DIF microscopy. The current study shows that the performance of detection of u-serrated patterns by CNNs is at an expert level, and the automated model may act as a reliable reference standard for u-serrated patterns and a computer-aided diagnosis of EBA. Another utility of the automated model would be of a supportive role for training of classification of u-serrated patterns.

Bullous systemic lupus erythematosus may exhibit a u-serrated pattern similar to EBA. However, such a deposition would be observed in patients with confirmed systemic lupus erythematosus based on other immunoserologic test results and clinical criteria, distinguishing those patients from those with EBA. This study did not include patients with bullous systemic lupus erythematosus due to the rarity of the disease.

That an automatic tool may perform better than a human rater is an interesting development indeed. A human rater performs serration pattern analysis using a microscope based on the first impression and searches for the confirmation of the pattern. The automatic tool analyzes all available data of image patches of the pattern and makes the prediction. The duration of pattern recognition and classification has not been assessed before, but it can be argued that the automatic tool will always be faster.

The DIF images in the data set are labeled only as u-serrated or non−u-serrated patterns. Methods using CNNs for recognition of multiple patterns, such as granular patterns, n-serrated patterns, and negative images, can be developed for future application.

Conclusions

This study shows that CNNs are able to recognize the u-serrated immunodeposition patterns in DIF images of patients with EBA with a high accuracy. The proposed approach can assist clinicians and pathologists in recognizing u-serrated patterns in DIF images and facilitate diagnosis of EBA.

References

1. Prost-Squarcioni C, Caux F, Schmidt E, Jonkman MF, Vassileva S, Kim SC, Iranzo P, Daneshpazhooh M, Terra J, Bauer J, Fairley J, Hall R, Hertl M, Lehman JS, Marinovic B, Patsatsi A, Zillikens D; International Bullous Diseases Group, Werth V, Woodley DT, Murrell DF: International Bullous Diseases Group: consensus on diagnostic criteria for epidermolysis bullosa acquisita. Br J Dermatol 2018, 179:30−41
2. Vodegel RM, Jonkman MF, Pas HH, de Jong MCJM: U-serrated immunodeposition pattern differentiates type VII collagen targeting bullous diseases from other subepidermal bullous autoimmune diseases. Br J Dermatol 2004, 151:112−118
3. Terra JB, Meijer JM, Jonkman MF, Diercks GFH: The n- vs. u-serration is a learnable criterion to differentiate pemphigoid from epidermolysis bullosa acquisita in direct immunofluorescence serration pattern analysis. Br J Dermatol 2013, 169:100−105
4. Meijer JM, Atefi I, Diercks GFH, Vorobyev A, Zuiderveen J, Meijer HJ, Pas HH, Zillikens D, Schmidt E, Jonkman MF: Serration pattern analysis for differentiating epidermolysis bullosa acquisita from other pemphigoid diseases. J Am Acad Dermatol 2018, 78:754−759.e6
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S: Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017, 542:115−118
6. Han SS, Moon IJ, Lim W, Suh IS, Lee SY, Na J-I, Kim SH, Chang SE: Keratinocytic skin cancer detection on the face using region-based convolutional neural network. JAMA Dermatol 2020, 156:29−37
7. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. Edited by International Conference on Learning Representations (ICLR). Cornell University, 2015, arXiv: 1409−1556
8. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. Edited by IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. pp. 770−778
9. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A: Going deeper with convolutions. Edited by IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015. pp. 1−9
10. Huang G, Liu Z, Maaten L, Weinberger K: Densely connected convolutional networks. Edited by IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. pp. 2261−2269
11. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L: MobileNetV2: inverted residuals and linear bottlenecks. Edited by IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018. pp. 4510−4520
12. Chen Y, Li J, Xiao H, Jin X, Yan S, Feng J: Dual path networks. Edited by Conference on Neural Information Processing Systems (NIPS); 2017. pp. 4467−4475
13. Hu J, Shen L, Sun G: Squeeze-and-excitation networks. Edited by IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018. pp. 7132−7141
14. He K, Zhang X, Ren S, Sun J: Identity mappings in deep residual networks. Edited by European Conference on Computer Vision (ECCV); 2016. pp. 630−645

15. Xie S, Girshick R, Dollar P, Tu Z, He K: Aggregated residual transformations for deep neural networks. Edited by IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. pp. 5987−5995
16. Winkler JK, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Thomas L, Lallas A, Blum A, Stolz W, Haenssle HA: Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA Dermatol 2019, 155: 1135−1141
17. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B, van der Laak J: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep 2016, 6:26286