 doi:10.20944/preprints202205.0302.v1

Visualizing CoAtNet Predictions for Aiding Melanoma Detection

Daniel Kvak, MSc.*
Carebot s.r.o.
Prague, Czech Republic

ORCID: 0000-0001-7808-7773

Abstract

Melanoma is considered to be the most aggressive form of skin cancer. Due to the similar shape of malignant and benign cancerous lesions, doctors spend considerably more time when diagnosing these findings. At present, the evaluation of malignancy is performed primarily by invasive histological examination of the suspicious lesion. Developing an accurate classifier for early and efficient detection can minimize and monitor the harmful effects of skin cancer and increase patient survival rates. This paper proposes a multi-class classification task using the CoAtNet architecture, a hybrid model that combines the depthwise convolution matrix operation of traditional convolutional neural networks with the strengths of Transformer models and self-attention mechanics to achieve better generalization and capacity. The proposed multi-class classifier achieves an overall precision of 0.901, recall 0.895, and AP 0.923, indicating high performance compared to other state-of-the-art networks.

Keywords: skin cancer; melanoma; computer-aided diagnostics; image classification; CoAtNet; convolutional neural networks; deep learning.

*Corresponding author: daniel.kvak@carebot.com

 doi:10.20944/preprints202205.0302.v1

1 Introduction

Artificial intelligence (AI) is emerging to assist healthcare professionals with routine tasks such as removing noise, analysing images or reading medical reports. (Hamet & Tremblay 2017) In deep learning, currently the most widely adopted AI technique, computer algorithms learn using backpropagation to predict outcomes based on large data sets. (Albawi et al. 2017) The efficiency of these methods has improved dramatically in recent years and can now be found in areas ranging from computer-aided diagnostics (CADx) to online shopping to autonomous vehicles. However, deep learning tools also raise troubling questions because they solve problems in ways that humans cannot always observe. (Wang et al. 2020, Holzinger et al. 2019) There is a growing call among researchers and institutions to clarify the basis on which artificial intelligence makes decisions. (Kvak et al. 2022, Amann et al. 2020, Samek & Müller 2019)

The US Food and Drug Administration (FDA) recently outlined ten guiding principles that should be the cornerstone for the development of clinically applicable artificial intelligence. (FDA 2021) These guiding principles can help support the introduction of objective, safe and effective medical devices to the market. Beyond monitoring or defining the correct use, the core principles include many practices that have proven successful in other sectors; however, the greatest emphasis is on the so-called explainability of predictions (XAI, explainable artificial intelligence), which limits the risk of clinical bias. (Ghassemi et al. 2021)

2 Background

One of the most common methods used to identify melanoma is the ABCD rule which was introduced in 1985. (Nachbar et al. 1994) The acronym stands for Asymmetry, Borderline Irregularity, Changes in Color and Diameter. In 2004, the letter E was added to the ABCD acronym to stand for Evolving. (Jensen & Elewski 2015) Each criterion has certain features that are recognized to distinguish between benign and malignant melanoma. In addition, the method failed to recognize certain malignant nevi in their early stages. (Carli et al. 2002, Liu et al. 2005)

Melanoma is less common than other types, but it is the most dangerous form of skin cancer because it can spread quickly to other parts of the body. (Coit et al. 2009) It results from neoplastic proliferation of melanocytes. Malignant melanoma predominantly affects the skin, but can also affect eyes, ears, leptomeninges, and the mucous membranes of the mouth or genital tract. (Bastian 2014) The incidence of melanoma is increasing, affecting mainly the light skin population. (Matthews et al. 2017, Rigel et al. 1996) The pathophysiology of melanoma development is not yet clearly understood. (Hida et al. 2020) Multiple pathogenetic mechanisms of melanoma development are hypothesized. Melanoma develops not only on sun-exposed skin, where UV radiation is the main pathogenetic factor, but also in body parts that are relatively protected

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0302.v1

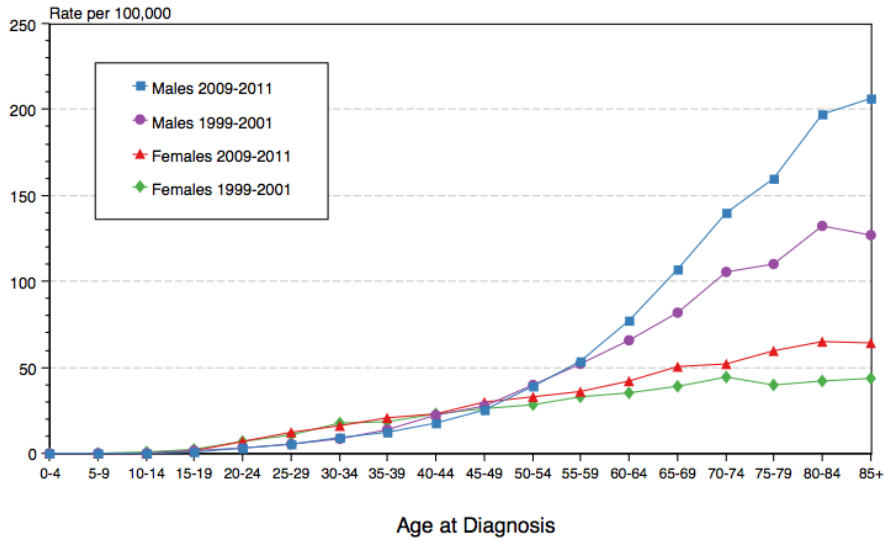


Figure 1: The delay-adjusted incidence and observed incidence of melanoma by age and gender in the United States between 1975 and 2011.

from radiation. (Apalla et al. 2017, Coit et al. 2009) When melanoma is suspected, it is important to biopsy the suspicious lesion on the skin or mucosa (excision with a 1-3 mm margin of tissue) and subsequent histological examination. (Bastian 2014)

3 Computational approach

CADx approaches based on deep learning and computer vision may represent an effective and, above all, affordable alternative to invasive histological examination. (Kassani & Kassani 2019) Applications based on convolutional neural networks (CNN) show promising results in medical image detection, classification and segmentation. (Li et al. 2014, Anwar et al. 2018) High accuracy is now achieved in interstitial lung disease classification (Shen et al. 2015) or in the detecion of colorectal adenomas and neoplastic lesions. (Yu et al. 2016) Many attempts have been made in the literature to improve the performance of CNN, either by using optimization methods to select significant features or by using image preprocessing techniques before the classification step. (Thoma 2017)

3.1 Proposed model arcitecture

CoAtNet offers a unique combination of **depthwise convolutions** (1) and **self-attention** (2) to allow fast and accurate advancement for large-scale image recognition and classification. The proposed architecture is based on the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0302.v1

observation that CNNs tend to exhibit improved generalization (i.e., the difference in performance between training and testing) due to their inductive bias, whereas self-attention models tend to show greater capacity (i.e., the ability to fit large-scale training data). (Dai et al. 2021)

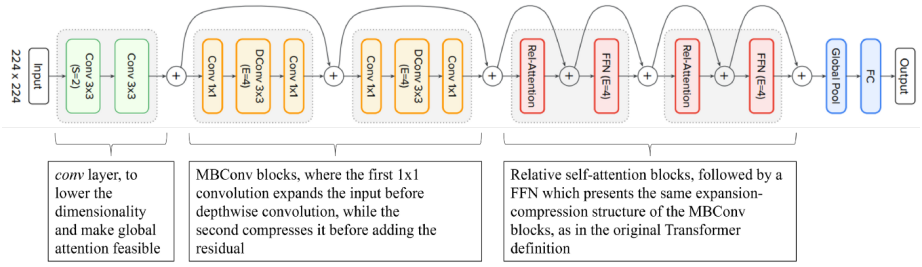


Figure 2: Overview of the used CoAtNet model.

(1) is a type of convolution operation where we use one convolution filter for each input channel. (Tan & Le 2019) Unlike spatially separable convolutions, depthwise convolutions work with kernels that cannot be split. (Guo et al. 2019) In a conventional 2D convolution performed over multiple input channels, the filter is as deep as the input and allows us to arbitrarily mix channels to generate individual features in the output. (Chang & Sha 2016) In contrast, depthwise convolutions maintain each channel separately. We can express this with the formula below:

$$y_i = \sum_{j \in \mathcal{L}(i)} w_{i-j} \odot x_j \tag{1}$$

(2) has become widespread technique adopted in natural language processing (NLP), with the fully-attentional Transformer model having largely replaced recurrenr neural networks (RNN) and being used in state-of-the-art language understanding models such as GPT, BERT, and XLNet. This technique allows the receptive field to be entire spatial locations, and computes weights based on renormalized pairwise similarity between pairs: if each pixel in the feature map is treated as a random variable and paring covariances are calculated, the value of each predicted pixel can be enhanced or weakened based on its similarity to other pixels in the image. The participating target pixels are the weighted sum of the values of all pixels, where the weights represent the correlation between each pixel and the target pixel. This can be represented by the following formula:

$$y_i = \sum_{j \in \mathcal{G}} \underbrace{\frac{\exp\left(x_i^\top x_j\right)}{\sum_{k \in \mathcal{G}} \exp\left(x_i^\top x_k\right)}}_{A_{i,j}} x_j \tag{2}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0302.v1

4 Dataset

The development of robust CADx systems for the automated diagnosis of skin lesions is hindered by the small size of clinically evaluated dermatoscopic image datasets available. (Garg et al. 2021) We assembled dermatoscopic images from various publicly available repositories while maintaining a representation of different populations, acquired and stored by different modalities.

The final dataset consists of 6,826 dermatoscopic images, representative of all important diagnostic categories in the field of various lesions: actinic keratoses, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, nevus, and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas, and hemorrhages). For a fraction of the images (∼50%), the ground truth was determined by histopathological examination, while in the remaining images the finding was decided by expert consensus or confirmed by in vivo confocal microscopy. A total of 300 images were extracted from the dataset as a test set (100 melanoma, 100 non-melanoma skin cancer, 100 benign skin lesions). The remaining 6,526 dermatology images were split between the training and validation set in an 80/20 ratio.

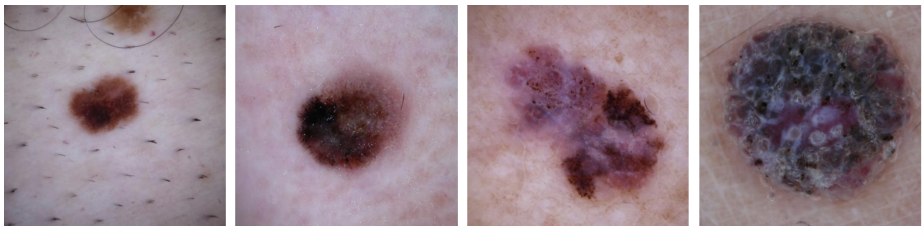


Figure 3: Examples of melanoma at different stages represented in the training set.

4.1 Data augmentation

Data augmentation increases the size of the input training data along with the regularization of the model, thus improving the generalization of the training model. (Mikołajczyk & Grochowski 2018) It also helps to create new train examples by randomly applying different transformations to the available dataset to reflect the noise in the real data. (Shorten & Khoshgoftaar 2019, Elgendi et al. 2021) In this study, we applied transformations involving random rotations (<= 0.25), modifications in contrast (0.9-1.1) and brightness (0.9-1.1), zoom (<= 0.25), and saturation (0.9-1.1). The extension of validation set was not investigated.

 doi:10.20944/preprints202205.0302.v1

Class	No. of images	Precision	Recall	AP
Average model performance	6,826	0.901	0.895	0.923
Actinic keratoses	332	0.786	0.821	0.772
Basal cell carcinoma	514	0.880	0.922	0.919
Benign keratosis-like lesion	1,099	0.894	0.877	0.903
Dermatofibroma	115	0.875	0.913	0.944
Melanoma	1,563	0.870	0.875	0.908
Nevus	3,061	0.935	0.913	0.958
Vascular lesions	142	1.000	0.931	0.995

Table 1: CoAtNet classifier performance on the used dataset.

5 Classifier performance

The classification performance of the proposed model for multi-class problem was evaluated for each component and the average classification performance of the model was calculated. Table 1 includes the precision (3) and recall (4) calculated based on the following equations below:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

For specific experiments and given that there is a class imbalance problem, the most reliable metric is the model average accuracy metric, while given that the accuracy is high, the second most important metric is the recall metric for individual classes. (Japkowicz & Stephen 2002) This is due to the importance of correctly identifying true cases that are malignant. AP (Average Precision) (5) summarizes a precision-recall curve as the weighted mean of precisions achieved at each threshold (Yilmaz & Aslam 2006), with the increase in recall from the previous threshold used as the weight:

$$\text{AP} = \sum_{n} (R_n - R_{n-1}) P_n \tag{5}$$

5.1 Visualizing model predictions

Despite the classifier showing impressive results on standard metrics, from a clinical perspective, it is important for us to determine whether features relevant to skin lesion detection and analysis were extracted during CoAtNet training using backpropagation. (Payer et al. 2019) As mentioned in the chapter 1 Introduction, medical devices should not serve as "black boxes" but need to provide additional information about how the model arrived at its predictions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0302.v1

(England & Cheng 2019, Amann et al. 2020, Samek & Müller 2019) Gradient-weighted Class Activation Mapping (Grad-CAM) is a method that uses gradient extraction from the last convolutional layer of a neural network to indicate the pixels that contribute most to the model output and the predicted probability of an image belonging to a predefined class. (Selvaraju et al. 2017) The resulting activation map can be plotted over the original image and can be interpreted as a visual tool to identify the regions that the model predicts whether an image belongs to a particular class. (Selvaraju et al. 2017, Panwar et al. 2020)

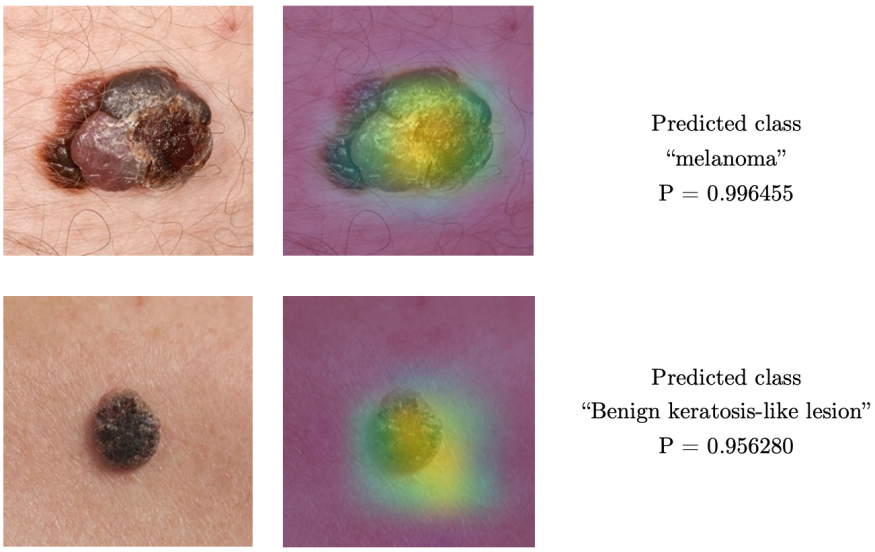


Figure 4: Grad-CAM activation heatmap visualization from CoAtNet model on real-world test data.

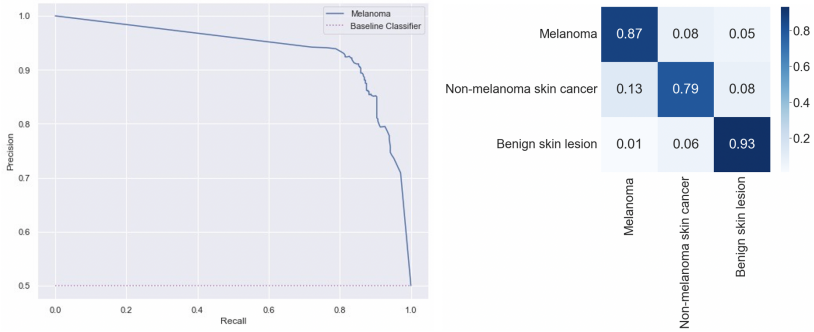
5.2 Model performance on test data

The precision-recall curve shows the trade-off between precision and recall for different thresholds. (Buckland & Gey 1994) A high area under the curve represents both high recall and high precision, with high precision associated with low False Positive cases and high recall associated with low False Negative cases. (Boyd et al. 2013) The combination of the Figure 4 and Figure 5 for the test set suggests that the model learned appropriate features for classification across malignant and benign lesions from a limited dataset.

6 Conclusions

In this study, we classified nine skin lesions with a particular focus on melanoma, which, although not as prevalent, is responsible for three-quarters of skin cancer

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0302.v1



(a) (b)

Figure 5: (a) Precision-Recall curve for Melanoma class. (b) Confusion matrix showing the results on the compiled 3-class test set.

related deaths. The classification of melanoma was performed using no lesion segmentation or complex image preprocessing. The proposed method is based on the state-of-the-art CoAtNet architecture, which incorporates the advantages of depthwise convolution and self-attention mechanism. Considering the necessity of large-scale data for efficient training, we applied data augmentation techniques to the existing dataset. Evidence from the exploratory analysis shows that the proposed approach significantly outperforms state-of-the-art models by achieving model average precision of 0.901, recall 0.895 and AP 0.923.

7 Authorship statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication.

8 Ethical procedure

The authors hereby declare that this research article meets all applicable standards with regards to the ethics of experimentation and research integrity. The authors also declare that the text of the article complies with ethical standards, the anonymity of the patients was respected.

 doi:10.20944/preprints202205.0302.v1

References

Albawi, S., Mohammed, T. A. & Al-Zawi, S. (2017), Understanding of a convolutional neural network, *in* '2017 international conference on engineering and technology (ICET)', Ieee, pp. 1–6.

Amann, J., Blasimme, A., Vayena, E., Frey, D. & Madai, V. I. (2020), 'Explainability for artificial intelligence in healthcare: a multidisciplinary perspective', *BMC Medical Informatics and Decision Making* **20**(1), 1–9.

Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M. & Khan, M. K. (2018), 'Medical image analysis using convolutional neural networks: a review', *Journal of medical systems* **42**(11), 1–13.

Apalla, Z., Nashan, D., Weller, R. B. & Castellsagué, X. (2017), 'Skin cancer: epidemiology, disease burden, pathophysiology, diagnosis, and therapeutic approaches', *Dermatology and therapy* **7**(1), 5–19.

Bastian, B. C. (2014), 'The molecular pathology of melanoma: an integrated taxonomy of melanocytic neoplasia', *Annual Review of Pathology: Mechanisms of Disease* **9**, 239–271.

Boyd, K., Eng, K. H. & Page, C. D. (2013), Area under the precision-recall curve: point estimates and confidence intervals, *in* 'Joint European conference on machine learning and knowledge discovery in databases', Springer, pp. 451–466.

Buckland, M. & Gey, F. (1994), 'The relationship between recall and precision', *Journal of the American society for information science* **45**(1), 12–19.

Carli, P., Massi, D., de Giorgi, V. & Giannotti, B. (2002), 'Clinically and dermoscopically featureless melanoma: when prevention fails', *Journal of the American Academy of Dermatology* **46**(6), 957–959.

Chang, J. & Sha, J. (2016), 'An efficient implementation of 2d convolution in cnn', *IEICE Electronics Express* pp. 13–20161134.

Coit, D. G., Andtbacka, R., Bichakjian, C. K., Dilawari, R. A., DiMaio, D., Guild, V., Halpern, A. C., Hodi, F. S., Kashani-Sabet, M., Lange, J. R. et al. (2009), 'Melanoma', *Journal of the National Comprehensive Cancer Network* **7**(3), 250–275.

Dai, Z., Liu, H., Le, Q. V. & Tan, M. (2021), 'Coatnet: Marrying convolution and attention for all data sizes', *Advances in Neural Information Processing Systems* **34**, 3965–3977.

Elgendi, M., Nasir, M. U., Tang, Q., Smith, D., Grenier, J.-P., Batte, C., Spieler, B., Leslie, W. D., Menon, C., Fletcher, R. R. et al. (2021), 'The effectiveness of image augmentation in deep learning networks for detecting covid-19: A geometric transformation perspective', *Frontiers in Medicine* **8**.

 doi:10.20944/preprints202205.0302.v1

England, J. R. & Cheng, P. M. (2019), 'Artificial intelligence for medical image analysis: a guide for authors and reviewers', *American journal of roentgenology* **212**(3), 513–519.

FDA (2021), 'Good machine learning practice for medical device development: Guiding principles'.

Garg, R., Maheshwari, S. & Shukla, A. (2021), Decision support system for detection and classification of skin cancer using cnn, *in* 'Innovations in Computational Intelligence and Computer Vision', Springer, pp. 578–586.

Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. (2021), 'The false hope of current approaches to explainable artificial intelligence in health care', *The Lancet Digital Health* **3**(11), e745–e750.

Guo, Y., Li, Y., Wang, L. & Rosing, T. (2019), Depthwise convolution is all you need for learning multiple visual domains, *in* 'Proceedings of the AAAI Conference on Artificial Intelligence', Vol. 33, pp. 8368–8375.

Hamet, P. & Tremblay, J. (2017), 'Artificial intelligence in medicine', *Metabolism* **69**, S36–S40.

Hida, T., Kamiya, T., Kawakami, A., Ogino, J., Sohma, H., Uhara, H. & Jimbow, K. (2020), 'Elucidation of melanogenesis cascade for identifying pathophysiology and therapeutic approach of pigmentary disorders and melanoma', *International Journal of Molecular Sciences* **21**(17), 6129.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K. & Müller, H. (2019), 'Causability and explainability of artificial intelligence in medicine', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **9**(4), e1312.

Japkowicz, N. & Stephen, S. (2002), 'The class imbalance problem: A systematic study', *Intelligent data analysis* **6**(5), 429–449.

Jensen, J. D. & Elewski, B. E. (2015), 'The abcdef rule: combining the "abcde rule" and the "ugly duckling sign" in an effort to improve patient self-screening examinations', *The Journal of clinical and aesthetic dermatology* **8**(2), 15.

Kassani, S. H. & Kassani, P. H. (2019), 'A comparative study of deep learning architectures on melanoma detection', *Tissue and Cell* **58**, 76–83.

Kvak, D., Bendik, M. & Chromcova, A. (2022), 'Towards clinical practice: Design and implementation of convolutional neural network-based assistive diagnosis system for covid-19 case detection from chest x-ray images', *arXiv preprint arXiv:2203.10596* .

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D. & Chen, M. (2014), Medical image classification with convolutional neural network, *in* '2014 13th international conference on control automation robotics & vision (ICARCV)', IEEE, pp. 844–848.

 doi:10.20944/preprints202205.0302.v1

Liu, W., Hill, D., Gibbs, A. F., Tempany, M., Howe, C., Borland, R., Morand, M. & Kelly, J. W. (2005), 'What features do patients notice that help to distinguish between benign pigmented lesions and melanomas?: the abcd (e) rule versus the seven-point checklist', *Melanoma research* **15**(6), 549–554.

Matthews, N. H., Li, W.-Q., Qureshi, A. A., Weinstock, M. A. & Cho, E. (2017), 'Epidemiology of melanoma', *Exon Publications* pp. 3–22.

Mikołajczyk, A. & Grochowski, M. (2018), Data augmentation for improving deep learning in image classification problem, *in* '2018 international interdisciplinary PhD workshop (IIPhDW)', IEEE, pp. 117–122.

Nachbar, F., Stolz, W., Merkle, T., Cognetta, A. B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O. & Plewig, G. (1994), 'The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions', *Journal of the American Academy of Dermatology* **30**(4), 551–559.

Panwar, H., Gupta, P., Siddiqui, M. K., Morales-Menendez, R., Bhardwaj, P. & Singh, V. (2020), 'A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images', *Chaos, Solitons & Fractals* **140**, 110190.

Payer, C., Štern, D., Bischof, H. & Urschler, M. (2019), 'Integrating spatial configuration into heatmap regression based cnns for landmark localization', *Medical image analysis* **54**, 207–219.

Rigel, D. S., Friedman, R. J. & Kopf, A. W. (1996), 'The incidence of malignant melanoma in the united states: issues as we approach the 21st century', *Journal of the American Academy of Dermatology* **34**(5), 839–847.

Samek, W. & Müller, K.-R. (2019), Towards explainable artificial intelligence, *in* 'Explainable AI: interpreting, explaining and visualizing deep learning', Springer, pp. 5–22.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2017), Grad-cam: Visual explanations from deep networks via gradient-based localization, *in* 'Proceedings of the IEEE international conference on computer vision', pp. 618–626.

Shen, W., Zhou, M., Yang, F., Yang, C. & Tian, J. (2015), Multi-scale convolutional neural networks for lung nodule classification, *in* 'International conference on information processing in medical imaging', Springer, pp. 588–599.

Shorten, C. & Khoshgoftaar, T. M. (2019), 'A survey on image data augmentation for deep learning', *Journal of big data* **6**(1), 1–48.

Tan, M. & Le, Q. V. (2019), 'Mixconv: Mixed depthwise convolutional kernels', *arXiv preprint arXiv:1907.09595* .

 doi:10.20944/preprints202205.0302.v1

Thoma, M. (2017), 'Analysis and optimization of convolutional neural network architectures', *arXiv preprint arXiv:1707.09725* .

Wang, F., Kaushal, R. & Khullar, D. (2020), 'Should health care demand interpretable artificial intelligence or accept "black box" medicine?'.

Yilmaz, E. & Aslam, J. A. (2006), Estimating average precision with incomplete and imperfect judgments, *in* 'Proceedings of the 15th ACM international conference on Information and knowledge management', pp. 102–111.

Yu, L., Chen, H., Dou, Q., Qin, J. & Heng, P. A. (2016), 'Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos', *IEEE journal of biomedical and health informatics* **21**(1), 65–75.