Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems

Automated Diagnosis of Skin Lesions

Fábio Santos, Filipe Silva and Petia Georgieva
Department of Electronics, Telecomunications and Informatics / IEETA
University of Aveiro
Aveiro 3810-193, Portugal
Emails: {fmts, fmsilva, petia}@ua.pt

***Abstract*—Recent years have seen significant progress in the automatic diagnosis of pigmented skin lesions, including advances in self-surveillance technologies accessible to patients and computer-aided diagnosis (CAD) tools for dermatologists. Rapid advances in mobile technologies and applications are playing a central role in providing educational aids and self-surveillance tools for patient use. At the same time, machine learning, specifically, deep learning is a fast-growing field that is being used for multiple medical imaging related problems, such as skin lesions classification. Recent studies based on deep networks produced promising results which have the potential to change the landscape of skin lesion diagnosis. Systems created based on these new advancements aim to provide support for both dermatologists in the decision making process and for patients that do not have access to skin professionals. This paper focuses on the current state of automated skin lesion diagnosis, while also providing a comprehensive view into the challenges and opportunities in dermatology care.**

***Keywords*—skin lesions; automated diagnosis; self-surveillance applications; deep learning classification**

I. Introduction

Skin cancer affects millions of people worldwide with an incidence rate that continues to increase. This represents a problem for the international health community in general. In the United states, one in five people develops skin cancer until the age of 70 [1]. In Europe, more than 100,000 people are diagnosed with melanoma annually, with 22,000 deaths due to this type of cancer [2]. Nevertheless, one of the most remarkable facts about skin cancer is that, when detected on late stages, there is a 23% chance of survival, but when detected early the 5 year survival rate rises to 99% [1]. Therefore, the early detection of skin cancer is a priority.

Skin cancer can be detected by dermatology professionals by simple visual examination of lesions. However, the difference between malignant and benign skin lesions can be negligible, making it a difficult task even for trained medical experts. As such, medical applications providing automated skin lesion diagnosis for decision support is a welcome addition to this field. Computer-aided diagnosis (CAD) systems have been extensively developed since the early 1990s as potential aids in the evaluation of melanocytic lesions. Initially, automated diagnosis was performed based on predefined techniques well known by dermatology professionals, such as the ABCD-rule [3], but often failed to either generalize to new cases or lacked the required accuracy. Since then, several software systems for the automated detection of melanoma in macroscopic and dermoscopic images appeared on market to provide a second opinion to expert clinicians and/or to give advanced training to new clinicians.

State-of-the-art surveys [4], [5], [6], [7] demonstrate the significant progress that has occurred in this field over the last few years, focusing on the computer vision algorithms typically used to analyse images and extract structural information. Most CAD systems begin with algorithms for lesion segmentation to determine the boundaries between the lesion and the surrounding skin. The lesion is then analysed for various diagnostic features such as asymmetry, border irregularity, color, dimensions and textures. Some systems provide quantification of such features for physician assessment, whereas others analyse the full constellation of features and provide a provisional diagnosis. Some have achieved a degree of sensitivity and specificity approaching expert dermatologists.

Modern digital technologies promise to have an increased impact on the way dermatology care will be delivered in next years. In particular, recent years have seen significant advances in both self-surveillance technologies accessible to patients and deep learning-based automated systems aimed to support the dermatologist in the decision making process. In the first case, mobile technologies and applications are playing a central role in providing educational and self-surveillance tools for patients use. In the second, convolutional neural networks (CNN) show remarkable performance for medical imaging diagnosis. In line with these developments, this paper provides a review of the latest research and literature related to the automated diagnosis of pigmented skin lesions, focusing on self-surveillance systems and those based on the recent trend of deep learning (DL).

The remainder of the paper is organised as follows. Section 2 overviews significant developments in self-surveillance applications and skin lesions classification using deep learning. Section 3 highlights the main challenges in deep learning classification and how they have been addressed for automated diagnosis of skin lesions. Section 4 concludes the paper and addresses the contribution.

II. Literature Review

A. Mobile Applications for Skin Self-Surveillance

Online health, eHealth and mHealth applications represent a rapidly developing field of medicine that has the potential to become powerful tools in the diagnosis and management of

978-1-7281-5456-5/20/$31.00 ©2020 IEEE

skin diseases [8]. These applications aim to enhance clinical care, promote health, prevent diseases and, most importantly, provide medical support when it is not available at a particular location or time. Generally, the acceptance towards this type of systems in the medical community keeps growing, but it is dependent on factors such as performance, accessibility and ease of use, which poses challenges for their adoption.

Mobile applications have been leveraged by the growing availability of smartphone attachments capable of turning these devices into compact dermatoscopes. As a result, they offer the patient the ability to monitor their skin condition using images self-captured with a smartphone, while the capability for processing and displaying results is very variable. Over the past few years, several review articles on mobile applications for dermatology have appeared, taking into account a wide range of sources and the technological reality at each moment [9], [10], [11], [8].

In one of the most recent contributions, Zaidan et al. [11] reviewed the literature on smartphone applications for skin cancer diagnosis in the period from 2011 to 2016. A total of 89 articles were classified into four groups: development and design, analytical, evaluative and comparative, and review and survey studies. Out of the 89 articles, 43 focus on the development of various AI algorithms and applications for assisting in the prevention and early detection of malignant melanoma. A total of 20 articles involve analytical studies on the incidence of skin cancer, the classification of malignant cancer or benign cancer and methods for prevention and diagnosis. A total of 15 articles consist of studies that range from evaluation or comparison of mobile apps to the exploration of features designed for skin cancer detection. A total of 11 articles comprise reviews and surveys referring to actual applications or providing a general overview of the technology.

In the meantime, Jaworek-Korjakowska and Kleczek [8] reported the existence of about 45 mobile applications related to mole diagnosis available on Apples App Store in March 2017. Most of them offered only educational information on melanoma. Nearly half of them allowed the user to take photos of their moles and to track changes over time using simple visual comparison. Only four applications performed melanoma risk assessment or lesion classification based on image analysis. These applications are DermaCompare (risk assessment through image matching), Lbax (mole diagnosis through content-based image retrieval), MySkinApp (risk assessment) and SkinVision (risk assessment through fractal analysis). From these applications, SkinVision received the European CE Marking and DermaCompare was approved by the U.S. Food and Drug Administration.

One of the most popular eHealth applications for this purpose is the Metaoptima's Dermengine web application. The user can submit an image of a suspicious lesion that will be compared, in terms of color, shape and pattern, with dozens of similar annotated images [12]. The search for similarity between images is carried out using deep learning techniques. Another popular app is the SkinVision which classifies lesions as either low, medium or high risk of skin cancer by using an assessment algorithm based on gray-scale images of lesions and their associated fractal maps. It achieves the overall sensitivity of 73%, specificity of 83%, and accuracy of 81%. The positive and negative predictive values were 49% and 83%, respectively [8].

B. Skin Lesion Classification Using Deep Learning

Deep learning refers to computational models composed of multiple processing layers capable of learning representations of data with multiple levels of abstraction [13]. The initial impact of deep learning for medical imaging was revealed through a special issue published in 2016 at the IEEE Transactions on Medical Imaging [14]. The surveys by Hu et al. [15] and Litjens et al. [16] focus on the methodological approaches in medical imaging, the contributions of deep learning to organ segmentation, lesion detection and tumor classification, as well as the challenges and trends in different application areas.

The main advantage of deep learning over other machine learning algorithms is that it removes the need for feature engineering, a process that requires specific knowledge of the problem domain. Current successes of deep learning are mainly due to advances in unsupervised pre-training [17], [18] and new methods to prevent overfitting [19]. At the same time, the latest developments have been accelerated by the use of powerful GPU cards [20] and the development of high-level software frameworks such as Theano [21], Caffe [22], Keras [23], TensorFlow [24] and PyTorch [25].

A systematic review of skin lesion classification using deep network models can be found in Brinker et al. [26]. Three aspects can be highlighted for understanding the problem of skin lesion classification. First, convolutional neural networks (CNNs) are currently the dominant architecture, providing state-of-the-art results. Second, most of the studies transfer the weights from networks trained on ImageNet to the target task in order to speed up training. Third, it is difficult to make a comparison and replicate results found in the literature in view of the diversity of datasets (some of which are proprietary).

A landmark study based on CNNs was published in 2017 by Esteva et al.[27], aiming to diagnose keratinocyte and melanoma cancer. These authors used a knowledge transfer strategy based on the pre-trained InceptionV3 model on which they built a new classifier. The performance of the network was measured by comparing the results of the diagnosis with those obtained by 21 dermatologists, demonstrating a level of performance comparable to these experts. The network used a very large set of labelled images in order to achieve high accuracy. This dataset of 129,450 samples was a combination of multiple sources, both proprietary and publicly available.

In another remarkable study, published in Annals of Oncology, Haenssle and colleagues [28] compare the results of a computer algorithm with those obtained by 58 expert dermatologists in diagnosing melanoma. More than 100,000 images of benign and malignant skin cancers, along with the diagnosis for each image, were used to train the model. On average, the deep neural network correctly detected 95% of melanomas when working with dermoscopic images, against the 86% obtained by dermatologists.

In the meantime, several combined efforts from the academia and industry tend to promote the application of digital skin imaging technologies. One of the most important initiatives is the International Skin Imaging Collaboration (ISIC) project. The ISIC project aims to contribute to improving the detection of melanoma and reducing the associated mortality [29]. Since 2016, this initiative has conducted an annual challenge for developers of artificial intelligence (AI) algorithms in the diagnosis of melanoma [30]. At the same time, the ISIC's repository of dermoscopic images has become a fundamental support for the research community. As a result of this effort, submissions to ISIC challenges have also been exploiting new concepts that improve the skin lesion diagnosis model's performance.

Part 3 of the ISIC'2019 challenge is dedicated to the development of a classifier capable of distinguishing between 7 different types of skin cancer. The ranking was made based on their normalized multiclass accuracy. The top 3 submissions had balanced accuracies of about 88,5%, 88,2%, 87,1% respectively and were all submitted by Metaoptima (the company behind Dermengine) [31]. To train those models, they used the provided dataset along with proprietary data. Additionally, they augmented the training data by performing random horizontal flips, random rotations, changes in brightness, saturation, and contrast. They used transfer learning from several pre-trained models trained on ImageNet (such as InceptionV3 or ResNet) and then ensembled the best performing ones [31].

The 2019 version of the ISIC challenge asked participants to classify dermoscopic images among nine different diagnostic categories, one of the classes being unknown. Similarly to the 2018 version, participants could use their own data to improve the network's performance that was ranked based on a metric [30]. The results turned out to be quite promising, with the best submission, posted by Geesert et al. [32], scoring 92.6% accuracy. Gessert's et al. approach to preprocessing was to crop the images, perform image binarization, apply shades of gray color constancy and, finally, resize the images. Data augmentation is also applied by randomly changing brightness, contrast, rotation, scale, shear and flip. They used two different input strategies: the first takes a random crop from the preprocessed image, while the second randomly resizes and scales the image when taking a crop from the preprocessed one. A transfer learning approach is used relying on EfficientNets trained on a combination of multiple datasets, including the HAM10000 [33]. The final prediction is made using an ensemble which contained the best performing models.

Although transfer learning is the most common approach, end-to-end learning can make sense in specific contexts, such as when data and computational resources are not scarce. In 2019, Ly et al. [34], trained multiple models from scratch with the intention of deploying such models for offline usage in smartphones. They justified this decision by arguing that using pre-trained models with large neural network architectures requires a lot more parameters than models trained from scratch. Their best model attained 86% accuracy, significantly better than other transfer learning approaches, while being much more compact. However, they used a huge dataset titled "PHDB" which was composed of multiple datasets with 80,192 labeled images, which explains the high performance.

III. Challenges in Deep Learning Classification

Despite the results achieved for classification of skin lesions, there are many challenges when training deep models with many layers composed of a huge number of adaptive parameters. This section highlights how deep learning methods are tackling current challenges in skin lesion classification.

A. Convolutional Neural Networks

Deep learning architectures can be seen as an improvement of traditional artificial neural networks translated into the largest number of processing layers. This is the main characteristic that allows them to learn complex nonlinear representations of data at lower and higher levels of abstraction [13], avoiding the error-prone and time consuming part of manual feature engineering. Notably, convolutional neural networks (CNN), a class of feedforward ANN originally designed for computer vision problems, dominates with the best results on varying image classification tasks. They still retain the core concepts of ANNs, but add different concepts which distinguish them from conventional networks.

In a CNN, each neuron in the first hidden layer will be connected only to a small region of the input neurons. Local connectivity aims to explore spatial locality in images. A second difference is the fact that weights and biases are shared across the hidden neurons reducing the number of parameters. Finally, pooling layers simplify the information in the output from the convolutional layer by removing unnecessary information. A common pool layer is max pooling which provides a way to know if a given feature is found anywhere in a region of an image. As result, CNNs become well adapted to translations and rotations, contributing to the generalization of features.

Over the years, several CNN architectures have been developed and tested against state-of-the-art benchmark challenges, such as the Imagenet Large Scale Visual Recognition Challenge (ILSVRC) [35]. In 2012, Krizhevsky et al. [36] submitted, for the first time, a CNN architecture (AlexNet) which outperformed hand-crafted feature learning on the ImageNet. It contained eight neural network layers, five convolutional and three fully-connected layers. This laid the foundation for traditional CNNs: a convolutional layer followed by an activation function and a max pooling operation.

Following AlexNet main ideas, the VGGNet[37] was created and became quite popular by winning the 2014s ILSVR. This architecture proved that representation depth is beneficial for the classification accuracy. This CNN architecture uses the traditional convolutional network, but with an increased depth along with smaller receptive fields. The variation of this network with 16 weight layers (VGG16) is composed by multiple sets of convolutional layers followed by pooling layers that build more abstract features. At the end, a fully connected structure converts the results of the convolution into a label. In line with this, Google introduced the concept of branching within a layer through inception blocks which allows abstraction of features at different spatial scales [38].

The advances made with CNNs are associated with the exploration of interesting ideas such as the use of different activation and loss functions, hyperparameter optimization and regularization. However, the latest advances in the ability to automatically learn representations from data has been the result of architectural innovations. From 2015, residual neural networks (ResNets) gained great popularity. This is an ANN inspired by the characteristics of pyramidal cells in the cerebral cortex [39]. ResNets use the concept of skip connections for the training of deep CNNs. Ongoing improvements to CNNs aim to make them scalable to large, heterogeneous and multiclass problems. Therefore, the focus moved away from parameter optimization to improved architectures. A more complete discussion of the different architectures, algorithms and applications can be found elsewhere [40], [41], [42], [43].

B. Data and Computational Resources

One important challenge in applying CNN to medical imaging is the large amount of labeled data required in order to achieve a performance that surpasses the shallow neural networks. For general image recognition problems, datasets such as ImageNet [44], containing over 14 million samples with over 20,000 classes, are usually used and serve as a benchmark. However, access to most medical data presents several difficulties, such as the fact that the data are often proprietary, not properly annotated and privacy concerns exist.

In addition to the number of samples, insufficient diversity of the available datasets, imbalanced datasets and different types of bias are important challenges that can hinder progress. For example, in the field of skin lesion classification the melanoma disease is often under represented, while the healthy label is over represented. If not taken into account properly, the class imbalance would bias a model to predict the healthy-label. Furthermore, situations where the diversity is insufficient may lead to redundant and uninformative training samples.

The contributes of the International Skin Imaging Collaboration (ISIC) have been relevant in providing an open source public access archive of skin images. This publicly available archive has been used for teaching purposes and for the development of automated skin lesion diagnosis systems [29]. At the same time, the challenges proposed annually around a specific dataset facilitate better comparisons between new and existing solutions by standardizing evaluation criteria. In fact, the difficulties in monitoring and comparing progress in the field are the result of several factors, among which stand out non-standardized evaluation metrics, the use of different datasets, and differences in how learning tasks are framed [26].

Two additional problems need to be addressed in future research. First, an automated diagnosis may require more than just a medical image. Demographics and prior medical history are examples of data used in medical practices. However, gaining access to these data and linking them to the images will require additional research efforts. Second, training deep learning models usually requires high computational requirements. This is the reason why most of deep learning frameworks take advantage of GPUs through the CUDA platform. However, for many research groups such computational power might be inaccessible. In such cases, it could be better to take advantage of cloud services to train the models.

C. Transfer Learning

As previously mentioned, supervised learning using deep neural networks requires large amounts of data and computational power. These requirements apply mainly when a deep model is trained from scratch. Even with an appropriate dataset and high computational resources, the training process can take a long time, especially while debugging the network to determine a good model fit.

In this context, transfer learning [45], [46], [47], [48] strategies are usually used to solve image classification problems [34], while minimizing the effects of the aforementioned problems. Transfer learning is a method of reusing a pre-trained model's knowledge for another related task. This means to carry the parameters from a pre-trained model based on the popular ImageNet dataset and using them to re-train the model with a different purpose.

In CNNs, hidden layers closer to the input output generic features like shapes and curves, while hidden layers closer to the output layers build more abstract features such as, for example, faces. In order to adapt the pre-trained models to a different task, it is usual to extract, from the pre-trained model, the parameters up to some layer while freezing some or no portion of those layers. As the layers near the input layer output generic features, their parameters are usually extracted and potentially frozen, while hidden layers near the output layer are usually not extracted and not frozen, because they output more abstract problem specific features. Additionally, one can expand the original pre-trained model architecture with their own classifier on top of it, in order to adapt the model into the specific task.

D. Training Deep Networks

During the training phase, it is important to fine tune the model aiming to achieve both accurate predictions from the training data, while providing good generalization to new data. The bias and variance trade-off is a well known problem in deep learning. On the one hand, the bias of a model is the error caused by the assumptions made to approximate the model to the true predictions. On the other hand, the variance of a model measures the sensitivity to small fluctuations in the training set. A good trade-off between bias and variance is required in order to avoid both underfitting and overfitting.

An underfit model does not perform well even on the training data, and, therefore, it has a high bias and a low variance. A common problem is to produce a model that performs well on the training data, but that generalizes poorly to new data [49]. In this case, it can be said that the model overfits and, therefore, has a low bias and a high variance. In order to evaluate whether a model is underfitting or overfitting, state-of-the art metrics may be used to help describe what is happening during training. In what concerns overfitting, different solutions to overfitting have been proposed and tested over the years. One common way of dealing with this problem is through regularization techniques, i.e., any technique used to improve model generalization. For example, L1 and L2 regularizations attempt to create less complex models [50], while techniques such as dropout reduce complex co-adaptations among neurons [51].

Another very effective method to minimize the overfitting problem is through a concept called data augmentation [52], [53]. The main idea behind this concept is to expand the training data by applying operations that reflect the real-world variation. The simpler approach is to apply geometric transformations, such as translations, rotations or flips to existing samples to create diversity and to increase the dataset's size. Another more complex approach is to synthetically create new images based on some original dataset (generative models) through methods such as generative adversarial networks (GANs) [54], [55].

In deep learning, the performance of a model is highly dependent on its training dataset. Datasets can easily cause the network to overfit whenever they do not provide proper real world examples, the required diversity and number of examples. There are several dermatological datasets and atlases available with annotation information about skin lesions, such as PH2 [56], HAM10000 [33] and BCN20000 [57] are among the publicly available datasets most commonly used by the scientific community. The HAM10000 dataset is considered a benchmark for the domain skin lesion classification.

E. Research Directions

Rapid developments in computing infrastructures and novel ways to execute collaborative applications promise to change the way medical care will be provided in near future. Bearing this in mind, Fig. 1 illustrates the research trend behind the future generation of automated diagnosis systems for dermatology. The proposed system consists of two software applications: the one on the patient side is based on a client-server architecture in which the main functionality of the client (mobile app) is to take photos of skin moles and send them to the server. The tasks performed by the server include the medical data archiving and the calculation of feature parameters associated with the diagnosis of suspicious lesions, keeping track of changes to existing moles and detection of new moles. The proposed design assumes regular visits to the dermatologist in order to perform a full skin examination and to allow an updated identification of the most worrisome lesions (i.e., suspicious moles). Then, the examiner tags these moles as being those that should merit more attention when the patient is to perform the skin self-examination.

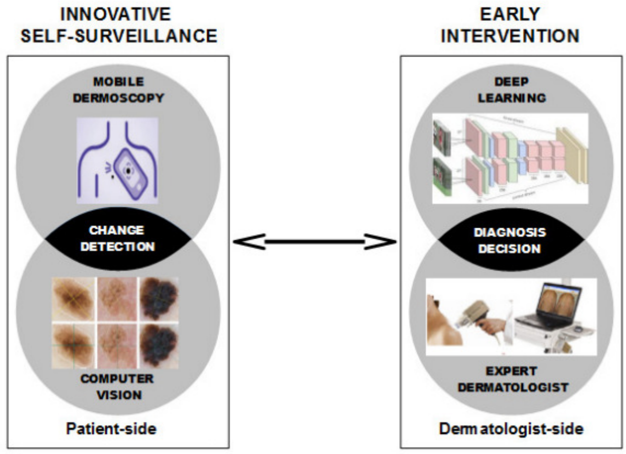

Fig. 1. The proposed computer-aided diagnosis system consists of an innovative self-surveillance application on the patient side and deep learning-based decision support tools on the dermatologist side.

IV. Conclusion

As the incidence of skin cancer rises, there is a clear need for skin lesion diagnosis tools integrated within eHealth applications that provide support for patients and health professionals. At the same time, new advancements on deep learning methods allow near dermatologist performance with high margin for improvement which overshadow other methods. Challenges such as the requirement for large datasets or the high computational requirements hamper the performance of models and need to be addressed before deploying such tools into production. However, promising techniques such as transfer learning and data augmentation prove to minimize the effects of such factors. Finally, it is expected that these issues will become less relevant as more labeled skin lesion data becomes publicly available.

Acknowledgment

This work was funded by National Funds through the FCT - Foundation for Science and Technology, in the context of the project UID/CEC/00127/2019.

References

[1] "Skin Cancer Facts & Statistics - The Skin Cancer Foundation," 2019. [Online]. Available: https://skincancer.org/skin-cancer-information/skin-cancer-facts/

[2] F. Bray, J. Ferlay, and et al., "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[3] W. Stolz, A. Riemann, and et al., "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," *European Journal of Dermatology*, vol. 4, pp. 521–527, 1994.

[4] J. Grus, *Dermoscopy image analysis*. CRC Press, 2015.

[5] N. Mishra and M. Celebi, “An overview of melanoma detection in dermoscopy images using image processing and machine learning,” 2016. [Online]. Available: http://arxiv.org/abs/1601.07843
[6] E. Okur and M. Turkan, “A survey on automated melanoma detection,” *Eng Appl Artif Intel*, vol. 73, no. 2018, pp. 50–67, 2018.
[7] S. Pathan, K. Prabhu, and P. Siddalingaswamy, “Techniques and algorithms for computer aided diagnosis of pigmented skin lesions A review,” *Biomed Signal Proces*, vol. 39, no. 2018, pp. 237–262.
[8] J. Jaworek-Korjakowska and P. Kleczek, “ESkin: Study on the smartphone application for early detection of malignant melanoma,” *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
[9] A. Brewer, D. Endly, and et al., “Mobile applications in dermatology,” *JAMA Dermatology*, vol. 149, no. 11, pp. 1300–1304, 2013.
[10] A. Kassianos, J. Emery, and et al., “Smartphone applications for melanoma detection by community, patient and generalist clinician users: A review,” *Br J Dermatol*, vol. 172, no. 6, pp. 1507–1518, 2015.
[11] A. Zaidan, B. Zaidan, and et al., “A review on smartphone skin cancer diagnosis apps in evaluation and benchmarking: coherent taxonomy, open issues and recommendations pathway solution,” *Health and Technology*, vol. 8, no. 4, pp. 223–238, 2018.
[12] “DermEngine — Visual Search.” [Online]. Available: https://www.dermengine.com/en-ca/visual-search
[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[14] H. Greenspan, B. Van Ginneken, and R. M. Summers, “Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
[15] Z. Hu, J. Tang, and et al., “Deep learning for image-based cancer detection and diagnosis A survey,” *Pattern Recognition*, vol. 83, pp. 134–149, 2018.
[16] G. Litjens, “A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, pp. 60–88, 2017.
[17] D. Erhan, Y. Bengio, and et al., “Why does unsupervised pre-training help deep learning?” *J Mach Learn Res*, vol. 15, pp. 1929–1958, 2010.
[18] Y. Bengio, “Learning deep architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
[19] N. Srivastava, G. Hinton, and A. Krishevsky, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
[20] R. Raina, A. Madhavan, and A. Ng, “Large-scale deep unsupervised learning using graphics processors,” in *Proceedings of the International Conference on Machine Learning*, 2009, pp. 873–880.
[21] F. Bastien, L. Pascal, and et al., “Theano: new features and speed improvements,” 2012. [Online]. Available: http://arxiv.org/abs/1211.5590v1
[22] Y. Jia, E. Shelhamer, and et al., “Caffe: convolutional architecture for fast feature embedding,” 2014. [Online]. Available: http://arxiv.org/abs/1408.5093
[23] F. Chollet and Others, “Keras,” \url{https://github.com/fchollet/keras}, 2015.
[24] M. Abadi, A. Agarwal, and et al., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” 2016. [Online]. Available: http://arxiv.org/abs/1603.04467
[25] A. Paszke, S. Gross, and et al., “Pytorch: An imperative style, high-performance deep learning library,” in *Neural Information Processing Systems Conference (NIPS)*, 2019, pp. 8024–8035.
[26] T. J. Brinker, A. Hekler, and et al., “Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review.” *Journal of medical Internet research*, vol. 20, no. 10, p. e11936, 2018.
[27] A. Esteva, B. Kuprel, and et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
[28] H. A. Haenssle, C. Fink, and et al., “Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists,” *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.
[29] “ISIC Archive.” [Online]. Available: https://www.isic-archive.com/
[30] “ISIC 2019.” [Online]. Available: https://challenge2019.isic-archive.com/
[31] A. Nozdryn-Plotnicki, J. Yap, and W. Yolland, in *International Skin Imaging Collaboration Challenge, MICCAI*.
[32] N. Gessert, M. Nielsen, and et al., in *International Skin Imaging Collaboration Challenge, MICCAI*.
[33] P. Tschandl, C. Rosendahl, and H. Kitler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific Data*, vol. 5, no. 180161, 2018.
[34] P. Ly, D. Bein, and A. Verma, “New compact deep learning model for skin cancer recognition,” in *IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, 2019.
[35] O. Russakovsky, J. Deng, and et al., “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, jun 2017.
[37] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks For Large-Scale Image Recogntinion,” 2015. [Online]. Available: https://arxiv.org/abs/1409.1556
[38] C. Szegedy, W. Liu, and et al., “Going deeper with convolutions,” in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 57, no. 3, 2015, pp. 1–9.
[39] K. He, X. Zhang, and et al., “Deep residual learning for image recognition,” in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[40] W. Liu, Z. Wang, and et al., “A survey of deep neural network architectures and their applications,” *Neurocomputing*, vol. 234, pp. 11–26, 2017.
[41] J. Gu, “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
[42] Q. Zhang, M. Zhang, and et al., “Recent advances in convolutional neural network acceleration,” *Neurocomputing*, vol. 323, pp. 37–51, 2019.
[43] A. Khan, A. Sohail, and et al., “A survey of the recent architectures of deep convolutional neural networks,” 2019. [Online]. Available: http://arxiv.org/abs/1901.06032
[44] J. Deng, W. Dong, and et al., “ImageNet: A large-scale hierarchical image database,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
[45] J. Yosinski, J. Clune, and et al., “How transferable are features in deep neural networks?” in *Neural Information Processing Systems Conference (NIPS)*, 2014.
[46] N. Tajbakhsh, J. Shin, and et al., “Convolutional neural networks for medical image analysis: full training or fine tuning?” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
[47] A. Menegola, M. Fornaciali, and et al., “Knowledge transfer for melanoma screening with deep learning,” in *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, 2017.
[48] A. Menegola, M. Fornaciali, R. Pires, and et al., “Towards automated melanoma screening exploring transfer learning schemes,” 2016. [Online]. Available: http://dx.doi.org/10.1145/3097983.3098021
[49] J. Grus, *Data Science From Scratch*. O’Reilly Media, 2015.
[50] A. Ng, “Feature selection, L1 vs. L2 regularization, and rotational invariance,” in *Proceedings of the International Conference on Machine Learning*, 2004.
[51] G. E. Hinton, N. Srivastava, and et al., “Improving neural networks by preventing co-adaptation of feature detectors,” 2012. [Online]. Available: http://arxiv.org/abs/1207.0580
[52] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” 2017. [Online]. Available: https://arxiv.org/abs/1712.04621
[53] T. C. Pham, C. M. Luong, and et al., “Deep CNN and Data Augmentation for Skin Lesion Classification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10752 LNAI. Springer Verlag, 2018, pp. 573–582.
[54] I. Goodfellow, J. Pouget-Abadie, and et al., “Deep residual learning for image recognition,” in *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, pp. 2672–2680.
[55] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: a review,” *Med Image Anal*, vol. 58, p. 101552, 2019.
[56] T. Mendonca, P. Ferreira, and et al., “Ph2 - a public database for the analysis of dermoscopic images,” in *Dermoscopy Image Analysis*, M. Celebi, T. Mendonca, and J. Marques, Eds. CRC Press, 2015.
[57] M. Combalia, N. Condella, and et al., “Bcn20000: Dermoscopic lesions in the wild,” 2019.