

Skin Lesion Diagnosis Using Convolutional Neural Networks

Daniel Alonso Villanueva Nunez ^a Yongmin Li ^a

^a*Brunel University London, United Kingdom*

Abstract

Cancerous skin lesions are the most common malignancy detected in humans, which, if not detected at early stage, may lead to death. Therefore, it is extremely crucial to have access to accurate results at early stages to optimize the chances of survival. Unfortunately, accurate results are only obtained by highly trained dermatologists which are not accessible to most people, especially in low-income and middle-income countries. Artificial Intelligence (AI) seems to tackle this problem because it has proven to provide equal or better diagnosis than healthcare professionals.

This project collects state-of-the-art techniques for the task of image classification from other fields and implements them in this project. Some of these techniques include mixup, presizing, test time augmentation, among others. These techniques were implemented in three architectures: DenseNet121, VGG16 with batch normalization, and ResNet50. Models are built with two purposes. First to classify images into seven categories – melanocytic nevus, melanoma, benign keratosis like lesions, basal cell carcinoma, actinic keratoses and intraepithelial carcinoma, vascular lesions, and dermatofibroma. Second to classify images into benign or malignant. Models were trained using 8012 images and its performance was evaluated using 2003 images. This model is trained end-to-end from image to labels and does not require the extraction of hand-crafted features.

Key words: Skin cancer, skin lesion, medical imaging, deep learning, Convolutional Neural Networks, DenseNet, ResNet, VGG

1 Introduction

Skin Cancer is the most common malignancy detected in humans. Some categories of cancerous skin lesions are dangerous because they can potentially spread all over the human skin, and lead to death if not cure in time. One example is melanoma, which is the most dangerous type of skin cancer, and it is expected to appear in 97,920 people in the United States by the end of

2022 according to the U.S. Cancer Statistics Centre [43]. 98% are the chances of a 5-year survival but it reduces to 18% once it reaches other organs. For all these reasons, it is important to diagnose malignant skin lesions at an early stage, to permit treatment before metastasis occurs. Malignant skin lesions are first diagnosed by visual inspection and potentially continued by dermoscopic analysis, a biopsy and histopathological examination [19]. Dermoscopic is a non-invasive skin imaging approach which permits visualize zoomed versions of the subdermal structures and surface of the skin [36]. Non-invasive procedures are always preferred because they do not destroy the lesion and increases the possibility to monitor its evolution [31]. Examination of these images are time-consuming, tedious and require domain-knowledge. Furthermore, the precision and reproducibility of the diagnosis is highly dependent on the dermatologist's experience. Some studies show that diagnostic's precision on dermoscopic imaging decreases when they are used by inexperienced physicians [3]. To help doctors make more accurate decisions, computer-aided diagnosis (CAD) systems for skin lesions are used [37]. These CAD systems are able to help doctors make better decisions by detecting lesion borders, removing noise, among others. However, the process of making an appointment to see the doctor for the skin lesion examination, then proceeding with the dermoscopic procedure, and waiting for the CAD system to extract features for assisting the dermatologists on examining the type of lesion, is long. What if there could exist a way to access highly experienced dermatologists from the commodity of your home that can provide highly accurate results by just using your phone?

Artificial intelligence has proven to surpassed human capabilities. For instance, studies from University Hospitals Birmingham shown that AI systems are able to outperform healthcare professionals when it comes to diagnose diseases [30]. Furthermore, it is expected in 2023 that 4.3 billion people will own a smartphone (Alsop, n.d.). Therefore, being able to place these AI systems on people's phones will immensely benefit millions of people who do not have access to highly trained physicians. In 2017, Esteva et al. reported the first work on the usage of deep learning convolutional neural networks (CNNs) to classify malignant lesions on the skin that could outperform 21 board-certified dermatologists [19]. Esteva et al.'s work was innovative because it did not need extensive pre-processing on images such as segmentation or extraction of visual features. This study presents a system, built using state-of-the-art techniques in deep learning, and that does not require the extraction of hand-crafted features, that is able to classify seven types of skin lesions as well as classify lesions into benign and malignant.

The aim of this project is to build a skin lesion image classifier that is able to identify seven types of skin lesions and whether it is benign or malignant, using state-of-the-art techniques and that does not require hand-crafted features. The rest of the paper is organised as follows: A literature review is provided

in Section 2. The methods are described in Section 3. Experiments and results analysis are presented in Section 4. Conclusions are drawn in Section 5.

2 Background

Artificial intelligence has revolutionized the world of healthcare and medical imaging applications, where various methods such as probabilistic modelling [29,27,28], graph cut [42,39,38,41,40], level-set [46,45,47,49,48,13,8–10,12,14,6,11,7] have been investigated. Recently, convolutional neural networks (CNN), a technique used in deep learning, have been found to bring the most progress on the automation of medical imaging diagnostic. A plethora of studies have shown that CNN models match or exceed doctors’ diagnostic performance [33–35].

For example, Esteva et al. built a model which performance was on par with 21 board-certified dermatologists. All skin images were biopsy-proven and tested in two binary classification cases: keratinocyte carcinoma versus benign seborrheic keratoses; and melanoma versus melanocytic nevi. The model obtained an AUC of 0.96 for carcinoma and an AUC of 0.94 for melanoma. To build this model, Esteva et al. collected images from three places: the Edinburgh Dermofit Library, the Stanford Hospital, and the ISIC Dermoscopic Archive [19]. In the test and validation dataset, blurry images and long-distance images were deleted, yet they were kept during training. The InceptionV3 architecture with pre-trained weights from ImageNet was used. During training, a global learning rate equal to 0.001, and a decay factor of 16 every 30 epochs was implemented. RMSProp was the function used to optimize the model. Images were randomly rotated between 0° and 359°. Random cropping and vertically flip with a probability of 0.5 were also used during data augmentation. After augmentation, images increased by a factor of 720.

Haenssle et al. [22] trained a model which performance was compared with 58 international dermatologists. The model was built to classify images into melanocytic nevi or melanoma. The studied shown that the CNN model obtained better performance than most physicians. The deep learning model obtained a specificity of 82.5%, which is 6.8% higher than the average obtained by the dermatologists . Google’s InceptionV4 was the architecture used to train this model. Han et al. [23] used the ResNet-152 architecture to build a model to detect 12 different skin diseases: dermatofibroma, pyogenic granuloma, melanocytic nevus, seborrheic keratosis, intraepithelial carcinoma, basal cell carcinoma, wart, hemangioma, lentigo, malignant melanoma, actinic keratosis, and squamous cell carcinoma. The model was compared with 16 dermatologists and obtained an AUC of 0.96 for melanoma.

Brinker et al. [4] utilized a pre-trained on ImageNet ResNet50 architecture to train images from the ISIC dataset for a binary classification task to predict melanoma versus nevus. 92.8% sensitivity and 61.1% specificity were achieved as compared to 89.4% mean sensitivity and 64.4% mean specificity obtained by 145 board-certified dermatologists from 12 German university hospitals. To train the model, two datasets were used: the HAM10000 dataset and the public ISIC image archive. High learning rates were used at the end of the architecture while small learning rates were used at the beginning. This technique is called differential learning rates. The scheduler used to train the model was the cosine annealing method. The experiment consisted in 13 epochs.

Fujisawa et al. [20] built a model that outperform 13 board-certified dermatologists. The model was built to classify skin lesions into malignant and benign. The accuracy of the model was 92.4%, the sensitivity was 96.3% and the specificity was 89.9%. Meanwhile, the dermatologists obtained an overall accuracy of 85.3%. The architecture chosen was GoogLeNet using pre-trained weights. The dataset utilized belongs to the University of Tsukuba Hospital. Images were trimmed to 1000x1000 pixels from the centre. Each image was randomly rotated by 15 degrees leading and saved as a different figure. This led to an increase of figures by a factor of 24. When the images were passed through the CNN, blur filters and changes in brightness were applied.

Other studies include that Eltayef et al. [18,15–17] reported various methods in combination with particle swarm optimization and markov random field for lesion segmentation in dermoscopy images.

As it can be seen, there are a plethora of studies that show that deep learning can match or outperform physicians when it comes to the detection of skin lesions. However, not all studies on the usage of CNNs for skin disease detection compared their results with physicians. Nonetheless, they can provide useful information when it comes to the techniques used to train the models.

Majtner et al. [32] achieved 0.801 accuracy using VGG16, 0.797% accuracy using GoogleNet, and 0.815% accuracy by assembling both of them. Models were built to classify the same seven types of skin lesions that this project is attempting to do. The ISIC 2018 dataset was used. For both architectures, only the last layers were trained. To deal with the unbalanced data, each image was horizontally flipped, which led dataset to increase by a factor of two. Then, a rotation factor was assigned to each image to balance the data.

Xie et al. [50] obtained 94.17% accuracy using the ISIC 2018 dataset to classify images into benign or malignant which is considered as a binary classification task. The best AUC reported was 0.970. The data to train this model was obtained from the General Hospital of the Air Force of the Chinese People's Liberation Army. Different methods to extract features from the images were

used. For instance, the authors built a self-generating neural network (SGNN) segmentation model to remove hairs from the image. Shape features were not used since many images are not complete. Five colour features were used: RGB features, LUV histogram distances, colour diversity, centroidal distances. Five statistical texture descriptors were computed from a grey level co-occurrence matrix – based texture. They were regional energy, entropy, contrast, correlation, and inverse difference moment. Finally, 6 concavity features were calculated using the lesion convex hull which describes degrees of concavities. To train the model, three artificial neural networks were ensembled. The experiment run with a max of 1000 epochs, sigmoid activation functions were used, and a learning rate of 0.7 was set. Bisla et al. obtained 81.6% accuracy when using the ISIC 2017 dataset to predict three classes of skin lesion: melanoma, nevus, and seborrheic keratosis [5]. A hair removal algorithm was implemented for cleaning images. U-Net algorithm was utilized to segment the lesion area from the skin. To tackle the unbalanced problem, generative adversarial networks (GANs) were utilized to generate more images from the labels with less instances. Horizontal flipping, vertical flipping and random cropping were the methods used for data augmentation. The ResNet50 pre-trained on ImageNet architecture was used.

Almaraz-Damian et al. [2] used transfer learning with different architectures for a binary classification model (benign vs malignant) utilizing the HAM10000 dataset with the architectures of VGG16, VGG19, Mobilenet v1, Mobilenet v2, ResNet50, DenseNet-201, Inception V3, Xception respectively. A mix between handcrafted and deep learning features were employed. To balance the data, a SMOTE oversampling technique was implemented.

Jojoa et al. [26] trained the pre-trained ResNet152 using the dataset for ISIC 2017 to predict skin cancer into benign or malignant. Images were pre-processed by extracting the lesion from the image using Mask R-CNN. From all the experiments done tweaking the learning rates and number of epochs, the best model obtained had an accuracy of 90.4%.

Aljohani et al. [1] used transfer learning to predict melanoma or no melanoma on dataset ISIC 2019 by experimenting different architectures including DenseNet201, MobileNetV2, ResNet50V2, ResNet152V2, Xception, VGG16, VGG19, and GoogleNet. All experiments were run with 50 epochs, but early stopping was used. To overcome the imbalanced dataset, the ImageDataGenerator from TensorFlow was implemented. The best architecture was the GoogleNet with 76% accuracy with 29 epochs.

3 Methods

In this project, we have selected three different types of network architectures to perform the task of skin lesion recognition: the VGG, ResNet and DenseNet.

3.1 VGG

VGG gives reference to a group in Oxford University named Visual Geometry Group [44]. This architecture was built with the aim to understand the relationship between the depth of an architecture and its accuracy. In order to allow the usage of a long structure and avoid using a large quantity of parameters, small 3x3 convolutional kernels were used. At the end of the architecture, three fully connected layers are placed. They are two VGGs normally used: VGG16 which has 16 layers and VGG19 which has 19 layers. VGG16 has 138 million trainable parameters. This project will be using a variant of VGG16 which is VGG16 with batch normalization because it allows faster training and makes the layers' inputs more stable by re-centring and re-scaling them.

3.2 ResNet

When deep networks are able to commence converging, a degradation problem appears. With the architecture depth expanding, accuracy saturates leading to fast degradation. Surprisingly, overfitting is not the root of this degradation, and attaching extra layers to an appropriate deep model results to higher training error [24]. Nevertheless, this problem is tackled by the introduction of deep residual learning framework which is the building block of the ResNet architecture. Instead of passing an input x through a block $F(x)$ and then passing it through a ReLU activation function, the output of $F(x)$ gets added with the input before it gets passed to the next ReLU Activation Function. There are many types of ResNet. Some of its variants are ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152. The number refers to the number of residual blocks stacked.

This project will be using both ResNet18 and ResNet50. ResNet18 has around 11 million trainable parameters whereas ResNet50 has over 23 million.

3.3 DenseNet

Improving model architecture is not as easy as just stacking new layers because they normally lead to the vanishing or exploding gradient problem [21]. Nonetheless, DenseNet solves this problem by simplifying the connection between layers which allows the no learning of redundant feature maps. Furthermore, feature reuse is achieved by taking full advantage of this connectivity, this is known as collective knowledge [25]. The difference between DenseNet and ResNets is that rather than adding the input with the output of the feature maps, they get concatenated. Just like ResNet is composed of residuals blocks, DenseNet is composed of dense blocks.

It exists four variants of DenseNet which are DenseNet121, DenseNet161, DenseNet169, and DenseNet201. For this project, DenseNet121 will be used which is composed of 120 convolutions, four average pooling layers, and 8 million trainable parameters.

4 Experiments

4.1 Dataset

For this project, the HAM10000 dataset is used (<https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000>). The dataset consists of 10,015 dermoscopic images belonging to seven categories which can be seen in Table 1.

The downloaded file from Kaggle contains all the images in two folders named “HAM10000_images_part_1” and “HAM10000_images_part_2”. It also contains a csv file named “HAM10000_metadata.csv” which maps the name of each image file to its label. The csv file has metadata for each image such as the region of the lesion (e.g., ear, head), age of the patient, and sex. An empty folder named “classes” was created containing subfolders with the name of each label. Then, the images from “HAM10000_images_part_1” and “HAM10000_images_part_2” were map to the subfolders from the folder “classes”. This was done with help of the csv file “HAM10000_metadata.csv”. This was done such way because that is the format FastAI, which is the framework used for this project, uses.

The data was split into 80% training and 20% validation set which means 8012 images for training and 2003 images for validation. A seed equal to 101096 was used to make this work reproducible. The dataset was only separated in

Label	Description	Amount
nv	melanocytic nevi	6,705
mel	melanoma	1,113
bkl	benign keratosis – like lesions (solar lentigines / seborrheic keratoses and lichen – planus like keratoses)	1,099
bcc	basal cell carcinoma	514
akiec	actinic keratoses and intraepithelial carcinoma / Bowen’s disease	327
vasc	vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and haemorrhage)	142
df	dermatofibroma	115

Table 1
Skin Lesion Labels and Amounts

training and validation because some classes have too little images for having a third split for the testing set.

4.2 Data Augmentation

Many deep learning practitioners apply data augmentation operations after resizing down the images to a standard form. However, this practice could lead to spurious empty zones, degraded data, or both of them. One example is when an image is rotated 30 degrees and the corners are filled with emptiness which bring not added information to the model. This can be seen in figure 19. Therefore, to tackle this issue, presizing adopts two steps. During the first step, presizing resizes the images to much larger dimensions compared to the training target dimensions. This allows images to have spare margin to permit more augmentation transforms on the inner regions without generating empty zones. During the second step, it joins all common augmentation operations (including a resize to the final target dimensions) and computes them on the GPU (Gugger Howard, 2020).

For this work, presizing will be used for data augmentation. First, all images will be expanded to a size of 460x460 pixels in RAM. Then, in the GPU, for each batch, first different augmentations will be stacked to the images. Second, images will be resized to 224x224 pixels and then normalized. That

means that for every batch, the computer is seeing a different combination of transformations. Nonetheless, these images disappear at the end of each batch. Then, in the next batch, a different combination of transformations is applied. This process is performed repeatedly until the training ends. Therefore, if a batch size equal to 64 is used to run 8012 images for 100 epochs, then the model will be seeing approximately 12,500 images, however, no new images are stored in RAM or in memory.

4.3 Results and Analysis

The classification results measured by precision, recall and F1-score for the three network architectures of DenseNet, VGG and ResNet are shown in Figure 1. Also, the confusion matrices of all the skin lesion labels for each individual network models are presented in Figure 2.

DenseNet121			
Classification Report			
	Precision	Recall	F1-Score
akiec	0.90	0.72	0.80
bcc	0.91	0.87	0.89
bkl	0.88	0.75	0.81
df	0.90	0.82	0.86
mel	0.83	0.69	0.75
nv	0.93	0.98	0.95
vasc	0.96	0.96	0.96
Accuracy			0.91

VGG16			
Classification Report			
	Precision	Recall	F1-Score
akiec	0.93	0.66	0.77
bcc	0.92	0.89	0.91
bkl	0.89	0.80	0.84
df	0.88	0.68	0.77
mel	0.83	0.67	0.74
nv	0.92	0.99	0.96
vasc	1.00	0.96	0.98
Accuracy			0.91

ResNet50			
Classification Report			
	Precision	Recall	F1-Score
akiec	0.93	0.67	0.78
bcc	0.93	0.91	0.92
bkl	0.86	0.81	0.83
df	0.90	0.82	0.86
mel	0.81	0.74	0.78
nv	0.94	0.98	0.96
vasc	0.93	1.00	0.96
Accuracy			0.92

Fig. 1. Classification results measured by precision, recall and F1-score for the three network architectures of DenseNet, VGG and ResNet.

5 Conclusions

In this project, we have built skin lesion image classifiers to identify seven types of skin lesions and whether it is benign or malignant. This task was accomplished by using state-of-the-art techniques in convolutional neural networks for image classification. The best model is assessed by the one who obtains the best recalls for the labels akiec, bcc and mel, and the best sensitivity when the model is tweaked for binary classification. The best model found was ResNet50 which achieved a recall for label akiec of 0.67, a recall for label bcc of 0.91, and a recall for label mel of 0.74. When the model was tweaked for binary classification, ResNet50 at a threshold of 0.06 was the one which obtained the best results achieving a sensitivity of 0.9235. In addition, all model's performances were boosted by using test time augmentation. Here, ResNet50 again obtained the best results with recall 0.69 for label akiec, recall 0.93 for label bcc, and a recall 0.76 for label mel. However, when the model

		DenseNet121						
		Confusion Matrix						
True Labels	akiec	46	1	10	1	2	4	0
	bcc	1	96	1	1	1	9	1
	bkl	3	3	163	0	18	29	0
	df	0	1	0	18	0	3	0
	mel	1	1	5	0	150	61	0
	nv	0	4	7	0	10	1326	0
	vasc	0	0	0	0	0	1	25
			akiec	bcc	bkl	df	mel	nv
		Predicted						

		VGG16						
		Confusion Matrix						
True Labels	akiec	42	2	11	2	3	4	0
	bcc	1	96	1	0	4	6	0
	bkl	2	2	173	0	13	26	0
	df	0	1	0	15	0	6	0
	mel	0	0	7	0	146	65	0
	nv	0	3	3	0	10	1331	0
	vasc	0	0	0	0	0	1	25
			akiec	bcc	bkl	df	mel	nv
		Predicted						

		ResNet50						
		Confusion Matrix						
True Labels	akiec	43	5	7	1	5	3	0
	bcc	0	100	1	0	2	6	1
	bkl	2	1	175	0	17	21	0
	df	0	0	1	18	0	3	0
	mel	0	0	10	0	162	45	1
	nv	1	2	10	1	14	1319	0
	vasc	0	0	0	0	0	0	26
			akiec	bcc	bkl	df	mel	nv
		Predicted						

Fig. 2. Confusion matrices of all the skin lesion labels for each of the three network architectures of DenseNet, VGG and ResNet.

was tweaked for binary classification the model that achieved the best results was VGG16 at threshold 0.05 with a sensitivity of 0.9540.

References

- [1] K. Aljohani and T. Turki. Automatic classification of melanoma skin cancer with deep convolutional neural networks. *AI*, pages 512–525, 2022.
- [2] J.-A. Almaraz-Damian, V. Ponomaryov, S. Sadovnychiy, and H. Castillejos-Fernandez. Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. *Entropy*, 22:1–23, 2020.
- [3] M. Binder, H. Soyer, H. Kittler, A. Binder, I. Wolf, and H. Kerl. Epiluminescence microscopy. a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *JAMA Dermatology*, 131(9):1160–1168, 1995.
- [4] T. J. Binker, Y. Peng, H. Liu, J. Wang, V. Mahler, I. Vujic, A. Rabinovich, S. W. Dusza, M. A. Marchetti, H. Kittler, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *European Journal of Cancer*, pages 148–154, 2019.

- [5] D. Bisla, K. Hata, P. Reiter, T. Schlegl, A. Uhl, B. Wimmer, and S. Kopp. Towards automated melanoma detection with deep learning: Data purification and augmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16–20, 2019.
- [6] B. Dodo, Y. Li, K. Eltayef, and X. Liu. Graph-cut segmentation of retinal layers from oct images. In *International Conference on Bioimaging*, 2018.
- [7] B. Dodo, Y. Li, K. Eltayef, and X. Liu. Min-cut segmentation of retinal oct images. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 86–99. Springer, 2018.
- [8] B. Dodo, Y. Li, K. Eltayef, and X. Liu. Automatic annotation of retinal layers in optical coherence tomography images. *Journal of Medical Systems.*, 2019.
- [9] B. Dodo, Y. Li, K. Eltayef, and X. Liu. Min-cut segmentation of retinal oct images. In *Cliquet Jr. A. et al. (eds) Biomedical Engineering Systems and Technologies*. Springer., 2019.
- [10] B. Dodo, Y. Li, and X. Liu. Retinal oct image segmentation using fuzzy histogram hyperbolization and continuous max-flow. In *IEEE International Symposium on Computer-Based Medical Systems*, 2017.
- [11] B. Dodo, Y. Li, X. Liu, and M. Dodo. Level set segmentation of retinal oct images. In *International Conference on Bioimaging. Czech Republic.*, 2019.
- [12] B. Dodo, Y. Li, A. Tucker, D. Kaba, and X. Liu. Retinal oct segmentation using fuzzy region competition and level set methods. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 93–98. IEEE, 2019.
- [13] B. I. Dodo, Y. Li, M. I. Dodo, and X. Liu. Simultaneous segmentation of retinal oct images using level set. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 120–136. Springer, Cham, 2020.
- [14] B. I. Dodo, Y. Li, D. Kaba, and X. Liu. Retinal layer segmentation in optical coherence tomography images. *IEEE Access*, 7:152388–152398, 2019.
- [15] K. Eltayef, Y. Li, B. Dodo, and X. Liu. Skin cancer detection in dermoscopy images using sub-region features. In *International Symposium on Intelligent Data Analysis*, pages 75–86. Springer, Cham, 2017.
- [16] K. Eltayef, Y. Li, and X. Liu. Detection of melanoma skin cancer in dermoscopy images. In *In Proc. International Conference on Communication, Image and Signal Processing. Dubai, UAE, .*, 2016.
- [17] K. Eltayef, Y. Li, and X. Liu. Detection of pigment networks in dermoscopy images. In *In Proc. International Conference on Communication, Image and Signal Processing. Dubai, UAE, .*, 2016.
- [18] K. Eltayef, Y. Li, and X. Liu. Lesion segmentation in dermoscopy images using particle swarm optimization and markov random field. In *IEEE International Symposium on Computer-Based Medical Systems*, 2017.

- [19] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542:115–118, 2017.
- [20] Y. Fujisawa, H. Koga, T. Kido, T. Kuwayama, A. Fukunaga, S. Yoshida, H. Matsue, Y. Shirakata, H. Tagami, and M. Tanaka. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *British Journal of Dermatology*, pages 247–248, 2019.
- [21] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.
- [22] H. A. Haenssle, C. Fink, P. Hepp, M. Glaß, M. Sütterlin, M. A. Marchetti, H. Kittler, J. Eberle, E. Schopf, K. Schütze, 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*, 29(8):1836–1842, 2018.
- [23] S. S. Han, H. J. Kim, S. H. Moon, E. J. Kim, S. H. Lee, H. S. Kim, J. W. Yoon, D. H. Kim, S. I. Park, M. G. Kim, et al. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, 138(7):1529–1538, 2018.
- [24] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [25] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. *IEEE*, 2018.
- [26] M. F. Jojoa Acosta, L. Y. Caballero Tovar, B. Garcia-Zapirain, and P. W. Spencer. Melanoma diagnosis using deep learning techniques on dermoscopic images. *BMC Medical Imaging*, pages 1–11, 2021.
- [27] D. Kaba, A. G. Salazar-Gonzalez, Y. Li, X. Liu, and A. Serag. Segmentation of retinal blood vessels using gaussian mixture models and expectation maximisation. In *International Conference on Health Information Science*, pages 105–112, 2013.
- [28] D. Kaba, C. Wang, Y. Li, A. Salazar-Gonzalez, X. Liu, and A. Serag. Retinal blood vessels extraction using probabilistic modelling. *Health Information Science and Systems*, 2(1):2, 2014.
- [29] D. Kaba, Y. Wang, C. Wang, X. Liu, H. Zhu, A. G. Salazar-Gonzalez, and Y. Li. Retina layer segmentation using kernel graph cuts and continuous max-flow. *Optics Express*, 23(6):7366–7384, 2015.
- [30] X. Liu, H.-W. Chih, H. Liu, Y. Yang, X. Gao, H. Xie, J. Chen, Y. Yang, Y. Yang, and Y. Han. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*, 1:271–296, 2019.

- [31] M. Maarouf, M. Cordova, A. Sahu, and S. R. Feldman. In vivo reflectance confocal microscopy: Emerging role in noninvasive diagnosis and monitoring of eczematous dermatoses. *Actas Dermo-Sifiliográficas*, pages 626–636, 2019.
- [32] T. Majtner, J. J. Navarro, J. D. Uribe, J. Montesinos, J. A. Lozano, and J. Pineda. Ensemble of convolutional neural networks for dermoscopic images classification. In *2018 15th International Symposium on Medical Information Processing and Analysis (SIPAIM)*, pages 1–8. IEEE, 2018.
- [33] N. McConnell, A. Miron, Z. Wang, and Y. Li. Integrating residual, dense, and inception blocks into the nnunet. In *IEEE 35th International Symposium on Computer Based Medical Systems*, 2022.
- [34] N. Ndipenoch, A. Miron, Z. Wang, and Y. Li. Simultaneous segmentation of layers and fluids in retinal oct images. In *IEEE Conference on Image and Signal Processing, BioMedical Engineering and Informatics*, 2022.
- [35] N. Ndipenoch, A. Miron, Z. Wang, and Y. Li. Retinal image segmentation with small datasets. In *10th International Conference on Bioimaging*, 2023.
- [36] A. Oakley. dermnetnz.org, 2004.
- [37] F. Ruffano, G. Argenziano, G. Pellacani, F. Lacarrubba, A. Lallas, C. Longo, I. Zalaudek, Z. Apalla, I. Zalaudek, S. Puig, I. Zalaudek, A. Di Stefani, and R. Hofmann-Wellenhof. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. *Cochrane Library*, pages 1–296, 2018.
- [38] A. Salazar-Gonzalez, D. Kaba, Y. Li, and X. Liu. Segmentation of the blood vessels and optic disk in retinal images. *IEEE journal of biomedical and health informatics*, 18(6):1874–1886, 2014.
- [39] A. Salazar-Gonzalez, Y. Li, and D. Kaba. MRF reconstruction of retinal images for the optic disc segmentation. In *International Conference on Health Information Science*, pages 88–99, 2012.
- [40] A. Salazar-Gonzalez, Y. Li, and X. Liu. Automatic graph cut based segmentation of retinal optic disc by incorporating blood vessel compensation. *Journal of Artificial Intelligence and Soft Computing Research*, 2(3):235–245, 2012.
- [41] A. G. Salazar-Gonzalez, Y. Li, and X. Liu. Retinal blood vessel segmentation via graph cut. In *International Conference on Control Automation Robotics & Vision*, pages 225–230, 2010.
- [42] A. G. Salazar-Gonzalez, Y. Li, and X. Liu. Optic disc segmentation by incorporating blood vessel compensation. In *IEEE Third International Workshop On Computational Intelligence In Medical Imaging*, pages 1–8, 2011.
- [43] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1):7–33, 2022.

- [44] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.
- [45] C. Wang, D. Kaba, and Y. Li. Level set segmentation of optic discs from retinal images. *Journal of Medical Systems*, 4(3):213–220, 2015.
- [46] C. Wang and Y. Li. Blood vessel segmentation from retinal images. In *The 20th IEEE International Conference on BioInformatics And BioEngineering*, 2020.
- [47] C. Wang, Y. Wang, D. Kaba, Z. Wang, X. Liu, and Y. Li. Automated layer segmentation of 3d macular images using hybrid methods. In *Proc. International Conference on Image and Graphics. Tianjing, China.*, volume 9217, pages 614–628, 2015.
- [48] C. Wang, Y. Wang, D. Kaba, H. Zhu, Z. Wang, X. Liu, and Y. Li. Segmentation of intra-retinal layers in 3d optic nerve head images. In *Proc. International Conference on Image and Graphics. Tianjing.*, volume 9219, pages 321–332, 2015.
- [49] C. Wang, Y. Wang, and Y. Li. Automatic choroidal layer segmentation using markov random field and level set method. *IEEE journal of biomedical and health informatics*, 2017.
- [50] F. Xie, J. Shen, X. Li, L. Tong, H. Zhang, J. Wang, and W. Li. Melanoma classification on dermoscopy images using a neural network ensembled model. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 36(3):849–858, 2017.