Received: 30 January 2020 | Revised: 18 May 2020 | Accepted: 21 June 2020

DOI: 10.1002/ett.4080

SPECIAL ISSUE ARTICLE

WILEY

Deep learning in medical imaging: A brief review

Sertan Serte[1,2] | **Ali Serener**[1,2] | **Fadi Al-Turjman**[3]

[1]Electrical and Electronic Engineering, Near East University, Nicosia, Turkey

[2]Artificial Intelligence Department, Near East University, Nicosia, Turkey

[3]Research center for AI and IoT, Near East University, Nicosia, Turkey

Correspondence
Sertan Serte, Electrical and Electronic Engineering, Near East University, Nicosia, North Cyprus via Mersin 10, Turkey.
Email: sertan.serte@neu.edu.tr

Abstract

Researchers have used deep learning methods for a human level or better disease identification and detection. This paper reports, in brief, the recent work in deep learning identification of diseases occurring at three unique parts of the human body: the skin, the thorax, and the eye. While earlier reviews reported on the theory, applications, and challenges of such research, what distinguishes this work from the others is the reporting and comprehensive analysis of the key results. In doing so, the paper not only summarizes the major conclusions drawn from them but also emphasizes their weaknesses. The hope is to help the researchers see the big picture in deep learning classification of the diseases of the skin, the thorax and the eye, and guide them to find the right future research direction.

1 | INTRODUCTION

Physicians use different imaging modality techniques, such as MRI, CT, ultrasound, and X-ray to help them detect, diagnose, and cure serious diseases. With help from radiologists, they can interpret a vast number of diseases occurring at distinct parts of a human body, ranging from brain to lungs to skin. Still, the tremendous diversity in disease pathology can make human errors possible.

In search of computer methods to aid in disease detection and recognition, researchers have experimented on convolutional neural networks (CNN) since the mid-90s.[1] Used for hand-written digit recognition at first,[2] CNN's popularity increased with the discovery of deep CNNs in 2012.[3-8] Since then, a growing number of researchers are using CNNs for medical image analysis (such as References 9-17). Some even show that CNNs can give comparable performance to medical doctors.[18]

We may use the Internet of Things (IoT) to connect medical devices in hospitals for the collection of big data. Each hospital can join a central unit this way for the automated detection of diseases on images using CNN. IoT is a crucial tool for reaching data in an instant, and artificial intelligence is essential for modeling for disease detection on images. Together they may form what we call The Internet of Medical Things (IoMT).[19]

CNN's distinct architectures and vast application areas (such as this and References 20-22) have likewise led the research groups to publish a few review papers on using deep CNNs in medical imaging (eg, References 23-27). While these papers focus on the theory, applications, and challenges of deep CNNs, they do not detail the results discovered by the researchers.

This survey paper, yet, aims to fill that gap by reporting the recent advances of deep learning use in medical imaging and analyzing the key results. In doing that, it hopes to offer the reader an opportunity to see the big picture, including any strengths and weaknesses that such research might have. The intention is not to cover the entire application areas in medical imaging but focus on only a few parts of the human body and be thorough. Hence, this survey paper includes

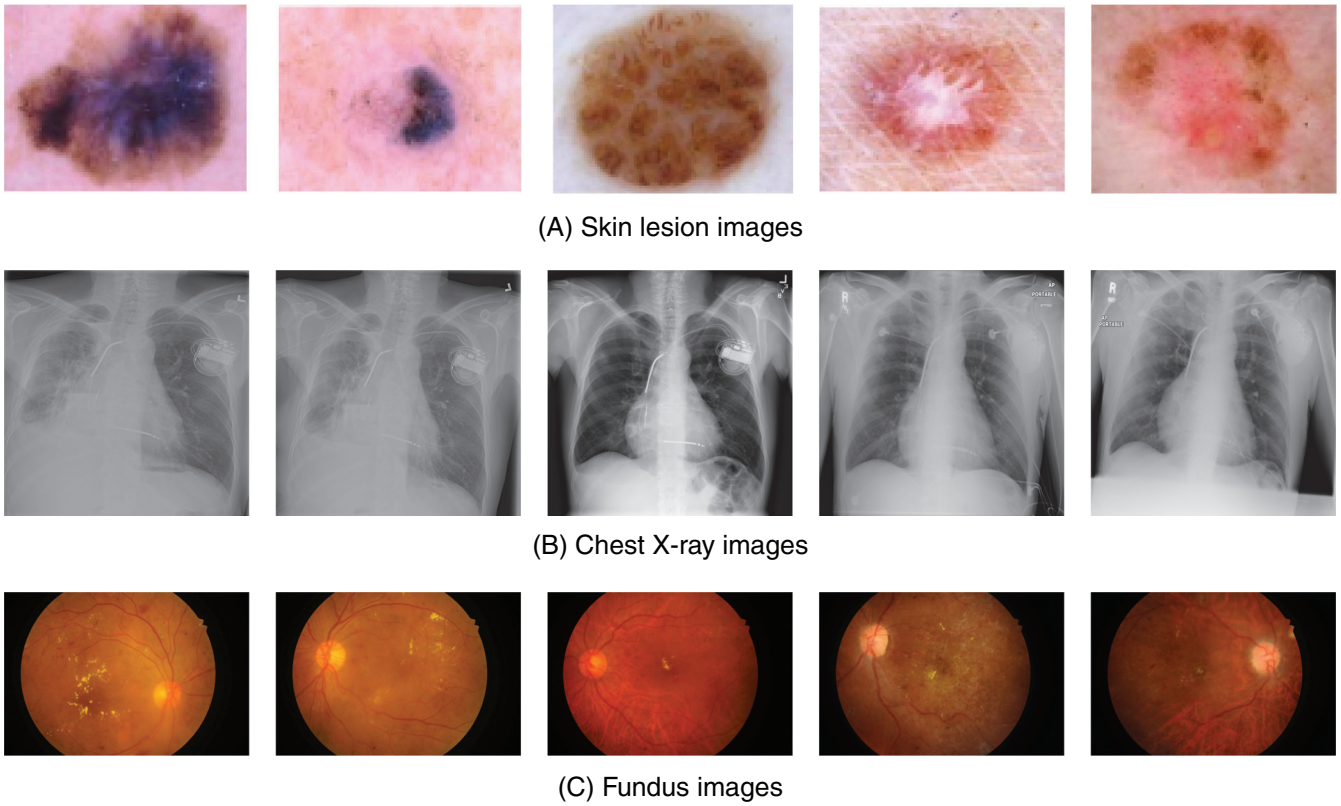

FIGURE 1 Sample images from each medical image category

the performance review of the deep learning methods for skin lesion classification, listing the earlier approaches, and comparing their accuracies. It reports the performances of the deep learning methods for the detection of throat diseases, again, listing the earlier deep learning approaches and comparing their performances. At last, it details the glaucoma recognition using deep learning methods and compares their performances.

Figure 1 shows the medical images of the three unique parts of the human body this paper is focusing on. The first is the human skin lesions. Figure 1A displays a variety of malignant and benign skin lesions. The second focus is on the human chest. Figure 1B displays a variety of chest X-ray images. The third and last focus is the human eye. Figure 1C shows distinct fundus images of the eye, in particular those that have glaucoma.

The organization of this review is in Section 2, we explain in brief a few CNN methods. In Section 3, we review the deep learning approaches for skin lesion classification and single, ensemble, and feature-based deep CNN models for skin lesion detection. Sections 4 and 5 summarize the deep learning methods for throat disease classification and report on the approaches for glaucoma classification. Section 6 gives the details of popular datasets used during the deep learning classification of skin, throat, and glaucoma diseases. Sections 7 and 8 report, analyze, and discuss the key results of researchers when they classified the three aforementioned diseases with deep learning. Section 9 concludes the review.

2 | DEEP LEARNING METHODS

We describe a few deep learning methods in this section.

2.1 | AlexNet deep network

AlexNet[3] consists of eight layers. Input images go through five convolutional layers and then three fully connected layers. Images are convolved with 96 filters of size 11×11 in the first convolution layer. After, the output of the first convolutional

layer is used as input to the second convolutional layer. Then the pooling is performed on the output of the second convolutional layer. This input is filtered with 256 kernels of size 5×5 in the second convolutional layer. This convolution process is done for third and fourth and fifth convolutional layers. After, fully connected layers are constructed.

2.2 | VGG-16 deep network

VGG Net[4] consists of 16 and 19 layers. The input images pass through two 64 filters with 3×3 sizes. Then, another two 128 sets of filters of 3×3 sizes are utilized to process the output of the first filtering. This convolution process is utilized with 256 and 512 sets of filters. Finally, fully connected layers are created.

2.3 | GoogleNet deep network

GoogleNet[5,6] consists of 20 layers. This deep learning model is based on inception modules. Inception modules allow parallel filtering of the layers. This is network is much deeper than VGG and AlexNet deep learning models.

2.4 | ResNet deep network

ResNet-18, ResNet-32, ResNet-50, and ResNet-152[7] are the mainly used models for medical image classification. These models are based on the residual learning method. Residual learning does not allow error accumulation on the convolution layers but enables a better representation of the content in the convolution layers.

3 | DEEP LEARNING IN SKIN LESION RECOGNITION

There are three different classification approaches for skin lesion classification. They are single deep learning methods, ensemble deep learning methods, and feature-based deep learning methods. The following sections review some of the recent research covering these methods.

3.1 | Single deep learning methods

Esteva et al[28] used GoogleNet Inception v3 deep learning model for skin lesion classification. They trained the deep learning method using 129 450 clinical images to recognize keratinocyte carcinomas and seborrheic keratoses, malignant melanomas, and benign nevi. The authors did two classification tasks. First, they performed keratinocyte carcinoma and seborrheic keratoses skin lesions classification. Second, they classified malignant melanomas and benign nevi detection lesions.

Han et al[29] deploy another model for twelve skin lesion classification. They based their proposed model on a fine-tuned ResNet-152 deep learning model. The created model allows basal cell carcinoma, squamous cell carcinoma, intraepithelial carcinoma, actinic keratosis, seborrheic keratosis, malignant melanoma, melanocytic nevus, lentigo, pyogenic granuloma, hemangioma, dermatofibroma, and warts classification.

Menegola et al[30] made use of ResNet-101 and Inception-v4 deep learning models for melanoma detection. They recognized melanoma among seborrheic keratoses and benign nevi skin lesions. The authors used several external datasets for creating deep learning models.

Zhang et al[31] proposed attention-based residual convolution neural networks for melanoma recognition. This new deep network is ResNet-50 deep network modification. They enhanced residual learning by attention modeling. The performance evaluation of this model outperforms the classic ResNet-50 residual-based model.

Fujisawa et al[32] used GoogLeNet CNN model for skin lesion classification. First, this method grouped malignant and benign types of skin lesions. Second, the method further divided malignant and benign types into epithelial and melanocytic groups. Likewise, they further divided each group into more subcategories. As a result, they could recognize twenty-one classes of skin lesions.

Rezvantalab et al[33] used DenseNet-210, ResNet-152, Inception v3, Inception ResNet v2 deep CNNs for melanoma classification. These models detect melanoma, basal cell carcinoma, actinic keratosis, benign keratosis, melanocytic nevi, dermatofibroma, vascular lesions, and atypical nevi. The ResNet-152 models outperform other models for melanoma and melanocytic nevi detection. The Inception v3 model outperforms the other models for actinic keratosis and benign keratosis classification. DenseNet-210 provided the best results for basal cell carcinoma and dermatofibroma classification.

Khamparia et al[19] researched a framework driven by deep learning Internet of Health Things (IoHT) for skin lesion classification. As part of it, they used transfer learning and pretrained VGG19, Inception V3, ResNet50, and SqueezeNet models for benign and malignant skin lesion classification.

Sujitha et al[22] presented a new image compression method for remote sensing using CNN. To balance image quality and the compression efficiency, they used two CNN, one at the encoding and the other at the decoding side.

3.2 | Ensemble deep learning methods

Matsunaga et al[34] used an ensemble of ResNet-50 CNNs for recognizing melanoma and seborrheic keratosis skin lesions. They based this approach on generating several ResNet-50 deep models. Each ResNet-50 model builds on different augmentation data. They fused each of the model probabilities for melanoma and seborrheic keratosis classes for skin class identification.

Harangi[35] used an ensemble of AlexNet, VGG, GoogleNet, and ResNet CNNs and joint model output probabilities for melanoma classification. Four different techniques lead the final probabilistic skin class prediction. The fusion strategies are the sum, product, maximal probabilities, and majority voting. This ensemble approach outperforms single AlexNet, VGGNet, GoogLeNet, and ResNet deep learning methods.

Nyiri and Kiss[36] used DenseNet, ResNet, VGG16, VGG19, Inception V3, and Xception CNNs for skin lesion classification. They predicted skin lesions using each of these CNN and again estimated lesion types by fusing output estimates of each of the models. The study presents two training strategies. One of them is to train the models using original images. Another is to train the models using cropped skin images around the region of interest. Joint CNNs outperform the single CNN for skin lesion recognition.

Nozdryn-Plotnicki et al[37] generated DenseNet, ResNet, Inception V3, and SEResNet deep learning models and fused these models for skin lesion prediction. The author's fusion approach led to higher accuracy of prediction than single CNN models.

Li and Shen[38] joint two ResNet models for melanoma detection. The authors based their approach on convolving images using two models and then combining saliency maps of the two models. At last, they estimated the skin lesions using joint output saliency maps.

Bi et al[39] used two ResNet models for recognizing melanoma among seborrheic keratosis and nevus. First, the authors predict melanoma using each of the ResNet models. Then they combine two ResNet architectures for the melanoma prediction. The performance results have shown that an ensemble of two ResNet models outperforms each of the single ResNet models.

3.3 | Feature-based deep learning methods

Mahbod et al[40] used AlexNet, VGGNet, and ResNet deep networks for feature extraction and then classified these futures using support vector machines. The authors likewise fused extracted features for skin lesion recognition. This study shows that fused features for lesion prediction outperform the characteristics of each of the deep learning models.

Hagerty et al[41] used ResNet-50 network together with several feature images for melanoma recognition. Handcraft features and deep features of ResNet are joint for melanoma identification. The results have shown that fusing handcraft and deep features provided higher melanoma accuracy.

Serte and Demirel[10] proposed Gabor wavelet-based deep learning to recognize malignant melanoma and seborrheic keratosis skin lesions. This method decomposed the skin image into seven directional Gabor band and then modeled each of the representations using AlexNet and ResNet-18 deep learning models.

Serte and Demirel[11] used a wavelet-based technique for melanoma classification. They based their approach on getting skin image into wavelet band and then modeling each image band using deep networks.

4 | DEEP LEARNING IN CHEST X-RAY RECOGNITION

Wang et al[42] used AlexNet, GoogleNet, VGG-16, and ResNet-50 deep CNNs. Authors generated these models by ImageNet models kept on chest X-ray models. This work provided eight thorax disease classification and showed that the ResNet-50 deep learning model outperforms other deep learning models.

Yao et al[43] proposed multi-resolution-based multi-deep learning models. Authors use the ResNet model to reduce chest X-ray images and then use the DenseNet model to classify chest X-ray images in different resolutions.

Wang and Xia[44] used classification and attention-based deep learning models. The authors named this model a ChestNet model. The classification part of the ChestNet model contained the ResNet-152 CNN model. The output of the classification map of this model is further modeled using convolutions for more exact classification. This work compared the proposed model with the classic ResNet model on the Chest X-ray 14 dataset. The performance evaluation of the ChestNet model showed that ChestNet outperforms the ResNet model.

Gundel et al[45] used the DenseNet-121 CNN for twelve chest x-ray image classification. The authors created this network by adapting the ImageNet pretrained DenseNet-121 model to chest images. The proposed method proved higher overall accuracy than AlexNet, GoogleNet, VGG-16, and ResNet-50 deep networks.

Rajpurkar et al[46] applied the DenseNet-121 CNN for the detection of pneumonia. The authors likewise used this model for the detection of 12 thorax diseases. Authors trained the DenseNet-121 model using Chest X-ray 14 dataset.[42]

Li et al[47] used an attention-based ResNet CNN. Their approach applies to U-Net deep learning segmentation model and they classify the local regions of the X-ray images using the ResNet model.

Kermany et al[48] proposed transfer learning-based CNNs for pneumonia detection.

Varshni et al[49] joint DenseNet-169 model and support vector machines SVM. They represented chest X-ray appearances and related extracted feature vectors using DenseNet-169 deep networks. They got the image representations bypassing the architecture comprising 169 layers. Then they classified image appearances using SVM for health and pneumonia images.

Ayan and Unver[50] used Xception and VGG-16 deep networks pneumonia classification and healthy X-ray images. They base their approach on retraining existing models for X-ray images and use transfer learning-based models for detection pneumonia among pneumonia and healthy X-ray images. The VGG-16 model outperforms the Xception model for the classification task.

5 | DEEP LEARNING IN GLAUCOMA RECOGNITION

The following sections review in brief the earlier glaucoma research that used deep learning and other classification methods.

5.1 | Glaucoma classification using deep learning methods

Various experts have studied deep learning to detect glaucoma. Among them, Chen et al[51] proposed a six-layer deep learning architecture for better characterization of glaucoma hidden patterns. In a separate work, Chen et al[52] showed an automated deep learning feature learning method to detect glaucoma. For automated glaucoma detection, Orlando et al[53] used preprocessing and pretrained CNNs. Chai[54] joint domain knowledge with fundus images and deep learning for automatic glaucoma detection. Zilly[55] segmented optic cup and optic disc to detect glaucoma by using a convolutional learning architecture with ensemble learning and entropy sampling. Raghavendra et al[56] classified fundus images into normal and glaucoma classes using an 18-layer CNN.

5.2 | Glaucoma classification using other methods

Likewise, researchers have used a few other classifiers to detect glaucoma: sequential minimal optimization,[57] support vector machine (SVM),[58-60] k-nearest neighbor,[61] naïve Bayes (NB),[62] least square-support vector machine (LS-SVM)[63,64] and artificial neural network (ANN).[65,66]

6 | DATASETS

This section describes the databases related to skin lesions, chest X-rays, and glaucoma imaging. Table 1 shows the number of images of skin databases. Table 2 shows the number of images of the different thorax diseases included in the databases, and Table 3 lists the number of fundus images in the databases related to types of glaucoma. Table 4 lists the abbreviations used in the datasets for each disease.

6.1 | Skin lesion datasets

6.1.1 | PH2

PH2[67] dataset contains images of two types of skin lesions: nevi and melanoma. Common nevus is observed on 80 of nevi skin lesion images and atypical nevus is observed on another 80. Malignant melanoma is observed on 40 skin lesion images.

6.1.2 | 2016 ISIC challenge

2016 International Skin Imaging Collaboration (ISIC) Skin Lesion Analysis Towards Melanoma challenge dataset[68] contains images of two skin lesion types: benign and malignant melanoma. Benign melanoma is observed on 388 skin lesion images and malignant melanoma is observed on 891.

6.1.3 | 2017 ISIC challenge

2017 ISIC challenge dataset[69] has images of three types of skin lesions: melanoma, seborrheic keratosis, and benign nevi. Of these images, melanoma is observed on 521 images, seborrheic keratosis on 386, and benign nevi on 1843 images.

TABLE 1 Skin lesion datasets and the number of images in each dataset category

Dataset	MEL	NV	BCC	AK	SCC	SK	BK	DF	VASC
PH2[67]	40	160	—	—	—	—	—	—	—
2016 ISIC[68]	1279	—	—	—	—	—	—	—	—
2017 ISIC[69]	521	1843	—	—	—	386	—	—	—
2018 ISIC[70]	1113	6705	514	327	—	—	1099	115	142
2019 ISIC[69-71]	4522	12875	3323	867	628	—	2624	239	253

TABLE 2 Chest X-ray datasets and the number of images in each dataset category

Dataset	NR	ATEL	CARD	EFF	INF	MS	NOD	PNA	PNX	CON	ED	EM	FB
Chest X-ray 8[42]	84 312	5789	1010	6331	10 317	6046	1971	1062	2793	—	—	—	—
Chest X-ray 14[42]	—	11 167	12 071	7646	16 316	142	6480	9836	4693	4544	8875	170	10 380

TABLE 3 Glaucoma datasets and the number of images in each dataset category

Dataset	Normal (Healthy)	Early Glaucoma	Glaucoma	Advanced Glaucoma	Total
HRF[72]	18	—	27	—	45
Drishti-GS1[73,74]	31	—	70	—	101
RIM-ONE[75]	261	—	194	—	455
sjchoi86-HRF[76]	300	—	101	—	401
ACRIMA[77]	309	—	396	—	705
Dataset of[78]	786	289	—	467	1542

TABLE 4 Abbreviations for diseases.

Imaging	Abbreviation	Disease
Skin	MEL	Melanoma
Skin	NV	Melanocytic Nevi
Skin	BCC	Basal Cell Carcinoma
Skin	SCC	Squamous Cell Carcinoma
Skin	BK	Benign Keratosis
Skin	DF	Dermatofibroma
Skin	VASC	Vascular Skin Lesions
Chest	NR	Normal
Chest	ATEL	Atelectasis
Chest	CARD	Cardiomegaly
Chest	EFF	Effusion
Chest	INF	Infiltration
Chest	MS	Mass
Chest	NOD	Nodule
Chest	PPNA	Pneumonia
Chest	PNX	Pneumothorax
Chest	CON	Consolidation
Chest	ED	Edema
Chest	EM	Emphysema
Chest	FB	Fibrosis
Chest	PL	Pleural Thickening
Chest	HE	Hernia

6.1.4 | 2018 ISIC challenge

2018 ISIC challenge dataset[70] contains images of seven skin lesions. Melanoma is observed on 1113 images, melanocytic nevus on 6705, basal cell carcinoma on 514, actinic keratosis on 327, benign keratosis on 1099, dermatofibroma on 115 and vascular lesion on 42 images.

6.1.5 | 2019 ISIC challenge

This dataset[69-71] has images of eight different skin lesions. Melanoma is observed on 4522 images, melanocytic nevus on 12 875, basal cell carcinoma on 3323, actinic keratosis on 867, benign keratosis on 2624, dermatofibroma on 239, vascular lesion on 253 and squamous cell carcinoma images on 628 images.

6.2 | Chest X-ray datasets

6.2.1 | Chest X-ray 8 and chest X-ray 14

Chest X-ray 14 dataset[42] is one of the largest datasets related to chest X-ray images. This dataset has images of 14 thorax diseases, including atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax,

consolidation, edema, emphysema, fibrosis, pleural thickening, hernia. The dataset also contains images that are observed to have no thorax diseases.

Chest X-ray 8 dataset, on the other hand, contains images of eight different thorax diseases.

6.3 | Glaucoma datasets

6.3.1 | HRF

In this dataset,[72] there are 45 fundus images. Diabetic retinopathy is seen on 15 images and glaucoma is seen on another 15. The rest of the images contain no eye disease.

6.3.2 | Drishti-GS1

This dataset[73,74] contains 101 fundus images. Seventy-one of these images are glaucomatous and 30 images contain no eye disease.

6.3.3 | RIM-ONE

RIM-ONE r2 dataset[75] has 455 fundus images. The sign or suspicion of glaucoma is observed on 200 these images. The rest of the images contain no eye disease.

6.3.4 | sjchoi86-HRF

sjchoi86-HRF dataset[76] contains 401 fundus images. Of these, 101 are glaucomatous. Three hundred images contain no sign of eye disease.

6.3.5 | ACRIMA

This dataset[77] has 705 fundus images. Glaucoma is observed on 396 of them. There is no eye disease on 309 images.

7 | PERFORMANCE EVALUATION

The area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity (SP) performance merits are used to test the accuracy of the methods. We can describe accuracy, sensitivity, and specificity as:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}, \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}, \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}}, \tag{3}$$

where true positive, positive, true negative, false positive, and false negative are denoted as TP, TN, FP, and FN, respectively.

TABLE 5 AUC values of recently studied deep learning models for skin lesion detection

Authors	Method	Network	Dataset	MEL	SK	NV	BCC	DF	AK	VASC
Serte and Demirel[11]	Wavelet trans.	ResNet-18	2017 ISIC	**0.96**	0.81	—	—	—	—	—
Serte and Demirel[11]	Wavelet trans.	ResNet-50	2017 ISIC	0.74	0.80	—	—	—	—	—
Serte and Demirel[11]	Wavelet coeff.	ResNet-18	2017 ISIC	0.82	0.82	—	—	—	—	—
Serte and Demirel[11]	Wavelet coeff.	ResNet-50	2017 ISIC	**0.96**	0.87	—	—	—	—	—
Serte and Demirel[11]	Wavelet coeff.	ResNet-50-18	2017 ISIC	0.91	0.87	—	—	—	—	—
Serte and Demirel[10]	Gabor	AlexNet	2017 ISIC	0.85	0.70	—	—	—	—	—
Serte and Demirel[10]	Gabor	AlexNet	2017 ISIC	**0.95**	0.80	—	—	—	—	—
Serte and Demirel[10]	Gabor	ResNet-18	2017 ISIC	0.84	0.73	—	—	—	—	—
Serte and Demirel[10]	Gabor	ResNet-18	2017 ISIC	**0.96**	0.86	—	—	—	—	—
Harangi[35]	Images	Ensemble	2017 ISIC	0.85	0.93	—	—	—	—	—
Zhang et al[31]	Images	Attention	2017 ISIC	0.86	**0.95**	—	—	—	—	—
Li and Shen[47]	Lesion Indexing	Ensemble	2017 ISIC	0.91	—	—	—	—	—	—
Matsunaga et al[34]	Images	Ensemble	2017 ISIC	0.87	**0.95**	—	—	—	—	—
Diaz[79]	Images	Ensemble	2017 ISIC	0.85	**0.96**	—	—	—	—	—
Menegola et al[30]	Images	Ensemble	2017 ISIC	0.87	0.94	—	—	—	—	—
Bi et al[39]	Images	Ensemble	2017 ISIC	0.87	0.92	—	—	—	—	—
Mahbod et al[40]	Deep features	Ensemble	2017 ISIC	0.87	**0.95**	—	—	—	—	—
Mahbod et al.[40]	Images and features	Ensemble	2017 ISIC	0.84	**0.97**	—	—	—	—	—
Hagerty et al[41]	Features	Ensemble	2017 ISIC	0.90	—	—	—	—	—	—
Hagerty et al[41]	Images and features	Ensemble	2017 ISIC	**0.94**	—	—	—	—	—	—
Rezvantalab et al[33]	Images	DenseNet-210	2018 ISIC	0.938	—	0.97	**0.993**	**0.999**	0.981	**1**
Rezvantalab et al[33]	Images	ResNet-152	2018 ISIC	**0.944**	—	**0.973**	0.991	0.998	0.976	**1**
Rezvantalab et al[33]	Images	Inception-v3	2018 ISIC	0.934	—	0.97	0.986	0.99	**0.984**	**1**
Rezvantalab et al[33]	Images	Inception-ResNet-v2	2018 ISIC	0.932	—	0.957	0.986	0.993	0.968	0.963

7.1 | Skin lesion classification

Table 5 shows the performance comparison of some of the newest research on skin lesion detection. The highest AUC values in each category are indicated in bold.

These recent studies have focused on the 2017 and 2018 ISIC challenge databases. The results in bold of studies that used the 2017 ISIC challenge database show that frequency-domain use of the images has improved the AUC performance for melanoma skin lesion classification. Using an ensemble of CNN has given a better AUC performance than image frequency domain use for seborrheic keratosis classification.

Rezvantalab et al[33] have used the 2018 ISIC challenge dataset for skin lesion classification. The results in bold of Table 5 show that all four CNN models used have almost equally shown high skin lesion detection results, with ResNet-152 architecture being the best of them.

7.2 | Chest X-ray classification

Table 6 reports the AUC performance comparison of unique CNN architectures on Chest X-ray 8 and Chest X-ray 14 datasets. Again, the highest AUC values are given in bold. We see from them that for both datasets the higher-layered CNN architectures have performed the best in chest X-ray classifications.

TABLE 6 Area under the receiver operating characteristic curve values of recently studied deep learning models for thorax disease detection

Authors	Network	Dataset	ATEL	CARD	EFF	INF	MS	NOD	PNA	PNX	CON	ED	EM	FB	PL	HE
Wang et al[42]	AlexNet	Chest X-ray 8	0.645	0.692	0.664	0.604	0.560	0.648	0.549	0.742	—	—	—	—	-	-
Wang et al[42]	GoogleNet	Chest X-ray 8	0.630	0.705	0.688	0.608	0.536	0.557	0.599	0.782	—	—	—	—	-	-
Wang et al[42]	VGG-16	Chest X-ray 8	0.628	0.708	0.650	0.589	0.510	0.655	0.510	0.751	—	—	—	—	-	-
Wang et al[42]	ResNet-50	Chest X-ray 8	**0.706**	**0.814**	**0.736**	**0.612**	**0.560**	**0.716**	**0.633**	**0.789**	—	—	—	—	-	-
Wang et al[44]	ResNet-152	Chest X-ray 14	0.743	0.874	0.811	0.677	0.783	0.697	0.695	0.809	0.725	0.832	0.822	0.804	0.751	0.899
Yao et al[80]	DenseNet	Chest X-ray 14	**0.772**	**0.904**	**0.859**	0.695	0.792	0.717	0.713	0.841	**0.788**	**0.882**	0.829	0.767	**0.765**	**0.914**
Yao et al[43]	ResNet-64	Chest X-ray 14	0.733	0.858	0.806	0.675	0.778	0.727	0.690	0.805	0.717	0.806	0.842	0.757	0.724	0.871
Gundel et al[45]	DenseNet-121	Chest X-ray 14	0.767	0.883	0.828	**0.709**	**0.821**	**0.758**	**0.731**	**0.846**	0.745	0.835	**0.895**	**0.818**	0.761	0.896

TABLE 7 Area under the receiver operating characteristic curve (AUC) values of recently studied deep learning models for glaucoma classification

Method	Network	Dataset	AUC
Diaz-Pinto et al77, [77]	Xception	HRF	0.83
Diaz-Pinto et al[77]	Xception	Drishti-GS1	0.80
Diaz-Pinto et al[77]	Xception	RIM-ONE	0.85
Diaz-Pinto et al[77]	Xception	sjchoi86-HRF	0.77
Diaz-Pinto et al[77]	Xception	ACRIMA	0.76
Ahn et al[78]	Average CNN	78	**0.94**
Ahn et al[78]	Inception v3	78	**0.93**
Chen et al[51]	6 layer CNN	ORIGA	0.831
Shibata et al[81]	ResNet	Private	**0.965**
Christopher et al[82]	ResNet-50	Private	**0.91**
Li et al[83]	AlexNet	ORIGA	0.8384
Li et al[84]	Inception v3	Private	**0.986**

7.3 | Glaucoma classification

Table 7 shows the AUC comparison of different recent studies on glaucoma classification. The best performing single CNN architectures for this classification are the results in bold. We see from them that ResNet and Inception v3 are among the best performing architectures for glaucoma classifications.

8 | DISCUSSIONS

The results have shown us that for improved performance the studies have preferred to use an ensemble of CNNs or frequency domain representations on skin lesion images rather than single CNN architectures. However, the observation is that no unique way of improving the performance exists, that is, a method that works for the detection of one skin lesion might not work for the detection of the others. What we have further observed is that the researchers have not used ensemble or frequency domain methods so far for the detection of skin lesions included in the 2018 ISIC challenge dataset.

The results of thorax disease detection on two different datasets have indicated that high layered CNN architecture use has resulted in the best performance. To see how generalizable this observation is, we need further research that applies the thorax disease detection to other different datasets.

One observation of glaucoma classification is that if the training and testing datasets are different, it affects the performance (see Diaz Diaz-Pinto et al[77] vs the others in Table 6). Moreover, the researchers need to check if their results are generalizable by applying the detection architectures to people of different ethnic backgrounds.

9 | CONCLUSION

In this paper, our goal has been to review the newest research on deep learning methods for skin, chest, and eye disease identification. Unlike previous research, we have aimed to report, analyze, and discuss the results provided on these topics. To accomplish this, we have compared the performances of different deep learning methods used for the detection of skin lesions and chest and eye diseases. First, we described the disease datasets, and then we listed the number of images included in these datasets. After, we provided the performance evaluation of proposed deep learning methods that used these datasets.

Our analysis has shown that the performances of deep learning methods depend on their architectures and the number of layers. We likewise observed that for skin classification, the researchers need to apply ensemble or frequency domain methods to newer datasets. For thorax disease detection, they also need to apply higher layered CNN models to other datasets and for glaucoma classification, they should consider a more diverse population of patients.

ORCID


Sertan Serte https://orcid.org/0000-0001-7130-0172

REFERENCES

1. Lo S-C, Lou S-L, Lin J-S, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Trans Med Imaging*. 1995;14:711-718.
2. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278-2324.
3. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM*. 2017;60:84-90.
4. K. Simonyan, A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. International Conference on Learning Representations; 2015.
5. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, *Going Deeper with Convolutions*. Boston, MA, USA: IEEE Conference on Computer Vision and Pattern Recognition; 2015.
6. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, *Rethinking the Inception Architecture for Computer Vision*. Las Vegas, NV, USA: IEEE Conference on Computer Vision and Pattern Recognition; 2016.
7. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA; 2016:770-778.
8. G. Huang, Z. Liu, L. Maaten, K. Weinberger, *Densely Connected Convolutional Networks*. Honolulu, HI, USA: IEEE Conference on Computer Vision and Pattern Recognition; 2017.
9. Kaymak S, Esmaili P, Serener A. Deep learning for two-step classification of malignant pigmented skin lesions. Paper presented at: Proceedings of the 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, Serbia; 2018:1-6.
10. Serte S, Demirel H. Gabor wavelet-based deep learning for skin lesion classification. *Comput Biol Med*. 2019a;113:103423.
11. Serte S, Demirel H. Wavelet-based deep learning for skin lesion classification. *IET Image Process*. 2019b.14(4):720– 726.
12. Kaymak S, Serener A. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning, Paper presented at: Proceedings of the IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP); 2018:265-269; Cluj-Napoca, Romania: IEEE.
13. Serener A, Serte S. Dry and wet age-related macular degeneration classification using oct images and deep learning. Paper presented at: Proceedings of the Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science; 2019a:1-4; Istanbul, Turkey: IEEE.
14. Serener A, Serte S. Transfer learning for early and advanced glaucoma detection with convolutional neural networks. Paper presented at: Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO); 2019b:1-4; Izmir, Turkey: IEEE.
15. Serener A, Serte S. Keratinocyte carcinoma detection via convolutional neural networks. Paper presented at: Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); 2019c:1-5; Ankara, Turkey: IEEE.
16. Serte S, Serener A. A generalized deep learning model for glaucoma detection Paper presented at: Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); 2019:1-5; Ankara, Turkey: IEEE.
17. Serener A, Serte S. Geographic variation and ethnicity in diabetic retinopathy detection via deep learning. *Turk J Electr Eng Comput Sci*. 2020;28:664-678.
18. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
19. Khamparia A, Singh PK, Rani P, Samanta D, Khanna A, Bhushan B. An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning. *Trans Emerg Telecommun Technol*. 2020;e3963.n/a
20. Turgut Z, Üstebay S, Ali Aydın M, Gürkaş Aydın GZ, Sertbaş A. Performance analysis of machine learning and deep learning classification methods for indoor localization in internet of things environment. *Trans Emerg Telecommun Technol*. 2019;30:e3705.
21. Ma W. Analysis of anomaly detection method for internet of things based on deep learning. *Trans Emerg Telecommun Technol*. 2020;e3893.n/a
22. Sujitha B, Parvathy VS, Lydia EL, Rani P, Polkowski Z, Shankar K. Optimal deep learning based image compression technique for data transmission on industrial internet of things applications. *Trans Emerg Telecommun Technol*. 2020;e3976.n/a

23. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys*. 2019;29:86-101.
24. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
25. Kim J, Hong J, Park H. Prospects of deep learning for medical imaging. *Precision and Future Medicine*. 2018;2 (2):37–52. http://dx.doi.org/10.23838/pfm.2018.00030.
26. Altaf F, Islam SM, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges, and future directions. *IEEE Access*. 2019;7:99540-99572.
27. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24-29.
28. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
29. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Investigat Dermatol*. 2018;138:1529–1538.
30. Menegola A, Tavares J, Fornaciali M, Li LT, de Avila SEF, Valle E. RECOD titans at ISIC challenge 2017. *CoRR*. 2017.
31. Zhang J, Xie Y, Xia Y, Shen C. Attention residual learning for skin lesion classification. *IEEE Trans Med Imag*. 2019;38:2092-2103.
32. Fujisawa Y, Otomo Y, Ogata Y, et al. Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol*. 2019;180:373-381.
33. Rezvantalab A, Safigholi H, Karimijeshni S. Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. *CoRR*. 2018;abs/1810.10348.
34. Matsunaga K, Hamada A, Minagawa A, Koga H. Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. *CoRR*. 2017;abs/1703.03108.
35. Harangi B. Skin lesion classification with ensembles of deep convolutional neural networks. *J Biomed Inform*. 2018;86:25-32.
36. Nyíri T, Kiss A. Novel ensembling methods for dermatological image classification. Paper presented at: Proceedings of the International Conference on Theory and Practice of Natural Computing; 2018:438-448; Springer, New York, NY.
37. Nozdryn-Plotnicki A, Yap J, Yolland W. Ensembling convolutional neural networks for skin cancer classification. International Skin Imaging Collaboration (ISIC) Challenge on Skin Image Analysis for Melanoma Detection. MICCAI; 2018.
38. Li Y, Shen L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*. 2018;18(2):556.
39. Bi L, Kim J, Ahn E, Feng D. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. *CoRR*. 2017;abs/1703.04197.
40. Mahbod A, Schaefer G, Ellinger I, Ecker R, Pitiot A, Wang C. Fusing fine-tuned deep features for skin lesion classification. *Comput Med Imag Graph*. 2019;71:19-29.
41. Hagerty J, Stanley R, Almubarak H, et al. Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images. *IEEE J Biomed Health Inform*. 2019;23(4):1385–1391.
42. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Paper presented at: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii; 2017.
43. Yao L, Prosky J, Poblenz E, Covington B, Lyman K. Weakly supervised medical diagnosis and localization from multiple resolutions; 2018. arXiv:1803.07703.
44. Wang H, Xia Y. Chestnet: A deep neural network for classification of thoracic diseases on chest radiography. *CoRR*. 2018;abs/1807.03058.
45. Guendel S, Grbic S, Georgescu B, et al. Learning to recognize abnormalities in chest x-rays with location-aware dense networks; 2018. arXiv:1803.04565.
46. Rajpurkar P, J. Irvin, K. Zhu, et al. Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning; 2017. arXiv:1711.05225.
47. Li B, Kang G, Cheng K, Zhang N. Attention-guided convolutional neural network for detecting pneumonia on chest x-rays, Paper presented at: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany; 2019:4851-4854.
48. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172:1122-1131.
49. Varshni D, Thakral K, Agarwal L, Nijhawan R, Mittal A. Pneumonia detection using cnn based feature extraction. Paper presented at: Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India; 2019:1-7.
50. Ayan E, Ünver HM. Diagnosis of pneumonia from chest x-ray images using deep learning. Paper presented at: Proceedings of the 2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey; 2019:1-5.
51. Chen X, Xu Y, Wong DWK, Wong TY, Liu J. Glaucoma detection based on deep convolutional neural network. Paper presented at: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy; 2015a:715-718.
52. Chen X, Xu Y, Yan S, Wong DWK, Wong TY, Liu J. Automatic feature learning for glaucoma detection based on deep learning. Paper presented at: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany; 2015b:669-677; Springer.
53. Orlando JI, Prokofyeva E, del Fresno M, Blaschko MB. Convolutional neural network transfer for automated glaucoma identification. Paper presented at: Proceedings of the 12th International Symposium on Medical Information Processing and Analysis, volume 10160, International Society for Optics and Photonics, Tandil, Argentina; 2017:101600U.

54. Chai Y, Liu H, Xu J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. *Knowl-Based Syst*. 2018;161:147-156.
55. Zilly J, Buhmann JM, Mahapatra D. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Comput Med Imag Graph*. 2017;55:28-41.
56. Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Inf Sci*. 2018;441:41-49.
57. Dua S, Acharya UR, Chowriappa P, Sree SV. Wavelet-based energy features for glaucomatous image classification. *IEEE Trans Inf Technol Biomed*. 2011;16:80-87.
58. Acharya UR, Dua S, Du X, Chua CK, et al. Automated diagnosis of glaucoma using texture and higher order spectra features. *IEEE Trans Inf Technol Biomed*. 2011;15:449-455.
59. Acharya UR, Ng E, Eugene LWJ, et al. Decision support system for the glaucoma using Gabor transformation. *Biomed Signal Process Control*. 2015;15:18-26.
60. Issac A, Sarathi MP, Dutta MK. An adaptive threshold based image processing technique for improved glaucoma detection and classification. *Comput Methods Prog Biomed*. 2015;122:229-244.
61. Acharya UR, Bhat S, Koh JE, Bhandary SV, Adeli H. A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images. *Comput Biol Med*. 2017;88:72-83.
62. Noronha KP, Acharya UR, Nayak KP, Martis RJ, Bhandary SV. Automated classification of glaucoma stages using higher order cumulant features. *Biomed Signal Process Control*. 2014;10:174-183.
63. Maheshwari S, Pachori RB, Acharya UR. Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. *IEEE J Biomed Health Inform*. 2016;21:803-813.
64. Maheshwari S, Pachori RB, Kanhangad V, Bhandary SV, Acharya UR. Iterative variational mode decomposition based automated detection of glaucoma using fundus images. *Comput Biol Med*. 2017;88:142-149.
65. Kausu T, Gopi VP, Wahid KA, Doma W, Niwas SI. Combination of clinical and multi-resolution features for glaucoma detection and its classification using fundus images. *Biocybern Biomed Eng*. 2018;38:329-341.
66. Nayak J, Acharya R, Bhat PS, Shetty N, Lim T-C. Automated diagnosis of glaucoma using digital fundus images. *J Med Syst*. 2009;33:337.
67. Mendonça T, Ferreira PM, Marques JS, Marcal AR, Rozeira J. Ph 2-a dermoscopic image database for research and benchmarking. Paper presented at: Proceedings of the 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC); 2013:5437-5440; Osaka, Japan: IEEE.
68. Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N, Halpern A. Skin lesion analysis toward melanoma detection. a challenge at the international symposium on biomedical imaging (isbi) Paper presented at: Proceedings of 2016, Hosted by the International Skin Imaging Collaboration (isic); 2016. arXiv preprint arXiv:1605.01397.
69. Codella NC, Gutman D, Celebi ME, et al., Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), Paper presented at: Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018:168-172; IEEE.
70. Tschandl P, Rosendahl C, Kittler H. The ham10000 dataset: a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. 2018;14(5):180161
71. Combalia M, Codella NC, Rotemberg V, et al., Bcn20000: dermoscopic lesions in the wild; 2019. arXiv preprint arXiv:1908.02288.
72. Budai A, Bock R, Maier A, Hornegger J, Michelson G. Robust vessel segmentation in fundus images. *Int J Biomed Imag*. 2013;2013:1 –11. https://doi.org/10.1155/2013/154860.
73. Sivaswamy J, Krishnadas S, Chakravarty A, et al. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomed Imag Data Pap*. 2015;2:1004.
74. Sivaswamy J, Krishnadas S, Joshi GD, Jain M, Tabish AUS. Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation. Paper presented at: Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); 2014:53-56; IEEE.
75. Fumero F, Alayón S, Sanchez JL, Sigut J, Gonzalez-Hernandez M. Rim-one: An open retinal image database for optic nerve evaluation, Paper presented at: Proceedings of the 2011 24th international symposium on computer-based medical systems (CBMS); 2011:1-6; IEEE.
76. Choi S. sjchoi86-HRF database; 2016. https://github.com/yiweichen04/retina_dataset. Accessed September 09, 2019.
77. Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A. Cnns for automatic glaucoma assessment using fundus images: an extensive validation. *Biomed Eng Online*. 2019;18:29.
78. Ahn JM, Kim S, Ahn K-S, Cho S-H, Lee KB, Kim US. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS One*. 2018;13:e0207982.
79. Díaz IG. Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions; 2017. arXiv:1703.01976.
80. Yao L, Poblenz E, Dagunts D, Covington B, Bernard D, Lyman K. Learning to diagnose from scratch by exploiting dependencies among labels, 2017. arXiv:1710.10501.
81. Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep*. 2018;8:14665.
82. Christopher M, Belghith A, Bowd C, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep*. 2018;8:16685.

83. Li A, Cheng J, Wong DWK, Liu J. Integrating holistic and local deep features for glaucoma classification. Paper presented at: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016:1328-1331; Orlando, FL, USA: IEEE.
84. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125:1199-1206.

How to cite this article: Serte S, Serener A, Al-Turjman F. Deep learning in medical imaging: A brief review. *Trans Emerging Tel Tech*. 2020;e4080. https://doi.org/10.1002/ett.4080