Rom J Morphol Embryol 2021, 62(4):1017–1028 ISSN (print) 1220–0522, ISSN (online) 2066–8279 doi: 10.47162/RJME.62.4.14

ORIGINAL PAPER

RJME Romanian Journal of Morphology & Embryology http://www.rjme.ro/

Deep learning with transfer learning in pathology. Case study: classification of basal cell carcinoma

RALUCA MARIA BUNGĂRDEAN[1)], MIRCEA-SEBASTIAN ŞERBĂNESCU[2)], COSTIN TEODOR STREBA[3)], MARIA CRIŞAN[4)]

[1)]*Department of Pathology, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania*

[2)]*Department of Medical Informatics and Biostatistics, University of Medicine and Pharmacy of Craiova, Romania*

[3)]*Department of Pulmonology, University of Medicine and Pharmacy of Craiova, Romania*

[4)]*Department of Histology, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania*

Abstract

Establishing basal cell carcinoma (BCC) subtype is sometimes challenging for pathologists. Deep-learning (DL) algorithms are an emerging approach in image classification due to their performance, accompanied by a new concept – transfer learning, which implies replacing the final layers of a trained network and retraining it for a new task, while keeping the weights from the imported layers. A DL convolution-based software, capable of classifying 10 subtypes of BCC, was designed. Transfer learning from three general-purpose image classification networks (AlexNet, GoogLeNet, and ResNet-18) was used. Three pathologists independently labeled 2249 patches. Ninety percent of data was used for training and 10% for testing on 100 independent training sequences. Each of the resulted networks independently labeled the whole dataset. Mean and standard deviation (SD) accuracy (ACC) [%]/sensitivity (SN) [%]/specificity (SP) [%]/area under the curve (AUC) for all the networks was 82.53±2.63/72.52±3.63/97.94±0.3/0.99. The software was validated on another 50-image dataset, and its results are comparable with the result of three pathologists in terms of agreement. All networks had similar classification accuracies, which demonstrated that they reached a maximum classification rate on the dataset. The software shows promising results, and with further development can be successfully used on histological images, assisting pathologists' diagnosis and teaching.

Keywords: basal cell carcinoma, BCC, deep learning, transfer learning, classification, subtypes.

Introduction

Basal cell carcinoma (BCC) is the most encountered skin malignancy with a rising incidence rate worldwide, even in younger patients [1–4]. Although BCCs rarely metastasize, they are locally destructive and have recurrences risk which impacts the quality of life [2], therefore a sharp diagnosis could forewarn and prevent this. Currently, the "gold standard" for diagnosis is histopathology [1–3], hence the purpose of our paper is to come in aid of the pathologist, by providing a reliable image analysis software.

The latest *World Health Organization (WHO) Classification of Skin Tumors* recognizes 10 different subtypes of BCC, as follows: nodular (with three different variants: keratotic, nodulocystic and adenoid), superficial (which can sometimes appear multifocal), micronodular, infiltrating, sclerosing, basosquamous, pigmented, with sarcomatoid differentiation, with adnexal differentiation and fibroepithelial [5].

The *National Comprehensive Cancer Network (NCCN) Guidelines* [6] as well as multidisciplinary experts from the *European Dermatology Forum*, *European Organization of Research and Treatment of Cancer* and the *European Association of Dermato-Oncology* [1] recommend the division of BCC subtypes into low-risk (LR) and high-risk (HR) groups based on recurrence risk or into "easy-to-treat" and "difficult-to-treat" groups based on specific treatment management problems. Therefore, the latest *WHO Classification of Skin Tumors* highlights a table that shows which subtypes are categorized as LR and HR. The LR subtypes are nodular, superficial, pigmented, infundibulocystic (a variant with adnexal differentiation) and fibroepithelial [5]. The HR subtypes are basosquamous, sclerosing, infiltrating, micronodular and sarcomatoid differentiation [5]. Although superficial BCC belongs to the LR group, a special remark must be made for the multifocal variant, because these tumors have an increased risk of incomplete excision with a higher recurrence rate and subclinical spread, therefore many authors consider multifocal superficial subtype as HR [7–9], which we also considered in the present study.

Due to whole-slide imaging (WSI) acquisitions, digital pathology has become more popular in the past years [10] and with its different automatic image analysis methods appear [11]. Using distinct pattern recognition algorithms and analysis methods, these pieces of software main purpose is capturing and interpreting morphological and architectural characteristic of different tissues and tumors, taking into consideration the diverse combination of visual patterns from histological slides which have variable morphology and architectural arrangements of cells [10, 12].

In computer science, artificial intelligence (AI) is the simulation of human intelligence processes by machines,

with the purpose of maximizing its chance of achieving its goal [13]. From all the types of AI, deep neural networks (DNNs) are a particular type of neural networks (NNs) that showed outstanding performance on public benchmark data sets [14–16] and are extensively used in image classification. Testing the performance of a classifier implies the computation of several performance descriptors, namely accuracy (ACC), sensitivity (SN), specificity (SP), and area under the curve (AUC).

Having settled the HR and LR concepts in BCC, while using an AI system, we run in a new concept called explainable AI (XAI). Since the DNNs seem to work out like a black-box, the concept of XAI emerged pretty soon after their creation and originates in the argument-driven decision-making of the human mind. Samek *et al.* [17] stratifies the most important arguments of XAI in verification of the system, improvement of the system, learning from the system, and compliance to legislation, while Hagras [18] resumes that the explainability sits at the intersection of several areas of active research in AI: transparency, causality, bias, fairness, safety. In the end, Montavon *et al.* [19] summarizes the methods of XAI, as follows: (*i*) functional approaches [20], where the explanation results from the local analysis of the prediction function (*e.g.*, SN analysis or Taylor series expansion), and (*ii*) message passing approaches [21, 22] that view the prediction as the output of a computational graph, and where the explanation is obtained by running a backward pass in that graph. Without further detailing the concept of XAI, we need to understand that for the transition from research to real life application, this concept must be, at least in part, applied.

The use of deep-learning (DL) convolutional networks in BCC diagnosis has been studied by other researchers as follows. van Zon *et al.* researched their use on margin control in Mohs surgery with the purpose of alleviating the workload of manually evaluating each slide by determining whether a WSI has BCC or not [23]. To do so, they created a two-model framework: one used pixel segmentation by applying a trained U-Net over the WSI and the other used a convolutional NN.

Using convolutional NN, Albahar propose a classification system for skin lesions that distinguishes between benign and malignant lesions such as seborrheic keratosis *vs.* BCC or nevus *vs.* melanoma and reports a maximum 97.49% ACC rate for the classification [24].

WSIs [25] or slides patches were analyzed using DL NN frameworks for accurate BCC detection on digitally scanned histopathology images. Jiang *et al.* proposes an inquiry on smartphone-captured microscopic ocular images [25] and achieves comparable performance on detecting BCC as the previously mentioned methods (WSI and patches from slides [26]). They developed two DL frameworks: a 'cascade' and a 'segmentation' framework [25]. The 'cascade' structure included a classification model for identifying difficult cases (pictures with low prediction confidence) as well as a segmentation model for more in-depth examination of the difficult [25]. All photos were directly segmented and categorized using the 'segmentation' framework.

Cruz-Roa *et al.* used DL for BCC detection on 1417 image patches of 300×300 pixels and obtained the best ACC (0.921±0.031), precision (0.901±0.041) and prediction model (0.79 to 0.96) on digital stained images where cancer regions were outputed in red and non-tumoral tissue in blue [12]. The issue encountered in their study was the fact that the software recognized large dark nuclei proliferations, thus being exposed to confusion with non-tumoral structures, such as epidermis or various glands [12].

Another approach to the use of DL methods was performed by Kimeswenger *et al.* who proposed an artificial NN screening of BCC WSI with the purpose of tumor region identification and swift classification [27]. In their research, the artificial NN's diagnosis-relevant regions were compared to pathologists' areas of interest, as detected by eye-tracking techniques. Their proposed method recognized BCC tumor areas with a 95% SN [confidence interval (CI): 0.951–0.979] and a 95% SP (CI: 0.859–0.960), moreover they also observed that machine learning algorithms, when compared to pathologists' eye-tracking results, rely on markedly distinct recognition patterns for tumor diagnosis [27], which shows that the use of this kind of software improves the diagnostic quality of the human eye.

Andrew *et al.* propose a different use of machine learning algorithms: an automatic decision-making algorithm that predicts multidisciplinary team recommendations for nasal BCC with the purpose of reducing caseload and providing automatic decisions [28]. Their results show a 37.5% of patients may be confidently predicted to undergo Mohs micrographic surgery (based on age and tumor location) likewise, the teams' choice of traditional treatment (surgical excision or radiotherapy) could be consistently predicted (depending on age and on phenotype and size of the tumor), resulting in an accurate prediction in 45.1% of cases [28].

Other researchers, such as Campanella *et al.*, researched BCC slides, along with other types of tumors, such as prostatic cancer and axillary lymph node with breast cancer metastasis, by using DL convolutional networks [29].

The state-of-the-art results on BCC image classification are summarized in Table 1. Network architecture, dataset overview, classifications target, and performance assessment of each study are presented.

Table 1 – ***State of the art results on BCC image classification/segmentation. The default target is classification unless otherwise specified***

Paper	Network architecture	Dataset	Target	Performance assessment
van Zon *et al.* (2021) [23]	U-Net for segmentation Proprietary for classification	171 WSI	Segmentation & classification Two classes: ▪ BCC; ▪ Normal.	Dice score 0.66 AUC 90%
Campanella *et al.* (2022) [29]	ResNet34	9962 WSI	Two classes: ▪ BCC; ▪ Normal.	AUC 98%

Paper	Network architecture	Dataset	Target	Performance assessment
Jiang *et al.* (2020) [25]	GoogleNet inception v3	6610 images	Two classes: ▪ BCC; ▪ Normal.	The 'cascade' framework SN 93% SP 91% The 'segmentation' framework SN 97% SP 94% AUC 98%
Cruz-Roa *et al.* (2013) [12]	Proprietary	1417 image patches	Two classes: ▪ BCC; ▪ Normal.	ACC 90% Precision 0.876 SN 86% SP 92% *F*-measure 0.872 Balanced ACC 89%
Santilli *et al.* (2020) [30]	Proprietary	190 iKnife scans	Two classes: ▪ BCC; ▪ Normal.	ACC 96.62% SN 100% SP 95%
Kimeswenger *et al.* (2021) [27]	VGG11	820 WSI	Segmentation Two classes: ▪ BCC; ▪ Normal.	SN 95% SP 95%
Putten *et al.* (2018) [31]	Proprietary ResNet-based	*ISIC* Archive	Two classes: ▪ BCC; ▪ Normal.	SN 96% SP 89%
Arevalo *et al.* (2015) [10]	Proprietary	1417 images	Segmentation Two classes: ▪ BCC; ▪ Normal.	AUC 0.981
Esteva *et al.* (2017) [32]	Inception v3	129 450 dermatoscopy images	Three classes	ACC 72%
			Nine classes	ACC 55%
Hosny *et al.* (2020) [33]	AlexNet	100 015 dermatoscopy images	Seven classes	ACC 98.70% SN 95.60% SP 99.27% Precision 95.06%
Jinnai *et al.* (2020) [34]	VGG-16	5846 clinical images	Seven classes: ▪ Malignant melanoma; ▪ BCC; ▪ Nevus; ▪ Seborrheic keratosis; ▪ Senile lentigo; ▪ Hematoma; ▪ Hemangioma.	ACC 86.2%
			Two classes: ▪ Malignant; ▪ Normal.	ACC 91.5% SN 83.3% SP 94.5%
Thomas *et al.* (2021) [35]	ResNet50	290 WSI	Four classes: ▪ BCC; ▪ Squamous cell carcinoma; ▪ Intraepidermal carcinoma; ▪ Normal.	ACC 93.6%
			Two classes: ▪ Malignant; ▪ Normal.	ACC 97.9%

ACC: Accuracy; AUC: Area under the curve; BCC: Basal cell carcinoma; ISIC: *International Skin Imaging Collaboration*; SN: Sensitivity; SP: Specificity; WSI: Whole-slide imaging.

Our research tackles the difficult topic of histopathology image analysis, namely the diagnosis of BCC subtypes and variants. The aim of the work is to develop a novel, reliable, and free application which can be used by pathologists for a first or second opinion and can be used by young trainees for learning. The reason why of all malignancies, we chose to study BCCs lies in the fact that they are the most common skin malignancy [1–3] and therefore there is enough data to properly test the reliability of the software.

The novelty of the approach is that the classification application based on DL with transfer learning aims to classify images in subtypes and variants of BCC, rather than discriminating between BCC and normal tissue, or BCC and other malignancies as presented in Table 1.

Materials and Methods

Materials

The dataset consists of 102 consecutive cases of BCC presented at the Clinical Municipal Hospital, Cluj-Napoca, Romania. Data was collected after receiving approval from the Research Ethics Committee (Approval No. 7749/21.09.2021).

The surgically excised tissue was histologically processed using the standard procedure, and the slides were stained using the Hematoxylin–Eosin (HE) staining. One hundred and seventy-eight slides were scanned using Pannoramic SCAN II, 3DHISTECH (Budapest, Hungary), using 20× objectives. Crops of 643×643 with 24 red, green, blue (RGB) were selected by a trained pathologist and using a

panel of three pathologists only the crops where the diagnosis was fully agreed on were kept. From the 643×643 image, we only used 512×512 around the center of the image, the rest being additional pixels that permitted image rotation, for data augmentation.

In the end the dataset consisted in 2249 image crops, labeled as: HR infiltrating and sclerosing (384), HR micronodular (316), HR superficial multifocal (25), LR with adnexal differentiation (78), LR nodular (508), LR nodular adenoid variant (218), LR nodular keratotic variant (101), LR nodular nodulocystic variant (126), LR pigmented (366), LR superficial unifocal (127). Samples of each subtype are presented in Figure 1 (a–j).

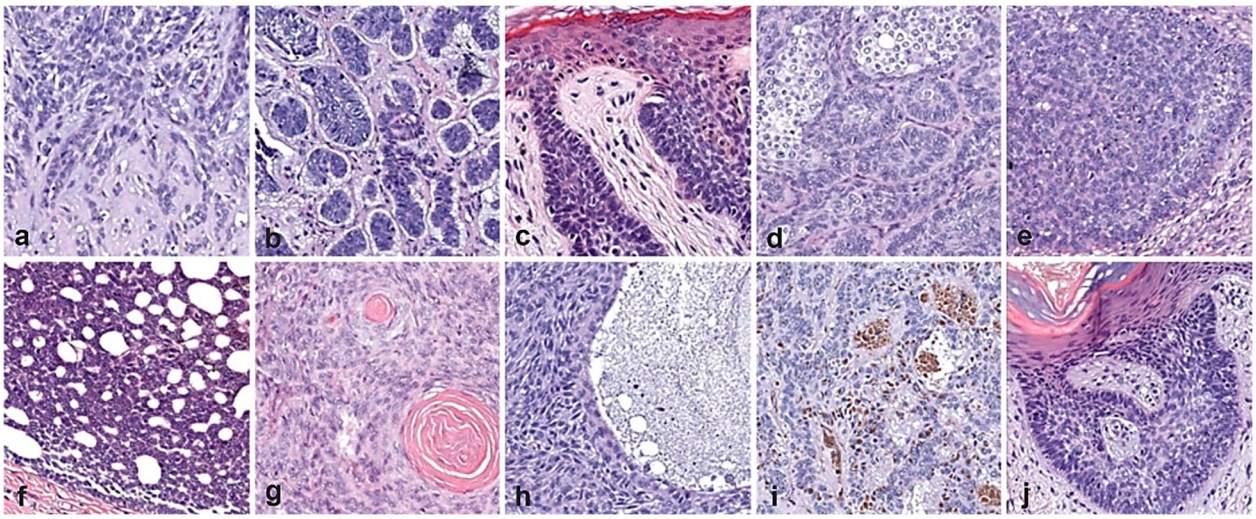

Figure 1 – ***Dataset samples: (a) HR infiltrating and sclerosing; (b) HR micronodular; (c) HR superficial multifocal; (d) LR with adnexal differentiation; (e) LR nodular; (f) LR nodular adenoid variant; (g) LR nodular keratotic variant; (h) LR nodular nodulocystic variant; (i) LR pigmented; (j) LR superficial unifocal. HE staining: (a–j) ×200. HE: Hematoxylin–Eosin; HR: High risk; LR: Low risk.***

Methods

Three DNNs, with different architecture, designed for the classification of real-world images in 1000 classes, trained on the ImageNet [36] were modified using the transfer learning concept [37] to classify BCC images in 10 different patterns, three considered as HR and seven as LR. The networks were modified, re-trained, and the results were statistically assessed. The best performer networks of each architecture were packed in a standalone application capable of classifying live images into the 10 subtypes of BCC named earlier. In the end, the application was used to predict the class of 50 new images and the results were compared with the opinion of the three pathologists' panel.

All the computations were done in Matlab® (MathWorks, USA).

Networks transfer, training, and evaluation

Three pre-trained networks were selected using previous work [38, 39] and a combination of good ACC and good prediction-time, as presented in [40].

The first network taken in consideration was AlexNet [14], a convolution-based NN that has five convolutional layers, in part followed by max-pooling layers, and two globally connected layers, that ends with 1000-way softmax. The network model can be downloaded for free from [41], and it is available in Matlab. The network was introduced in 2012.

The second model was GoogLeNet [16], a significantly different network and with non-linear architecture, having 22 layers, and introducing the architecture concept of "inception". The network model can be downloaded for free from [42], and it is also available in Matlab. The network was introduced in 2014.

The third model was ResNet-18 [43], also having a non-linear architecture, with 18 layers, and being introduced in 2016. The network model is also available in Matlab.

Each of the models underwent transformations to adapt to the new task. Instead of having to classify images from the real world (as they were intended for), all network's final layers have been replaced with new layers, thus the last fully connected layer was replaced, together with the output layer, for the new classification task. The output layer had 10 classes to match the labels of our dataset.

Next, the modified pre-trained networks, having the last layers reset and adapted for the new classification task, underwent the training procedure. The training was done on the new dataset, each time 10% of the images were kept for validation and 90% were used in the training itself in the so-called tenfold cross-validation methodology. The training hyper-parameters options were kept the same for all three architectures, and set empirically, with Mini Batch Size set to 100, the initial learning rate to 0.0001, and a validation patience of 4. All networks were trained using the stochastic gradient descent with momentum (SGDM) as optimizer.

The performance was assessed in terms of mean ACC, SN, SP, and AUC.

Standalone application

The best performing network from each model in all 100 runs were packed in a standalone application, designed and compiled in Matlab, that has two functionalities. The first functionality is an image-by-image classification approach where the user opens an image and the application predicts the class, showing the three network's outputs.

The second functionality is the live option, which acquires images from the cursor position and predicts the class, similar with the other functionality, but in a continuous manner, thus offering the possibility of using the application together with a live microscope camera or WSI image viewer.

Clinical evaluation

Fifty images from five consecutive BCCs were randomly selected from WSIs resulting in the same procedure as the original dataset was created. The images were labeled using the application described in the previous subchapter and independently by three pathologists. The agreement between the three best networks and the three pathologists was further assessed.

Statistical assessment

NNs neurons weight adjustment is dependent on the training and validation dataset and the way the data is presented to the network. This puts the NNs family algorithms in the larger family of stochastic algorithms. This in turn makes our DNN algorithms stochastic. Since we want robust and trustworthy results, the algorithms must be independently run several times with different input data and the evaluation result must be presented as a mean [together with standard deviation (SD)] result. To obtain a suitable statistical power – two-tailed type of null hypothesis with default statistical power goal $p \geq 95\%$ and type I error $\alpha=0.05$ – level of significance, all three models have been independently run 100 time with a tenfold cross-validation using random 90% of the data for training and the remainder 10% for validation. For each run, the training and testing data was randomly selected, but all three networks used the same data, so their performance can be directly compared.

Kolmogorov–Smirnov, Lilliefors, and the Shapiro–Wilk *W* tests were used to seek the normal distribution of data.

Agreement between the three networks and the three pathologists form the panel has been measured using Cohen's *kappa* (κ) coefficient [44]. The statistic is a more robust measure than the simple percent agreement because it includes the possibility of the agreement to appear by chance. The *p*-value of *kappa* is not reported since it is irrelevant to its meaning because even low values of *kappa* can be significantly different from zero. Thus, magnitude guidelines for *kappa* coefficient have been set, however they are not universally accepted. A common interpretation [45] characterize values <0 as indicating no agreement and 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. Different authors [46] characterize *kappas* over 0.75 as excellent, 0.40 to 0.75 as fair to good, and less than 0.40 as poor. Since there is no evidence to support any interpretation, and the intervals are set rather empirical, for interpretation Landis & Koch [45] guidelines were used.

Results

A total of 300 trained networks resulted, 100 having AlexNet architecture, 100 having GoogLeNet architecture, and the last 100 having ResNet-18 architecture. Mean and SD for ACC together with normality assessment are presented in Table 2.

Table 2 – ***Network ACC [%] performance assessment. Kolmogorov–Smirnov with p-level, Lilliefors p, Shapiro–Wilk W with p-level***

	AlexNet	GoogLeNet	ResNet-18
ACC [%]	82.16±2.60	83.03±2.67	82.42±2.61
Kolmogorov–Smirnov	0.12	0.10	0.09
p-level	<0.01	<0.01	0.03
Lilliefors *p*	<0.01	0.01	0.04
Shapiro–Wilk *W*	0.97	0.98	0.99
p-level	0.02	0.16	0.36

ACC: Accuracy.

Though the normality tests indicated mixed Gaussian and non-Gaussian distribution of the samples, since the sample size is large (n=100) the normality of the data can be assumed, thus parametric tests can be used. One-way analysis of variance (ANOVA) showed sum of squares (SS)=39, with two degrees of freedom, mean square (MS)=19.63, F=2.83, F_{crit}=3.02 and a p-value of 0.06 meaning there are no significant differences between the mean ACC of the three models. *Post-hoc* Tukey's test confirms the results and individual p-values are presented in Table 3.

Table 3 – ***Statistical assessment of mean ACC. Post-hoc Tukey's test p-value***

	GoogLeNet	ResNet-18
AlexNet	0.05	0.77
GoogLeNet		0.23

ACC: Accuracy.

Mean and SD for SN are presented in Table 4.

Table 4 – ***Network SN [%] performance assessment***

	AlexNet	GoogLeNet	ResNet-18
SN [%]	72.74±3.74	72.90±3.29	71.93±3.88

SN: Sensitivity.

One-way ANOVA showed SS=55, with two degrees of freedom, MS=27.45, F=2.06, F_{crit}=3.02 and a p-value of 0.129 meaning there are no significant differences between the mean SN of the three models. *Post-hoc* Tukey's test confirms the result and individual p-values are presented in Table 5.

Table 5 – ***Statistical assessment of mean SN. Post-hoc Tukey's test p-value***

	GoogLeNet	ResNet-18
AlexNet	0.89	0.25
GoogLeNet		0.14

SN: Sensitivity.

Mean and SD for SP are presented in Table 6.

Table 6 – ***Network SP [%] performance assessment***

	AlexNet	GoogLeNet	ResNet-18
SP [%]	97.90±0.30	98.01±0.31	97.93±0.31

SP: Specificity.

Though the normality tests indicated mixed Gaussian and non-Gaussian distribution of the samples, since the sample size is large (n=100) the normality of the data can

be assumed, thus parametric tests can be used. One-way ANOVA showed *SS*=0.66, with two degrees of freedom, *MS*=0.33, *F*=3.41, F_{crit}=3.02 and a *p*-value of 0.03 suggesting there are significant differences between the SP of the three models. *Post-hoc* Tukey's test confirms the result and individual *p*-values are presented in Table 7.

Table 7 – *Statistical assessment of mean SP. Post-hoc Tukey's test p-value*

	GoogLeNet	ResNet-18
AlexNet	0.03	0.76
GoogLeNet		0.15

SP: Specificity.

Both GoogLeNet and AlexNet outperformed ResNet-18 in the matter of AUC, as seen in Table 8.

Table 8 – *Network AUC performance assessment*

	AlexNet	GoogLeNet	ResNet-18
AUC	0.9946±0.0031	0.9956±0.0025	0.9931±0.0045

AUC: Area under the curve.

One-way ANOVA showed *SS*=0.000309, with two degrees of freedom, *MS*=0.00015, *F*=12.56, F_{crit}=3.02 and a *p*-value of <0.001 meaning there are significant differences between the mean AUC of the three models. *Post-hoc* Tukey's test confirms the results and individual *p*-values are presented in Table 9.

Table 9 – *Statistical assessment of mean AUC. Post-hoc Tukey's test p-value*

	GoogLeNet	ResNet-18
AlexNet	0.08	0.01
GoogLeNet		0.001

AUC: Area under the curve.

Though there is no statistical difference in the mean AUC between AlexNet and GoogLeNet they both outperform ResNet-18.

Confusion matrices on all samples in the dataset from the best performing network of each architecture are presented in Tables 10–12, while HR and LR confusion matrices of the same run are presented in Table 13.

Table 10 – *Confusion matrix of the best performing model of AlexNet*

AlexNet		Target class									
		1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Output class	***1***	381	1	0	0	1	0	0	1	0	0
	2	1	309	0	3	3	0	0	0	0	0
	3	1	0	22	0	0	0	0	0	0	2
	4	0	1	0	75	1	1	0	0	0	0
	5	0	3	0	0	502	0	0	1	2	0
	6	0	3	0	3	5	204	0	2	1	0
	7	2	1	0	0	2	0	95	0	1	0
	8	0	1	0	0	10	2	0	110	3	0
	9	1	7	0	0	18	4	0	0	355	1
	10	0	0	1	0	0	0	0	0	1	125

Class codes are 1: HR infiltrating and sclerosing, 2: HR micronodular, 3: HR superficial multifocal, 4: LR adnexal, 5: LR nodular, 6: LR nodular adenoid, 7: LR nodular keratotic, 8: LR nodular nodulocystic, 9: LR pigmented, 10: LR superficial unifocal. HR: High risk; LR: Low risk.

Table 11 – *Confusion matrix of the best performing model of GoogLeNet*

GoogLeNet		Target class									
		1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Output class	***1***	383	0	0	0	0	0	0	0	1	0
	2	7	305	0	3	0	0	1	0	0	0
	3	0	1	13	0	1	0	0	0	0	10
	4	0	4	0	73	0	1	0	0	0	0
	5	2	3	0	0	496	0	0	1	5	1
	6	0	0	0	1	5	208	0	0	4	0
	7	3	1	0	0	1	1	95	0	0	0
	8	1	0	0	0	8	5	1	110	1	0
	9	2	5	0	1	9	10	0	2	234	3
	10	0	1	0	0	4	0	0	0	1	121

Class codes are 1: HR infiltrating and sclerosing, 2: HR micronodular, 3: HR superficial multifocal, 4: LR adnexal, 5: LR nodular, 6: LR nodular adenoid, 7: LR nodular keratotic, 8: LR nodular nodulocystic, 9: LR pigmented, 10: LR superficial unifocal. HR: High risk; LR: Low risk.

Table 12 – *Confusion matrix of the best performing model of ResNet-18*

ResNet-18		Target class									
		1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Output class	***1***	382	0	0	1	1	0	0	0	0	0
	2	4	310	0	0	1	0	1	0	0	0
	3	0	0	14	0	0	0	1	1	1	8
	4	3	0	0	74	1	0	0	0	0	0
	5	0	2	0	0	501	0	0	3	2	0
	6	0	0	0	1	2	211	0	1	3	0
	7	2	1	0	0	2	0	95	1	0	0
	8	0	0	0	0	6	1	0	118	1	0
	9	2	1	0	0	6	4	1	0	351	1
	10	1	0	0	0	0	0	0	0	0	126

Class codes are 1: HR infiltrating and sclerosing, 2: HR micronodular, 3: HR superficial multifocal, 4: LR adnexal, 5: LR nodular, 6: LR nodular adenoid, 7: LR nodular keratotic, 8: LR nodular nodulocystic, 9: LR pigmented, 10: LR superficial unifocal. HR: High risk; LR: Low risk.

Table 13 – *Confusion matrices of the best performing model of each architecture on HR and LR class prediction*

Output class	AlexNet		GoogLeNet		ResNet-18	
	HR	***LR***	***HR***	***LR***	***HR***	***LR***
HR	715	10	709	16	710	15
LR	20	1504	22	1502	12	1512
	Target class		**Target class**		**Target class**	

HR: High risk; LR: Low risk.

The best performing networks from each architecture were packed in a standalone application. AlexNet's best ACC [%]/SN [%]/SP [%]/AUC performance was 95.95/94.49/99.51/0.99, GoogLeNet's 95.06/90.28/99.42/0.99, and ResNet-18's 97.02/92.67/99.65/0.99 when tested on the whole dataset. A running application view can be seen in Figure 2.

The HR/LR ACC [%]/SN [%]/SP [%]/AUC, as resulted from Table 13 was 98.66/98.62/98.68/0.99 for AlexNet, 98.31/97.79/98.55/99.42/0.99 for GoogLeNet, and 98.79/97.93/99.21/0.99 for ResNet-18.

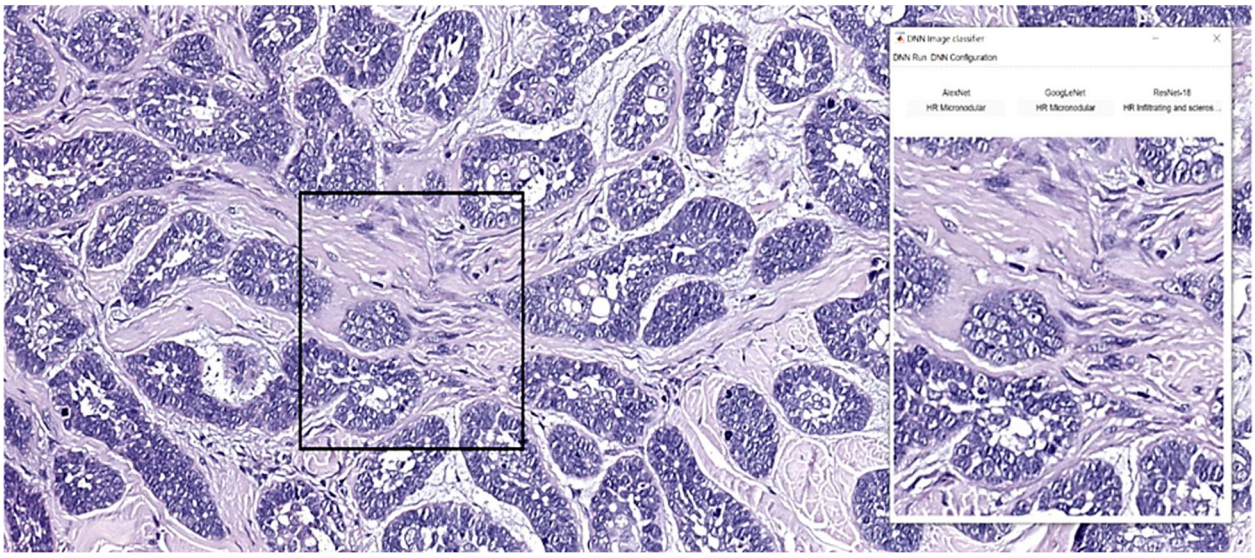

Figure 2 – ***Standalone application. Running the "Live" mode of a new image.***

The standalone application and the "*.MAT*"-file for the three pre-trained networks are offered for free on the following repository https://github.com/Mircea-Sebastian-Serbanescu/3-Deep-Neural-Networks-on-10-Subtypes-of-Basal-Cell-Carcinoma-AlexNet-GoogLeNe-ResNet-18-App.git.

Results of the clinical evaluation are presented in Table 14.

Table 14 – ***Clinical evaluation assessment: Cohen's kappa (κ) coefficient. P1–P3 represent the three panel pathologists***

	P1	P2	P3	AlexNet	GoogLeNet	ResNet-18
P1	1	0.7570	0.6815	0.7004	0.7098	0.6197
P2		1	0.7481	0.7331	0.7421	0.6817
P3			1	0.7214	0.6703	0.6068
AlexNet				1	0.8289	0.6837
GoogLeNet					1	0.7475
ResNet-18						1

All Cohen's *kappas* were above 0.6 (Table 14), showing substantial agreement. Overall, the networks agreed more (mean = 0.7533) than the pathologists (mean = 0.7288). The best agreement was 0.8289, being interpreted as almost perfect, and being registered between AlexNet and GoogLeNet. Average agreement between the pathologists and the networks was 0.6862, with GoogLeNet and P2 having the best agreement of 0.7421 and ResNet-18 showing all *kappas* smaller than 0.7 when compared with the pathologists.

Discussions

Medical fields such as radiology have been researching computational methods for a long time now [47–50]. Because in pathology practice, converting whole glass slides into digital slides using digital slide scanners has only emerged a few years ago, the use of computational methods in the field of pathology has lagged behind in the healthcare digital revolution, as Campanella *et al.* [26] also states when assessing DL methods on WSI. However, as scanner technology has advanced so has digital pathology with transfer learning, thus being able to improve assisted diagnosis, shorten diagnosis time and enable pathologists to work in a digital system. Because of the rapid increase of medical devices with digital recording systems and the vast amount of data provided by DL technologies, the healthcare system stands to gain a great deal. Fuchs & Buhmann [51], and Louis *et al.* [52] also agree that computational pathology has a promising path ahead.

Our study is based on DL convolutional networks. Other researchers also studied DL methods in various medical fields [53] for example ophthalmology [54–56], brain tumor segmentation [57], skin cancer [32, 58], radiology [48, 49, 59] or pathology [25, 60–62].

From a histopathological point of view, the various subtypes and their variants are all characterized by a proliferation of basaloid cells with hyperchromatic nuclei and scant cytoplasm with a characteristic palisade arrangement at the periphery [5, 9]. The surrounding stroma is fibro-myxoid and often creates artificial slits around the tumor islets [5, 9, 63].

Different architecture patterns are encountered depending on the subtype. Although most images were correctly labeled by our application, we will discuss in the following paragraph the biggest confusions made by each model. In our study, the best performing model of GoogLeNet confused seven images of micronodular subtype with in-filtrating and sclerosing subtypes. We think this happened because of their histological appearance which consist of small basaloid cell islets variable in size and shape [5, 9]. While basaloid islets may have a similar appearance, the same cannot be said about the surrounding stroma: the sclerosing subtype has a desmoplastic sclerotic stroma, the infiltrating subtype has an infiltrative invasion pattern and the micronodular subtype is surrounded by normal dermis [5, 9, 64]. We speculate tissue processing and staining may have played a role in the confusion matrix. In four instances, GoogLeNet also confused micronodular subtype with adnexal differentiation subtype. The later consists of basaloid cell islands with particular differentiation, such as hair matrix, follicular infundibulum, sebaceous, ductal, eccrine, or apocrine [5].

Another confusion made by GoogLeNet (10 images) was between unifocal and multifocal superficial BCC. By its definition, the superficial subtype is characterized by small basaloid cell lobules confined to the papillary dermis [5]. We think this confusion was made because the network saw a connection (which was not seen by pathologists) between multifocal lesions and interpreted them as a single tumor. Five images with adenoid variant of the nodular subtype were confused by GoogLeNet with the simple nodular subtype. The histopathology of the nodular subtype consists of solid islands of basaloid cells with typical peripheral palisading and disorganized cell arrangement in the center of the isle [5, 9]. Adenoid variant of the nodular subtype has a specific characteristic: cribriform nests [5]. We believe that in the misdiagnosed images, the solid nodular nest prevailed in quantity when compared to the cribriform nests, so they were classified by the network as nodular.

Following the same path, AlexNet also wrongly classified two images with nodular keratotic variant as simple nodular subtype and 10 images with nodulocystic variant as simple nodular. Both keratotic and nodulocystic are variants of the nodular subtype, the keratotic variant showing mature keratinization in the basal cell nests and the nodulocystic variant showing cystic degeneration [5, 9]. AlexNet confused 18 images with pigmented BCC and classified them as nodular subtype. The pigmented subtype is characterized by the presence of melanin pigment [5]. The issue with this definition is the fact that there is no specific required quantity of melanin pigment for diagnosis, therefore we consider that the error was closely related with the quantity of pigment present in the analyzed images.

Analyzing the confusion matrices, in the XAI acceptance of learning from the system, we can see where the networks mistakenly labeled images from one class with labels from another.

As a general observation on the confusion matrices, we concluded that ResNet-18 has a uniform confusion distribution, where both AlexNet and GoogLeNet have a more convergent confusion.

Research with AI in skin lesions went further and the use of *ex vivo* confocal microscopy obtaining a diagnostic SN and SP of 88% and 91%, respectively [61], using different AI algorithms for data preparation and decision. For data preparation they used a cycle-consistent generative adversarial network [63] to add color to the images and used a U-Net architecture [65] with transfer learning from EfcientNet-B0 [66], which had been previously pre-trained on ImageNet [36] for decision. Similar results were reported a reflectance confocal microscopy research [29].

Giving the vast applications of DL methods in various BCC topics, we find our research relevant to this matter and did not find any other research able to discriminate between 10 or more subtypes of BCC.

In the matter of XAI, our approach only explains its HR and LR decision by the BCC subtypes detected in the image. This is not entirely in accordance with the XAI concept, where the focus is on understanding the basic logics within the network. From the example in Figure 2, we can see the overall pattern of the BCC is micronodular, but the sub-sample that feeds the network contains more interstitial tissue tricking the ResNet-18 network into predicting infiltrating and sclerosing, while the other two DNNs got the right answer. Nevertheless, both subtypes are part of the HR class, so the overall output is correct. Augmenting the decision of ResNet-18 or the other two DNNs on this particular case goes further in the XAI concept, but with the current dataset it is not possible. One explainable approach would be using semantic segmentation that classifies each pixel, in both BCC patterns and surrounding elements, while the decision comes after. This approach could better implement the XAI concept and is the direction of future research.

By creating a two-model framework (pixel segmentation by a trained U-Net over the WSI and convolutional NN), van Zon *et al.* tried to determine whether a slide had BCC on it or not [23]. By comparison, in our study we also used convolutional NN, bringing up that we used three different network architecture, and we modified them using the transfer learning concept. Although the number of studied cases (70) and images (171) were smaller compared to ours (102 cases and 2249 images) and the approach different, van Zon *et al.* obtained promising results with an average dice score of 0.66 and an average area under the receiver operating curve (ROC) value of 0.90 [23]. In comparison, our study obtained best AUC values ranging from 0.9956 (GoogLeNet) to 0.9931 (ResNet-18).

Campanella *et al.* [26] and also used convolutional NN but had best results with trained ResNet-34 models, as opposed to us where AlexNet and ResNet-18 had similar performance.

As mentioned before, Jiang *et al.* [25] developed two DL frameworks (a 'cascade' and a 'segmentation' framework) that analyses smartphone-captured images from the microscope ocular. A classification model for detecting problematic cases as well as a segmentation model for a more in-depth investigation of the challenging scenarios were incorporated in the 'cascade' framework which attained 0.93 SN, 0.91 SP and 0.976 AUCs [26]. All photos were directly segmented and categorized using the 'segmentation' framework, which was more accurate than the 'cascade' one, achieving 0.97 SN, 0.94 SP and 0.987 AUC [25]. All thought Jiang *et al.* [25] obtained better SN and SP when compared to our study (SN between 72.9 and 71.93, SP between 98.01 and 97.90), the main focus of the software was different, theirs being BCC recognitions and ours being subtype classification.

Santilli *et al.* used DL framework to detect BCC from tissue burns on surgical margins using real-time rapid evaporative ionization mass spectrometry technology (REIMS) and recorded each burn as a peak in a chromatogram which they later analyzed using a DL framework and achieved an average ACC (SD) of 96.62% (1.35%) for the classification between burnt BCC and non-tumoral tissue [30]. Furthermore, they obtained a SN of 100.00% and SP of 95.00% for the two compared groups [30] which stands to prove the diversity of DL technology uses.

Our previous experience with transfer learning on prostate carcinoma image classification has shown no statistical differences on ACC between AlexNet and GoogLeNet, but when used on an external dataset, AlexNet had better results [38,67]. Similar to those findings, all three networks had comparable performance, GoogLeNet performed slightly better but with non-sustained significance.

Better results could have been obtained using a hardware color normalization before applying the algorithms.

As intended, we advocate that our application can assist pathologists by providing first or second opinions and shortening diagnosis time. Fernandez *et al.* also states that computational algorithms can minimize the strain of professionals and even provide a second opinion, boosting the number of accurate diagnoses [47]. According to Esteva *et al.*, DL systems could help doctors by providing second opinions and by indicating problematic areas in photos [68].

Digital pathology can be used in medical education for training purposes, to provide insight into cases or as an evaluation/testing strategy [69]. We think that DNN can add value to the previously stated benefits by providing students and young doctors with a self or remote-study option that can be accessed at any given time. Jahn *et al.* states that the first contact pathologists have with digital slides is for medical teaching purposes and specifies that things like image resolution or data storage are of secondary importance [70]. We envision the use of our application in medical universities and teaching hospitals and foresee it as a great asset for young practitioners.

Study limitations

An important limitation in the research comparing human and algorithms performance is a lack of clinical context. Computer algorithms do not take into consideration patients' clinical setting (such as disease history, family history, other paraclinical tests, patient anamnesis and so on). They only take into consideration the input image, thus creating a bias in the final diagnosis.

Another limitation is the size of the data set. To achieve the best possible image analysis system performance, large datasets are required. Our data set falls into the small dataset for a specific task category, and this is due to the limitation of cases of the hospital itself. More than that, the dataset is unbalanced due to the natural distribution of the cases. We must make a special remark in the context of the coronavirus disease 2019 (COVID-19) pandemic, which drastically reduced the number of operated BCC cases in our hospital. Without a doubt, the present COVID-19 pandemic, as well as the social distancing restrictions that have accompanied it, has also provided an advantage by boosting digital pathology.

Authors, such as Esteva *et al.*, also highlight the need for larger labelled datasets and state that smaller datasets have an easier data collection process but tend to perform poorly than larger ones [68].

Nevertheless, the cost of digital slide storage, secured servers and the cost of image analysis systems development and maintenance must be taken into consideration and interpreted in a cost-effective manner when assessing its use in pathological diagnosis. In a comprehensive study, Farahani *et al.* [71] evaluates the advantages and limitations of WSI and concludes that it will transform pathologists practice and serve as beneficial for the patients' diagnosis. Another broad research was done by Jahn *et al.* the pros, cons and emerging perspectives have been researched and the issue of cost-effectiveness has been tackled, although a definitive verdict is yet to be given, due to the scant number of studies in this domain and to the often-estimated costs derived from single-center experiences that are difficult to generalize [70].

Conclusions

The paper shows the concept, implementation, and validation of a DL approach on BCC subtype/variant image classification, based on the concept of transfer learning and trained on a dataset of 2249 images labeled with 10 subtypes and grouped in HR and LR. The NNs were obtained from AlexNet, GoogLeNet and ResNet-18 and achieved an average ACC/SN/SP/AUC of around 82%/72%/98%/0.99, with a slightly better performance form GoogLeNet. In the matter of agreement (Cohen's *kappa*), the networks had higher agreement performance than the pathologists. The best agreement was registered between AlexNet and GoogLeNet. A standalone application packed the best performing networks of each architecture, and together with the three pre-trained NN "*.MAT*"-files was offered for free on a public repository.

Conflict of interests

The authors declare no conflict of interests.

Informed consent statement

Informed consent was obtained from all subjects involved in the study, and the research was approved (No. 7749/21.09.2021) by the Research Ethics Committee of the Clinical Municipal Hospital, Cluj-Napoca, Romania.

Acknowledgments

This research was funded by the University of Medicine and Pharmacy of Craiova, grant number 26/24c/13.07.2021.

We are grateful to Dr. Cătălina Ileana Bungărdean and Dr. Diana Maria Zamfirescu, from the Clinical Municipal Hospital, Cluj-Napoca, for their support on image labeling.

References

[1] Peris K, Fargnoli MC, Garbe C, Kaufmann R, Bastholt L, Seguin NB, Bataille V, Marmol VD, Dummer R, Harwood CA, Hauschild A, Höller C, Haedersdal M, Malvehy J, Middleton MR, Morton CA, Nagore E, Stratigos AJ, Szeimies RM, Tagliaferri L, Trakatelli M, Zalaudek I, Eggermont A, Grob JJ; European Dermatology Forum (EDF), the European Association of Dermato-Oncology (EADO) and the European Organization for Research and Treatment of Cancer (EORTC). Diagnosis and treatment of basal cell carcinoma: European consensus-based interdisciplinary guidelines. Eur J Cancer, 2019, 118:10–34. https://doi.org/10.1016/j.ejca.2019.06.003 PMID: 31288208

[2] Trakatelli M, Morton C, Nagore E, Ulrich C, Del Marmol V, Peris K, Basset-Seguin N; BCC Subcommittee of the Guidelines Committee of the European Dermatology Forum. Update of the European guidelines for basal cell carcinoma management. Eur J Dermatol, 2014, 24(3):312–329. https://doi.org/10.1684/ejd.2014.2271 PMID: 24723647

[3] Marzuka AG, Book SE. Basal cell carcinoma: pathogenesis, epidemiology, clinical features, diagnosis, histopathology, and management. Yale J Biol Med, 2015, 88(2):167–179. PMID: 26029015 PMCID: PMC4445438

[4] Wortsman X, Carreño L, Morales C. Skin cancer: the primary tumors. In: Wortsman X (ed). Dermatologic ultrasound with clinical and histologic correlations. Springer, New York, NY, USA, 2013, 249–282. https://doi.org/10.1007/978-1-4614-7184-4_9 https://doi.org/10.1007/978-1-4614-7184-4 https://link.springer.com/chapter/10.1007/978-1-4614-7184-4_9

[5] Scolyer RA. Keratinocytic/epidermal tumours. In: Elder DE, Massi D, Scolyer RA, Willemze R (eds). World Health Organization (WHO) Classification of skin tumours. 4th edition, vol. 11, International Agency for Research on Cancer (IARC) Press, Lyon, France, 2018, 23–63. https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/WHO-Classification-Of-Skin-Tumours-2018

[6] Bichakjian CK, Olencki T, Aasi SZ, Alam M, Andersen JS, Berg D, Bowen GM, Cheney RT, Daniels GA, Glass LF,

Grekin RC, Grossman K, Higgins SA, Ho AL, Lewis KD, Lydiatt DD, Nehal KS, Nghiem P, Olsen EA, Schmults CD, Sekulic A, Shaha AR, Thorstad WL, Tuli M, Urist MM, Wang TS, Wong SL, Zic JA, Hoffmann KG, Engh A. Basal cell skin cancer, version 1.2016, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw, 2016, 14(5):574–597. https://doi.org/10.6004/jnccn.2016.0065 PMID: 27160235

[7] Sexton M, Jones DB, Maloney ME. Histologic pattern analysis of basal cell carcinoma. Study of a series of 1039 consecutive neoplasms. J Am Acad Dermatol, 1990, 23(6 Pt 1):1118–1126. https://doi.org/10.1016/0190-9622(90)70344-h PMID: 2273112

[8] Saldanha G, Fletcher A, Slater DN. Basal cell carcinoma: a dermatopathological and molecular biological update. Br J Dermatol, 2003, 148(2):195–202. https://doi.org/10.1046/j.1365-2133.2003.05151.x PMID: 12588368

[9] Vantuchová Y, Čuřík R. Histological types of basal cell carcinoma. Scripta Medica (Brno), 2006, 79(5–6):261–270. https://www2.med.muni.cz/biomedjournal/pdf/2006/05/261_270.pdf

[10] Arevalo J, Cruz-Roa A, Arias V, Romero E, González FA. An unsupervised feature learning framework for basal cell carcinoma image analysis. Artif Intell Med, 2015, 64(2):131–145. https://doi.org/10.1016/j.artmed.2015.04.004 PMID: 25976208

[11] He L, Long LR, Antani S, Thoma GR. Histology image analysis for carcinoma detection and grading. Comput Methods Programs Biomed, 2012, 107(3):538–556. https://doi.org/10.1016/j.cmpb.2011.12.007 PMID: 22436890 PMCID: PMC3587978

[12] Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. Med Image Comput Comput Assist Interv, 2013, 16(Pt 2):403–410. https://doi.org/10.1007/978-3-642-40763-5_50 PMID: 24579166

[13] ***. Computational intelligence: a logical approach. Choice Reviews Online, 1998, 35. https://doi.org/10.5860/choice.35-5701 https://www.choice360.org/features/

[14] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ (eds). Advances in Neural Information Processing Systems (NIPS) 25, 26th Annual Conference on NIPS 2012, Lake Tahoe, Nevada, USA, December 3–6, 2012. Curran Associates, Inc., New York, USA, 2013, 1097–1105. https://www.proceedings.com/content/017/017576webtoc.pdf https://papers.nips.cc/paper/2012

[15] Ciresan D, Giusti A, Gambardella L, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ (eds). Advances in Neural Information Processing Systems (NIPS) 25, 26th Annual Conference on NIPS 2012, Lake Tahoe, Nevada, USA, December 3–6, 2012. Curran Associates, Inc., New York, USA, 2013, 2843–2851. https://www.proceedings.com/content/017/017576webtoc.pdf https://papers.nips.cc/paper/2012

[16] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. 2015 Institute of Electrical and Electronics Engineers (IEEE) Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015, 1–9. https://doi.org/10.1109/CVPR.2015.7298594 https://ieeexplore.ieee.org/document/7298594/authors#authors

[17] Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv Preprint, 2017 Aug 28, arXiv:1708.08296. https://doi.org/10.48550/arXiv.1708.08296 https://arxiv.org/abs/1708.08296

[18] Hagras H. Toward human-understandable, explainable AI. Computer, 2018, 51(9):28–36. https://doi.org/10.1109/MC.2018.3620965 https://ieeexplore.ieee.org/document/8481251

[19] Montavon G, Lapuschkin S, Binder A, Samek W, Müller KR. Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recognit, 2017, 65:211–222. https://doi.org/10.1016/j.patcog.2016.11.008 https://www.sciencedirect.com/science/article/pii/S0031320316303582?via%3Dihub

[20] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. Workshop at International Conference on Learning Representations 2014, arXiv Preprint, 2013 Dec 20, arXiv:1312.6034. https://doi.org/10.48550/arxiv.1312.6034 https://arxiv.org/abs/1312.6034

[21] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European Conference on Computer Vision (ECCV) 2014, arXiv Preprint, 2013 Nov 12, arXiv:1311.2901. https://doi.org/10.48550/arxiv.1311.2901 https://arxiv.org/abs/1311.2901

[22] Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One, 2015, 10(7):e0130140. https://doi.org/10.1371/journal.pone.0130140 PMID: 26161953 PMCID: PMC4498753

[23] van Zon MCM, van der Waa JD, Veta M, Krekels GAM. Whole-slide margin control through deep learning in Mohs micrographic surgery for basal cell carcinoma. Exp Dermatol, 2021, 30(5):733–738. https://doi.org/10.1111/exd.14306 PMID: 33656186

[24] Albahar MA. Skin lesion classification using convolutional neural network with novel regularizer. IEEE Access, 2019, 7:38306–38313. https://doi.org/10.1109/ACCESS.2019.2906241 https://ieeexplore.ieee.org/document/8669763

[25] Jiang YQ, Xiong JH, Li HY, Yang XH, Yu WT, Gao M, Zhao X, Ma YP, Zhang W, Guan YF, Gu H, Sun JF. Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network. Br J Dermatol, 2020, 182(3):754–762. https://doi.org/10.1111/bjd.18026 PMID: 31017653

[26] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med, 2019, 25(8):1301–1309. https://doi.org/10.1038/s41591-019-0508-1 PMID: 31308507 PMCID: PMC7418463

[27] Kimeswenger S, Tschandl P, Noack P, Hofmarcher M, Rumetshofer E, Kindermann H, Silye R, Hochreiter S, Kaltenbrunner M, Guenova E, Klambauer G, Hoetzenecker W. Artificial neural networks and pathologists recognize basal cell carcinomas based on different histological patterns. Mod Pathol, 2021, 34(5):895–903. https://doi.org/10.1038/s41379-020-00712-7 PMID: 33184470

[28] Andrew TW, Hamnett N, Roy I, Garioch J, Nobes J, Moncrieff MD. Machine-learning algorithm to predict multidisciplinary team treatment recommendations in the management of basal cell carcinoma. Br J Cancer, 2022, 126(4):562–568. https://doi.org/10.1038/s41416-021-01506-7 PMID: 34471257 PMCID: PMC8854628

[29] Campanella G, Navarrete-Dechent C, Liopyris K, Monnier J, Aleissa S, Minhas B, Scope A, Longo C, Guitera P, Pellacani G, Kose K, Halpern AC, Fuchs TJ, Jain M. Deep learning for basal cell carcinoma detection for reflectance confocal microscopy. J Invest Dermatol, 2022, 142(1):97–103. https://doi.org/10.1016/j.jid.2021.06.015 PMID: 34265329

[30] Santilli AML, Jamzad A, Janssen NNY, Kaufmann M, Connolly L, Vanderbeck K, Wang A, McKay D, Rudan JF, Fichtinger G, Mousavi P. Perioperative margin detection in basal cell carcinoma using a deep learning framework: a feasibility study. Int J Comput Assist Radiol Surg, 2020, 15(5):887–896. https://doi.org/10.1007/s11548-020-02152-9 PMID: 32323209

[31] Putten EV, Kambod A, Kambod M. Deep residual neural networks for automated basal cell carcinoma detection. In: ***. 2018 Institute of Electrical and Electronics Engineers (IEEE) Engineering in Medicine and Biology Society (EMBS) International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 4–7 March 2018, 337–340. https://doi.org/10.1109/BHI.2018.8333437 https://ieeexplore.ieee.org/document/8333437

[32] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017, 542(7639):115–118. https://doi.org/10.1038/nature21056 PMID: 28117445 PMCID: PMC8382232

[33] Hosny KM, Kassem MA, Fouad MM. Classification of skin lesions into seven classes using transfer learning with AlexNet. J Digit Imaging, 2020, 33(5):1325–1334. https://doi.org/10.1007/s10278-020-00371-9 PMID: 32607904 PMCID: PMC7573031

[34] Jinnai S, Yamazaki N, Hirano Y, Sugawara Y, Ohe Y, Hamamoto R. The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules, 2020, 10(8):1123. https://doi.org/10.3390/biom10081123 PMID: 32751349 PMCID: PMC7465007

[35] Thomas SM, Lefevre JG, Baxter G, Hamilton NA. Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. Med Image Anal, 2021, 68:101915. https://doi.org/10.1016/j.media.2020.101915 PMID: 33260112
[36] ***. ImageNet, 2021. ImageNet [online], Stanford Vision Lab, Stanford University, Princeton University, accessed: July 7, 2021. https://image-net.org/
[37] Bozinovski S. Reminder of the first paper on transfer learning in neural networks, 1976. Informatica, 2020, 44(3):291–302. https://doi.org/10.31449/inf.v44i3.2828 https://www.informatica.si/index.php/informatica/article/view/2828
[38] Şerbănescu MS, Manea NC, Streba L, Belciug S, Pleşea IE, Pirici I, Bungărdean RM, Pleşea RM. Automated Gleason grading of prostate cancer using transfer learning from general-purpose deep-learning networks. Rom J Morphol Embryol, 2020, 61(1):149–155. https://doi.org/10.47162/RJME.61.1.17 PMID: 32747906 PMCID: PMC7728132
[39] Nica RE, Şerbănescu MS, Florescu LM, Camen GC, Streba CT, Gheonea IA. Deep learning: a promising method for histological class prediction of breast tumors in mammography. J Digit Imaging, 2021, 34(5):1190–1198. https://doi.org/10.1007/s10278-021-00508-4 PMID: 34505960 PMCID: PMC8554900
[40] ***. MATLAB, 2021: Pretrained deep neural networks. MATLAB & Simulink [online], MathWorks, Inc., accessed: July 7, 2021. https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html
[41] ***. GitHub, 2021: BVLC AlexNet model [online]. GitHub, Inc., accessed: July 7, 2021. https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet
[42] ***. GitHub, 2021: BVLC GoogLeNet model [online]. GitHub, Inc., accessed: July 7, 2021. https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet
[43] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: ***. 2016 Institute of Electrical and Electronics Engineers (IEEE) Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016, 770–778. https://doi.org/10.1109/CVPR.2016.90 https://ieeexplore.ieee.org/document/7780459
[44] McHugh ML. Interrater reliability: the *kappa* statistic. Biochem Med (Zagreb), 2012, 22(3):276–282. PMID: 23092060 PMCID: PMC3900052
[45] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics, 1977, 33(1):159–174. PMID: 843571
[46] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 3rd edition, Book Series: Wiley Series in Probability and Statistics, Wiley–Interscience, John Wiley & Sons, Inc., New York, USA, 2003. https://doi.org/10.1002/0471445428 https://onlinelibrary.wiley.com/doi/book/10.1002/0471445428
[47] Fernandes V, Junior GB, de Paiva AC, Silva AC, Gattass M. Bayesian convolutional neural network estimation for pediatric pneumonia detection and diagnosis. Comput Methods Programs Biomed, 2021, 208:106259. https://doi.org/10.1016/j.cmpb.2021.106259 PMID: 34273674
[48] Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, Huang CS, Shen D, Chen CM. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. Sci Rep, 2016, 6:24454. https://doi.org/10.1038/srep24454 PMID: 27079888 PMCID: PMC4832199
[49] Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, Barfett J. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Invest Radiol, 2017, 52(5):281–287. https://doi.org/10.1097/RLI.0000000000000341 PMID: 27922974
[50] Şerbănescu MS, Pleşea IE. A hardware approach for histological and histopathological digital image stain normalization. Rom J Morphol Embryol, 2015, 56(2 Suppl):735–741. PMID: 26429166
[51] Fuchs TJ, Buhmann JM. Computational pathology: challenges and promises for tissue analysis. Comput Med Imaging Graph, 2011, 35(7–8):515–530. https://doi.org/10.1016/j.compmedimag.2011.02.006 PMID: 21481567
[52] Louis DN, Feldman M, Carter AB, Dighe AS, Pfeifer JD, Bry L, Almeida JS, Saltz J, Braun J, Tomaszewski JE, Gilbertson JR, Sinard JH, Gerber GK, Galli SJ, Golden JA, Becich MJ. Computational pathology: a path ahead. Arch Pathol Lab Med, 2016, 140(1):41–50. https://doi.org/10.5858/arpa.2015-0093-SA PMID: 26098131 PMCID: PMC4996078
[53] Belciug S. Artificial intelligence in cancer: diagnostic to tailored treatment. 1st edition, Elsevier–Academic Press, San Diego, USA, 2020. https://doi.org/10.1016/C2019-0-00740-6 https://www.sciencedirect.com/book/9780128202012/artificial-intelligence-in-cancer?via=ihub=
[54] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 2016, 316(22):2402–2410. https://doi.org/10.1001/jama.2016.17216 PMID: 27898976
[55] Cerentini A, Welfer D, Cordeiro d'Ornellas M, Pereira Haygert CJ, Dotto GN. Automatic identification of glaucoma using deep learning methods. Stud Health Technol Inform, 2017, 245:318–321. PMID: 29295107
[56] Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs *via* deep learning. Nat Biomed Eng, 2018, 2(3):158–164. https://doi.org/10.1038/s41551-018-0195-0 PMID: 31015713
[57] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with deep neural networks. Med Image Anal, 2017, 35:18–31. https://doi.org/10.1016/j.media.2016.05.004 PMID: 27310171
[58] Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hassen ABH, Thomas L, Enk A, Uhlmann L; Reader study level-I and level-II Groups; Alt C, Arenbergerova M, Bakos R, Baltzer A, Bertlich I, Blum A, Bokor-Billmann T, Bowling J, Braghiroli N, Braun R, Buder-Bakhaya K, Buhl T, Cabo H, Cabrijan L, Cevic N, Classen A, Deltgen D, Fink C, Georgieva I, Hakim-Meibodi LE, Hanner S, Hartmann F, Hartmann J, Haus G, Hoxha E, Karls R, Koga H, Kreusch J, Lallas A, Majenka P, Marghoob A, Massone C, Mekokishvili L, Mestel D, Meyer V, Neuberger A, Nielsen K, Oliviero M, Pampena R, Paoli J, Pawlik E, Rao B, Rendon A, Russo T, Sadek A, Samhaber K, Schneiderbauer R, Schweizer A, Toberer F, Trennheuser L, Vlahova L, Wald A, Winkler J, Wölbing P, Zalaudek I. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol, 2018, 29(8):1836–1842. https://doi.org/10.1093/annonc/mdy166 PMID: 29846502
[59] Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal, 2017, 35:303–312. https://doi.org/10.1016/j.media.2016.07.007 PMID: 27497072
[60] Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Comput Assist Interv, 2013, 16(Pt 2):411–418. https://doi.org/10.1007/978-3-642-40763-5_51 PMID: 24579167
[61] Combalia M, Garcia S, Malvehy J, Puig S, Mülberger AG, Browning J, Garcet S, Krueger JG, Lish SR, Lax R, Ren J, Stevenson M, Doudican N, Carucci JA, Jain M, White K, Rakos J, Gareau DS. Deep learning automated pathology in *ex vivo* microscopy. Biomed Opt Express, 2021, 12(6):3103–3116. https://doi.org/10.1364/BOE.422168 PMID: 34221648 PMCID: PMC8221965
[62] Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, Venugopalan S, Timofeev A, Nelson PQ, Corrado GS, Hipp JD, Peng L, Stumpe MC. Detecting cancer metastases on gigapixel pathology images. arXiv Preprint, 2017 Mar 3, arXiv:1703.02442. https://doi.org/10.48550/arXiv.1703.02442 https://arxiv.org/abs/1703.02442
[63] Moisejenko-Golubovica J, Volkov O, Ivanova A, Groma V. Analysis of the occurrence and distribution of primary and recurrent basal cell carcinoma of head and neck coupled to the assessment of tumor microenvironment and Sonic hedgehog signaling. Rom J Morphol Embryol, 2020, 61(3):821–831. https://doi.org/10.47162/RJME.61.3.20 PMID: 33817723 PMCID: PMC8112792
[64] Mercuţ IM, Simionescu CE, Stepan AE, Andreiana BC, Ciurea AM, Mercuţ R, Ciurea ME. The immunoexpression of

MMP-1 and MMP-13 in eyelid basal cell carcinoma. Rom J Morphol Embryol, 2020, 61(4):1221–1226. https://doi.org/10.47162/RJME.61.4.23 PMID: 34171070 PMCID: PMC8343475

[65] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds). Medical image computing and computer-assisted intervention – MICCAI 2015. Proceedings of the 18th International Conference, Munich, Germany, October 5–9, 2015, Part III. Book Series: Lecture Notes in Computer Science (LNCS), vol. 9351, Springer, Cham, Switzerland, 2015, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28 https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28 https://doi.org/10.1007/978-3-319-24574-4

[66] Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. 36th International Conference on Machine Learning (ICML) 2019. arXiv Preprint, 2019 May 28, arXiv:1905.11946. https://doi.org/10.48550/arxiv.1905.11946 https://arxiv.org/abs/1905.11946

[67] Şerbănescu MS, Oancea CN, Streba CT, Pleşea IE, Pirici D, Streba L, Pleşea RM. Agreement of two pre-trained deep-learning neural networks built with transfer learning with six pathologists on 6000 patches of prostate cancer from Gleason2019 Challenge. Rom J Morphol Embryol, 2020, 61(2):513–519. https://doi.org/10.47162/RJME.61.2.21 PMID: 33544803 PMCID: PMC7864291

[68] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nat Med, 2019, 25(1):24–29. https://doi.org/10.1038/s41591-018-0316-z PMID:30617335

[69] Boyce BF. Whole slide imaging: uses and limitations for surgical pathology and teaching. Biotech Histochem, 2015, 90(5):321–330. https://doi.org/10.3109/10520295.2015.1033463 PMID: 25901738

[70] Jahn SW, Plass M, Moinfar F. Digital pathology: advantages, limitations and emerging perspectives. J Clin Med, 2020, 9(11): 3697. https://doi.org/10.3390/jcm9113697 PMID: 33217963 PMCID: PMC7698715

[71] Farahani N, Parwani AV, Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathol Lab Med Int, 2015, 7:23–33. https://doi.org/10.2147/PLMI.S59826 https://www.dovepress.com/whole-slide-imaging-in-pathology-advantages-limitations-and-emerging-p-peer-reviewed-fulltext-article-PLMI

Corresponding author

Mircea-Sebastian Şerbănescu, Associate Professor, MD, PhD, Department of Medical Informatics and Biostatistics, University of Medicine and Pharmacy of Craiova, 2 Petru Rareş Street, 200349 Craiova, Dolj County, Romania; Phone +40745–766 610, e-mail: mircea_serbanescu@yahoo.com

Received: March 10, 2022

Accepted: June 6, 2022