Deep Convolutional Neural Networks on Automatic Classification for Skin Tumour Images

SVETLANA SIMIĆ[*], *Faculty of Medicine, University of Novi Sad, Hajduk Veljkova 1–9, 21000 Novi Sad, Serbia.*

SVETISLAV D. SIMIĆ[**], *Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia.*

ZORANA BANKOVIĆ[†], *Frontiers Media SA, Paseo de Castellana 77, Madrid, Spain.*

MILANA IVKOV-SIMIĆ[††], *Faculty of Medicine, University of Novi Sad, Hajduk Veljkova 1–9, 21000 Novi Sad, Serbia.*

JOSÉ R. VILLAR[§], *University of Oviedo, Campus de Llamaquique, 33005 Oviedo, Spain.*

DRAGAN SIMIĆ[§§], *Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia.*

Abstract

The skin, uniquely positioned at the interface between the human body and the external world, plays a multifaceted immunologic role in human life. In medical practice, early accurate detection of all types of skin tumours is essential to guide appropriate management and improve patients' survival. The most important issue is to differentiate between malignant skin tumours and benign lesions. The aim of this research is the classification of skin tumours by analysing medical skin tumour dermoscopy images. This paper is focused on a new strategy based on deep convolutional neural networks which have recently shown a state-of-the-art performance to define strategy to automatic classification for skin tumour images. The proposed system is tested on well-known *HAM10000* data set. For experimental results, verification is performed and the results are compared with similar researches.

Keywords: automatic classification, dermoscopy images, deep learning, convolutional neural networks

[*]E-mail: svetlana.simic@mf.uns.ac.rs
[**]E-mail: simicsvetislav@uns.ac.rs
[†]E-mail: zbankovic@gmail.com
[††]E-mail: milana.ivkov-simic@mf.uns.ac.rs
[§]E-mail: villarjose@uniovi.es
[§§]E-mail: dsimic@eunet.rs

Vol. 30, No. 4, © The Author(s) 2021. Published by Oxford University Press.

Advance Access published 4 March 2021 https://doi.org/10.1093/jigpal/jzab009

1 Introduction

The skin is uniquely positioned at the interface between the human body and the external world. It has an important immunologic role and, for that reason, the correlation between skin tumour development and immunologic mechanisms is intensely studied. Non-immunologic risk factors like individual predisposition, sun and environmental exposure are all contributing to the skin neoplasia incidence [1]. Early accurate detection of all skin tumour types is essential for patients' morbidity or survival.

Generally, in medical practice, the most important skin tumours could be classified either as malignant tumours or benign lesions [2]. The aim of this research is the classification of skin tumours on malignant tumours and benign lesions for skin tumour dermoscopy images. In this research, only malignant melanomas will be analysed. Melanomas are the most malignant skin tumours. They grow in melanocytes, the cells responsible for pigmentation. This type of skin cancer is rapidly increasing, though its related mortality rate is increasing more modest. The critical factor in the assessment of a patient prognosis in skin cancer is early diagnosis.

As melanoma incidence has risen rapidly in many parts of the world, so has melanoma mortality, albeit at a slower pace. For example, in the USA, melanoma incidence increased by about fivefold between 1950 and 1990, whereas mortality increased by slightly less than twofold during the same period [3]. Similar observations have been made in western and northern Europe [4, 5]. More than 60,000 people in the USA were diagnosed with invasive melanoma in 2000, and more than 8,000 died of the disease [6]. In 2011, an estimated 70,230 adults in the USA were diagnosed with melanoma. It is estimated that 8,790 deaths from melanoma will occur within a year. Almost 10,130 deaths from melanoma were recorded in 2019 [7]. In Australia, an analysis conducted in the mid-1990s demonstrated that melanoma mortality was still climbing overall; however, among the youngest birth cohorts, there were early signs of a decline in mortality. The most recent data [8] suggest that between 1989 and 2002, overall melanoma mortality stabilized in Australian males and declined in females (−0.8% per annum). Its frequency is rising in many other countries; for example, 10 cases were reported each year in Algeria [9].

Currently, experienced dermatologists use dermoscopy analysis and observation based on dermatology criteria to pose a working diagnosis of a type of a skin tumour. The availability of an objective system able to classify moles helps physician dermatologists in diagnosis and early detection of melanoma. In the development of such a system it is not necessary to classify every mole, but rather to achieve such a precision that all malignant tumours are classified as dangerous and none is overlooked. The evaluation of the clinical exams' accuracy showed that without the use of a dermatoscope, dermatologists can detect only 65–80% of cases of melanoma. The dermoscopy increases the diagnostic accuracy rate for about 10–27% [10].

The aim of this research is the classification of skin tumours by analysing medical skin tumour dermoscopy images. This paper is focused on a new strategy based in general deep learning (DL) and its application on deep convolutional neural networks (DCNNs) which combines mathematics and artificial intelligence techniques. DCNNs have the highest level of vision tasks, such as image classification and object detection, and recently have shown a state-of-the-art performance to define strategy to automatic classification for skin tumour images. The implemented model is tested on the part of well-known skin images of *HAM10000* data set. This paper is an extension of our previous research [11] and continues the authors' previous researches in computer-assisted diagnosis methods presented in [12, 13, 14, 15].

The rest of the paper is organized in the following way: Section 2 provides an overview of the basic idea on image classification and related work. Section 3 presents modelling the automatic

classification for skin tumour. The preliminary experimental results tested with the well-known *HAM10000* data set are presented in Section 4. Section 5 provides conclusions and some points for future work.

2 Image classification and related work

Medical data collection can necessitate an incredible amount of time and effort; however, once collected, the data/information can be utilized in several ways: (i) to improve early detection, diagnosis and treatment; (ii) to predict patient diagnosis; (iii) to speck early warning symptoms and mobilize resources to proactively address care; (iv) to increase interoperability and interconnectivity of healthcare; (v) to enhance patient care via mobile health, telemedicine, as well as self-tracking and/or home devices [16]. Classification is used mostly as a supervised learning method, and the goal of the classification is predictive. General references regarding data classification are presented in [17], and the very good contemporary hybrid classification techniques can be found in the textbook [18].

2.1 Image skin tumour classification—classical methods

In the past three decades, many approaches, which can be called c*lassical methods*, have been proposed to solve the skin tumour image classification problem to help physicians make decisions regarding this particular illness and future patient treatments. Several dermatoscopic rules in these methods, for the automatic classification systems of melanoma in order to generate new high-level features, are based on the golden rule following the dermatology criteria named: *ABCDE* (***A****symmetry*, ***B****order*, ***C****olour*, ***D****imension*, ***E****volution*); 7-point checklist; Menzies method; CASH (***C****olour*, ***A****rchitecture*, ***S****ymmetry*, ***H****omogeneity*) algorithm—allowing semantic analysis depicted in [11].

The proposed flow diagram for modelling the automatic classification for skin tumour images based on classical methods is presented in Figure 1 [11].

2.2 Image skin tumour classification—DL methods

The paper [19] presents a unified method for histopathology image representation learning, visual analysis interpretation and automatic classification of skin histopathology images as either having basal cell carcinoma or not. The novel approach is inspired by ideas from image feature representation learning and DL and yields a DL architecture that combines an autoencoder learning layer, a convolutional layer. A softmax classifier for cancer detection, a generalization of a logistic regression classifier [20], and visual analysis interpretation are used.

The research paper [21] presents the state-of-the-art approach for solving the automatic skin tumour image classification based on DL. The DCNN Inception-v3, which has been previously trained *end-to-end* to ImageNet, a hierarchical image database, is used. The CNN is additionally *fine-tuned* using the dataset of 129,450 skin lesion clinical images consisting of 2,032 different diseases. The proposed algorithm is validated by comparing experimental results provided by the neural network in respect of the diagnosis by two dermatologists, wherein the skin disorders were grouped into nine different classes. The CNN in this situation achieves *precision* of 55.4%, which is better than the results achieved by dermatologists on a subset of the test data set, since their *precision* was 53.4% and 55.0%. And, finally, the performance on a binary classification of benign and malignant changes was measured, achieving the *accuracy* of 96%.

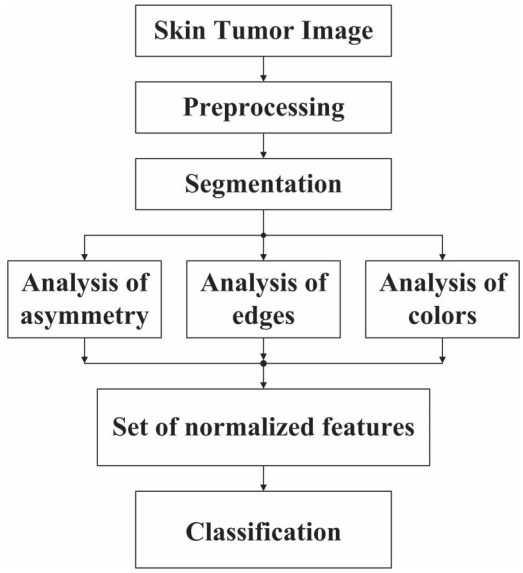

FIGURE 1. Flow diagram for automatic classification images based on classical methods [11].

Another state-of-the-art DL approach is presented in paper [22]. That paper uses DL to classify clinical images of skin lesions in 12 different classes of skin diseases: (i) basal cell carcinoma, (ii) squamous cell carcinoma, (iii) intraepithelial carcinoma, (iv) actinic keratosis, (v) seborrheic keratosis, (vi) malignant melanoma, (vii) melanocytic nevus, (viii) lentigo, (ix) pyogenic granuloma, (x) hemangioma, (xi) dermatofibroma and (xii) wart. The CNN (Microsoft ResNet-152 model; Microsoft Research Asia, Beijing, China) was fine-tuned with images from the training portion of the Asan dataset, MED-NODE dataset, and atlas site images, 19,398 images in total. The trained model was validated with the testing portion of the Asan, Hallym and Edinburgh datasets. With the Asan dataset, *the area under the curve* for the diagnosis of basal cell carcinoma, squamous cell carcinoma, intraepithelial carcinoma and melanoma was 0.96 ± 0.01, 0.83 ± 0.01, 0.82 ± 0.02 and 0.96 ± 0.00, respectively. With the Edinburgh dataset, the area under the curve for the corresponding diseases was 0.90 ± 0.01, 0.91 ± 0.01, 0.83 ± 0.01 and 0.88 ± 0.01, respectively. With the Hallym dataset, the sensitivity for basal cell carcinoma diagnosis was $87.1\% \pm 6.0\%$. The tested algorithm performance with 480 Asan and Edinburgh images was comparable to that of 16 dermatologists. And finally, the performance of CNN could be improved with additional images with a broader range of ages and ethnicities. The accuracy of the system in the classification of the image to one of 12 categories was between 55.7% and 57.3%, depending on the data set used for validation, while differentiating melanoma from other diseases was correct in 83% and 96% of cases, depending on the dataset. These results are comparable to those presented in paper [21], and it can be noted that these approaches are considered today as state-of-the-art.

Smartphone applications are readily accessible and potentially offer an instant risk assessment of the likelihood of malignancy, so that people could immediately seek further medical attention from a clinician for more detailed assessment of the lesion. There is, however, a risk that melanomas might be missed and treatment delayed if the application reassures the user that their lesion is low risk [23].

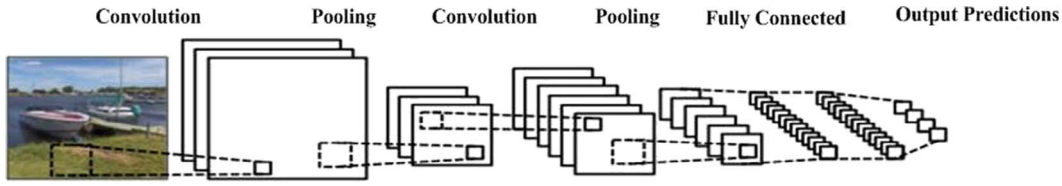

FIGURE 2. Typical convolutional neural network architecture [30].

SkinVision [24] presents a mobile application available for Android [25] and iOS [26] devices for which it is claimed that it allows users to self-check their moles. The application uses CNN to analyse the created image. In the studies from 2014, an older version of this application presented the accuracy of 81% in identifying melanoma [27, 28]. In the new study, from 2019 [24], the accuracy is improved and it now stands at 95%. Developers of this applications claim that these applications helped diagnose more than 35,000 cancer patients.

3 Modelling the automatic classification for skin tumour

The problem with the classical artificial neural networks is in increasing the number of neurons in layers; therefore, training the network becomes rather computationally demanding, since the number of connections increases quadratic with the number of neurons in the layers. Likewise, fully connected layers cause the network becoming inclined towards overfitting. Dropout has brought significant advances to modern neural networks and it is being considered as one of the most powerful techniques to avoid overfitting.

3.1 Convolutional neural network—basic approach

DL is a branch of machine learning architectures that attempts to model high-level abstractions in data using multiple processing layers. One of the DL models, the CNN, is used to recognize cursive numbers and has demonstrated to be useful in object recognition [29]. It was published more than 20 years ago. CNNs are commonly used for processing and analysing the image because they are designed in a special way that allows them to recognize the two-dimensional shape with a high invariant degree of the translation, scaling, distortion and other forms of distortion (Figure 2) [30].

CNNs have emerged as a powerful classification tool and are consistently used in object classification competitions. Several factors have contributed to the success of research using neural networks, including: (i) the acquisition of sufficiently large volumes of data required for the training of neural network models through the internet; (ii) improvements in graphic processing unit performance and the development of methods using the graphic processing unit for computation; and (iii) the advancement of various DL methods such as rectified linear unit (ReLU), dropout and batch normalization [31, 32]. Despite these technological advances, however, the lack of a valid clinical dataset has limited the applications of DL research in medicine.

3.2 Approach to transfer learning

The end-to-end learning paradigm and design neural network architecture is a basic approach to transfer learning (TL). TL approach for the tasks of image-sentence matching, image captioning tasks and image region-annotation is used (Figure 3).

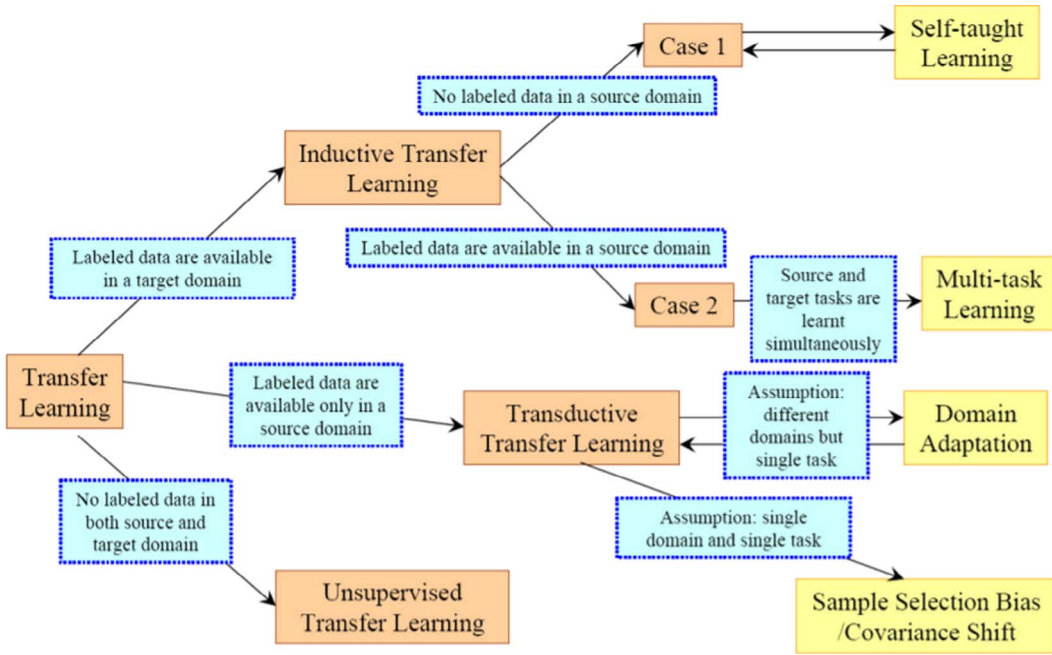

FIGURE 3. TL approach—'*end-to-end*' learning paradigm [33].

The objective of TL is to take advantage of data from the first setting to extract information that may be useful when learning or even when directly making predictions in the second setting. Due to the benefits of end-to-end training approach and TL, these are quality experimental results from large-scale related datasets such as ImageNet [33].

3.3 Basic architecture for the automatic classification system

The automatic classification system for skin tumour is based generally on deep architecture and partially on the use of CNN with the TL techniques. The basic architecture is based on typical CNN architecture (Figure 1) and the inputs in the system are: skin tumour images and related meta-data. On the images, no previous corrections have been performed. The related meta-data consist of main information about images and information about the class image to which it belongs. The whole process is divided into two phases: (i) testing and (ii) verifications. In the testing phase, the input in the system is the image only, while the output appears as a prediction of the class picture, precisely the probability that the image belongs to one or another class.

VGG-16 network architecture with fine-tuned weights of the ImageNet data set is used. This architecture is first-time presented and described in details in the research paper [34]. ImageNet data set consists of 14,000 images divided into 1,000 classes. VGG-16 architecture on this data-set has the accuracy of 92.7%, and 'true prediction' is in the top five predictions considered. VGG-16 network architecture consists of 16 layers divided into six segments. All segments except the last segment contain few convolution layers: first two segments have two convolution layers each, while other segments have three convolution layers each. The last layer in VGG-16 architecture is the max pooling layer. It consists of three fully connected layers, where first two layers contain 4,096 neurons

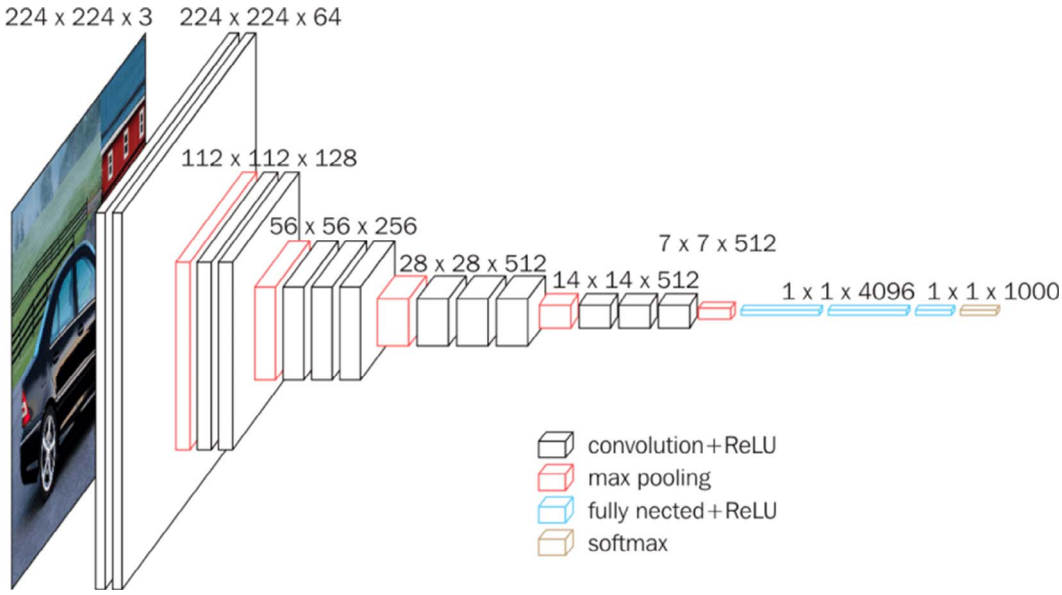

FIGURE 4. VGG-16 neural network architecture [34].

and the last layer, which is the output layer, contains 1,000 neurons, with the number of classes in IamageNet data set. The image size in the input layer is 224 × 224 RGB pixels. The schema of VGG-16 CNN architecture is presented in Figure 4.

3.4 Architecture of the automatic classification system for skin tumour images

The aim of this research is the automatic classification system for skin tumour obtained by analysing medical skin tumour dermoscopy images. Particularly, the focus of this research is the automatic classification system for skin tumours obtained by analysing medical skin tumour dermoscopy images for the most important classes: malignant skin tumours—melanomas and benign lesions. Therefore, some modifications of the basic architecture for the automatic classification system, presented in sub-section 3.3, have been developed for the automatic skin tumour classification system.

First of all, the last layer contains only two neurons. There are many possible classes that exist—melanomas and benign lesions. Using the *softmax* activation function, the output of the neurons transforms the probability of the classes and the sum is one. In this network architecture model, all weighted parameters are fixed in every layer, except in the last layer. Of a total of 138,268,738 parameters, 8,194 of them are trained.

The second tested method applied, next to classical training, where the modified architecture is tested, is the approach in which the images apply *augmentation* techniques on the input data set before training. These *augmentation* techniques involve the following: to rotate up to 20 degrees, to zoom up to 15%, to change the width and height up to 20%, to shear for up to 15 degrees and finally to perform a horizontal flip. It means that five different techniques are used on the input data set; so finally, the input data set consists of 10,670 skin images for this VGG-16 architecture.

The third tested method applied modified all fully connected layers, so that two layers had 128 neurons, while the last layer, again, had to have two of them. In this architecture, the values of network parameter values are learning in the last three completely connected layers. Then, out of the total of 17,942,850 network parameters, 3,228,162 of them are trained. The training was held in the mini-batch mode through 50 epochs.

3.5 Implementation of the system

The system is completely implemented in clean, efficient Python programming language and with fully open-source libraries. As a development environment, *Spyder* [35] was used. In the initial stages of development solutions, as well as during subsequent testing of different combinations of processing and image analysis, *Jupyter Notebook* was used [36]. Publicly available implementation of *VGG-16 network* is used as well [37].

In the optimization process, *Root Mean Square Propagation* (*RMSProp*) optimizer is used. Geoffrey Hinton from the University of Toronto, the most important promoter of this iterative method for optimizing, did not publish *RMSprop Optimization Algorithm for Gradient Descent with Neural Networks* in a formal academic paper, but it still is one of the most popular gradient descent optimization algorithms for DL, and he presented it in [38]. The gist of *RMSprop* algorithm is as follows: (i) maintain a moving (discounted) average of the square of gradients and (ii) divide the gradient by the root of this average.

In this experiment, the following value of learning parameters: *learning_rate* = 0.001; *rho* = 0.9; *momentum* = 0.0; *epsilon* = 1e-07; *centered* = False; are used in the optimization process. The discounting factor for the history/coming gradient is presented by the parameter *rho*. When the parameter *centered* = True, gradients are normalized by the estimated variance of the gradient; on the other side, when *centered* = False, they are normalized by the uncentered second moment. Setting this parameter to True may be helpful in the training process, but it is more expensive in terms of computation time, consuming memory and CPU usage.

4 Experimental results and discussion

The proposed automatic classification system for skin tumour images is based on DCNNs, and in our research, it was tested on tumour images from the well-known repository *HAM10000* data set [39]. This repository includes representative collection of important diagnostic categories in pigmented lesions.

4.1 Data set

HAM10000 data set ('***H****uman* ***A****gainst* ***M****achine with 10000 training images*') is a collection of dermatoscopic images from different populations, acquired and stored by different modalities; nevertheless, this data set does not contain clinical dimensions of moles on the images, nor any other data that could help determine the size of the mole.

The content of the repository *HAM10000* data set which consists of 10,015 dermatoscopic images divided into seven classes is presented in Table 1. From the content of the repository *HAM10000* data set, the first images that are eliminated are images with vignette (that shades off gradually into the surrounding photograph), since vignette makes skin image very asymmetric, and it is difficult to recognize differences between skin tumour and vignette. Then, following *HAM10000* metadata, with approximately same number of benign and malignant skin tumours, 971 cases of benign skin

TABLE 1. Repository *HAM10000* data set, number of classes and instances.

Classes	Instances
Bowen's disease	327
Basal cell carcinoma	514
Benign keratosis	1,099
Dermatofibroma	115
Melanoma	1,113
Melanocytic, benign nevus	6,705
Vascular lesions	142

TABLE 2. Comparison of clinical diagnostics performance for three network approaches.

Performance		Network approach		
		PS	PS + AUG	ALL
Accuracy	*min*	0.957845	0.957845	0.967213
	mean	0.964871	0.964871	**0.973302**
	max	0.967213	0.969555	0.978923

tumour (*Positives*) and 1,163 cases of malignant skin tumours (*Negatives*) are selected, which means that the total of 2,134 skin tumour images are used in this research.

4.2 Experimental results

The data set is divided in two classes, 80% for training and 20% for testing classifiers. Therefore, there are 1,711 training instances and 428 testing instances. Five trainings and evaluations have been performed, which are declared in advance to be a representative sample of all three approaches described in the sub-section 3.4: (i) modification of only the last layer without augmentation, (ii) modification of the last layer with augmentation and (iii) modification of fully connected layers. Training is conducted using *mini-batch* processing in 50 epochs. The proposed technique performance evaluated in term of accuracy is defined in the following way: *Accuracy*—the probability that the diagnostic test is performed correctly:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN),$$

where, TP (*True Positives*)—Correctly classified positive cases; TN (*True Negatives*)—Correctly classified negative cases; FP (*False Positives*)—Incorrectly classified negative cases; FN (*False Negatives*)—Incorrectly classified positive cases.

The accuracy of different system approaches based on DL approaches is presented in Table 2. There, one can observe the minimum, maximum and mean value of the accuracy. Columns which are the descriptions of the modifications of the network have the following meanings: PS—modification of only the last layer without augmentation; PS + AUG—modification of the last layer with augmentation; and ALL– modification of fully connected layers.

It can be seen that the experimental results demonstrate that the measured performance for every three approaches is very similar. The difference between the best and the worst individual accuracy is 2.1%, and when the average accuracy is compared, the difference is only 0.84%. The difference in the value of mean accuracy in approaches differs according to whether or not the augmentation data

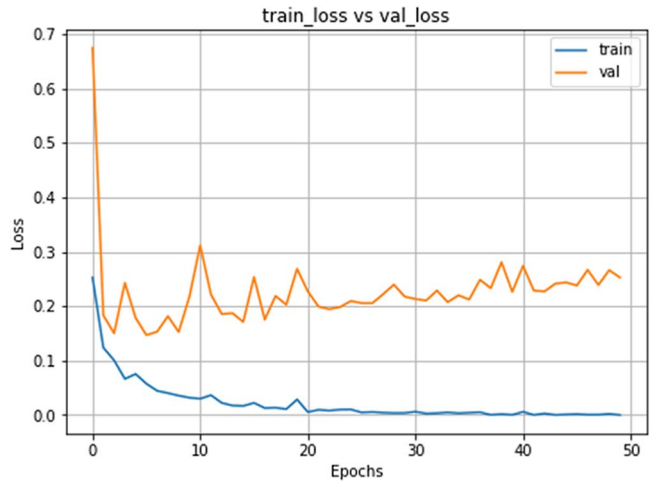

FIGURE 5. Function '*Loss*' per epochs for PS approach.

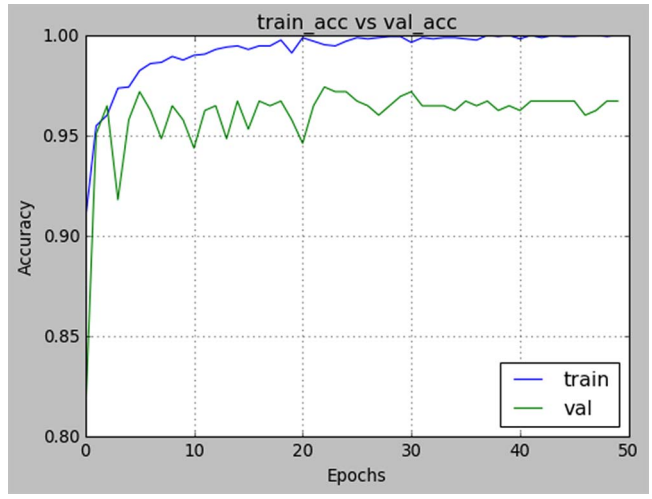

FIGURE 6. Accuracy per epochs for PS approach.

set was carried, although the approach the last layer with augmentation achieved better maximum result. Improving the accuracy is not much better, but it should be noted, in a situation where the training takes place with all fully connected layers.

As already mentioned, the training phase could not be performed in less periods, for example 30 epochs, because the accuracy stabilizes relatively quickly and further training can actually cause overfitting. This can be seen in Figure 5, with a graph of change in the value of the *Loss* function per epochs for training and test data set, and in Figure 6, which presents a graph of accuracy of training and test data set per epoch. Both figures are created after a training approach that modifies only the last layer of the network.

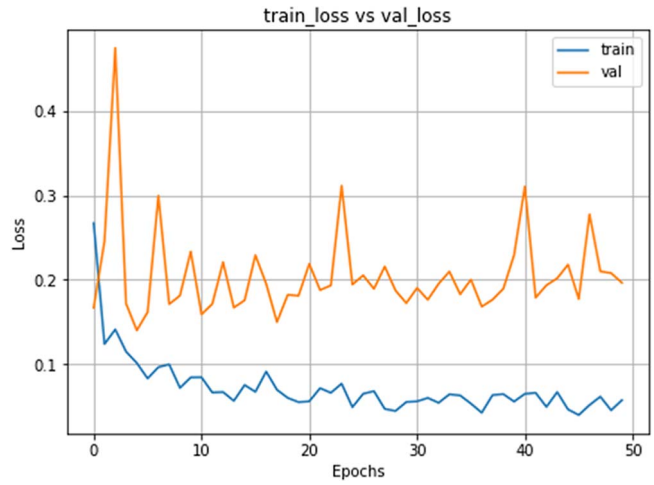

FIGURE 7. Function '*Loss*' per epochs for PS + AUG approach.

In the graphs, it is evident that in both cases, globally, after 20 epochs, the error (*Loss*) starts rising slowly, indicating a slight overfitting, and this specifies the accuracy of the second graph that is very close to 100%, although it does not have great influences on the accuracy of the test data set.

In Figures 7 and 8, same types of graphics and images as in Figures 5 and 6 are presented; however, these graphs are generated after the utilization of the training model for an approach modification of the last layer with augmentation of the input data set. In Figure 7, the function '*Loss*' per epoch for the modification of the last layer with augmentation of the input data set is presented, while Figure 8 depicts the accuracy per epoch for the same approach. In these graphs (Figures 6 and 8), it can be observed that the accuracy of the training set does not reach 100% and curves on the graphs have larger fluctuations. This is a result which, at the input data set, always arises images that have been modified in some way: the rotation is up to 20 degrees, the zoom is up to 15%, change of the width and height is up to 20%, shear is up to 15 degrees and finally there is a horizontal flip; thus, the new model can never learn all the information 'by heart'.

4.3 Discussion

These experimental results could be compared with some other skin tumour image analyses. In research presented in [21], the DCNN Inception-v3 has been trained to ImageNet image database. The DCNN is additionally fine-tuned by the dataset of 129,450 of the skin lesions clinical images consisting of up to 2,032 different diseases. Finally, the performance on a binary classification of benign and malignant changes was measured and it achieved the accuracy of 96%.

The research presented in the paper [22] used DL to classify clinical images of skin lesions from four datasets: MED-NODE, Asan, Hallym and Edinburgh datasets. It depended on the data set used for validation; the comparison of different melanoma from other diseases was correct in between 83% and 96% of cases. The study presented in the research paper [24], where 35,000 cancer patients were tested, had the accuracy value 95%.

The study [40] depicted different DCNN architectures: DenseNet 201, ResNet 152, Inception v3 and InceptionResNet v2. These were used and applied on 10,135 dermoscopy skin images in

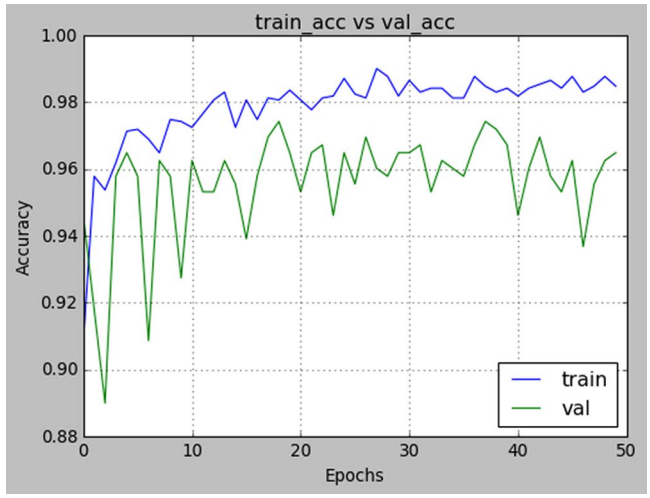

FIGURE 8. *Accuracy* per epochs for PS + AUG approach.

total: *HAM10000* data set which contains 10,015 instances and *PH2* data set which consists of 120 instances. The best accuracy values for comparing melanoma and basal cell carcinoma are 94.40% for ResNet 152 architecture and 99.30% for DenseNet 201 architecture, versus 82.26% and 88.82% by dermatologists, respectively. Overall, the mean accuracy results display that all DL models outperformed dermatologists by at least 11%.

5 Conclusion and future work

The aim of this paper is the automatic classification of skin tumours on malignant tumours and benign lesions for skin tumour dermoscopy images. Only malignant melanomas are analysed, because melanomas are the most malignant skin tumours. The automatic classification system is based generally on deep architecture and partially on the use of DCNN with the use of TL techniques.

The system is completely implemented in Python programming language and its fully open-source libraries. Publicly available implementation of DCNN architecture VGG-16 network is used. The system is tested on 2,143 selected tumour images from repository *HAM10000* data set, and 10,760 modified tumour images, with the applied *augmentation* techniques, which totals 12,903 tumour images. Preliminary experimental results encourage further research by the authors because the proposed strategy on experimental data set has *Accuracy* between 96.5% and 97.5%.

Our research presents the automatic classification system for skin tumour images based on DCNN and three different VGG-16 architectures: (i) the last layer without augmentation, (ii) last layer with augmentation on the input data set and (iii) modification of fully connected layers. In is important to notice that in the first and third DCNN architectures, the number of input skin images is 2,134 from repository *HAM10000* data set, and on the other hand, in the second architecture, the number of input skin images is 10,670, which is the consequence of five different augmentation techniques applied on the input data set. The mean accuracy for three different implemented DCNN architectures are 96.48%, 96.49% and 97.33%, respectively. The experimental results could be compared with other researches; when different DCNN and different data sets are used, experimental results are very

close. All presented researches have mean accuracy between 96% and 99%, which is very close to the results in our research.

Nevertheless, improvements are possible in many aspects. Our future research will focus on the following: (i) precise analysis values of the parameters and their effect on accuracy; (ii) testing complex CNN architecture such as *ResNet*; (iii) broadening the data set with new classes which exist in the meta data set; and(iv) creating new DL model combined with novel artificial intelligence techniques which will efficiently solve real-world skin tumours data sets.

Acknowledgment

This research has been funded by the Spanish Ministry of Science and Innovation under project MINECO-TIN2017-84804-R and by the FCGRUPIN-IDI/2018/000226 project from the Asturias Regional Government and has been supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, through the project no. 451-03-68/2020-14/200156: 'Innovative scientific and artistic research from the FTS (activity) domain'.

References

[1] S. Rangwala and K. Y. Tsai. Roles of the immune system in skin cancer. *British Journal of Dermatology*, **165**, 953–965, 2011. doi: 10.1111/j.1365-2133.2011.10507.x.

[2] N. Rajaram, J. S. Reichenberg, M. R. Migden, T. H. Nguyen and J. W. Tunnell. Pilot clinical study for quantitative spectral diagnosis of non-melanoma skin cancer. *Lasers in Surgery and Medicine*, **42**, 716–727, 2010. doi: 10.1002/lsm.21009.

[3] A. Jemal, S. S. Devesa, T. R. Fears and P. Hartge. Cancer surveillance series: changing patterns of cutaneous malignant melanoma mortality rates among whites in the United States. *Journal of the National Cancer Institute*, **92**, 811–818, 2000. doi: 10.1093/jnci/92.10.811.

[4] deE. Vries, F. I. Bray, J. W. Coebergh and D. M. Parkin. Changing epidemiology of malignant cutaneous melanoma in Europe 1953–1997: rising trends in incidence and mortality but recent stabilizations in western Europe and decreases in Scandinavia. *International Journal of Cancer*, **107**, 119–126, 2003. doi: 10.1002/ijc.11360.

[5] de and European Network of Cancer RegistriesE. Vries, F. I. Bray, A. M. Eggermont and J. W. Coebergh. Monitoring stage-specific trends in melanoma incidence across Europe reveals the need for more complete information on diagnostic characteristics. *European Journal of Cancer Prevention*, **13**, 387–395, 2004. doi: 10.1097/00008469-200410000-00006.

[6] M. Elgamal. Automatic skin cancer images classification. *International Journal of Advanced Computer Science and Applications*, **4**, 287–294, 2013. doi: 10.14569/IJACSA.2013.040342.

[7] S. M. Swetter, H. Tsao, C. K. Bichakjian, C. Curiel-Lewandrowski, D. E. Elder, J. E. Gershenwald, V. Guild, J. M. Grant-Kels, A. C. Halpern, T. M. Johnson, A. J. Sober, J. A. Thompson, O. J. Wisco, S. Wyatt, S. Hu and T. Lamina. Guidelines of care for the management of primary cutaneous melanoma. *Journal of the American Academy of Dermatology*, **80**, 208–250, 2019. doi: 10.1016/j.jaad.2018.08.055.

[8] P. Baade and M. Coory. Trends in melanoma mortality in Australia: 1950–2002 and their implications for melanoma control. *Australian and New Zealand Journal for Public Health*, **29**, 383–386, 2005. doi: 10.1111/j.1467-842X.2005.tb00211.x.

[9] M. Messadi, A. Bessaid and A. Taleb-Ahmed. Extraction of specific parameters for skin tumour classification. *Journal of Medical Engineering and Technology*, **33**, 288–295, 2009. doi: 10.1080/03091900802451315.

[10] W. Abbes and D. Sellami. High-level features for automatic skin lesions neural network based classification. IEEE IPAS'16: International Image Processing, Application and Systems Conference, Hammamet, Tunisia, 2016. doi: 10.1109/IPAS.2016.7880148.

[11] S. Simić, S. D. Simić, Z. Banković, M. Ivkov-Simić, J. R. Villar and D. Simić. A hybrid automatic classification model for skin tumour images. In *Hybrid Artificial Intelligent Systems*, vol. 11734, pp. 722–733. Springer, Berlin, Heidelberg, 2019.

[12] B. Krawczyk, D. Simić, S. Simić and M. Woźniak. Automatic diagnosis of primary headaches by machine learning methods. *Open Medicine*, **8**, 157–165, 2013. doi: 10.2478/s11536-012-0098-5.

[13] S. Simić, Z. Banković, D. Simić and S. D. Simić. A hybrid clustering approach for diagnosing medical diseases. In *Hybrid Artificial Intelligent System*, vol. 10870, pp. 741–775. Springer, Berlin, Heidelberg, 2018.

[14] S. Simić, D. Simić, P. Slankamenac and M. Simić-Ivkov. Computer-assisted diagnosis of primary headaches Hybrid Artificial Intelligence Systems. In *Third International Workshop, HAIS 2008*, Burgos, Spain, September 24–26. vol 5271, pp. 314–321, Springer, Berlin, 2008. doi: 10.1007/978-3-540-87656-4_39.

[15] S. Simić, D. Simić, P. Slankamenac and M. Simić-Ivkov. *Rule-based fuzzy logic system for diagnosing migraine Artificial Intelligence: Theories, Models and Applications*, Vol. 5138 of Lecture Notes in Computer Science, pp. 383–388, Springer, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-87881-0_37.

[16] D. G. Kumar. *Machine Learning Techniques for Improved Business Analytics*. Business Science Reference, IGI Global, Hershey, Pennsylvania, United States, 2018.

[17] C. C. Aggarwal. *Data Classification: Algorithms and Applications*. Chapman and Hall/CRC - Taylor & Francis Group, Abingdon-on-Thames, United Kingdom, 2014.

[18] M. Wozniak. *Hybrid Classifiers: Methods of Data, Knowledge, and Classifier Combination*. Springer, Berlin Heidelberg, 2016.

[19] A. A. Cruz-Roa, J. E. Arevalo Ovalle, A. Madabhushi and F. A. González Osorio. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. *Medical Image Computing and Computer-Assisted Intervention*, **16**, 403–410, 2013. doi: 10.1007/978-3-642-40763-5_50.

[20] A. Krizhevsky, I. Sutskever and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, **60**, 84–90, 2017. doi: 10.1145/3065386.

[21] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, **542**, 115–118, 2017. doi: 10.1038/nature21056.

[22] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park and S. E. Chang. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, **138**, 1529–1538, 2018. doi: 10.1016/j.jid.2018.01.028.

[23] and Cochrane Skin Cancer Diagnostic Test Accuracy GroupN. Chuchu, Y. Takwoingi, J. Dinnes, R. N. Matin, O. Bassett, J. F. Moreau, S. E. Bayliss, C. Davenport, K. Godfrey, S. O'Connell, A. Jain, F. M. Walter, J. J. Deeks and H. C. Williams. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. *Cochrane Database of Systematic Reviews*, **12**, CD013192, 2018. doi: 10.1002/14651858.CD013192.

[24] A. Capritto. https://www.cnet.com/health/how-to-use-your-smartphone-to-detect-skin-cancer/ (accessed 1 October 2019).

[25] Kraanspoor. https://play.google.com/store/apps/details?id=com.rubytribe.skinvision.ac (accessed 1 October 2019).

[26] A. Bartlett. https://apps.apple.com/eg/app/skinvision-detect-skincancer/id545293136 (accessed 1 October 2019).
[27] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou and D. Ioannides. Epidemiological trends in skin cancer. *Dermatology Practical & Conceptual*, **7**, 1–6, 2017. doi: 10.5826/dpc.0702a01.
[28] MelNet. Skin cancer primary prevention and early detection steering committee New Zealand skin cancer primary prevention and early detection strategy, 2014 to 2017. *MelNet*, 2014. ISBN: 978-0-478-44901-3 (Online).
[29] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**, 2278–2324, 1998. doi: 10.1109/5.726791.
[30] S. O. Haykin. *Neural Networks and Learning Machines*, 3rd edn. Pearson, London, United Kingdom, 2008.
[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA. 1–9, 2015. doi: 10.1109/CVPR.2015.7298594.
[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, **15**, 1929–1958, 2014.
[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and F.-F. Li. ImageNet: a large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.
[34] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. Proceedings of 3rd International Conference on Learning Representations, ICLR, arXiv:1409.1556, San Diego, CA, USA, 2015.
[35] P. Raybaut. https://www.spyder-ide.org/ (accessed 1 October 2019).
[36] Project Jupyter. https://jupyter-notebook.readthedocs.io/en/stable/ (accessed 1 October 2019).
[37] M. Loaiciga and A. Shah. https://github.com/anujshah1003/Transfer-Learning-in-keras---customdata/blob/master/vgg16.py (accessed 1 October 2019).
[38] G. Hinton, N. Srivastava and K. Swersky. http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf (accessed 1 October 2019).
[39] P. Tschandl. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T (accessed 1 October 2019).
[40] A. Rezvantalab, H. Safigholi and S. Karimijeshni. Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. CoRR, Cornell University, Ithaca, New York, US. arXiv:1810.10348, 2018.

Received 10 February 2020