Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/13681
DOI URL: http://dx.doi.org/10.21474/IJAR01/13681

RESEARCH ARTICLE

SKIN LESION CLASSIFICATION FROM DERMOSCOPY AND CLINICAL IMAGES WITH A DEEP LEARNING APPROACH

Aigli Korfiati[1], Giorgos Livanos[1], Christos Konstandinou[1], Sophia Georgiou[2] and George Sakellaropoulos[1]

1. Department of Medical Physics, School of Medicine, University of Patras, Greece.
2. Department of Dermatology, School of Medicine, University of Patras, Greece.

Manuscript Info

Manuscript History
Received: 31 August 2021
Final Accepted: 30 September 2021
Published: October 2021

Key words:-
Skin Lesion Classification, Deep Learning, Integrated Data, Dermoscopy, Imaging, Melanoma

Abstract

Computer-aided diagnosis (CAD) systems based on deep learning approaches are now feasible due to the availability of big data and the availability of powerful computational resources.The medical image-based CAD systems are of great interest in numerous diseases, but especially for skin cancer diagnosis, deep learning models have been mostly developed for dermoscopy images. Models for clinical images are few, mainly due to the unavailability of big volumes of relevant data. However, CAD systems able to classify skin lesions from clinical images would be of great valueboth for the population and clinicians as an initial early screening of lesions that would leadpatients to visiting a dermatologist in case of suspicious lesions. This is even more pronounced in areas where there is lack of dermoscopy instruments. Thus, in this paper, we aimed to build a classifier based on bothdermoscopy and clinical images able to discriminate skin cancer from skin lesions. The classification is made among three benign and two malignant categories, which include Nevus, Benign but not nevus, Benign but suspicious for malignancy, Melanoma and Non-Melanocytic Carcinoma.The proposed deep learning classifier achieves an Area Under Curve ranging between 0.75 and 0.9 for the five examined categories.

Introduction:-

Skin cancer is increasingly evident worldwide (Goyal et al., 2020) with the most common forms of skin cancer being non-melanocytic, such as Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC). Cutaneous melanoma, on the other side, although being rare, has been associated with high mortality rates (Rebecca et al., 2020). Early detection of skin cancers is of crutial importance in an effort to reduce mortality. Thus, researchers have shown interest in building automated frameworks to aid skin cancer diagnosis, starting from 1985 when the ABCD rule (Ali et al., 2020) was devised. More sophisticated computer aided diagnosis (CAD) systems using dermoscopy or clinical images started subsequently to emerge (Barata et al., 2018). The use ofmachine learning and deep learning algorithms in CAD systems made them even more powerful, but numerous limitations are still present (Goyal et al., 2020) and have to be faced before such systems start to be part of the clinical practice.

Medical image-based CAD systems for skin cancer diagnosis with deep learning approaches have been mostly developed for dermoscopy images. Models for clinical images are few, mainly due to the unavailability of big

Corresponding Author:- Aigli Korfiati
Address:- Department of Medical Physics, School of Medicine, University of Patras, Greece.

volumes of relevant data. This poses limitations to deep learning methods, because they depend on data availability, since they learn from the data in order to be able to perform predictions.However, CAD systems able to classify skin lesions from clinical images would be of great value both for the population and clinicians as an initial early screening of lesions that would lead patients to visiting a dermatologist in case of suspicious lesions. This is even more pronounced in areas where there is lack of dermoscopy instruments.

In the present paper, we build a classifier based on both dermoscopy and clinical images able to discriminate skin cancer from skin lesions. Dermoscopy and clinical images for building and testing the classifier were retrieved from ebioMelDB (Korfiati et al., 2021), a collection of skin cancers multimodal data and PAD-UFES-20 (Pacheco et al., 2020),a recently published dataset which contains clinical images.The classification is made among three benign and two malignant categories, which include Nevus, Benign but not nevus, Benign but suspicious for malignancy, Melanoma and Non-Melanocytic Carcinoma. Performance metrics are measured in a validation set of images. The proposed deep learning classifier achieves an Area Under Curve ranging between 0.75 and 0.9 for the five examined categories in the validation set.Melanoma is predicted with an AUC of 0.78, while non-melanocytic carcinomas with an AUC of 0.9.

Materials And Methods:-

Data

Training and Validation data

Dermoscopy and clinical images for training and validating the classifier were retrieved from ebioMelDB (Korfiati et al., 2021), a collection of skin cancers multimodal data and PAD-UFES-20 (Pacheco et al., 2020), a recently published dataset which contains clinical images. The majority of the dermoscopic data is part of the International Skin Imaging Collaboration (ISIC) Archive, a collection of dermoscopy images which was released during the challenges hosted from ISIC organization from 2016 to 2020. To avoid duplicates, only ISIC2019 (Combalia et al., 2019) and ISIC 2020 Challenge (Rotemberg et al., 2021) training data were included in the analysis. Another source of dermoscopy images is the PH^2 database (Mendonça et al., 2013). The 7-pt dataset(Kawahara et al.,2018) hosts both dermoscopy and clinical images. Another source of clinical images is MED-NODE (Giotis et al., 2015). The aforementioned images were used for training the model, besides a 10% that were kept out for validation purposes.

The total count of images used for training is38647. Among them 35177 are dermoscopy and 2816 are clinical images. For validation 4162 images were used, of which 3631dermoscopy and 531 clinical.

Data classes

Maintaining the categorization of the images suggested in ebioMelDB our data are organized into five categories: NEV (Nevus), MEL (Melanoma), NNV (Benign non-nevus), NMC (Non-melanocytic carcinoma) and SUS (Benign but suspicious for malignancy). The benign categories are NEV, NNV and SUS. The nevus category NEV includes simple, common, blue, clark, combined, con-genital, dermal, recurrent, reed or spitz nevus. The benign non-nevus category NNV includes dermatofibroma, lentigo, melanosis, miscellaneous, seborrheic keratosis, vascular lesion, benign keratosis, cafe-au-lait macule, lentigo NOS and lichenoid keratosis. The benign, but suspicious for malignancy category SUS includes actinic keratosis, atypical melanocytic proliferation and atypical nevus. The malignant melanoma category MEL includes melanoma and melanoma metastases images. Finally, the category of non-melanocytic carcinomas includes basal cell carcinoma and squamous cell carcinoma.

Deep learning model

Model description

Learning good representations, along with a deep enough deep learning architecture, requests the underlying training data to be massive. Training from scratch such a model can be deterrent in certain domains due to the lack of labeled data. In order to minimize the negative impact of data scarcity, we applied a network-based transfer learning (Tan et al.,2018) and fine tuning along with data augmentation. The proposed model for lesion classification into the five classes is presented in Figure 1.

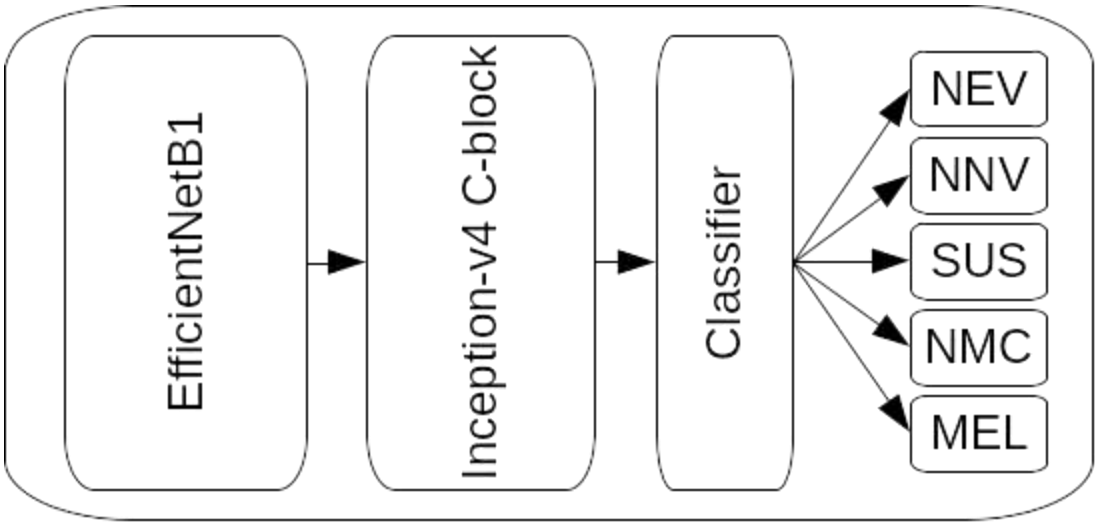

Figure 1:- Proposed model architecturegeneral description. The input images are resized to 224x224 pixels and the intensities are normalized to the range 0-1. The classifier consists of 3 dense layers where the first two layers contain 32 nodes each and the final dense layer has 5 outputs which are the five prediction classes.

During hyper-parameter optimization, we deployed an EfficientNet-B0 (Tan et al.,2019) pre-trained on the ImageNet (Deng et al.,2009)dataset for faster iterations. From this pre-trained model, we removed the classification layer and the feature maps of the last Convolutional layer were passed to a thin Inception-v4(Szegedyet al.,2018) C-block. This block consists of three parallel layers, which enable learning of various scale patterns. The output of this block is flattened through a global average layer (Lin et al.,2014). After optimizing the hyperparameters, we changed the EfficientNet-B0 to an EfficientNet-B1 pretrained on ImageNet dataset which we fine-tuned by adjusting all layers' weights for our task. A final fully connected layer with softmax activation function was used in order to form the final output as class probabilities. We used the Swish (Ramachandran et al, 2017) activation function for all layers. The Inception-v4 C-block is presented in Figure 2.

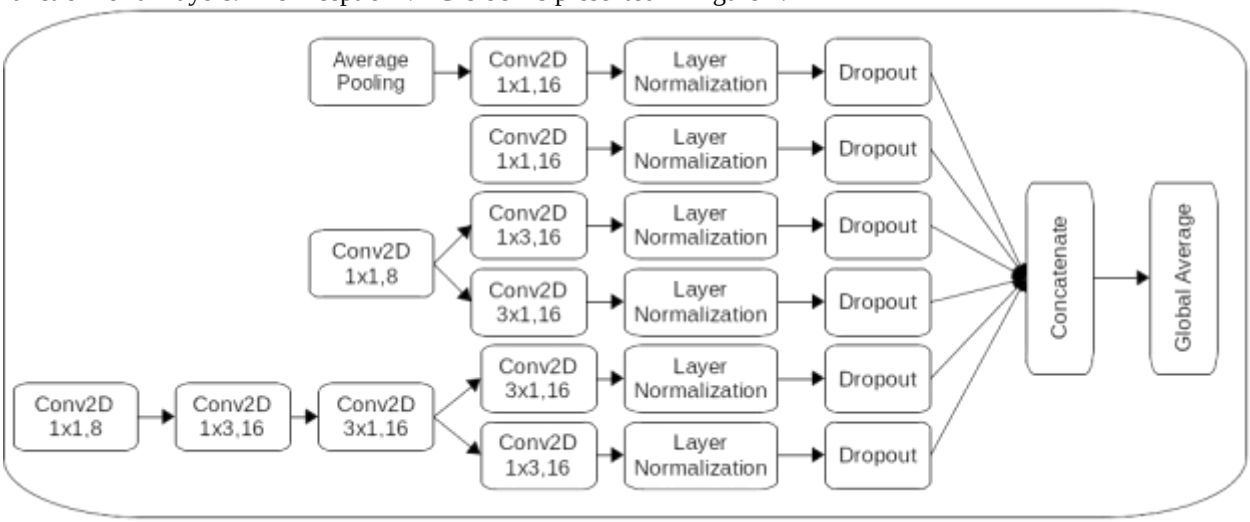

Figure 2: The Inception-v4 C-block architecture as initially described in(Szegedyet al.,2018). In order to avoid overfitting, we implemented a narrower version with 8 and 16 nodes instead.

Training

In order to train the model, we used heavy augmentation techniques on the available images. The underrepresented classes on each task were oversampled considering also the image type of each sample. We used cross entropy objective function and AdaMax(Kingmaet al.,2014) optimization algorithm with initial learning rate $lr = 10^{-5}$, $\beta_1 = 0{,}9$ and $\beta_2 = 0{,}999$. During training, if the loss didn't improve for 5 consecutive epochs, the learning rate decreased by a factor of 0.1 and if the loss value didn't improve for 10 consecutive epochs, we stopped training. We

fine-tuned the best performing model by applying weight updates to the pretrained model with an initial $lr_{fine} = 10^{-6}$ and applying the same early stopping strategy.

Setup

We implemented the proposed method with python 3.8 and Tensorflow framework(Abadi et al.,2016). The data augmentation was implemented with tensorflowkeras image preprocessing and tensorflow-addons image preprocessing modules. For the input pipeline we used the tensorflow dataset API. This work was supported by computational time granted from the National Infrastructures for Research and Technology S.A. (GRNET S.A.) in the National HPC facility – ARIS -under project ID pa210303. The development environment was on Redhat/Centos 6.7 OS and the provided hardware of each GPU node used was 2x Haswell - Intel(R) Xeon(R) E5-2660v3 CPU, 64 GByte RAM and 2x NVIDIA Tesla K40 GPU.

Results:-

In order to evaluate the performance of the proposed model we run it with the following parameters.

	Learning Rate	Batch Size	Optimizer	Pre-trained model	Image Size	Activation function	Dropout	Loss Function
Values	0.00001	16	Adamax	EfficientNetB1	224	Swish	0.25	Cross Entropy

Table 1:- Selected parameters for the deep learning model.

In the validation set we computed the precision, recall and F1 score metrics as shown below and the receiver operating characteristic - ROC area under the curve - AUC.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

The results for these metrics are presented in Table 2.

Validation	Precision	Recall	F1 score	AUC
NEV	0.890	0.605	0.72	0.876
NNV	0.249	0.449	0.32	0.753
SUS	0.326	0.544	0.408	0.862
NMC	0.557	0.567	0.562	0.9
MEL	0.371	0.550	0.443	0.78

Table 2:- Performance metrics: Precision, Recall, F1 score and receiver operating characteristic area under the curve - AUC calculated in the validation set.

The respective ROC curves for each class are also presented in Figure 3.

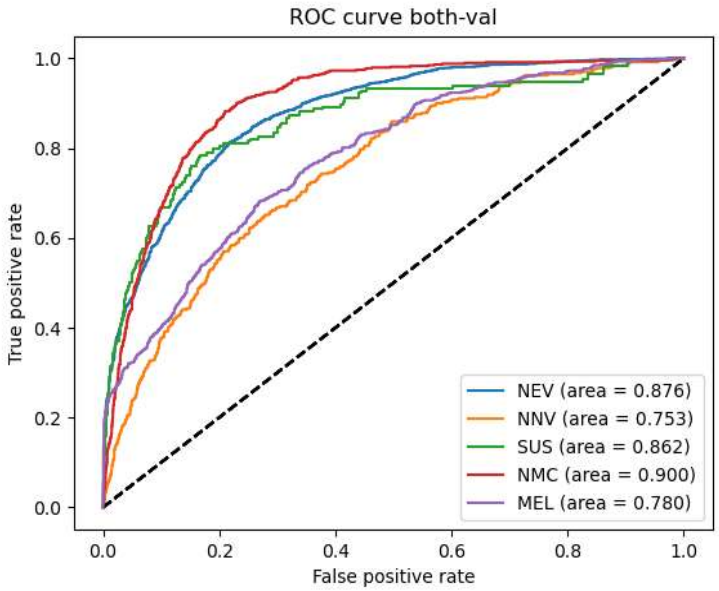

Figure 3:- ROC curves for each of the five classes: NEV - Nevus, MEL - Melanoma, NNV - Benign non-nevus, NMC - Non-melanocytic carcinoma and SUS - Benign but suspicious for malignancy.

In the following matrix, we can observe how images from the five actual different classes were classified in the model's predictions.

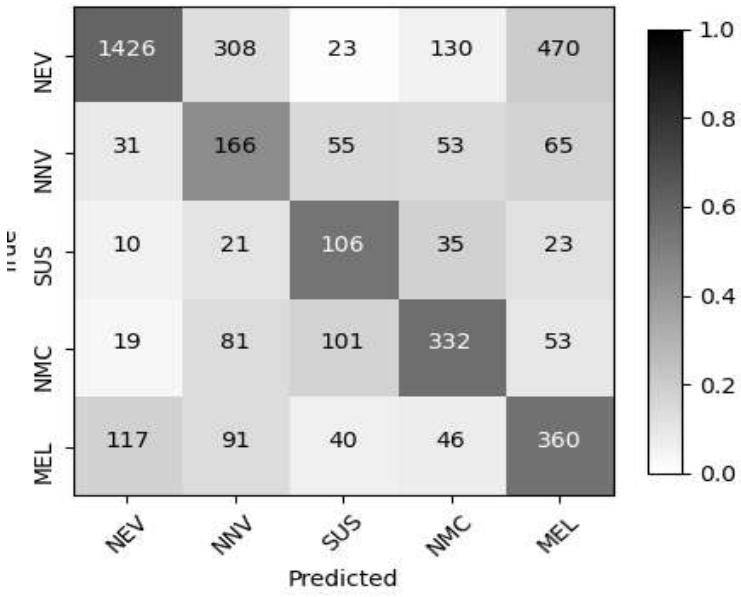

Figure 4:- Confusion matrix of true versus predicted classes:NEV - Nevus, MEL - Melanoma, NNV - Benign non-nevus, NMC - Non-melanocytic carcinoma and SUS - Benign but suspicious for malignancy.

Discussion:-

The proposed multiclass classification scheme gives quite encouraging results in the classification of skin lesions and the detection of skin cancers. The benign class nevus is predicted correctly with a precision of 0.89 and a recall of 0.61. The second-best classification performance is achieved for the malignant class non melanocytic carcinoma followed by melanoma as the third best predicted class. This can be attributed to two main reasons. The remaining two benign categories include benign lesions with a) high diversity and at the same time b) few images for each lesion type. Thus, having not enough data for some classes leads to the class imbalance problem which does not allow deep learning algorithms to perform well.

Additionally, from the results it can be observed that benign lesions of one of the three categories are misclassified as one of the other benign lesions categories. As a future direction, we plan to treat the problem as a binary problem trying to discriminate skin cancers from benign lesions. Another interesting observation is that melanoma is mostly misclassified as nevus and vice versa. Again, we also plan to create a model able to distinguish melanoma from nevus. Another interesting future direction would be the integration of patient clinical data together with the clinical and dermoscopy images.

Improvement of the classification performance could also be achieved with improvements in the deep learning model. Dealing with the class imbalance problem, experimenting with different data augmentation techniques and preprocessing of the images in order to remove hair and other marks are potential paths for a better classification model.

Conclusion:-

Computer-aided diagnosis systems for skin cancer have been of great interest for a number of years, but only recently with the availability of big volumes of data and of powerful computational resources, deep learning approaches were enabled to emerge in skin cancer related studies. Such approaches have been mostly developed for dermoscopy images because the unavailability of big volumes of clinical images does not ease the development of clinical images based CAD systems. However, these would be of great value both for the population and clinicians as an initial early screening of lesions alarming patients to visiting a dermatologist in case of suspicious lesions. This is even more pronounced in areas where there is lack of dermoscopy instruments.

Herein, we described a deep learning model based on both dermoscopy and clinical images able to discriminate skin cancer from skin lesions. The classification is made among three benign and two malignant categories, which include Nevus, Benign but not nevus, Benign but suspicious for malignancy, Melanoma and Non-Melanocytic Carcinoma. The proposed deep learning classifier achieved encouraging results giving at the same time numerous future directions for improvement.

Acknowledgement:-

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning 2014-2020» in the context of the project “Biomarkers extraction from imaging and molecular biology data using computational models to assist malignant melanoma diagnosis, prognosis and treatment” (MIS 5047174).

References:-

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.
2. Ali AR, Li J, Yang G. Automating the ABCD rule for melanoma detection: a survey. IEEE Access. 2020 Apr 28;8:83333-46.
3. Barata C, Celebi ME, Marques JS. A survey of feature extraction in dermoscopy image analysis of skin cancer. IEEE journal of biomedical and health informatics. 2018 Jun 11;23(3):1096-109.
4. Combalia M, Codella NC, Rotemberg V, Helba B, Vilaplana V, Reiter O, Carrera C, Barreiro A, Halpern AC, Puig S, Malvehy J. BCN20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288. 2019 Aug 6.
5. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.
6. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N. MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. Expert systems with applications. 2015 Nov 1;42(19):6578-85.
7. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification for diagnosis of skin cancer: Challenges and opportunities. Computers in Biology and Medicine. 2020 Oct 27:104065.

8. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE journal of biomedical and health informatics. 2018 Apr 9;23(2):538-46.
9. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
10. Korfiati A, Livanos G, Konstantinou C, Georgiou S, Sakellaropoulos G. ebioMelDB: Multi-modal Database for Melanoma and Its Application on Estimating Patient Prognosis. InIFIP International Conference on Artificial Intelligence Applications and Innovations 2021 Jun 25 (pp. 33-44). Springer, Cham.
11. Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXiv preprint arXiv:1312.4400.
12. Mendonça T, Ferreira PM, Marques JS, Marcal AR, Rozeira J. PH 2-A dermoscopic image database for research and benchmarking. In2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC) 2013 Jul 3 (pp. 5437-5440). IEEE.
13. Pacheco AG, Lima GR, Salomão AS, Krohling B, Biral IP, de Angelo GG, Alves Jr FC, Esgario JG, Simora AC, Castro PB, Rodrigues FB. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in brief. 2020 Oct 1;32:106221.
14. Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.
15. Rebecca VW, Somasundaram R, Herlyn M. Pre-clinical modeling of cutaneous melanoma. Nature communications. 2020 Jun 5;11(1):1-9.
16. Rotemberg V, Kurtansky N, Betz-Stablein B, Caffery L, Chousakos E, Codella N, Combalia M, Dusza S, Guitera P, Gutman D, Halpern A. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific data. 2021 Jan 28;8(1):1-8.
17. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.
18. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018, October). A survey on deep transfer learning. In International conference on artificial neural networks (pp. 270-279). Springer, Cham.
19. Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105-6114). PMLR.