Melanoma Classification Using Deep Learning

dept. HealthCare Engineering and mangment (to Dr.Inas)
Cairo University
By: Ahmed Moataz 1180236
Ahmed Mahmoud Hussein 1180364
Yumna Hamdy 3190001

***Abstract*— Skin cancer is common cancer type and despite beign mostly non malignant, due to high case numbers it's pretty serious diasease and can lead serious cases if not detected, treated in time. It's usually diagnosed by eye for primarily and followed by further clinical analysis if needed. Even though the rares outcome is called melanoma it's the deadliest one, so early detection is pretty important. For this task using computer aided diagnosis might be helpful for primarily steps and early detections. Better detection might save thousands of lives.**

Keywords*— CNN, EfficientNet, VGG, Skin lesion classification, Deep learning, Melanoma

INTRODUCTION

Skin cancer is the most prevalent type of cancer. Melanoma, specifically, is responsible for 75% of skin cancer deaths, despite being the least common skin cancer. The American Cancer Society estimates over 100,000 new melanoma cases will be diagnosed in 2020. It's also expected that almost 7,000 people will die from the disease. As with other cancers, early and accurate detection—potentially aided by data science—can make treatment more effective.

Currently, dermatologists evaluate every one of a patient's moles to identify outlier lesions or “ugly ducklings” that are most likely to be melanoma. Existing AI approaches have not adequately considered this clinical frame of reference. Dermatologists could enhance their diagnostic accuracy if detection algorithms consider “contextual” images within the same patient to determine which images represent a melanoma. If successful, classifiers would be more accurate and could better support dermatological clinic work

As the leading healthcare organization for informatics in medical imaging, the Society for Imaging Informatics in Medicine (SIIM)'s mission is to advance medical imaging informatics through education, research, and innovation in a multi-disciplinary community. SIIM is joined by the International Skin Imaging Collaboration (ISIC), an international effort to improve melanoma diagnosis. The ISIC Archive contains the largest publicly available collection of quality-controlled dermoscopic images of skin lesions

The system get image of skin and meta data as an input and return binary classification 0 for benign & 1 for malignant.

RELATED WORKS

Skin cancer has been widely studied by the AI research community. The methods for automated identification and analysis of images can be categorized as Traditional methods (conventional machine learning models) and convolutional neural networks. Related works in each category are described below.

A. Traditional method

Traditional machine learning was performed by Ballerini [13] using a hierarchical classification system based on the K-Nearest Neighbors (K-NN) model used color and texture features extracted from skin lesion images. The method archived an overall classification accuracy of 74 % over five common classes of skin lesions, including two non-melanoma cancer types. Barata C. et. al. [14] examining the role played by color features only, by texture features only, and by combining both of them in the final classification. The research concluded that over a dataset of 176 dermoscopy images from Hospital Pedro Hispano, Matosinhos, color features outperform texture features when used alone. Dreiseitl et al [15] compared a number of traditional machine learning approaches on the task of classifying pigmented skin lesions to conclude that conventional ANNs and SVMs performed on about the same level, with k-nearest neighbors and decision trees performing worse. Classification accuracies were improved with an ensemble of four classifiers namely, support vector machine, random forest, logistic model tree, and hidden naive Bayes applied on a set of 289 dermoscopy images (114 malignant, 175 benign) [16] The method achieved an accuracy of 91.26% and area under the curve value of 0.937 when 23 features were used.

Kawahara et al [17] combined traditional with convolutional approaches by training linear classification models with features extracted from convolutional neural networks. The method achieved an accuracy of 81.8% over the entire 10-class dataset of 1300 images captured from a standard (non-dermoscopic) camera. The features used for traditional methods require segmenting the lesions. A systematic overview of recent border detection methods is shown by Celebi ME et al [18] indicates the feasibility of the approach and the problems faced while applying the discussed methods. Iyotomi et al [19] discuss web services designed using a highly accurate dermatologist-like tumor area extraction algorithm. The system achieved a sensitivity of 85.9% and a specificity of 86.0% on a set of 1258 dermoscopy images. Celebi et al [20] describe a segmentation method to segregate the lesion from the background skin. Using color and texture related features, the image is divided into various clinically significant regions using the Euclidean distance transform and finally, optimal features are selected using an optimization framework. The method achieved a specificity of 92.34% and a sensitivity of 93.33% on a set of 564 images

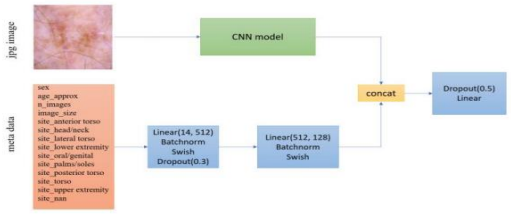
B. Convolution neural network method

Early convolutional networks including Lopez et al VGG model [21] and Simonyan [22] using RMSProp optimizer trained with 3 different training methods and comparison between the proposed methods concluded that the fine-tuning

method worked the best. The models applied on datasets from the ISIC archive [23] achieved 78.66% sensitivity. In other research by Milton et al [24] PNASNet5-Large, InceptionResNetV2, SENet154, InceptionV4 models trained on dermoscopic images post preprocessing and augmentation over the 2018 ISIC challenge dataset [23] were compared. The research concluded that the PNASNet-5-Large model performed better than other models scoring 0.76 on the dataset. Liao [25] investigated the feasibility of a universal skin disease diagnosis system using deep convolutional neural networks (CNN) by further back-propagating. The system achieved 73.1% Top-1accuracy and 91.0% Top-5 accuracy when testing on the Dermnet dataset. On the OLE dataset, the system achieved Top-1 and Top-5 accuracies as 31.1% and 69.5% respectively. Codella et al [26] studied segmentation and classification approaches in ensembles to show these performed better than human graders in terms of accuracy and specificity with similar sensitivity using the dataset of ISIC 2016 (ISBI 016). El-Khatib et al [27] suggests a global fusion-based decision system that uses the results obtained by three different methods to establish the fusion weights. Method 1 used a neural network for classification. Method 2 used fine-tuned CNN and method 3 used SVM. The fusion method achieved an accuracy of 95% on the PH2 database and on the ISIC 2019 database accuracy of 93%. Research by Baghersalimi et al [28] proposed DermoNet, which can reuse information from preceding layers to ensure high accuracy in later layers using densely connected convolutional blocks and skip connections. similar to Densenet [29]. The method was evaluated on the ISBI 2016, ISBI 2017, and the PH2 dataset, and in runtime performance of DermoNet with two other related architectures, that are fully convolutional networks and U-Net, Dermonet turned out to be faster and well suitable for practical application.

Li et al [30] proposed methods to tackle all three tasks of ISIC 2017 i.e. lesion segmentation (task 1), lesion dermoscopic feature extraction, and lesion classification. The researchers proposed a deep learning framework consisting of two fully-convolutional residual networks (FCRN) to simultaneously produce the segmentation result and the coarse classification result. The classifier is further refined using a lesion index calculation unit (LICU) and a straightforward CNN is proposed for the dermoscopic feature extraction task. The method achieved 0.718 for segmentation, 0.833 for feature extraction, and 0.823 for lesion classification. In research by Unver et al [31], a pipeline for skin lesion segmentation in dermoscopic images combining a deep convolutional neural network named as You Only Look Once (YOLO)[32] and the Grab Cut algorithm is explained. These methods achieved a 90% sensitivity rate on the ISBI 2017 dataset. There has been successful attempt at classifying skin lesions from HAM10000 dataset using a simple CNN model with modified Adam optimizer that gave 78% accuracy [33].

The standard approach is to take a deep CNN model (such as the most popular EfficientNet) trained on ImageNet, replace the last layer so that the output dimension equals the target's dimension, and fine tune it on the specific dataset.
That was a competition, but their target was pretty different that ours. They aimed to classify the data into benign-malignant (i.e., no melanoma) or malignant (i.e., melanoma). They noticed that the target information is included in the diagnosis column: target is malignant if and only if diagnosis is melanoma. But diagnosis column is more granular when an image is benign. They believed using diagnosis as target to train the model could give the model more information.
They added to CNN model is the addition of 14 metadata features in some models: sex, age approx, 10 one-hot encoded anatom site general challenge features, image size in bytes and n images, where n_images is the number of all images of that patient in the data. The metadata go through two fully connected layers before being concatenated with the CNN features. In three of four metadata models, the two hidden layers have dimensions 512 and 128. The addition of metadata models in the ensemble provides good model diversity.[4]



1) Strengths

1. CNNs are very accurate architectures in the image detection field.
2. The winning architecture used diagnosis as target and adding metadata in some models.[4]
3. The addition of metadata models in the ensemble provides good model diversity.[4]
4. No manual feature extraction is required in deep learning algorithms.

2) Weaknesses

1. The integration of attention mechanisms increases the number of parameters of the deep learning models, thus increasing their complexity.[12]
2. CNNs are time consuming and needs high computational power.

DATASET AND FEATURES:

A. Dataset Description:

We will use a dataset created by SIIM & ISIC datasets available on this link https://www.kaggle.com/competitions/siim-isic-melanoma-classification/overview The dataset was generated by the International Skin Imaging Collaboration (ISIC) and images are from the following sources: Hospital Clínic de Barcelona, Medical University of Vienna, Memorial Sloan Kettering Cancer Center, Melanoma Institute Australia, The University of Queensland, and the University of Athens Medical School. The Train set has 33,126 rows while the Test set has 10,982 rows. The images are provided in DICOM format. This can be accessed using commonly available libraries like pydicom and contains both image and metadata. It is a commonly used medical imaging data format. Images are also provided in JPEG and TFRecord format (in the jpeg and tfrecords directories, respectively). Images in TFRecord format have been resized to a uniform 1024x1024. Metadata is also provided outside of the

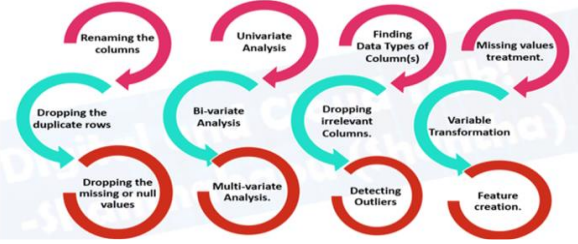
DICOM format, in CSV files which include the following data as in Table I below.

B. The flow of melanoma classification

includes **Data preparation** (the preprocessing techniques also include methods such as image size adjustment, normalization, a contrast enhancement and intensity adjustment, space correction, binarization, morphological operations, gray-scaling, and noise reduction. At this stage, noise and other artifacts are removed from images. **Model structure** (which involves defining data input and dimensions, as well as network core modules, classifiers, and loss function and network output); **Training the model** (which involves choosing backbone, defining parameters, and constructing and performing training); and **testing and applying the model**. We can also roughly divide the process into four parts: Input, network, training, and output. When we try to improve the effect of model training, we can optimize these four aspects. In the past few years, there has been an increasing tendency, not only to develop and use different modern CNN backbones to solve complex real-world problems, but also to apply advanced techniques for achieving better training of these models like transfer learning techniques, data augmentation methods, and the development of ensembles of CNNs.

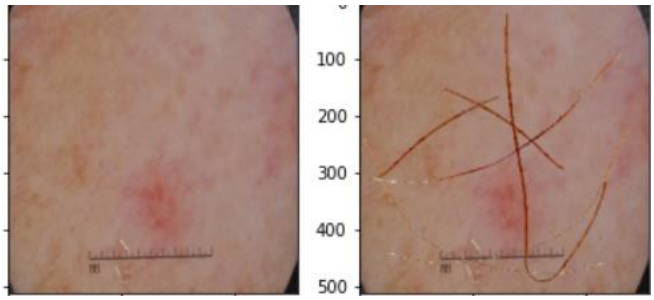
C. Preprocessing & augmentation:

The dataset images have different sizes; in order to be used in neural networks, they must have the same dimensions. Therefore, resize dataset images to the input dimension of the defined neural network architecture. In addition, normalize pixel values to the range between 0 and 1.[7] There were



(fig1 Data preprocessing)

missing values in the metadata. So, we managed to fill them using aggregate functions. We also checked the presence of duplicates and made sure that there aren't any duplicates. Most of the malignant results are found around first 20 scans, it's more likely to be malignant result if there are more scan images. So, there many images for every single patient. That's why we calculated the number of images per patients. We transformed all categorical features un numerical to numbers by using "Label Encoding**". For the augmentation**, we added more hair to the images, in addition to some flipping, shifting, rotation and shearing.



(image before and after adding hair)

D. Data Down-Sampling

ISIC 2020 skin cancer dataset is severe imbalanced which contains only 584 confirmed melanoma cases and 32,542 benign cases. By randomly replicating some of the melanoma photos, a random down-sampling strategy is utilized to decrease the size of the major class (benign cases) [8] When the dataset is imbalanced, the number of the used images

2067 for train and 517 for validation. Antother reason why we did downsampling is to save memory due to limited computational resourses.

E. Metadata Information

Metadata consists of data on lesion, an improvement of approximately 7% in balanced ACC when applying metadata information on DL models [6]. Therefore, Feature encoding (turning values into numbers). A machine learning model requires all values to be numerical) for example, Feature encoding is applied to 'sex', 'anatomy', 'diagnosis' and in addition feature normalization by normalizing the age variable to the range between 0 and 1. [7]

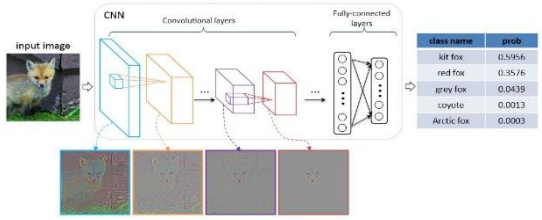
METHODS:

We managed to make two models, VGG16 and EfficientnetB0 and we got 75% accuracy. The best model was VGG net and we were able to get accuracy of 82% from it. we fine-tuned the parameters of the architecture. VGG is a famous CNN architecture that first showed it was possible to do accurate image recognition with a deep network and small convolutional filters. It proved this to the world scoring highly in the ImageNet Challenge. The VGG network is constructed with very small convolutional filters The VGG16 consists of 13 convolutional layers and three fully connected layers.

Convolution layers: VGG's convolutional layers leverage a minimal receptive field, i.e., 3×3, the smallest possible size that still captures up/down and left/right. Moreover, there are also 1×1 convolution filters acting as a linear transformation of the input.

Hidden layers: All the hidden layers in the VGG network use ReLU. VGG does not usually leverage Local Response Normalization (LRN) as it increases memory consumption and training time. Moreover, it makes no improvements to overall accuracy.

Fully connected: The VGGNet has three fully connected layers. Out of the three layers, the first two have 4096 channels each, and the third has 1000 channels, 1 for each class.

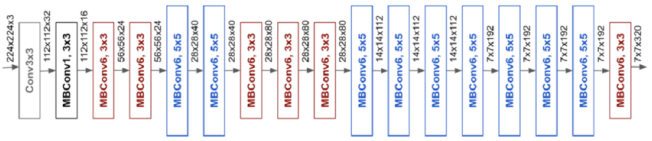


The used loss function "Binary cross entropy". Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1. It then calculates the score that penalizes the probabilities based on the distance from the expected value. That means how close or far from the actual value.

$$logloss = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij}\log(p_{ij})$$

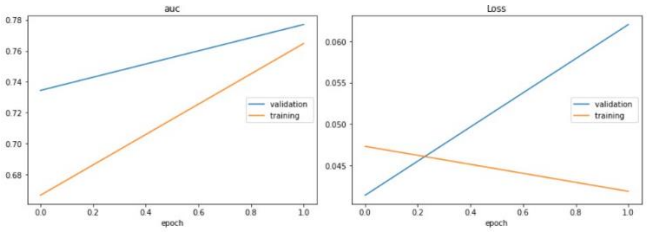
We used "Adam optimizer", This optimization algorithm is a further extension of stochastic gradient descent to update network weights during training. Unlike maintaining a single learning rate through training in SGD, Adam optimizer updates the learning rate for each network weight individually.

The second algorithm used was "**EfficientnetB0**", EfficientNet-b0 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network can classify images into 1000 object categories.



EXPERIMENTS/RESULTS

For gradient descent we used minibatch with 8 batches as batches ranging from 8 to 32 gave the best results as the key advantage of using minibatch as opposed to the full dataset goes back to the fundamental idea of stochastic gradient descent1. In batch gradient descent, you compute the gradient over the entire dataset, averaging over potentially a vast amount of information. It takes lots of memory to do that. In addition Stable Convergence is another advantage, the more stable converge towards the global minimum since we calculate an average gradient over n samples that results in less noise. Furthermore, To get the training steps we divided the train shape rows by the batch size which gave us 258 when 8 batchsize was used. In addition we included metadata and built baseline model using it and ensembled it with the cnn features obtained. For the metrics we used accuracy and obtained 82.12% and used AUC and obtained the following graph



CONCLUSION/FUTURE WORK:

We performed EDA to understand the data then used the preprocessing and augmentation techniques to treat the data and get better results and in the end included metadata features and ensembled it with the CNN features learned by the model we used. We are planning to use more complex algorithms by obtaining more computational resources to try to get higher accuracies, by using many architectures. Also, we can use other augmentation and preprocessing ways to be able to higher our models ' performance. Finally, we are looking forward to use more melanoma images like using previous years competitions to better detect melanoma.

REFERENCES

[1] https://www.kaggle.com/c/siim-isic-melanoma-classification

[2] https://doi.org/10.1038/s41597-021-00815-z

[3] Y. Zhang and C. Wang, "SIIM-ISIC Melanoma Classification With DenseNet," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 14-17, doi: 10.1109/ICBAIE52039.2021.9389983.

[4] https://arxiv.org/pdf/2010.05351.pdf

[5] https://dl.acm.org/doi/abs/10.1145/3437378.3437396

[6] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9858890

[7] https://bibliotecadigital.udea.edu.co/bitstream/10495/24997/4/ArbelaezNestor_2021_MelanomaClassificationModel.pdf

[8] https://link.springer.com/content/pdf/10.1007/978-3-030-91103-4.pdf?pdf=button

[9] https://link.springer.com/content/pdf/10.1007/978-3-031-08530-7.pdf?pdf=button%20sticky

[10] https://www.analyticsvidhya.com/blog/2021/04/steps-to-complete-a-machine-learning-project/

[11] https://sci-hub.ru/https://doi.org/10.1145/3437378.3437396

[12] https://arxiv.org/pdf/2204.12406v1.pdf

[13] Ballerini L, Fisher RB, Aldridge B, Rees J. A Color and Texture Based Hierarchical K-NN Approach to the Classification of Non-melanoma Skin Lesions. Color Medical Image Analysis. Springer, Dordrecht; 2013. pp. 63–86

[14] Barata, C., Ruela, M., Francisco, M., Mendonça, T., & Marques, J. S. (2013). Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal, 8(3), 965-979.

[15] Dreiseitl, S., Ohno-Machado, L., Kittler, H., Vinterbo, S., Billhardt, H., & Binder, M. (2001). A comparison of machine learning methods for the diagnosis of pigmented skin lesions. Journal of biomedical informatics, 34(1), 28-36..

[16] Garnavi, R., Aldeen, M., & Bailey, J. (2012). Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. IEEE Transactions on Information Technology in Biomedicine, 16(6), 1239-1252.

[17] Kawahara, J., BenTaieb, A., & Hamarneh, G. (2016, April). Deep features to classify skin lesions. In 2016 IEEE 13th international symposium on biomedical imaging (ISBI) (pp. 1397-1400). IEEE Conference Publication. [cited 4 Sep 2020]. Available: https://ieeexplore.ieee.org/document/7493528

[18] Celebi ME, Iyatomi H, Schaefer G, Stoecker WV. Lesion Border Detection in Dermoscopy Images. 2010. doi:10.1016/j.compmedimag.2008.11.002

[19] Iyatomi, H., Oka, H., Celebi, M. E., Hashimoto, M., Hagiwara, M., Tanaka, M., & Ogawa, K. (2008). An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Computerized Medical Imaging and Graphics, 32(7), 566-579.

[20] Celebi M, Kingravi HA, Uddin B, Iyatomi H, Alp Aslandogan Y, Stoecker WV, et al. A methodological approach to the classification of dermoscopy images. Comput Med Imaging Graph. 2007;31: 362.

[21] Lopez, A. R., Giro-i-Nieto, X., Burdick, J., & Marques, O. (2017, February). Skin lesion classification from dermoscopic images using deep learning techniques. In 2017 13th IASTED international conference on biomedical engineering (BioMed) (pp. 49-54). https://ieeexplore.ieee.org/abstract/document/7893267

[22] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://arxiv.org/ pdf/1409.1556.pdf

[23] ISIC Archive. [cited 4 Sep 2020]. Available: https://www.isic-archive.com/

[24] Milton MAA. Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge. 2019. Available: http://arxiv.org/abs/1901.10802

[25] Liao, H. (2016). A deep learning approach to universal skin disease classification. University of Rochester Department of Computer Science, CSC.. Available: https: //pdfs.semanticscholar.org/af34/fc0aebff011b56ede8f46ca0787cfb1324ac.pdf

[26] Codella N, Nguyen Q-B, Pankanti S, Gutman D, Helba B, Halpern A, et al. Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images. 2016. Available: http://arxiv.org/abs/1610.04662

[27] El-Khatib H, Popescu D, Ichim L. Deep Learning–Based Methods for Automatic Diagnosis of Skin Lesions. Sensors . 2020;20. doi:10.3390/s20061753

[28] Baghersalimi S, Bozorgtabar B, Schmid-Saugeon P, Ekenel HK, Thiran J-P. DermoNet: densely linked convolutional neural network for efficient skin lesion segmentation. EURASIP Journal on Image and Video Processing. 2019;2019: 1–10.

[29] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. 2016. Available: http://arxiv.org/abs/1608.06993

[30] Li Y, Shen L. Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network. Sensors . 2018;18: 556.

[31] Ünver HM, Ayan E. Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. Diagnostics (Basel, Switzerland). 2019;9. doi:10.3390/diagnostics9030072

[32] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2015. Available: http://arxiv.org/abs/1506.02640

[33] Prasad T. and Siddhivinayak K. (2020, June). Skin Lesion Classification : A CNN Way .IJEAT . ISSN : 2249-8958 , Volume-9 Issue-5, June 2020

CONTRIBUTION

Ahmed Mahmoud (33.33%) EDA & preproceesing
Yumna Hamdy (33.33%) Augmentation & Modeling
Ahmed Moatez (33.33%) Modeling & Ensembling