Skin Cancer Analysis using Deep Learning

Anonymous Author(s)
Affiliation
Address
`email`

1 Statement of the problem

The present study aims to contribute with the detection of melanoma, which is the major cause of death from skin cancer (1). Ultraviolet(UV) radiation exposure is considered a factor that increases the risk to develop melanoma. Unfortunately UV index measured in Peru during the summer are very high along its territory: it can go from 10 to 14 (2). According to the Instituto Nacional de Enfermedades Neoplasticas (INEN), the most important Hospital for cancer treatment in Peru, the number of cases during the period 2012-2017 showed an increase of 50.73% compared to 2006-2011 (3). This situation has led authorities to organize medical campaigns for the detection and prevention of this type of cancer looking forward to detecting the disease in an early stage which can contribute to decrease the death toll. Different techniques are currently used to identify melanoma on the skin. One of the most popular is the ABCDE (4), which is a self-skin examination that helps visually if the characteristics of spots on the skin correspond to melanoma. An automatic tool is considered helpful in this detection process especially in Peru where access to health institutions can be limited in some regions. Image analysis have been widely used to perform an automated test. Therefore, a Peruvian mobile application called MELapp was developed for this purpose, however, it delivers a result in 24 hours (5). Other more advanced applications such as SkinVision offer good alternative for prevention; however, costs might be inaccessible for a large population in Peru and the LatAm community. Hence, the need to develop in-house technology which has its beginning in the study of deep learning methods applied to cancer detection.

2 Methodology

The data used in this study was released for the ISIC 2018 challenge (6). Table 1 resumes the total amount of images used in this study, and the division into three different datasets (i.e. train, validate and test). From the three tasks presented in the challenge (i.e. lesion segmentation, lesion attribute detection, and disease classification), we gathered 13779 images and selected only those having a segmentation-labeled mask, and expert annotation. We took the 10% of the 10599 selected images for the test dataset, and we augmented the remaining 80% by performing 4 rotations (i.e. $\pm 30°$, $\pm 30°$) and 3 flips in the $x,\ y$ and xy axis. Finally, the remaining 80% was divided into a training dataset (80%) and validation dataset (20%).

Table 1: Number of images in original and augmented dataset.

Dataset	**Original dataset**	**Augmented dataset**
Train	7632	61056
Validation	1908	15264
Test	1059	1059

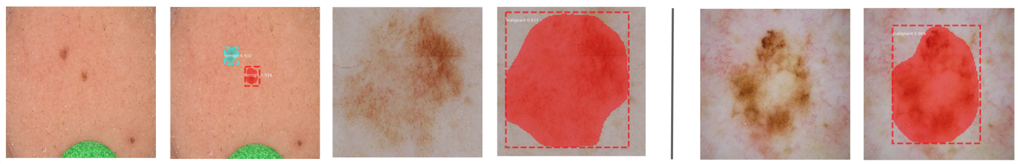
We used the Mask-RCNN architecture proposed by K. He, *et al.,* in (7; 8). To evaluate the performance of the model, we used the sensitivity, specificity and accuracy for the classification, and implemented

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

the IOU (Jaccard coefficient) for the segmentation stage of the network. We trained the network in a Nvidia GTX1070 GPU with 8GB of memory.

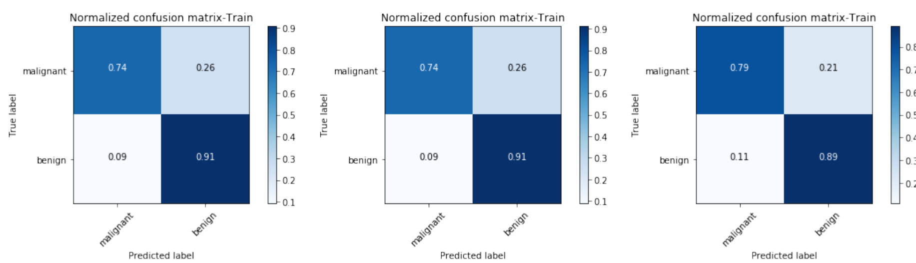
3 Results

Figures 1a-d show correct cases of segmentation and classification. In addition, Figures 1e,f and 2 show that there are several cases in which there is a misclassification of the lesion even when the image is segmented correctly. This is also reflected in the accuracy value being lower than the IOU as shown in Table 2. From the confusion matrices in Figure 2, we found higher values of specificity over sensitivity due to the unbalance of the training dataset.



(a) Image (b) Prediction (c) Image (d) Prediction (e) Image (f) Prediction

Figure 1: Left:benign tumor, right: malign tumor



(a) Train (b) Validation (c) Test

Figure 2: Confusion Matrix for train (a), validation (b), and test (c) dataset

Table 2: Metrics used to evaluate the performance of the Mask-RCNN.

Dataset	Sensitivity	Specificity	Accuracy	IOU
Train	0.749	0.913	0.855	0.948
Validation	0.742	0.911	0.853	0.949
Test	0.791	0.895	0.805	0.956

4 Conclusions and discussion

Results from the top 5 teams in the challenge showed specificity and sensitivity values higher than 0.9, and a Jaccard Index between 0.79 and 0.8. Although high sensitivity is required for clinical applications to be viable, our IOU results (i.e. 0.95 for test dataset) indicate high precision in generating a fair approximated mask and boundaries of the lesions in order to compute, at least, three of the five parameters of the standard medical practice: asymmetry, borders and diameter (from the ABCDE test for melanoma). Further work will focus on improving accuracy to provide the doctors with an additional second opinion for better diagnosis of the disease. Multiple class classification might be a solution as the network might learn new features from other types of lesions and improve the metrics. Other instance segmentation models for real time applications, such as YOLACT (9), will also be analyzed as Mask-RCNN has a higher computational footprint, and the goal is to deploy the model into a mobile app. Advantages of both approaches will be considered as the model can always be deployed as a web service; however, this involves bigger infrastructure which increases the cost of maintenance. On the other hand, regarding the image resolution (e.g. images taken from different cellphones), we take advantage of the mask-RCNN as it is able to classify instances regardless the input image size. Finally, results obtained in this trial run showed the potential of deep learning approaches to be embedded into useful medical tools which can benefit the LatAm region.

References

[1] W. H. Organization, “Health effects of uv radiation.” [Online]. Available: https://www.who.int/uv/health/uv_health2/en/index1.html

[2] Senamhi, “Verano 2018: Radiación uv registra nivel de riesgo “Muy Alto”.” [Online]. Available: https://www.senamhi.gob.pe/?&p=prensa&n=769

[3] INEN, “Casos nuevos de cancer registrados en inen, periodo 2000-2017.” [Online]. Available: https://portal.inen.sld.pe/indicadores-anuales-de-gestion-produccion-hospitalaria/

[4] D. Jensen and B. Eleeski, “The abcdef rule: Combining the “abcde rule” and the “ugly duckling sign” in an effort to improve patient self-screening examinations,” *The Journal of clinical and aesthetic dermatology*, vol. 8, no. 2, p. 15, Feb 2015.

[5] InnovatePeru, “Crean una app para detectar cancer de piel.” [Online]. Available: https://innovateperu.gob.pe/noticias/noticias/item/1615-crean-app-para-detectar-cancer-de-piel

[6] I. 2018, “Isic 2018: Skin lesion analysis towards melanoma detection.” [Online]. Available: https://www.isic-archive.com/#!/topWithHeader/tightContentTop/challenges

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2980–2988.

[8] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, “Detectron,” https://github.com/facebookresearch/detectron, 2018.

[9] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “Yolact: Real-time instance segmentation,” *arXiv*, 2019.