An app to detect melanoma using deep learning: An approach to handle imbalanced data based on evolutionary algorithms

Pedro B. C. Castro
Bio-inspired Computing Lab
LABCIN - UFES
Vitória, Brazil
pedrobccastro@gmail.com

Breno Krohling
Bio-inspired Computing Lab
LABCIN - UFES
Vitória, Brazil
breno.krohling@aluno.ufes.br

Andre G. C. Pacheco
Bio-inspired Computing Lab
LABCIN - UFES
PPGI - UFES
Vitória, Brazil
agcpacheco@inf.ufes.br

Renato A. Krohling
Bio-inspired Computing Lab
LABCIN - UFES
PPGI - UFES
Vitória, Brazil
rkrohling@inf.ufes.br

***Abstract*—Skin cancer is a quite common type of cancer. Its incidence is more often in caucasian people and melanoma is the most lethal one. In order to increase patient prognosis, developing tools to assist early diagnosis is quite important. In the last few years, several methods have been proposed to deal with automated melanoma detection. Nonetheless, most of them are based only in dermoscopy images and/or do not take into account lesion clinical information. In this paper, we developed an ease and accessible mobile tool to assist melanoma detection. The app is linked to a convolutional neural network (CNN) trained on images collected from smartphones and lesion clinical information. Since the occurrence of melanoma is much smaller than other skin lesions, most of the datasets for this problem are imbalanced. To deal with this issue, we present an approach based on evolutionary algorithm to balance datasets. The proposed approach obtained promising results in comparison with related works with a balanced accuracy of 92% and a recall of 94%. However, it is a preliminary study, therefore these metrics may vary when applied to other datasets.**

***Index Terms*—melanoma, skin cancer, app, deep learning, CNN, data balancing**

I. Introduction

The skin cancer occurrence, melanoma and non-melanoma, has increased over the last decades. Currently, the World Health Organization (WHO) estimates that 2-3 million non-melanoma cancers and 132,000 melanomas occur every year in the world [1]. According to the Brazilian Cancer National Institute (INCA), one in every three cancer diagnosis is a skin cancer [2]. The presence of skin cancer is strongly related to the incidence of ultraviolet radiation caused by sunlight exposure [3]. Due to the lack of pigmentation, caucasian people are under the highest risk [4]. Early detection is crucial to increase patient prognosis since melanoma spread is a life-threatening [5].

Several computer-aided diagnoses (CAD) have been proposed to automated skin cancer detection [6]–[15]. In the last few years, most approaches applied Convolutional Neural Networks (CNN) trained on dermoscopy images [6]–[12]. However, in emerging countries such as Brazil, in particular in the countryside [16], there are not enough dermatologists nor dermatoscopes[1]. However, smartphones may be useful to overcome this situation. According to the Ericsson report [17], in 2019 the total number of mobile subscriptions around the world was around 8 billion. In Brazil around 78% of the population have their own smartphone [18]. Therefore, a smartphone based application to assist doctors to diagnose melanoma during the screening process is very desired.

Pacheco and Krohling [19] approached the task of lesion detection considering the six most common skin diseases. Their work uses images collected from smartphones instead of dermoscopy ones. The proposed model also considers the lesion clinical information related to each image. The authors obtained an average improvement of around 7% in the balanced accuracy, which shows the impact of clinical information on automated skin cancer classification.

Dai *et al.* [20] proposed an embedded *iOS* mobile application for skin cancer detection using a CNN. The model was trained using 10,000 dermoscopy images clustered into 7 different types of skin lesions. As it is based on dermoscopy images, to use a smartphone app is necessary a dermatoscope attached to it. This is a limitation since such device is expensive and not often available in remote areas. Also, it does not use clinical information to improve performance. Phillips, Fosu and Jouny [21] focused on developing an *Android* application to detect melanoma using Support Vector Machine (SVM) trained on a dataset composed of 20 images of 3 types of skin lesions. Instead of an embedded solution in a smartphone device, the authors decided to use a server to perform the model. As the model was trained using few samples, its performance is limited.

Alquran *et al.* [14] also worked with melanoma detection using a dataset composed of dermoscopy images. The authors also applied SVM that achieved a balanced accuracy and recall of 84.01% and 83.33%, respectively.

In this paper, we extend Pacheco and Krohling [19] in the following points: 1) we propose an approach to deal

[1]medical device that magnifies the lesion for better visualization

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

with imbalanced data; 2) we link their model with a mobile application; 3) we expand the dataset including other skin lesions.

The remainder of this paper is organized as follows: in section 2, we describe the CNN approach and data balancing techniques. In section 3, we present the technologies used to develop the app. In section 4, we provide experimental results and discussions; and in section 5, we draw some conclusions.

II. An Approach to detect melanoma using deep learning

In this section, we describe an approach to combine clinical images and lesion clinical information using a CNN [19]. Next, we describe the proposed data balancing methods.

A. Deep model architecture to detect melanoma

Pacheco and Krohling [19] presented a dataset composed of clinical images and lesion clinical information, and an approach to combine both. Each sample in the dataset has a clinical diagnosis, an image, and eight clinical features: the patient's age, the part of the body where the lesion is located, if the lesion itches, bleeds or has bled, hurts, has recently increased, has changed its pattern, and if it has an elevation. The clinical information was encoded in 22 variables: 15 bits for region of the body, 1 integer for age, and 6 bits for the remaining features. These features are based on common questions that dermatologists ask patients during an appointment [19].

In order to combine clinical images and lesion clinical information, Pacheco and Krohling [19] proposed a straightforward mechanism to control the contribution of features extracted from images (FI) and clinical information (CI). We applied the same approach, but now for melanoma detection. Figure 1 shows a schematic diagram of the used approach.

It is possible to assign more importance for FI or CI by changing the number of features of each one. As the number of clinical information data N_{CI} is fixed, one can manipulate the number of features extracted from the image N_{FI}. In (1) is described how to calculate the N_{FI} given the N_{CI} and the contribution factor (λ) of N_{CI} from all the features.

$$N_{FI} = \frac{N_{CI}}{1-\lambda} - N_{CI},\ \lambda \in [0,1]. \tag{1}$$

B. Data Balancing

A dataset is imbalanced when the number of samples for each class is not uniform distributed among the classes. Classifiers tend to perform worse on imbalanced dataset since they are designed to generalize from data [22]. To deal with imbalanced data, we propose two new methods based on evolutionary algorithm: the Mixup Extrapolation Balancing (MUPEB) and the Differential Evolution (DE). In addition, We have also applied two standard data balancing techniques: oversampling and weighted loss function. The balancing techniques are described in the following.

1) Weighted loss function: This technique does not change the frequency of samples on datasets. It consists of using a weighted loss function based on a strategy that penalizes miss classification of minority classes. In this paper, we applied the weighted cross-entropy as a loss function. The weight assigned to each label is described by:

$$W_i = \frac{N}{n_i} \tag{2}$$

where N is the total of samples and n_i is the number of samples of class i.

2) Oversampling: This technique consists of making copies of the training dataset until the classes have the same number of samples [23]. Several works applied oversampling or some variation to tackle the imbalanced data problem [22], [24].

3) Mixup Extrapolation Balancing (MUPEB): This technique balances the dataset by applying a series of mixup [25] and extrapolation operations. From 2 images, the method generates 5 new ones. This is done until the dataset gets balanced. The possible combinations of 2 images, X_1 and X_2, are described by:

$$P_1 = 0.5X_1 + 0.5X_2 \tag{3}$$

$$P_2 = 1.5X_1 + 0.5X_2 \tag{4}$$

$$P_3 = 0.5X_1 + 1.5X_2 \tag{5}$$

$$P_4 = 1.5X_1 - 0.5X_2 \tag{6}$$

$$P_5 = 1.5X_2 - 0.5X_1 \tag{7}$$

where *P* represents each of the possible combinations applied to the images X_1 and X_2. Regarding the clinical data, it will be copied from one of the images in the following way: for the images resulting from (3), the clinical data is chosen randomly; from (4) and (6) the clinical data will come from the first image of the pair, X_1. Lastly, the images resulting from (5) and (7), the clinical data will come from the second image of the pair, X_2. It is worth mentioning that since samples from the same class share similar clinical information, this oversampling method does not impact significantly in the model training phase.

4) Differential Evolution (DE): inspired by the mutation operator from the differential evolution algorithm [26], which combines 3 images resulting in a new image. The operator is defined as follows:

$$X_4 = X_1 + \alpha(X_2 - X_3) \tag{8}$$

where X is a set of images and α is a factor ranging from -0.5 to 0.5, a new value for α is chosen in each combination according to a uniform probability distribution. Regarding clinical information used, for each combination generated, the

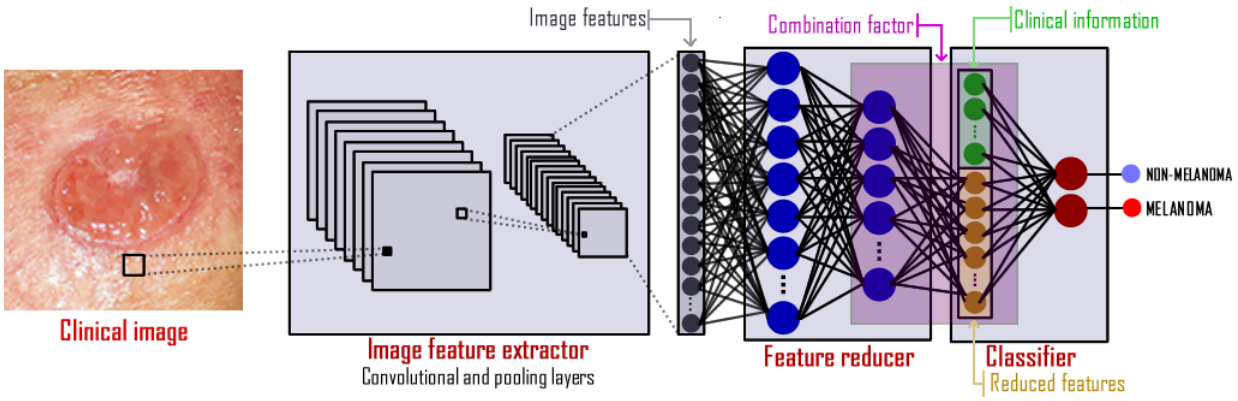


Fig. 1: The illustration of the model proposed by Pacheco and Krohling [19]. In this work, we modified the last layer for melanoma detection

clinical information is randomly chosen between one of the three base images. This technique is also applied only for data from the same class.

III. App development

In order to assist on melanoma detection, we developed a multi-platform smartphone application. The app's purpose is to assist clinicians who have no or low dermatological experience or do not have access to a dermatoscope. Using the app, clinicians may prioritize patients with possible skin cancer on screening process, leading them to a specialist.

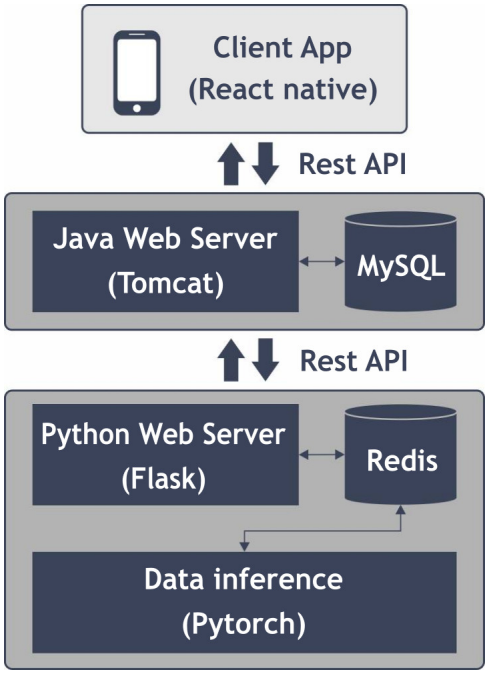


Fig. 2: Schematic diagram of the smartphone app to melanoma detection

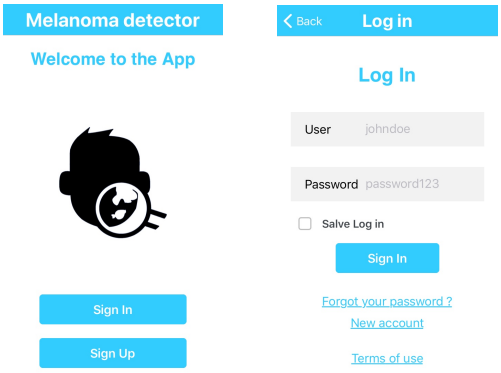
Embedding a CNN in a smartphone presents two main requirements: 1) the weight's size that can be too large and does not fit on the device's memory; and 2) the need of computational resource to perform the model. Since not all smartphones can fulfill these requirements, we decided to deploy the CNN model on a server. Figure 2 shows a schematic diagram of the developed system.

On the client side, we have a mobile application developed using *React Native*[2] framework, and *Expo SDK*[3]. The application sends skin lesion images along with their clinical information to the server. The server performs the CNN model and replies the diagnosis prediction. Finally the app displays it on the screen.

The first back-end layer is based on the java web server *Tomcat* that implements a *Rest API* to be consumed by the user as a service. All user information, for log purpose, is stored in a *MySQL* database. The second layer is based on *Flask*[4], a framework based on *Python* that is designed for micro applications. It makes a direct execution of the machine learning models, which were developed also in *Python*. Every request for processing a new clinical image with its clinical information that arrives at *Flask* is queued in *Redis* [5], a *NoSQL* key-based database. If no data is being processed then the first available data of the *Redis* queue is read and sent to a previous trained model. The result of the model is then stored on *Redis* that will be further consulted by the user.

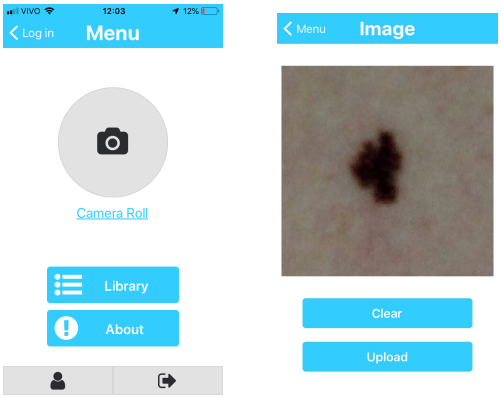
Next is presented screenshots of the application to illustrate its workflow. Figure 3 shows the main screen and the log in screen, respectively. Next, Figure 4 shows the menu and the image upload process, respectively. Last, Figure 5 shows the form to collect clinical information and the image of the lesion itself with the diagnosis prediction, respectively.

[2]https://facebook.github.io/react-native/docs/getting-started
[3]https://docs.expo.io/versions/latest/
[4]https://flask-doc.readthedocs.io/en/latest/
[5]https://redis.io/



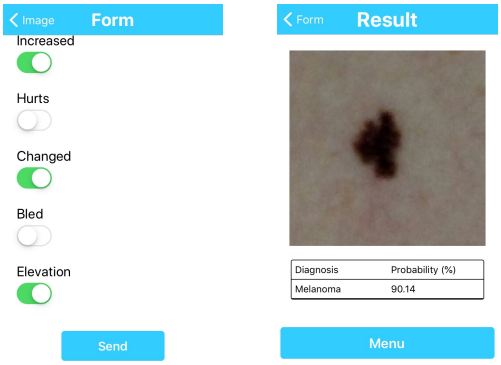
(a) App home screen (b) App log in

Fig. 3: App's home screen and log in screen



(a) App menu (b) App image acquisition

Fig. 4: App's menu screen and image acquisition's screen



(a) App clinical information (b) App result

Fig. 5: App's clinical information screen and result's screen

IV. Experiments and Results

In this section, we present the dataset used to train the CNN, a visualization of the features extracted from the images, the criteria used in the networks evaluation, and the results obtained from the simulations.

A. Dataset

The PAD-UFES dataset used in this work is a combination of the one introduced by Pacheco and Krohling [19], which is composed of the six most common skin diseases: Actinic Keratosis (ACK), Basal Cell Carcinoma (BCC), Melanoma (MEL), Nevus (NEV), Squamous Cell Carcinoma (SCC), and Seborrheic Keratosis (SEK). We added one more disease class labeled as Others, which includes lesions that were not represented within the previous 6 classes. Next, we split the data into melanoma and non-melanoma as presented in Table I.

TABLE I: The frequency of each label in PAD-UFES dataset. We included the OTHERS label into the original dataset proposed by Pacheco and Krohling [19].

Disease	Images	Diseases	Images
MEL	67	Melanoma	67
ACK	543	Non-melanoma	1990
BCC	442		
NEV	196		
SCC	149		
SEK	215		
OTHERS	445		
Total	2057		

B. Visualization

We applied the t-Distributed Stochastic Neighbor Embedding (t-SNE) [27], which is a visualization of high-dimensional data. In total, 2048 features were extracted from all dataset samples after the last ResNet50 convolutional layer. These features were reduced to two dimensions using t-SNE and shown in Figure 6, which we observe that some samples of melanoma are overlapped with non-melanoma ones.

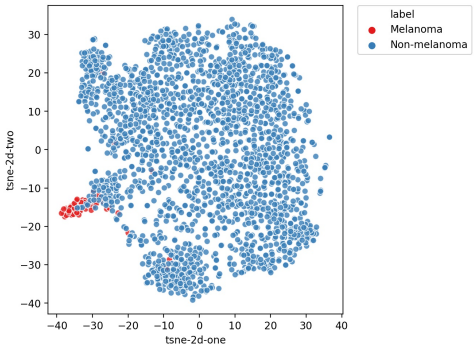


Fig. 6: Visualization of the features extracted by ResNet50 from all samples using t-SNE.

C. Evaluation criteria

As evaluation criteria, we aimed first at a high recall, followed by a high accuracy, and last for a high precision. This choice is justified since the recall is directly related to the number of false negative, i.e., the number of melanomas classified by the network as non-melanoma. A false negative is

the worst scenario for melanoma detection since the clinician assumes that the lesion is a non-melanoma. The precision is related to the number of false positive, meaning the number of non-melanoma lesions classified as melanomas by the network. In this case, although the patient would be worried, the clinician will send the patient to a specialist.

D. Results

The results are divided according to the type of simulation performed. First, we present a sensitivity study used to find the best setup for the model. Second, we review the impact of clinical information combined with image on the CAD performance. Finally, we investigated the impact of data balancing on the model's performance.

For all tests, a ResNet50 was trained using the architecture described in Pacheco and Krohling [19] combining features extracted from the images with lesion clinical information using a 5-fold cross-validation. ResNet50 was used due to its effective performance [19]. We performed the training phase for 100 epochs using Stochastic Gradient Descent (SGD) optimizer with a learning rate equal to 0.01 that decreases by a factor of 1/2 or 1/5 every 20 epochs, alternately. We applied a standard data augmentation [28] and used the presented techniques to deal with imbalanced dataset. All images were resized to 224×224×3. The evaluation metrics were: balanced accuracy (BACC), precision (PR), recall (REC), and F-measure. The use of F-measure was necessary because of the high imbalance among dataset classes.

E. Best Setup

In order to find the best setup, we used standard data augmentation and weighted loss to deal with the imbalanced dataset, as proposed in [19]. The goal is to find the network's best configuration of hyperparamters, so we can apply it to further experiments. In order to increase the system's recall, we also introduced the use of F-measure defined by:

$$F_\beta = (1+\beta^2)\frac{PR.REC}{(\beta^2.PR)+REC} \tag{9}$$

We carried out a sensitivity analysis, changing the value of β in (9) along with 5 combinations of features (C) extracted from the image (FI) and clinical information (CI). Table II presents all 5 combinations of FI and CI.

TABLE II: Sensitivity analysis taking into account the importance of the FI and CI

C	FI x CI	N_{FI}	N_{CI}
1	90% x 10%		198
2	80% x 20%		88
3	70% x 30%	22	51
4	60% x 40%		33
5	50% x 50%		22

From these experiments, we obtained the best setup with β equal to 7 and 70% of FI and 30% of CI. The metrics using this setup are a BACC of 89.00 $\pm$ 3.64, a precision of 15.65 $\pm$ 6.87 and a recall of 100.00 $\pm$ 0.

F. The impact of clinical information

For the study of the impact of clinical information, the number of features extracted from image is equal to the best results obtained in Sec. IV-E. Also, we simulated with the five values of beta used previously in order to find the best results. Table III presents the results. From Table III and Sec. IV-E, we can compare the performance of the network with and without clinical information. Table IV presents these results. From Table IV, we can notice that the use of clinical information provided an average increase in both BACC and recall of 1.26% and 11.43%, respectively. However, we obtained an average decrease in precision of 7.01%.

G. The impact of data balancing

The impact of balancing techniques is assessed by comparing the 4 balancing approaches, i.e., weighted loss function (WGT), oversampling (OVE), MUPEB, and DE. The setup used was the best found in Sec. IV-E. Table V presents the obtained results.

The results with the weighted loss function presented the best result in terms of recall. The proposed approach based on the mutation operator of DE provided the best result in terms of balanced accuracy. Both approaches, WGT and DE presented significantly better results when compared with those obtained in [14] with an increase of about 16% in recall and 6% in BACC.

V. Conclusion

In this paper, we presented a tool to support the diagnostic of melanoma using a multi platform app for smartphone using deep neural networks. The results obtained with clinical information presents an average balanced accuracy of 89% and a recall of 100%. It is important to mention that these are results from a preliminary study and therefore may not stand in real case applications. The study of the impact of clinical information has shown that clinical information is relevant to melanoma detection since it improved on average balanced accuracy and recall in about 1% and 11%, respectively. Regarding the methods proposed to deal with imbalanced dataset, both Mixup Extrapolation Balancing and Differential Evolution presented a lower performance compared to weighted loss function. However, results obtained with DE inspired mutation operator were very close to those obtained with weighted loss function. Since the dataset is small, we are working on new data augmentation methods to improve our results further.

TABLE III: F-measure regarding varying beta without clinical information

Beta	Metrics		
	BACC	***PR***	***REC***
1	77.37 $\pm$ 13.84	49.56 $\pm$ 23.21	57.14 $\pm$ 28.57
3	78.83 $\pm$ 14.62	31.34 $\pm$ 14.59	62.86 $\pm$ 30.77
5	82.92 $\pm$ 8.86	22.25 $\pm$ 9.75	77.14 $\pm$ 17.14
7	87.74 $\pm$ 4.03	22.66 $\pm$ 8.29	88.57 $\pm$ 10.69
10	87.41 $\pm$ 6.96	22.45 $\pm$ 4.41	85.71 $\pm$ 15.65

TABLE IV: Clinical information study

Beta	Metrics		
	BACC	***PR***	***REC***
7*	89.00 ± 3.64	15.65 ± 6.87	100.00 ± 0.00
7	87.74 ± 4.03	22.66 ± 8.29	88.57 ± 10.69

*With clinical information.

TABLE V: The impact of data balancing

BAL	Metrics		
	BACC	***PR***	***REC***
WGT	89.00 ± 3.64	15.65 ± 6.87	100.00 ± 0.00
OVE	84.34 ± 14.21	32.68 ± 8.93	74.28 ± 30.50
MPE	86.83 ± 12.36	29.9 ± 14.74	82.86 ± 27.70
DE	92.39 ± 4.31	35.67 ± 22.64	94.28 ± 7.00

VI. Acknowledgment

P.B.C. Castro, B. Krohling and R.A. Krohling would like to thank the financial support of the Fundação de Amparo a Pesquisa e Inovação do Espírito Santo (FAPES) - grant n. 575/2018. R.A. Krohling also thanks the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - grant n. 309729/2018-1 and H. Knidel who helped to conceptualize this project. We also thank all the members of the Dermatological Assistance Program (PAD-UFES), specially prof. P.L. Frasson, and the support of J.G.M. Esgario.

References

[1] WHO - World Health Organization. (2019) How common is the skin cancer? [Online]. Available: https://www.who.int/uv/faq/skincancer/en/index1.html

[2] INCA - Instituo Nacional do Câncer. (2019) The cancer incidence in Brazil. [Online]. Available: http://www1.inca.gov.br/estimativa/2018/estimativa-2018.pdf

[3] WHO - World Health Organization. (2019) Health effects of UV radiation. [Online]. Available: https://www.who.int/uv/health/uv_health2/en/index1.html

[4] WHO - World Health Organization . (2019) Who is most at risk of getting skin cancer? [Online]. Available: https://www.who.int/uv/faq/skincancer/en/index2.html

[5] D. Schadendorf, A. C. van Akkooi, C. Berking, K. G. Griewank, R. Gutzmer, A. Hauschild, A. Stang, A. Roesch, and S. Ugurel, "Melanoma," *The Lancet*, vol. 392, no. 10151, pp. 971–984, 2018.

[6] N. Zhang, Y.-X. Cai, Y.-Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," *Artificial Intelligence in Medicine*, vol. 102, p. 101756, 2020.

[7] A. Hekler, J. S. Utikal, A. H. Enk, A. Hauschild, M. Weichenthal, R. C. Maron, C. Berking, S. Haferkamp, J. Klode, D. Schadendorf *et al.*, "Superior skin cancer classification by the combination of human and artificial intelligence," *European Journal of Cancer*, vol. 120, pp. 114–121, 2019.

[8] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz *et al.*, "Deep neural networks are superior to dermatologists in melanoma image classification," *European Journal of Cancer*, vol. 119, pp. 11–17, 2019.

[9] T. J. Brinker, A. Hekler, A. Hauschild, C. Berking, B. Schilling, A. H. Enk, S. Haferkamp, A. Karoglan, C. von Kalle, M. Weichenthal *et al.*, "Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark," *European Journal of Cancer*, vol. 111, pp. 30–37, 2019.

[10] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk, and C. von Kalle, "Skin cancer classification using convolutional neural networks: systematic review," *Journal of Medical Internet Research*, vol. 20, no. 10, p. e11936, 2018.

[11] R. C. Maron, M. Weichenthal, J. S. Utikal, A. Hekler, C. Berking, A. Hauschild, A. H. Enk, S. Haferkamp, J. Klode, D. Schadendorf *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *European Journal of Cancer*, vol. 119, pp. 57–65, 2019.

[12] 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*, vol. 542, no. 7639, p. 115, 2017.

[13] P. Bumrungkun, K. Chamnongthai, and W. Patchoo, "Detection skin cancer using SVM and snake model," in *2018 International Workshop on Advanced Image Technology (IWAIT)*. IEEE, 2018, pp. 1–4.

[14] H. Alquran, I. A. Qasmieh, A. M. Alqudah, S. Alhammouri, E. Alawneh, A. Abughazaleh, and F. Hasayen, "The melanoma skin cancer detection and classification using support vector machine," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. IEEE, 2017, pp. 1–5.

[15] M. Q. Khan and et al., "Classification of melanoma and nevus in digital images for diagnosis of skin cancer," *IEEE Access*, vol. 7, pp. 90 132–90 144, 2019.

[16] H. Feng, J. Berk-Krauss, P. W. Feng, and J. A. Stein, "Comparison of dermatologist density between urban and rural counties in the united states," *JAMA Dermatology*, vol. 154, no. 11, pp. 1265–1271, 2018.

[17] Ericsson. (2019) Ericsson mobility report. [Online]. Available: https://www.ericsson.com/4aacd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf

[18] IBGE - Instituto Brasileiro de Geografia e Estatística. (2017) Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal 2017. [Online]. Available: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101631_informativo.pdf

[19] A. G. C. Pacheco and R. A. Krohling, "The impact of patient clinical information on automated skin cancer detection," *Computers in Biology and Medicine*, vol. 116, p. 103545, 2020.

[20] X. Dai, I. Spasić, B. Meyer, S. Chapman, and F. Andres, "Machine learning on mobile: An on-device inference app for skin cancer detection," in *2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC)*, 2019, pp. 301–305.

[21] O. F. K. Phillips and I. Jouny, "Mobile melanoma detection application for android smart phones," in *2015 41st Annual Northeast Biomedical Engineering Conference (NEBEC)*, 2015, pp. 1–2.

[22] R. Akbani, S. Kwek, and . Japkowicz, "Applying support vector machines to imbalanced datasets," in *Machine Learning: ECML 2004*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 39–50.

[23] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. of the Int'l Conf. on Artificial Intelligence*, 2000.

[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[25] H. Zhanga, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[26] R. Storn and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[27] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Rsesearch*, vol. 9, pp. 2579–2605, 2008.

[28] F. Perez and et.al., "Data augmentation for skin lesion analysis," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer International Publishing, 2018, pp. 303–311.