Telehealth Ecosystems in Practice
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI230738

Challenges in Early Diagnosis of Melanoma Using Tele-Dermatology

Adela-Vasilica GUDIU[a,1] and Lacramioara STOICU-TIVADAR[a]
[a] *University Politehnica Timisoara, Romania*

Abstract. Numerous classification systems have been developed over the years, systems which not only provide assistance to dermatologists, but also enable individuals, especially those living in areas with low medical access, to get a diagnosis. In this paper, a Machine Learning model, which performs a binary classification, and, which for the remainder of this paper will be abbreviated as ML model, is trained and tested, so as to evaluate its effectiveness in giving the right diagnosis, as well as to point out the limitations of the given method, which include, but are not limited to, the quality of smartphone images, and the lack of FAIR image datasets for model training. The results indicate that there are many measures to be taken and improvements to be made, if such a system were to become a reliable tool in real-life circumstances.

Keywords. Melanoma detection, Tele-dermatology, Clinical images, Dermoscopy, Remote diagnosis, Machine Learning, Nevus, Tele-health, Dermatology.

1. Introduction

Over the past few years, the number of people diagnosed with melanoma has been on the rise, particularly among those who are subject to excessive exposure to the sun; in Europe, the incidence was reported to be <25 cases per 100.000 population [1].

Melanoma is classified among the rare types of cancer. Furthermore, due to how difficult it is to diagnose in its early stages, it is also considered as one of the most lethal ones. Therefore, detection at an incipient stage is vital in order to enhance the probability of achieving positive results during treatment. To diagnose skin cancer, dermatologists screen the zone of interest considering the ABCDE signs, which are as follows: Asymmetry, Border irregularity, Color variation, Diameter and Evolution. In addition, they also take into account individual patient information, such as: age, lesion location. These parameters are important in coming up with a conclusion, however, for more accurate diagnoses the specialists use dermatoscopes, with which they visualize the subsurface of the skin [2].

To aid dermatologists in giving a correct verdict, computer-aided diagnosis (CAD) systems and methods, such as [3][4][5], were developed. A new area of interest, along with the growing preoccupation with tele-health, is contributing to the development of the tele-dermatology picture forwarding method, which helps those living in low-income or isolated areas get in touch with a dermatologist, as these communities are reported to have a low density of physicians, nurses, and midwives [6]. Sending a picture of the

[1] Corresponding Author: Adela-Vasilica GUDIU, Faculty of Automation & Computer Science, Bulevardul Vasile Parvan 2, Timisoara, E-mail: adela.gudiu@student.upt.ro.

lesion and getting a diagnosis can improve medical access statistics in the above-mentioned areas. Nonetheless, specialists may prove reluctant to embrace these approaches, as there are a number of limitations that can be distinguished in 5 main categories: clinical, economic, technological, legal, and ethical [7].

The present study focuses on evaluating the reliability, and stating the limitations of an ML model, which was created for physicians as a support for decision-making and was trained with feature data extracted from both dermoscopic images and mobile phone images.

2. Methods

The first step was finding two open-source public-access databases that could be used in the feature extraction. One of the datasets being used, which contains dermoscopic images with confirmed diagnoses, is from the International Skin Imaging Collaboration SIIM-ISIC 2020 Challenge Dataset [8], whose data collections are popular among researchers that are performing medical image analysis using machine learning [9]. A filtered search was performed on their gallery and only dermatoscopic images with melanoma and nevus formations were saved for feature extraction. The same filtered search was performed on the second dataset, PAD-UFES-20, which comprises images of nevus and melanoma formations that were taken with various smartphone cameras. That being the case, the latter are defined by their variety of sizes, resolutions, and lighting conditions [10].

After collecting the dermoscopic images, the next steps taken were, in order: Processing the images, Segmenting the images, and performing Feature Extraction; the extracted features were: Standard Deviation, Mean, L1 Norm, L2 Norm, Mean Absolute Deviation [11][12]. All these tasks were performed using MATLAB [13]. The extracted features were then split into a training dataset (a total of 1981 dermoscopic image features) and a testing dataset (a total of 250 dermoscopic image features) and given as input for training and testing to an ML model created in Visual Studio 2019 [14] using the ML.NET package, which enables us to create Machine Learning models trained for a predetermined interval of time specified by the user [15]. By giving it a previously labeled input dataset, it runs through a number of trainers and chooses the one with the highest accuracy degree, commonly known as best trainer. The training dataset labels were balanced to avoid biases in classification. That is to say, the features extracted from the images provided to the training dataset have been carefully chosen so as to ensure an approximately equal distribution of the labels (taking into consideration the fact that we are working within a binary classification system, the labels are 0 and 1, corresponding to nevus, respectively melanoma formations). **Figure 1** presents the workflow of the process.

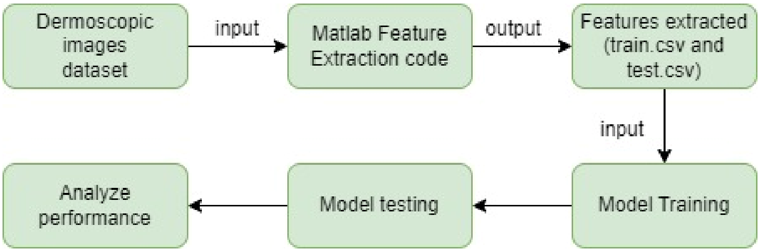

Figure 1. Workflow

3. Results

Subsequent to training, the accuracy on the training data was 84.66%. The test dataset had 250 input lines in total, from which 201 were dermoscopic image features, and 49 were smartphone image features. One quick glance **Table 1** shows how the model gave more accurate results on the dermoscopic images, with an overall accuracy of ≈68.65%. The model failed to correctly categorize most of the melanoma features from smartphone images, out of a total of 24 image features, only 4 were categorized with the corresponding label.

Figure 2 visually represents the melanoma test dataset (both smartphone and dermoscopic images) and the nevus test dataset (both smartphone and dermoscopic images), taking into consideration the total values of True Positive (TP) and False Negative (FN).

Table 1. Model accuracy on test data

Diagnosis	Dermoscopic images	Smartphone images
Melanoma	≈54.90%	≈16.66%
Nevus	≈82.82%	≈56%
Overall	≈68.65%	≈36.73%

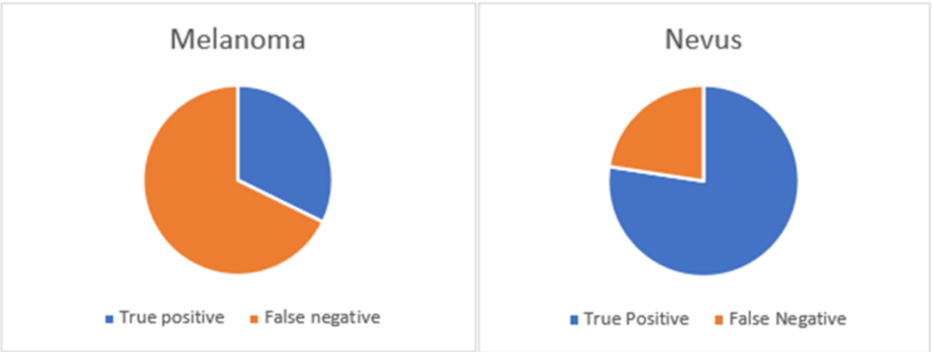

Figure 2. Melanoma and Nevus TP and FN.

4. Discussion and Conclusions

There are certain limitations to this study, some of which arise due to the fact that the ML model was not trained with a big data input set, as it is hard to find big open source databases that contain images of melanoma and nevus formations that were labeled following a thorough medical examination, as well as big datasets of smartphone images that were accordingly labeled. A more comprehensive and inclusive training dataset could give better results in smartphone image classification, given the fact that the smartphone images dataset was smaller compared to the dermoscopic images dataset. Also, an increase in the number of FAIR image databases should help with the improvement of the model, as FAIR principles assure the findability, accessibility, interoperability, and reusability of the datasets [16].

A different issue that has the potential to make classification difficult is the fact that the lesions vary in dimension, pattern, and site on the body [17]. Some body parts are more difficult to photograph without the help of another person, for example: the back side of the torso, hence the outcome picture can be of poor quality, and thus provide further challenge in the diagnostic process, as smartphone images are already of inferior quality compared to dermoscopic images.

Another difficult task is selecting the minimum number of feature sets that will give the best performance of the ML model, as there is no predefined list of standard features that will guarantee the best results [18].

As mentioned in [19], despite the claims that classification algorithms outperform dermatologists in giving diagnoses, there are far more obstacles that need to be overcome by these systems in order for them to be reliable, in view of the fact that the diagnosis flow in a real-life scenario depends on multiple factors, from skin color and number of nevus formations to the medical history of the patient. In this study, the vast majority of the images used for feature extraction were from light skinned individuals, which can affect the categorization of other skin color types due to a lack of data. As it is highlighted in [20], the incidence of melanoma is lower for people with skin of color compared to light skin individuals, but the morbidity and fatality is higher in case of the former. Therefore, there should be a great focus in diversifying the datasets that are used, so that they can be inclusive and not targeted to a certain category of people.

A close collaboration between the developers of these systems and physicians can help eliminate these shortcomings and can also reduce the skepticism regarding the validity of this classification method; it will enable knowledge sharing and the development of new approaches, as developers are provided with valuable input concerning needs and requirements. As melanoma is one of the lethal types of cancer, a noninvasive approach in early diagnosis that enables people in isolated areas to get access to medical care is a much-needed tool that can help increase the number of people treated, as well as significantly reduce the cost of treatments and medical visits.

References

[1] Conforti C, Zalaudek I. Epidemiology and Risk Factors of Melanoma: A Review. Dermatol Pract Concept. 2021 Jul 1;11(Suppl 1):e2021161S. doi: 10.5826/dpc.11S1a161S. PMID: 34447610; PMCID: PMC8366310.

[2] Pacheco A.G., Krohling R.A., The impact of patient clinical information on automated skin cancer detection, Comput. Biol. Med., 116 (2020), Article 103545

[3] Bakheet, S.; Al-Hamadi, A. Computer-Aided Diagnosis of Malignant Melanoma Using Gabor-Based Entropic Features and Multilevel Neural Networks. Diagnostics 2020, 10, 822. https://doi.org/10.3390/diagnostics10100822

[4] Kim, C.-I.; Hwang, S.-M.; Park, E.-B.; Won, C.-H.; Lee, J.-H. Computer-Aided Diagnosis Algorithm for Classification of Malignant Melanoma Using Deep Neural Networks. Sensors 2021, 21, 5551. https://doi.org/10.3390/s21165551

[5] Jojoa Acosta, M.F., Caballero Tovar, L.Y., Garcia-Zapirain, M.B. et al. Melanoma diagnosis using deep learning techniques on dermatoscopic images. BMC Med Imaging 21, 6 (2021). https://doi.org/10.1186/s12880-020-00534-8

[6] Winkelmann, J., Muench, U. & Maier, C.B. Time trends in the regional distribution of physicians, nurses and midwives in Europe. BMC Health Serv Res 20, 937 (2020). https://doi.org/10.1186/s12913-020-05760-y

[7] Pala, P., Bergler-Czop, B., and Gwiżdż, J. (2020). Teledermatology: idea, benefits and risks of modern age – a systematic review based on melanoma. Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii, 37(2), pp.159-167. https://doi.org/10.5114/ada.2020.94834

[8] International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset. International Skin Imaging Collaboration https://doi.org/10.34970/2020-ds01 (2020)

[9] Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, Moi Hoon Yap, Analysis of the ISIC image datasets: Usage, benchmarks and recommendations, Medical Image Analysis, Volume 75, 2022, 102305, ISSN 1361-8415, https://doi.org/10.1016/j.media.2021.102305.

[10] Andre G.C. Pacheco, Gustavo R. Lima, Amanda S. Salomão, Breno Krohling, Igor P. Biral, Gabriel G. de Angelo, Fábio C.R. Alves Jr, José G.M. Esgario, Alana C. Simora, Pedro B.C. Castro, Felipe B. Rodrigues, Patricia H.L. Frasson, Renato A. Krohling, Helder Knidel, Maria C.S. Santos, Rachel B. do Espírito Santo, Telma L.S.G. Macedo, Tania R.P. Canuto, Luíz F.S. de Barros, PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones, Data in Brief, Volume 32, 2020, 106221, ISSN 2352-3409, https://doi.org/10.1016/j.dib.2020.106221

[11] Jaleel, J.A., Salim, S. and Aswin, R.B., 2012. Artificial neural network based detection of skin cancer. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 1(3)

[12] Tumpa, P.P. and Kabir, M.A., 2021. An artificial neural network based detection and classification of melanoma skin cancer using hybrid texture features. Sensors International, 2, p.100128.

[13] The MathWorks Inc. (2022). MATLAB version: 9.14 (R2023a), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

[14] Visual Studio 2019. Microsoft Corporation, 2019

[15] ML.NET 2.0.0, Microsoft Corporation

[16] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guizzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffery, Rajaram Kaliyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Barbara Magagna, Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reisen, Philippe Rocca-Serra, Robert Pergl, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Schneider, George Strawn, Mark Thompson, Andra Waagmeester, Tobias Weigel, Mark D. Wilkinson, Egon L. Willighagen, Peter Wittenburg, Marco Roos, Barend Mons, Erik Schultes; FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence 2020; 2 (1-2): 10–29. doi: https://doi.org/10.1162/dint_r_00024

[17] Grignaffini, F.; Barbuto, F.; Piazzo, L.; Troiano, M.; Simeoni, P.; Mangini, F.; Pellacani, G.; Cantisani, C.; Frezza, F. Machine Learning Approaches for Skin Cancer Classification from Dermoscopic Images: A Systematic Review. Algorithms 2022, 15, 438. https://doi.org/10.3390/a15110438

[18] Javed, R., Rahim, M.S.M., Saba, T. et al. A comparative study of features selection for skin lesion detection from dermoscopic images. Netw Model Anal Health Inform Bioinforma 9, 4 (2020). https://doi.org/10.1007/s13721-019-0209-1

[19] Vocaturo, Eugenio, and Ester Zumpano. "Useful Datasets and Very Recent Approach for Melanoma Image Classification." SPAST Express, 2021

[20] Hugh M. Gloster, Kenneth Neal, Skin cancer in skin of color, Journal of the American Academy of Dermatology, Volume 55, Issue 5, 2006, Pages 741-760, ISSN 0190-9622, https://doi.org/10.1016/j.jaad.2005.08.063