Deep Learning based Melanoma Diagnosis Using Dermoscopic Images

Conor Wall, Fraser Young, Li Zhang, Emma-Jane Phillips

Department of Computer and Information Sciences, Faculty of Engineering and Environment, Northumbria University,
Newcastle, NE1 8ST, UK
[†]E-mail: li.zhang@northumbria.ac.uk

Richard Jiang* and Yonghong Yu[††]

**Computing and Communication, Lancaster University, Lancaster, UK*
[††]College of Tongda, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

The most common malignancies in the world are skin cancers, with melanomas being the most lethal. The emergence of Convolutional Neural Networks (CNNs) has provided a highly compelling method for medical diagnosis. This research therefore conducts transfer learning with grid search based hyper-parameter fine-tuning using six state-of-the-art CNN models for the classification of benign nevus and malignant melanomas, with the models then being exported, implemented, and tested on a proof-of-concept Android application. Evaluated using Dermofit Image Library and PH2 skin lesion data sets, the empirical results indicate that the ResNeXt50 model achieves the highest accuracy rate with fast execution time, and a relatively small model size. It compares favourably with other related methods for melanoma diagnosis reported in the literature.

Keywords: Melanoma Diagnosis; Convolutional Neural Network; Transfer Learning; Remote Healthcare.

1. Introduction

The most common malignancies in the world are skin cancers [1]. Skin cancers are broadly categorised into two groups which are melanomas and non-melanomas (benign and other types of skin cancers). Of the two, melanomas are by far the most fatal [2].

Melanomas are the fifth most common type of cancer in the UK. Since the 1990s, the incidence rate for melanomas has increased over 128% with around

15,000 incidences every year and the mortality rate in 2016 being approximately 15% [3]. The most investigated and well-known risk factor for developing melanomas is exposure to ultraviolet or UV light [4].

Early detection can often result from awareness measures such as ‘Skin Cancer Awareness Month’ as people are made wary of any potential warning signs they may have and thus perform a self-assessment and then determine if they need to see a dermatologist or otherwise [5]. However, a self-assessment is not reliable, and medical screenings by dermatologists can be expensive and not accessible to people living in remote areas. A study conducted by Carrera et al. [6] aimed at finding the overall effectiveness of a variety of methods for melanoma detection such as the ABCD (Asymmetry, Border, Colour, and Diameter) rule, CASH and Menzies methods. They indicated that these methods often had reduced inter-observer agreement levels and by the end of the study, they concluded that more needed to be done to improve dermoscopic technologies, algorithms and criteria.

Medical imaging using deep learning techniques has drawn significant attention owing to the compelling performance of the deep networks for image classification. Therefore in this research, we conduct transfer learning based on a number of state-of-the-art deep Convolutional Neural Networks (CNNs) for melanoma identification. To promote early and instant diagnosis, we deploy the CNN-based skin lesion classification system to smartphone platforms. The empirical results based on the evaluation using two well-known skin lesion data sets, i.e. Dermofit Image Library and PH2, have indicated the efficiency of the proposed method. Our approach also compares favorably with other existing studies in the literature. In short, it provides an accessible, low-cost, and highly accurate method of melanoma diagnosis through the use of CNNs implemented on smartphone devices.

2. Related Work

As computational technologies have advanced, deep learning models have become the most prominent techniques for medical imaging [7, 8, 9]. As an example, Esteva et al. [10] demonstrated the use of a CNN for the detection of malignant melanomas, through the use of the Google Inception-v3 CNN architecture, pre-trained on ImageNet and then trained again on a dataset consisting of over 130,000 images of benign and malignant skin lesions. Their system achieved a mean AUC of 0.95 for melanoma detection, indicating a superior level of performance even when compared to the dermatologists involved in this particular study, and thus showing the vast potential of using a

CNN for the purpose of melanoma detection. Tan et al. [8, 9] proposed deep CNN architecture generation and hyper-parameter fine-tuning for melanoma classification using evolutionary algorithms.

There are several CNN architectures that could potentially be implemented and modified, such as DenseNet [11], and ResNeXt [12]. These CNN architectures and concepts are some of the comparatively more recent and state-of-the-art CNN innovations, meaning that a performance comparison would need to be made in order to determine which would be best suited.

Regarding the current literature concerning deep learning melanoma detection, there is a clear gap regarding the implementation of a CNN model on smartphone applications for the purpose of melanoma detection. As stated by Esteva et al. [10], the ability to implement such a CNN on a mobile device would provide a highly accurate and accessible method of melanoma detection; it would be highly beneficial in terms of improving the rate of early detection and thus, reducing mortality rates. However, significant considerations will also need to be made regarding the chosen model so that performance and usability would not become a significant issue.

3. The Proposed Methodology

3.1. *Dataset*

We employ Dermofit Image Library and PH2 skin lesion datasets for system evaluation. A total of 484 images of two classes, i.e. 270 'benign nevi' and 214 'malignant melanoma' are extracted from the above databases, with 364 and 120 images extracted from Dermofit Image Library and PH2, respectively. All images were scaled to 100×100 and split into a ratio of 80:20 for forming the train and test sets, respectively. Figure 1 illustrates example images used in this research.

3.2. *Transfer Learning*

Transfer learning refers to the fine-tuning of a pre-trained CNN model through the retraining of the final layer in order to benefit from the internal biases and generalisations attained from larger models [13]. By utilising transfer learning, direct benefits were attained in multiple domains, with training times significantly reduced, and high model accuracy. Therefore, we conduct lesion classification using transfer learning based on a number of well-known deep architectures. A grid search technique is also employed to fine-tuning model hyper-parameters to enhance performance.

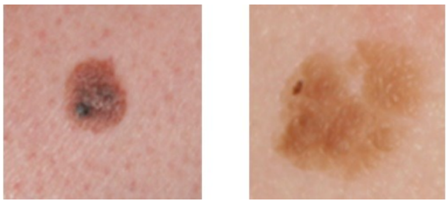

Fig. 1. Extracted example images for malignant melanoma (left) and benign nevus (right) respectively

3.3. *Training Configuration*

Multiple models were selected for testing in this research study, consisting of variations of ResNeXt, Wide-ResNet, MobileNetv2, and DenseNet. Training configurations and hyper-parameter customisation were evaluated using a grid search technique in conjunction with random selection for initialisation; this allowed for the autonomous testing of each model to save computational cost and to provide enhanced performance.

3.4. *Android Development*

Through utilising the Android ecosystem for development, we are able to target a significant user base if the system were to be moved into production. By enabling remote healthcare applications as such, we can potentially deploy the technology into deprived areas with limited access to healthcare around the globe.

By targeting Android Development API 18, *Android Studio* estimates that the application will run on approximately 95.9% of active Android devices globally, further ensuring that the software does not discriminate given outdated technology.

4. Evaluation

As observed in Table 1, ResNeXt50 provides the highest accuracy rate for the test set while still retaining an above-average execution time in classifying skin lesions from an Android smartphone. In conjunction with the small model size of 23.5mb, ResNeXt50 does not discriminate hardware, with most smartphones of recent times containing significantly higher amounts of storage, further establishing the non-exclusivity of the technology at hand.

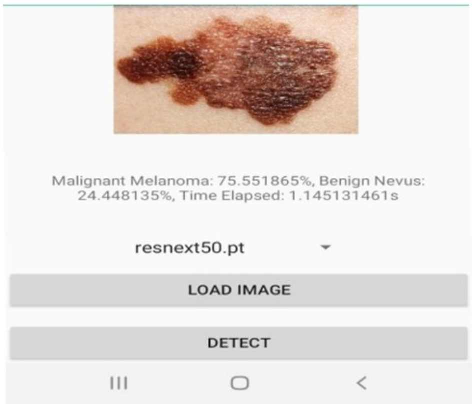

Fig. 2. Android application successfully detecting a 'malignant melanoma' with ResNeXt50

Table 1. The most optimal training configurations identified based on trial-and-error

Model	Epochs	Learning Rate	Avg. Execution Time (seconds)	Accuracy	Size (MB)
ResNeXt50	**16**	**0.0005**	**1.5653**	**0.9658**	**23.5**
ResNeXt101	16	0.001	5.1304	0.9582	87.2
Wide ResNet50	20	0.005	4.6034	0.9062	67.3
Wide ResNet101	16	0.0001	7.3067	0.9167	125.3
DenseNet201	16	0.0005	1.9820	0.9375	20.0
MobileNetV2	16	0.0005	0.6219	0.9063	3.5

Note: Avg. execution time refers to the execution of models from an Android device running 'Android 10'.

It should be seen, however, in the most extreme of cases where an individual may only have access to a particularly old device, utilising a lighter model with a lesser accuracy rate may be beneficial as to allow the prospective user to utilise the software. To provide this functionality, the Android application observed in Figure 2 contains a dropdown menu containing all trained models for tailoring to the specific device.

Table 2. Grid search hyper-parameters for ResNeXt50

Model	Epochs	Learning Rate	Accuracy
ResNeXt50	10	0.001	0.8422
	10	0.1	0.8037
	15	0.001	0.8820
	10	0.0005	0.9100
	16	**0.0005**	**0.9658**

Although outside of the specification of the current system, it is unclear as to whether remotely hosting a highly trained model in the cloud will perform better than the proposed system; despite the system requiring an internet connection in that scenario.

By utilising grid-search methodologies for hyper-parameter optimisation across all models, we were able to increase the performance of each model significantly with autonomy. Referring to Table 2 which illustrates the process of calculating the highest accuracy by evaluating the model on the test set with optimised parameters, our best configuration utilised 16 epochs and a learning rate of 0.0005 over the ResNeXt50 model.

5. Conclusion and Future Work

We employ transfer learning based on several state-of-the-art CNN models for skin lesion classification. A grid search method is used to conduct hyper-parameter fine-tuning to optimize performance. Based on the empirical results, the ResNeXt50 model in particular stands out as the most suitable technique for melanoma detection on Android devices. In terms of the average test accuracy rate, execution time and the model size, the ResNeXt50 excels in each category with the highest test accuracy rate of 96.58%, and the second-highest performing execution time of 1.5653(s). The proposed system also outperforms existing studies, such as Tan et al. [8], with an accuracy rate of 95.37% for melanoma classification using the same data set. Currently, the proposed system does not utilise any segmentation or masking techniques [14-16] in order to provide the user of the affected areas of skin; in future work, it will be beneficial to implement the system with a tested algorithm such as Mask R-CNN to further improve the functionality of the system. Evolutionary algorithms will also be utilized for optimal hyper-parameter selection [17-23] to enhance performance.

References

1. J. D'Orazio, S. Jarrett, A. Amaro-Ortiz, and T. Scott. (2013). UV Radiation and the Skin. *International Journal of Molecular Sciences*, 14(6), pp.12222-12248.
2. J. Gershenwald and G. Guy. (2015). Stemming the Rising Incidence of Melanoma: Calling Prevention to Action. *JNCI: Journal of the National Cancer Institute*, 108(1).
3. Cancer Research UK. (2019). Melanoma skin cancer incidence statistics. [online] Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/incidence#ref-0
4. K. Lacy and W. Alwan. (2013). Skin cancer. *Medicine*, 41(7), pp.402-405.
5. J. Robinson and A. Halpern. (2016). Cost-effective Melanoma Screening. *JAMA Dermatology*, 152(1), p.19.
6. C. Carrera, M. Marchetti, S. Dusza, G. Argenziano, R. Braun, A. Halpern, N. Jaimes, H. Kittler, J. Malvehy, S. Menzies, G. Pellacani, S. Puig, H. Rabinovitz, A. Scope, H. Soyer, W. Stolz, R. Hofmann-Wellenhof, I. Zalaudek and A. Marghoob. (2016). Validity and Reliability of Dermoscopic Criteria Used to Differentiate Nevi From Melanoma. *JAMA Dermatology*, 152(7), p.798.
7. K. Simonyan, A. Vedaldi and A. Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*. pp.1-2.
8. T. Tan, L. Zhang and C.P. Lim. (2020). Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks. *Knowledge-Based Systems*.
9. T. Tan, L. Zhang and C.P. Lim. (2019). Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models. *Applied Soft Computing*, p.105725.
10. A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau and S. Thrun. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), pp.115-118.

11. G. Huang, Z. Liu, L. Maaten and K. Weinberger. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
12. S. Xie, R. Girshick, P. Dollar, Z. Tu and K. He. (2017). Aggregated Residual Transformations for Deep Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
13. H. Shin, H. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. Summers. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5), pp.1285-1298.
14. T. Tan, L. Zhang, C.P. Lim, B. Fielding, Y. Yu, and E. Anderson. (2019). Evolving Ensemble Models for Image Segmentation Using Enhanced Particle Swarm Optimization, *IEEE Access*.
15. P. Kinghorn, L. Zhang and L. Shao. (2019). A Hierarchical and Regional Deep Learning Architecture for Image Description Generation. *Pattern Recognition Letters*.
16. P. Kinghorn, L. Zhang and L. Shao. (2018). A region-based image caption generator with refined descriptions. *Neurocomputing*. 272 (2018) 416-424.
17. T. Tan, L. Zhang, S.C. Neoh, and C.P. Lim. (2018). Intelligent Skin Cancer Detection Using Enhanced Particle Swarm Optimization. *Knowledge-Based Systems*.
18. D. Pandit, L. Zhang, S. Chattopadhyay, C.P. Lim and C. Liu. (2018). A Scattering and Repulsive Swarm Intelligence Algorithm for Solving Global Optimization Problems. *Knowledge-Based Systems*.
19. K. Mistry, L. Zhang, S.C. Neoh, C.P. Lim and B. Fielding. (2017). A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition. *IEEE Transactions on Cybernetics*. 47(6)1496–1509.
20. L. Zhang, K. Mistry, S.C. Neoh, and C.P. Lim. (2016). Intelligent facial emotion recognition using moth-firefly optimization. *Knowledge-Based Systems*. Volume 111, Nov. 2016, 248–267.
21. W. Srisukkham, L. Zhang, S.C. Neoh, S. Todryk and C.P. Lim. (2017) Intelligent Leukaemia Diagnosis with Bare-Bones PSO based Feature Optimization. *Applied Soft Computing*, 56. pp. 405-419.
22. H. Xie, L. Zhang, C.P. Lim, Y. Yu, C. Liu, H. Liu and J. Walters. (2019). Improving K-means clustering with enhanced Firefly Algorithms. *Applied Soft Computing*, 84, p.105763.
23. B. Fielding and L. Zhang. (2018). Evolving Image Classification Architectures with Enhanced Particle Swarm Optimisation. *IEEE Access*.