International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

Classification of Lesion Images Using Transfer Learning Approach

[1]U. M. Prakash,[2]Astha Gupta,[3]Sumanta Banerjee,[4]K.R. Gokul Anand

ABSTRACT—*Caused due to the uninhibited increase of abnormal skin cells, skin cancer is a result of unrepaired DNA damage to skin cells which in turn, leads to mutations, or genetic defects. These defects cause rapid multiplication of skin cells and they eventually formulate malignant tumours. Although skin cancer is one of the most lethal types of cancer, a fast diagnosis can lead to a very high chance of survival. The diagnosis of skin cancer is primarily performed using visual methods, usually an initial clinical screening. Dermoscopic analysis, a biopsy and histopathological examination consist of the conventional methods that follow the clinical screening. An automated classification of skin disease using images is a difficult job due to the microscopic variability of the appearance of different classes of skin lesions. Over the last few years, convolutional neural networks (CNN) have been increasingly employed for the task of automatic and semi-automatic image classification. Through this work, we aim to use a transfer learning-based deep learning approach to detect cancerous lesions in dermatological images. The process would involve pre-processing and data augmentation tasks being performed on the lesion images. Following this, a pre-trained transfer learning model would be fine-tuned and used for feature-selection and a classifier model would be added on top of it to classify the images of skin lesions into 'malignant' and 'benign' categories. The model was tested using standard evaluation metrics to evaluate its effectiveness. Our results show that a transfer learning approach can work as an effective screening tool to detect cancerous lesions.*

Keywords—Deep learning, skin lesions classification, convolutional neural networks, skin cancer, transfer learning

I. INTRODUCTION

Skin cancer is regarded as one of the most fatal of all cancer types. Primarily, two types of skin cancer are common- melanoma and carcinoma. Out-of-control growth of melanocytes leads to the development of melanoma cancer. Melanoma can grow and eventually mutate to the rest of the body and therefore, considered to be extremely aggressive.

Early diagnosis is of the essence in the treatment of skin cancer. If the tumour is caught early, it leads to a very high chance of complete and full recovery. However, more often than not, skin cancer is detected when cancer has spread leading to a difficult and long treatment process. Conventional methods to detect both melanoma and carcinoma consist of being treated manually with a physical exam, followed by a biopsy or imaging tests if required. These tests rely on human resources to be able to diagnose the skin lesions which makes the whole process time –consuming. Hence, automatic systems, which can classify skin lesions into their various malignant

[1]*CSE,SRM-IST, Chennai, India, prakashm3@srmist.edu.in.*

[2] *CSE, SRM-IST, Chennai, India, ag3843@srmist.edu.in.*

[3] *CSE,SRM-IST, Chennai, India, palashsumanta@gmail.com.*

[4] *CSE, SRM-IST Chennai, India, gokulnand.r@srmist.edu.in.*

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

and benign types, can prove to be very useful by working as initial screening tools or as supplementary safety-net expert systems.

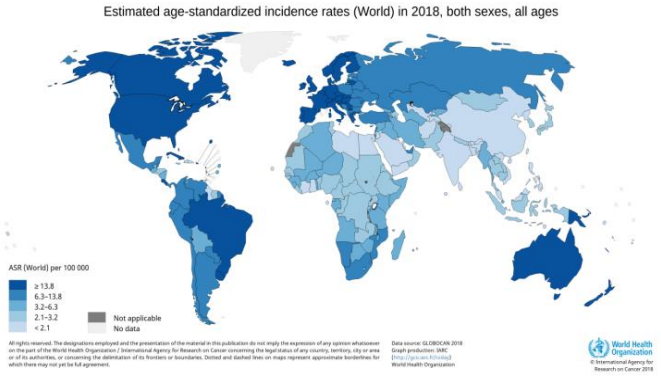

Figure 1: Incidence rates of skin cancer worldwide (2018) [1]

The first set of methods utilized for the automatic detection of skin cancer used many machine learning algorithms to perform classification tasks. However, in recent times, researchers have started utilizing deep learning as a method to perform image segmentation and classification of skin lesions in order to detect the presence of cancer.

Using deep learning, a computer model is trained in order for it to perform the task of classifying data into different categories directly from datasets consisting of images, sounds, or text. Deep learning-based models have proven to achieve a considerably high level of accuracy, sometimes, even leaving behind human performance. Owing to the high computational ability of neural networks to learn a multitude of features from the given datasets, they are being used in a wide variety of applications.

II. STATE OF THE ART (LITERATURE SURVEY)

Most deep learning-based classification tasks perform a multi-classification method in order to categorize the different types of skin lesions. However, the method proposed by Catarina Barata and Jorge S. Marques [2] uses the hierarchical approach used by dermatologists to perform the classification. The skin lesions are first classified as malignant and benign, and then the benign images are classified further. The data is pre-processed and augmented before being fed to the model. Three models are used-one multi-class and two hierarchical to perform comparative studies. The results indicate that the hierarchical approach combined with lesion segmentation led to higher classification accuracy.

SertanKaymak et al [3] propose a two-step classification method for the skin lesion images present in the ISIC 2018 Challenge dataset. Their approach is based on using three different models. The first one is used to distinguish between non-melanocytic from melanocytic classes. The next model classifies between melanocytic nevus and melanoma and lastly, benign classes and non-melanocytic malignant classes are modelled. All three models make use of the AlexNet architecture to showcase accuracies of 78%, 84% and 58% respectively.

The melanoma classification technique proposed by Le Thu Thao and Nguyen Hone Quang [4] uses a convolutional-deconvolutional network which uses both convolutional and de-convolutional layers for the image

segmentation task. For the classification, two different architectures are used- the first model uses a CNN architecture being trained on the augmented image dataset and the second one uses a transfer learning-based VGG model pre-trained on the ImageNet data and then fine-tuned to classify skin lesions. Two classification tasks are performed by each model. Firstly, images are classified as malignant and benign and then, they are classified as melanocytic and non-melanocytic. Better performance is achieved by the transfer learning approach for both classification tasks.

The importance of data augmentation as a pre-processing step has been showcased by EnesAyan and Halil Murat Unver though their research publication [5]. This research method aims to prove the supremacy of models which use data augmentation as a pre-processing step. The researchers use the same convolutional neural network architecture for two different datasets. The first one is the original dataset of the ISIC 2017 challenge and the second is an augmented dataset obtained by performing rotations, shifting, zooming and flipping the images of the original data. Using augmentation, it becomes possible to overcome the problem of having an unbalanced dataset and helps to avoid over-fitting. The results obtained clearly indicate that augmentation leads to better classification accuracy.

This research work proposed by Jeremy Kawahara et al [6] uses a linear classifier which was uses features extracted from a convolutional neural network that is pre-trained, to distinguish between ten categories of skin lesions. Further, unlike most of the conventional works, the proposed approach does not need segmentation or any difficult preprocessing. Per-image-mean subtracted images are used for multi-scale feature extraction. The task of pooling is performed throughout the augmented feature space to classify the images. The technique of transfer learning is applied to pre-train the neural networks which are then fine-tuned for the skin lesion categorization. The results achieved showcase a higher accuracy level than the existing cutting-edge methods. For the entire multi-class dataset containing 1300 images captured using a non-dermoscopic camera, the approach attains an accuracy of 81.8%.

III. PROPOSED WORK

The proposed methodology makes use of the technique of transfer learning to build a classifier that could distinguish between malignant and benign categories of skin lesion images.

A. Pre-Processing

The images in the dataset were first converted into a NumPy format for the purpose of training. For this, a data loader function was created and all the images were converted from an RGB format of (0,255) to binary format of (0,1). Following this, the labels were created for the image classes with 0 set as benign and 1 set as malignant. To introduce randomness, a shuffle was performed on the data and then the train, test and validation split was performed. The ratio for the split was 60:20:20.

B. Data Augmentation

In order to overcome the unavailability of a larger dataset, we have used data augmentation techniques like rotation, horizontal flipping, vertical flipping and zooming to enhance the data while training the model. It has

been shown that data augmentation has the ability to increase the performance of deep learning models where a large dataset is not available [5].

C. Architecture

One of the biggest issues, that affects the performance of a deep learning model is the unavailability of a large training dataset. In order to overcome this problem, the relatively novel technique of transfer learning can be employed.

The process of transfer learning involves training a base network on a much larger base dataset and task. The features learnt by the base network are then leveraged for the target network which trains on the target dataset for the target task. [7]

IV. IMPLEMENTATION

For the purpose of this study, we make use of a subset of the publicly available International Skin Image Collaboration (ISIC) dataset [8]. Our dataset consists of a total of 3297 images- 1800 images in the benign category and 1497 images in the malignant category. The dataset has been split into Train, Test and Validation categories using a 60:20:20 ratio. Therefore, the training data has 1080 benign images and 900 malignant images. Similarly, the validation dataset has 297 and 360 malignant and benign images respectively. Finally, the test dataset consists of 300 and 360 malignant and benign images respectively.

In the implementation of our proposed methodology, we have made use of a ResNet50 [9] model that has been trained previously on the ImageNet dataset [10] as the base model. This is used as a feature extractor for our model. In order to fine-tune the ResNet50 network, the fully connected topmost layers were removed. These were replaced by one global max-pooling layer, one fully connected dense layer of 512 units, one dropout layer for regularization and a sigmoid activation layer.

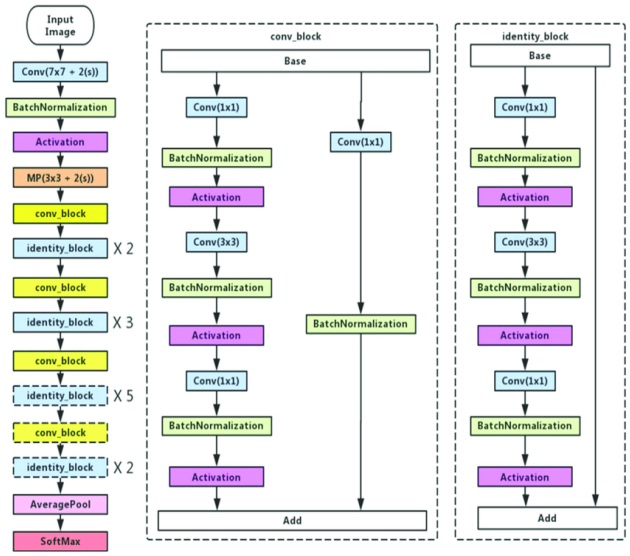

Figure 2: ResNet50 Architecture

Initially, all layers in the ResNet50 network were frozen, and feature extraction was performed for the new fully connected layers so that the weights for these layers would not be completely random and the gradient wouldnot be too large when we start fine-tuning. After 3 epochs of feature extraction, the final convolutional block of ResNet50was unfrozen and the fine-tuning process was started for the model for 30 epochs.

Adam optimizer was selected for the model and the learning rate was initialized at 0.001. Learning rate decay was also used so that the learning rate would be halved if the validation accuracy reached a plateau for 3 epochs. Binary cross-entropy was used to calculate the loss.

The reasoning behind training the model for a relatively low number of epochs was the fact that using ResNet50 Architecture as feature extractor there was going to be overfitting if the model was trained too much

V. RESULTS DISCUSSION

To evaluate the effectiveness of the transfer learning approach, the performance of our proposed model was compared to that of a baseline model that was trained-from-scratch on the same dataset with the same values for all the hyper-parameters.

The fine-tuned ResNet50 model proposed, obtained a validation accuracy of 83.7% after training on 30 epochs. The AUROC curve was 0.928 was obtained for the test dataset. On the other hand, the model trained from scratch reached a validation accuracy of 77.4% after 30 epochs and the AUROC curve equal to 0.916 was obtained.

TABLE 1: COMPARISON OF MODELS USED

Name	**Performance Metrics**		
	Validation Accuracy	***Test Accuracy***	***AUROC***
Baseline Model	77.4%	78.7%	0.916
Fine-tuned VGG 16 Model	82.1%	82.8%	0.921
Fine-tuned ResNet50 Model	85.8%	87.7%	0.957

The metrics clearly showcase that transfer learning performed better than a model trained from scratch. This finding indicates the supremacy of transfer learning over other available methods in dealing with a problem with a small dataset.

To determine the best network to act as the feature selector, ResNet50, VGG-16 Networks were trained as feature selectors and fine-tuned to obtain the best possible accuracy. In the case of Binary Classification, A is considered to be a very good metric to evaluate the model. The proposed model (Fine-Tuned ResNet 50) had the highest AUROC of 0.928 in comparison to the AUROC of other models.

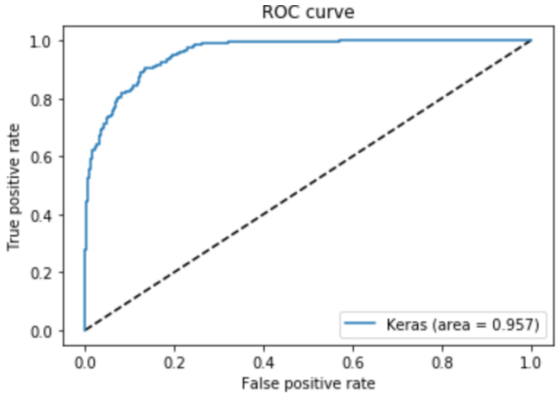

Figure 4: ROC Curve for ResNet50 as Feature Extractor

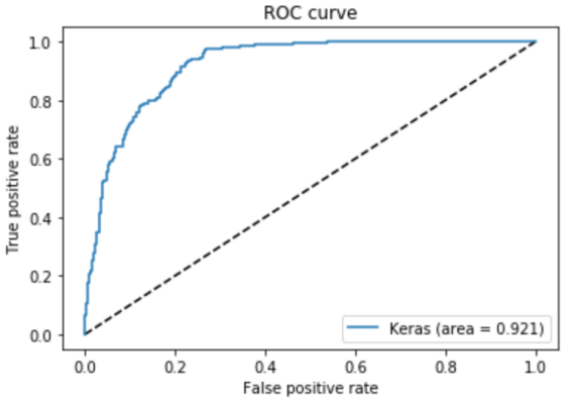

Figure 5: ROC Curve for VGG16 as Feature Extractor

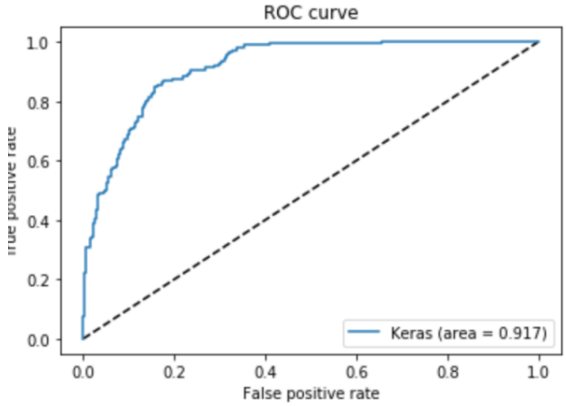

Figure 6: ROC Curve for Baseline Model

VI. CONCLUSION

The results achieved in this study are consistent with the other works that have been done with deep convolutional neural networks. The results obtained by our proposed model are in accordance with the current res and display that better results are obtained by models that are deeper.

The results verify the proposal that features learned by pre-trained models help to learn features for a completely different domain dataset, in our case the skin lesions dermoscopic images dataset. The use of transfer

learning models for feature extraction is a high performance-yielding technique for medical image analysis. Additionally, fine-tuning and data augmentation are important parameters for a better model.

The current technology can easily be used as an essential safety net feature in the field of healthcare. Future work in this field can lead to the development of affordable and accessible technologies that could help people even in the most remote parts of the world where they have limited access to medical facilities.

Deep convolutional neural networks-based models can help as an initial screening tool and can be utilized by clinicians to prevent the possibility of a false positive or a false negative result and can improve the overall quality of the results.

REFERENCES

1. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018, Available at dietandcancerreport.org .
2. C. Barata and J. S. Marques, "Deep Learning For Skin Cancer Diagnosis With Hierarchical Architectures," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 2019, pp. 841-845.doi: 10.1109/ISBI.2019.8759561
3. S. Kaymak, P. Esmaili and A. Serener, "Deep Learning for Two-Step Classification of Malignant Pigmented Skin Lesions," *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, Belgrade, 2018, pp. 1-6. doi: 10.1109/NEUREL.2018.8587019
4. L. T. Thao and N. H. Quang, "Automatic skin lesion analysis towards melanoma detection," *2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES)*, Hanoi, 2017, pp. 106-111.doi: 10.1109/IESYS.2017.8233570
5. E. Ayan and H. M. Ünver, "Data augmentation importance for classification of skin lesions via deep learning," *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Istanbul, 2018, pp. 1-4.doi: 10.1109/EBBT.2018.8391469
6. J. Kawahara, A. BenTaieb and G. Hamarneh, "Deep features to classify skin lesions," *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Prague, 2016, pp. 1397-1400.doi: 10.1109/ISBI.2016.7493528
7. Tschandl P., Rosendahl C. & Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)
8. Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern: “Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)”, 2017; arXiv:1710.05006.
9. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.doi: 10.1109/CVPR.2016.90

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

10. Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy: "BCN20000: Dermoscopic Lesions in the Wild", 2019; arXiv:1908.02288