JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Skin Lesion Classification using Deep Learning

Isha Mulajkar[1], Mr. R.L. Nandargi[2], Vasanti Mule[3], Bhatesh Patil[4]
Department of E&TC, SKNCOE, SPPU, Pune

Abstract- Skin cancer is one of the major types of cancers with an in-creasing incidence over the past decades. Accurately diagnosing skin lesions to discriminate between benign and malignant skin lesions is crucial to ensure appropriate patient treatment. While there are many computerised methods for skin lesion classification, convolutional neural networks (CNNs) have been shown to be superior over classical methods. In this work, we propose a fully automatic computerised method for skin lesion classification which employs optimised deep features from a number of well-established CNNs and from different abstraction levels. It trains the model and gives the accuracy of 96%. The proposed work helps early identification of skin disease and can be validated and treated appropriately by medical practitioners.

Keywords: Medical imaging, Deep Learning, CNN, lesion classification.

I. INTRODUCTION

Skin cancer is one of the most common cancer type worldwide. Skin cancer is the uncontrolled growth of abnormal skin cells. It occurs when DNA damage to skin cells (most often caused by ultraviolet radiation from sunshine or tanning beds) triggers mutations, or genetic defects, that lead the skin cells to multiply rapidly and form malignant tumors. Except in rare instances, most skin cancers arise from DNA mutations induced by ultraviolet light affecting cells of the epidermis. If detected early it can be cured through a simple excision while diagnosis at later stages is associated with a greater risk of death. In the proposed work, Convolutional Neural Networks (CNN) are used to accurately classify pigmented skin lesions in dermoscopic images to detect malignant skin lesions as early as feasible.

The HAM10000 dataset which consists of 10015 images has been used in the proposed work. The HAM10000 dataset is a vast collection of dermoscopic images of pigmented skin lesions which are very common from multiple sources. Each image and patient had seven features, namely, age of the patient, sex of the patient, lesion id which is a unique identifier for a particular type of lesion, image id which is a unique identification number for an image, dx type for technical validation, Skin lesion's geographical location, and a diagnostic skin lesion category which is a classification of skin lesions that can be used to diagnose a condition.

II. DATASET

The HAM10000 "Human against Machine with 10,000 training images" dataset is one of the largest datasets, which contains 10,015 total dermoscopy images, used for detecting pigmented skin lesions, that are publicly accessible through the ISIC repository. This dataset is grouped into seven different classes with a number of images, i.e., melanocytic nevus, actinic keratosis, dermatofibroma, basal cell carcinoma, vascular lesion, benign keratosis, and melanoma. The dataset contains 54% male and 45% female skin lesion images. It is a complex dataset with many skin lesion images having low inter-class and high intra-class variation issues, therefore the classification of these skin classes is not an easy task, and the chances of a high misclassification rate are significant.

III. IMPLEMENTATION

- EDA –

Exploratory Data Analysis is an approach to analyse the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

- Pre Processing -

A common first step in the deep learning workflow to prepare raw data in a format that the network can accept.

- Sampling –

Sampling is a method that allows us to get information about the population based on the statistics from a subset of the population (sample), without having to investigate every individual.

- Sequential layers-

This layer is the highest-level building block in deep learning. A layer is a container that usually receives weighted input, transforms it with a set of mostly non-linear functions and then passes these values as output to the next layer.

- Train –

Deep learning training is when a deep neural network (DNN) "learns" how to analyse a predetermined set of data and make predictions about what it means. It involves a lot of trial and error until the network is able to accurately draw conclusions based on the desired outcome.

- Validation–

The "validation data" is a set of data held separate from your training data. It's used during the training process to see how the network would perform on data it hasn't been directly trained on.

- Prediction-

"Prediction" refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome.

IV. RESULT

The CNN were trained and tested on a Windows 10 computer. The model was created with Python 3.7.9, with Keras, Imutils, and cv2Numpy as dependencies. The following are the graphs that show variation in model accuracy [fig1] and loss [fig2] respectively. It compares accuracy and loss of training and validation data.

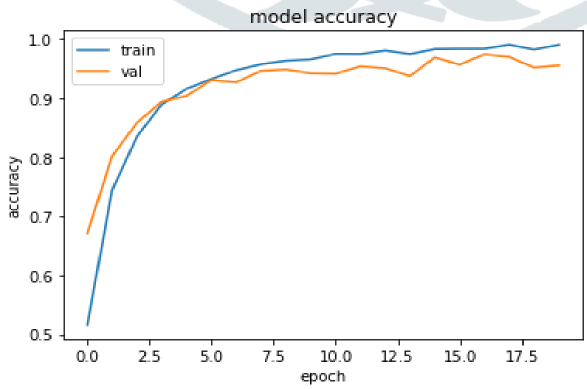

Fig1. Variation in accuracy of train and val

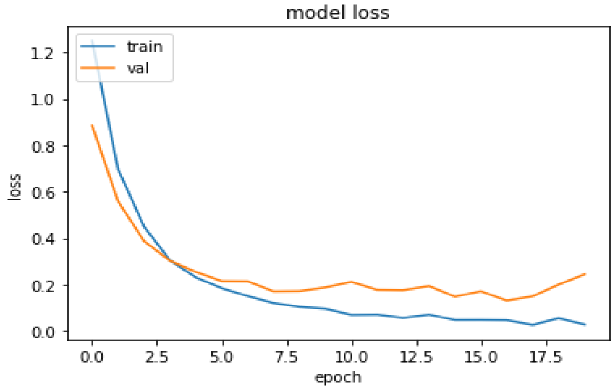

Fig2. Variation in loss of train and val

V. CONCLUSION

Deep learning has become a mature technology for the classification of image content and can achieve similar or superior accuracy as human experts in the classification of skin lesions. The use of deep learning applications that automatically evaluate clinical and dermoscopic images and classify skin lesions offer great potential for improving and implementing prevention and screening measures and increasing their efficiency. The proposed work has applied CNN techniques to classify the skin lesion images. The experiments were conducted on the HAM10000 dataset. The customized CNN techniques were evaluated after the experiments based on Accuracy. The results show that the customized CNN has obtained an accuracy of 96%. This suggests that the proposed CNN has a better classification performance for the HAM10000 data set.

VI. REFERENCES

1. Almeida M.A.M., Santos I.A.X. Classification Models for Skin Tumor Detection Using Texture Analysis in Medical Images. *J. Imaging.* 2020;6:51. doi: 10.3390/jimaging6060051. [CrossRef] [Google Scholar]

2. Ki V., Rotstein C. Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care. The Canadian journal of infectious diseases & medical microbiology. *Can. J. Infect. Dis. Med. Microbiol.* 2008;19:173–184. [PMC free article] [PubMed] [Google Scholar]

3. K.SCOTT.MADER https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000/discussion?datasetId=54339 [Kaggle]

4. Tschandl P., Rosendahl C., Kittler H.

The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. [Google Scholar]

5. Harangi B.Skin lesion classification with ensembles of deep convolutional neural networks. [Google Scholar]

6. Zhang N., Cai Y.-X., Wang Y.-Y., Tian Y.-T., Wang X.-L., Badami B. Skin cancer diagnosis based on optimized convolutional neural network. [Google Scholar]