9/8/2019

**Skin Disease identification by building the CNN model in keras TensorFlow**

**Skin Disease identification by building the CNN model in keras TensorFlow**

**Abstract:**

Skin diseases are a major global health problem associated with high number of people. With the passage of time as classification techniques become popular, different disease prediction can be done. The traditional diagnosis technique aims at improving the quality of existing diagnostic systems by proposing advanced feature extraction and classification methods. In this research paper we used dataset of skin diseases. These images are subjected to pre-processing the images using machine learning techniques.

1. **Introduction:**

Skin is the largest organ is our body. Skin is that, which helps to protects the body from infection, heat, injury, and any type of damage which is caused by ultraviolet (UV) radiation. The fast development of computer technology in present decades, the use of data mining technology plays a crucial role in the analysis of skin diseases. Researchers are constantly developing various prediction methods, but the largest researchers use only a few classification algorithms instead of ensemble methods.

Machine learning is modified form of Artificial intelligence. There are many types of machine learning techniques. To classify the available data set we used different classification techniques. Machine learning techniques and algorithms are very productive. ML techniques used for doing the data analysis and diagnose the disease. Different ML algorithms work differently for different diseases [2].

According to the previous related work that has been done on the Skin disease, many researches find the algorithm that gives maximum accuracy. According to researchers, different algorithm showing different level of accuracy.

In this research paper, we classified the skin diseases into seven classes and build CNN model. By training and testing the existing data set we found results that how many images that we used in our dataset are related to which class [3].

The motive of this paper is to find the accuracy percentage in prediction of skin disease using different data mining techniques.

1. **Literature Review:**
2. **Automated Skin Disease Identification using Deep Learning Algorithm**

Patnaik S. K et all presented [1] a new approach of deep learning through which various kind of skin diseases can be predicted. Mainly three systems V3, Inception Resnet V2 and Mobile Net used for image recognition system. In this paper [1] Author used logistic regression and random forest technique for training and testing purpose. According to results by using these techniques accuracy level is increased from 75% to 88%.

1. **A survey on applying machine learning techniques for management of diseases**

Enas M.F. El Houby presented a paper [2] in which they surveyed the Machine learning techniques for finding and preventing the most common diseases such as cancer, hepatitis and skin disease. Author used artificial neural network, K-nearest neighbor, decision tree and associative classification for determining the accuracy of these three diseases and he concluded that the achieved accuracy of numerous applications ranged from 70% to 100 % according to the disease, available data sets and used techniques.

1. **Prediction of Skin Diseases using Data Mining Techniques**

S. Reena Parvin1 et all presented a paper [3] in which they described about the machine learning techniques that they used for predicting the skin diseases. The diagnosis system involves two stages of process such as training and testing, the Features values of the training data set are compared to the testing data set of each type. Authors used three algorithms Multi-SVM classifier, K-NN and Naïve Bayesian classifier and according to results the overall accuracy of using Multi-SVM classifier is 97% to 98%.

**3. Methodology:**

In order to bring about execution and experiments we used the dataset of ISIC2018. This data set is about the images of different skin diseases. Dataset is about the images of the different skin diseases. We firstly Load our dataset into the google drive and used Google Collaboratory for processing the images. Then divide our data into training and testing. we split folders of our data in the ratio of 70 and 30. After splitting we design CNN model and train and test our data accordingly.

Google Collaboratory provides the great tool for improving your coding skills, but it also allows absolutely anyone to develop deep learning applications using popular libraries such as Py Torch, TensorFlow, Keras, and OpenCV. Collaboratory provides GPU and it's totally free. Everyone can use the resources of google Collaboratory.

**3.1 Performance Parameters:**

Performance parameters that are used in our paper are accuracy, correctly classifies instances, Incorrectly classified instances. When we preprocess our dataset on Collab these classifies on our dataset different performance parameters showed different accuracy percentages.

* 1. **Classification:**

Classification is a data mining function that allocates items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Classification divides data samples into target classes. The classification technique predicts the target class for each data points.

* 1. **Dense Layers:**

A dense layer is just a regular layer of neurons in a neural network. Each neuron receives input from all the neurons in the previous layer, thus densely connected. The layer has a weight matrix W, a bias vector b, and the activations of previous layer a.

**3.4 Pooling**:  
pooling is a concept in deep learning visual object recognition that goes hand-in-hand with convolution. The idea is that a convolution (or a local **neural network** feature detector) maps a region of an image to a feature map. For example, a 5x5 array of pixels could be mapped to oriented edge features.

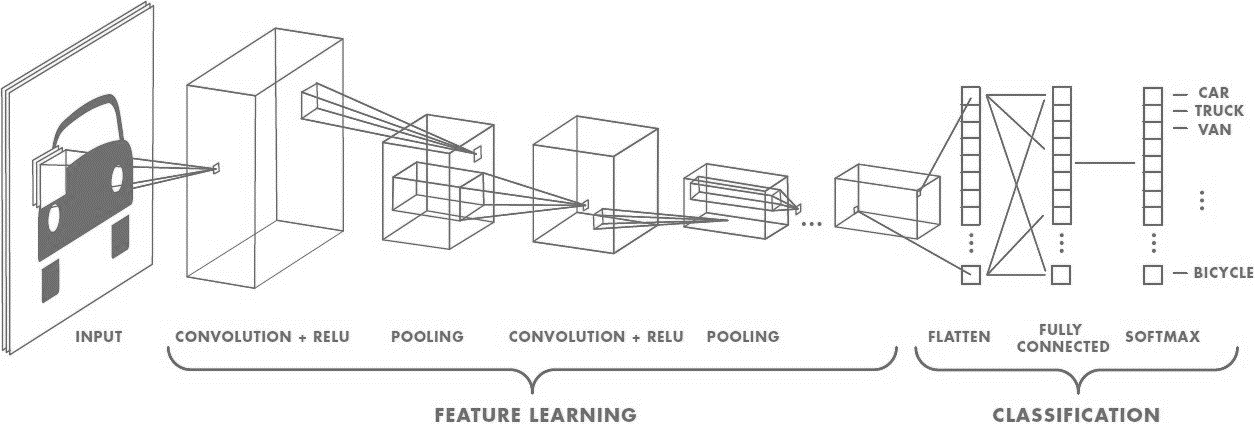


Figure 1 Feature learning and classification

Figure 1 explains about the how we train our dataset with the convolutional model. In convolutional Modeling the Relational Modeling where a mapping from MER to relations by means of rules of mapping is carried out. We build our model using ConV2 and the functions that are used here are Relu and SoftMax activation function. After creating the conventional model, we add dense layers on top. Figure 2 in experiments and result section is displaying the complete architecture of our model. Then we compile and train our model.

1. **Experiments and Results:**

We basically analyzed and predict data to recognize seven classes of skin disease.

**Patient Dataset:**

We used ISIC2018 dataset. This dataset consists of images of 7 different classes of skin disease.

The seven classes of skin disease are:

|  |  |
| --- | --- |
| 0 | Melanoma |
| 1 | Melanocytic Nevus |
| 2 | Basel cell carcinoma |
| 3 | Actinic keratosis |
| 4 | Benign Keratosis |
| 5 | Dermatofibroma |
| 6 | Vascular Icsion |

**Table 1 Skin Disease classes**

When we build CNN model and add dense layers on top, we found out the total parameters, trainable parameters and non-trainable parameters. By using sequential model out of 7,950,145

parameters, 7,949,953 parameters are trainable and only 190 parameters are non-trainable.

**Results:**

Table 2 represents the total number of parameters, Trainable parameters and non-trainable parameters.

|  |  |
| --- | --- |
| Total Parameters | 7,950,145 |
| Trainable Parameters | 7,949,953 |
| Non-trainable parameters | 190 |

**Table 2 Model summary**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv2d\_1 (Conv2D) (None, 94, 94, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

batch\_normalization\_1 (Batch (None, 94, 94, 32) 128

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling2d\_1 (MaxPooling2 (None, 47, 47, 32) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv2d\_2 (Conv2D) (None, 45, 45, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

batch\_normalization\_2 (Batch (None, 45, 45, 64) 256

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling2d\_2 (MaxPooling2 (None, 22, 22, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_1 (Flatten) (None, 30976) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 256) 7930112

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dropout\_1 (Dropout) (None, 256) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 1) 257

=================================================================

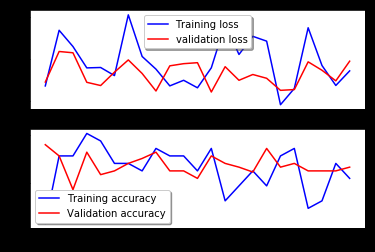
Total params: 7,950,145

Trainable params: 7,949,953

Non-trainable params: 190

**Figure 2**

The graph shown below in figure 3 showing the results of training accuracy and validation accuracy.



**Figure 3**

1. **Conclusion:**

In this system we proposed a technique through which we classify the skin diseases according to the domain. Out of 7,950,145 parameters 7,949,953 parameters are trainable only 190 parameters are non-trainable so accuracy achieved through this is not much higher. Although the accuracy achieved through this CNN model is not much higher accuracy but in future by doing changes in parameters, we can achieve much higher accuracy than the existing one.

**References:**

[1] Patnaik, S. K., Sidhu, M. S., Gehlot, Y., Sharma, B., & Muthu, P. (2018). Automated Skin Disease Identification using Deep Learning Algorithm. *Biomedical & Pharmacology Journal*, *11*(3), 1429.

[2] El Houby, E. M. (2018). A survey on applying machine learning techniques for management of diseases. *Journal of Applied Biomedicine*, *16*(3), 165-174.

[3] Pravin, S. R., & Jafar, O. A. M. (2017). Prediction of skin disease using data mining techniques. *IJARCCE*, *6*(7), 313-318.

[4] El Houby, E. M. (2018). A survey on applying machine learning techniques for management of diseases. Journal of Applied Biomedicine, 16(3), 165-174.

[5] Vijiyarani, S., & Sudha, S. (2013). Disease prediction in data mining technique–a survey. *International Journal of Computer Applications & Information Technology*, *2*(1), 17-21.