Detection of Skin Cancer using Deep CNN

Hemalatha N, Nausheeda B.S, Athul K.P, Navaneeth

Abstract: **D**evelopment of abnormal cells are the cause of skin cancer that have the ability to attack or spread to various parts of the body. The skin cancer signs may include mole that has varied in size, shape, color, and may also haveno –uniform edges, might be having multiple colours, and would itch orevn bleed in some cases. The exposure to the UV-rays from the sun is considered to be accountable for more than 90% of the Skin Cancer cases which are recorded.In this paper, the development of a classificiation system for skin cancer, is discussed, using Convolutional Neural Network which would help in classifying the cancer usingTensorFlow and Keras as Malignantor Benign. The collected images from the data set are fed into the system and it is processed to classify the skin cancer. After the implementation the accuracy of the Convolutional 2-D layer system developed is found to be 78%.

Keywords:Skin Cancer, convolutional neural network, keras, TensorFlow, Benign, Malignant.

I. INTRODUCTION

The out-of-control growth of the abnormal cells in the skin layers which are the outermostin the body, is the condition which is caused by unrepaired DNA damage that triggers mutation, and this is knwon as Skin Cancer. It begins when cells which are healthy changes and the above said happens leading to it, forming a tumor. A tumor which is so formed can be not cancerous, in other words benign or it could be cancerous, a cancerous tumor is called as malignant, and this can grow and spread to other parts of the body, whereas the benign tumor would continue growing in some cases, but not spread.

According to survey conducted byCancer Council in 2015, more than 921,000 people are infected with cancer and the rate of death was around 21,513 in the world. The skin cancer condition called malignant melanoma is curable if diagnosed in early stages and treated. Therefore, early diagnosing and detection is a crucial issue faced by the patients. However,only an experienced doctor may be able todifferentiate skin cancer from other skin diseases in the early stages. Thus, a skin cancer detection system which is computer-based is necessary,which would aid in givingsuggestion for non- specialized user[1]. Developing such a diagnosis system would be helping the patients to an extent. In this paper, an automated skin cancer detector for the classification of cancer images and thus differentiating the cancer into either benign or malignant melanoma is discussed using Convolutional Neural Networking. The images which are used for this purpose are digital photos which are collected fromwww.kaggle.com.

N Hemalatha, Associate Professor, Department of IT and Bioinformatics, AIMIT, St Aloysius College, Mangalore, India.

Nausheeda B S, Assistant Professor, Department of Information Technology, Aloysius Institute of Management and Information Technology, ST. Aloysius College, Mangalore, India.

Athul KP, pursuing, M.Sc.in Big Data, ST. Aloysius Institute of Management and Information Technology, Mangalore, India

Navaneeth, pursuing, MCA, ST. Aloysius Institute of Management and Information Technology, Mangalore, India.

II. LITERATURE REVIEW

Chen *et al*., in their paper, by collaborating the results at the finallayer of the deep convolutional neural network with a Conditional Random Field (CRF) which was fully connected [2], resolved the weak location property of deep networks. Their "Deep Lab" system was able to locate the boundaries of segments at an accuracy that goes beyond previous methods by attaining an IOU accuracy of 71.6% in their test set. They also have shown how these results could be attained efficiently by carefully re-purposing of the network and a applying the wavelet algorithm which would result in allowing neural net to be densely computated and produce responses at 8 fps on a GPU which is highend.

Yunzhu*et al*., in their study, proposed the improvement of the performance of the melanoma classification by using the Data Augmentation along with DeepCNN and exploring how the three classifier performances are influenced by the introduction of image augmentation [3]. Their contributions of the approach includes: (1) cutting-edge performance attained with the Area Under the Curve (89.2%), Average Precision (73.9%), and Pixel Point Volume (82.3%) in melanoma classification task on test dataset. (2) They examined the influence of skin lesion image deformations on performances of classifiers such as NN, SVM and RF. Observations were made that the performances of all three classifiers were improved and influenced by the data augmentation differently. While Nearest Neighbor classifier achieved the finest performance, Support Vector Machine and Random Forest classifier's performance had an inimical effect when theDAUGwas changed from 50 to 100. Eventhough, in this research, AUC achieved aprime performance, theaverage SEN compared to other top midels has to be noted and thus it needs more refinements and improvements.

RahimehRouhia, *et al*., in their paper have proposed few automated methods based on the improvemnet of the region growing and segmentations of CNN which would help inattaining an appropriate templates and an adaptive threshold, in order to preserve the information about the tumor boundary, which would be aiding in the process to diagnose malignancy and benign in mammograms. The first method discussed about the generation of an adaptive threshold in the growing stage of segmentation process by using ANN and intensity features [4]. Thesecond method was the determination of CNN templates by the application if learning algorithms with the aid of genetic algorithm, this method thus resulted in the improvement of the CNN layers and the region growing methods which was in breast tumor segmentations.

To determine the quality of the segmentationmethods which were proposed, after the preprosing of the data, a genetic algorithm was used to pick the correct or appropriate features. Finallydifferent classifierslikeSVM, MLP, random forest, KNN, and naive Bayes were used to perform the tumor classification. It was noted that MLP, in both the proposed methods, performed the best diagnosis.

III. MATERIALS AND METHODS

A. Datasets:- The dataset used in this implementation was Skin cancer: Malignant vs Benign collected from kaggle. com. A total of 3609 images collected are separated into two groups to train for the accuracy of each group. Group A contains 1197 digital malignant images; group B contains 1440 digital benign images and. Another set of 890 images were gathered for using as the testing set. The 890 digital images comprise of 360 digital melanoma images, 300 digital benign images as shown in Fig 1.

(a)   (b)

(c)   (d)
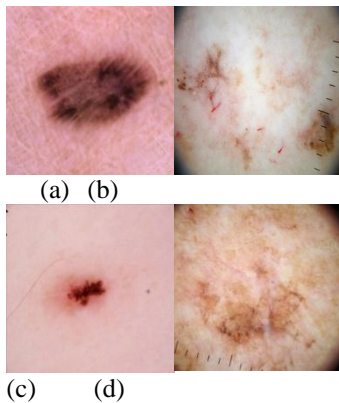
Fig. 1

(a)Benign train set sample (b) Malignant train set sample
(c) Benign test set sample (d) Malignant test set sample
(Courtesy: www.kaggle.com[5])

B. Algorithm: - Convolutional Neural Network (CNN) using Kerasis a Deep Learning algorithm capable of capturing an input image, assigning significance, i.e., learnable weights, to differentiate aspects or the objects which are present in the image and separating them from each other. The pre-processing required, compared to the other classification algorithms, is lower for the ConvNetalgorithm[6,7].

C. Methodology: - Here we explain the algorithm how the dataset is carried on with its calculations.

Step 1: Start

Step2: Packages are imported. (The package used here was keras, NumPy, sklearn, itertools and matplotlib)

Step 3: Initialization of Epochs, images, height, width etc., is done

Step 4: The dataset folder is extracted and converted turning the images into array using image data generator.

Step 5: The number of classes in the array has been extracted and printed in order the know the number of classes.

Step 6: The 2637 images are taken as the train data and 660 images are taken as the test data.

Step 7: A function is created to combine the images and the image labels and it is classified as [0,1] and [1,0].

Step 8: Sequential Model of keras is generated.

Step 9: The datasets is checked with the epochs accordingly and the desired loss, accuracy, value loss and value accuracy at each stage is calculated and the keras model is fitted.

Step 10: Using the model generated the test image is checked with the function created and labeled.

Step 11: Using the model, the test batches are predicted.

Step 12: Using the test labels and the predicted batches, a confusion matrix is generated.

Step 13: A function to plot the confusion matrix is created.

Step 14: In the confusion matrix created, the True label and the Predicted label is plotted.

Step 15: Stop.

IV. RESULT

The experiment was performed in python (jupyter notebook) with the input data which was classified as benign and malignant images. The software features are calculated by mapping the R, G, B components of the objects, then extracting the features of the sheets and comparing them with the corresponding values of the features that are recorded in the library.

Here we have noticed that the accuracy of the model kept changing when we changed the number of epochs, number of convolutional layer and max pooling layers.The results obtained with different Epochs and the number of convolutional network layers value is shown in the Table 1.

From the experiment conducted using skin-cancer-malignant-vs-benign dataset as shown in Table 1, we noticed that with 6 Epochs and Conv2D layers with 32 nodes attained an accuracy of 56.07 , at the same time with 12 Epochs and the same CNN layers model attained an accuracy of 72.83.

When the CNN layers was changed to 64 nodes,model acquired an accuracy of 78.44 with 12 Epochs. Referring to the results attained, we can conclude that with the increase in the Epochs and CNN layers, the accuracy keeps on increasing.

Table 1. Table showing acquired results with different epochs and different CNN layers applied

Epochs	CNN Layers	Step per epochs & validation steps	Accuracy	Loss
6	Conv2D(32, (3, 3)	Step per validation = 6 Validation Steps = 4	56.07	43.93
6	Conv2D(64, (3, 3)	Step per validation =6 Validation Steps =4	64.06	35.94
10	Conv2D(32, (3, 3)	Step per validation =6 Validation Steps =4	70.76	29.24
10	Conv2D(64, (3, 3)	Step per validation =6 Validation Steps =4	72.02	27.98
12	Conv2D(32, (3, 3)	Step per validation =6 Validation Steps =4	72.83	27.17
12	Conv2D(64, (3, 3)	Step per validation =6 Validation Steps =4	74.48	25.52
6	Conv2D(32, (3, 3)	Step per validation = 10 Validation Steps = 8	65.94	34.06
6	Conv2D(64, (3, 3)	Step per validation =10 Validation Steps =8	69.94	30.06
10	Conv2D(32, (3, 3)	Step per validation =10 Validation Steps =8	70.76	29.24
10	Conv2D(64, (3, 3)	Step per validation =10 Validation Steps =8	71.02	28.98
12	Conv2D(32, (3, 3)	Step per validation =10 Validation Steps =8	72.19	27.81
12	Conv2D(64, (3, 3)	Step per validation =10 Validation Steps =8	78.44	21.56

deep learning architecture of Sequential model ofkeras. The accuracy of the model was going on increasing with the number of epochs and the convolutional layers. In this paper with 12 epochs and 64 nodes and with 8 validation steps we got an accuracy of 78.44 , hence we conclude that model increases its accuracy when the numb roof epochs nodes and validation steps increases, in the future studies the accuracy of the model further can be worked upon by increasing the number of epochs , nodes and validation steps , which requires system with hardware requirements as well as the number of training dataset also can be further increased .

REFERENCES

1. Ho Tak Lau, Adel Al-Jumaily, "Early Detection of Skin Cancer: Study Based on Neural Network Classification".
2. L Chen. *et al*, "A. L. 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs"
3. Yunzhu Li, *et al*,"Skin Cancer Detection and Tracking"
4. RahimehRouhia, Mehdi Jafarib, ShohrehKasaeic, PeimanKeshavarziana, "Benign and malignant breast tumors classification based on region growing and CNN segmentation Skin Cancer Detectionand Tracking Using Data Synthesisand Deep Learning".
5. www.kaggle.com
6. Tri-Cong Pham, Chi-Mai Luong Muriel Visani, and Van-Dung Hoangwh "Deep CNN and Data Augmentation for Skin L`esion Classification".
7. Long, J, *et al*,"Fully convolutional networks for semantic segmentation".

AUTHORS PROFILE

N Hemalatha, PhD is an Associate Professor in the Department of IT and Bioinformatics, AIMIT, St Aloysius College, Mangalore, India. She earned her Ph.D in the area of Computational Biology using Machine Learning. Her research interest includes Bioinformatics, Pattern Recognition, Biomedical image processing, Machine learning and Deep learning.

Nausheeda B S is a part time researchscholarin the Department of Computer Science, Mangalore University, India. She is currently working as an Assistant Professor in the Information Technology Department, Aloysius Institute of Management and Information Technology, ST. Aloysius College, Mangalore, India. Her research interests are Pattern Recognition and Biomedical Image Processing

Athul KP, is currently pursuing his M.Sc.in Big Data Analytics from ST. Aloysius Institute of Management and Information Technology, Mangalore, India. His research interest are Big Data, Machine Learning, Deep learning and Biomedical Image Processing

Navaneeth, is currently pursuing his MCA from ST. Aloysius Institute of Management and Information Technology, Mangalore, India. His research interest are Machine Learning, Deep learning and BiomedicalImage Processing.

V. CONCLUSION

In this paper, Benign and malignant disease classification has been performed with the images from skin-cancer-malignant-vs-benign dataset using pretrained