

Skin Cancer Classification Using Deep Learning

Abstract : the aim of this project was to develop a deep learning model for the classification of skin cancer. We utilized a dataset consisting of 10,015 skin lesion images with associated clinical information. Preprocessing techniques were applied to standardize the images, and the dataset was divided into training and test sets. CNN (Convolutional Neural Network) architecture was designed and trained on the training data to classify skin lesions into different cancer types .

Approach :

I used the data set which was provided to me , "<https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000/versions/2>" This dataset was used in this project .

Datapreprocessing : names of the lesions were given in the dataset , so we 7 classes for each cell type

```
# lesion names are given in the description of the dataset
lesion_type_dict = {
    'nv': 'Melanocytic nevi',
    'mel': 'Melanoma',
    'bkl': 'Benign keratosis-like lesions ',
    'bcc': 'Basal cell carcinoma',
    'akiec': 'Actinic keratoses',
    'vasc': 'Vascular lesions',
    'df': 'Dermatofibroma'
}

lesion_ID_dict = {
    'nv': 0,
    'mel': 1,
    'bkl': 2,
    'bcc': 3,
    'akiec': 4,
    'vasc': 5,
    'df': 6
}

lesion_names = ['Melanocytic nevi', 'Melanoma', 'Benign keratosis-like lesions ',
                'Basal cell carcinoma', 'Actinic keratoses', 'Vascular lesions',
                'Dermatofibroma']

lesion_names_short = ['nv', 'mel', 'bkl', 'bcc', 'akiec', 'vasc', 'df']

df_skin['lesion_type'] = df_skin['dx'].map(lesion_type_dict)
df_skin['lesion_ID'] = df_skin['dx'].map(lesion_ID_dict)

print('Total number of images', len(df_skin))
print('The problem is unbalanced, since Melanocytic nevi is much more frequent than other labels')

df_skin['lesion_type'].value_counts()
```

And I also resized the image into 100x100 , and to increase the dimension of dataset I rotated the images and created multiple images using cv2 library .

Then I have created two lists X, Y . X stores the image features and Y stores the Labels corresponding to that image .

I have used 80 % of the data in training of the model and 20% part for testing . class weights were computed to address the unbalanced nature of the dataset .

Model : I have used a CNN model (Convolutional Neural Network) , I have used sequential() which allows me to create a linear stack of layers .

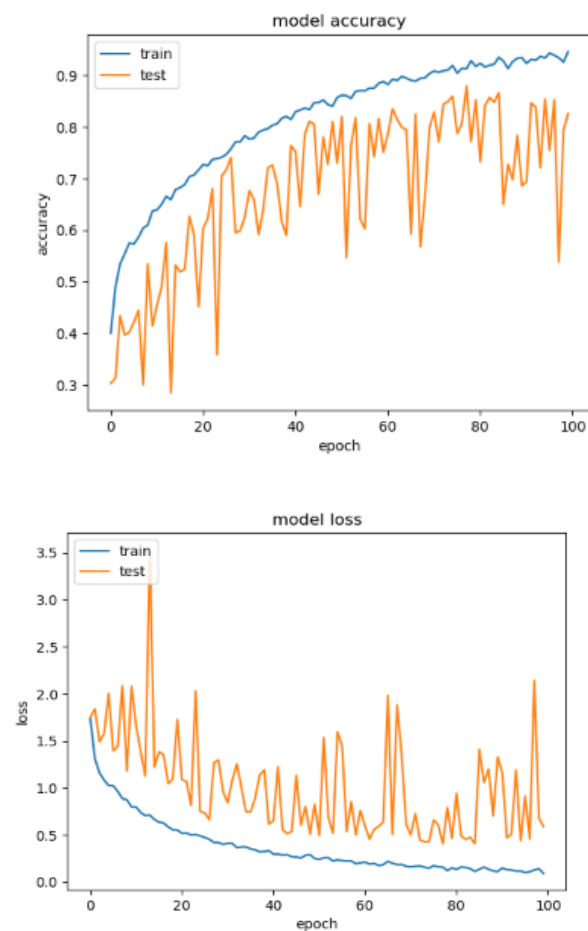
The model starts with a series of convolutional layers , first layer performs convolution with 96 filters , each having 11x11 kernel size and a stride of 4 in both dimensions , the input shape of the image is (100,100,3) . And MaxPool2D , performs max pooling with 3x3 pool size and stride of 2 in both dimensions .

The same pattern of convolution layers , batch normalization and max pooling is applied in 2nd and 3rd layer , In the 4th layer convolution layer has kernel size of 1x1 and stride equal to 1 to preserve the spatial dimensions.

And the 5th convolution layer is similar to the 4th layer and max pooling is applied . Flatten function is used to flatten the output of previous layers into a 1D vector . and the model continues with the two dense layers each of unit 4096 to mitigate overfitting . and then output layer is a dense layer of unit 7 corresponding to the 7 classes in the data set and it uses softmax activation to produce class probabilities .

Training : In training I have used the batch size of 32 and epoch of 100 , and adam optimizer is used with learning rate of 0.0001 and categorical_crossentropy function is used for loss function and accuracy metric is used to evaluate the accuracy of the model .

And we noticed the accuracy of 82.59%



On using test data I got following results :

Total-test-data; 8767, accurately-predicted-data: 7241, wrongly-predicted-data: 1526

And the accuracy for different cells are :

accuracy for Melanocytic nevi = 0.8
accuracy for Melanoma = 0.92
accuracy for Benign keratosis-like lesions = 0.87
accuracy for Basal cell carcinoma = 0.92
accuracy for Actinic keratoses = 0.87
accuracy for Vascular lesions = 0.99
accuracy for Dermatofibroma = 0.96