Skin Cancer Detection using CNN Algorithm

Roopa Sri Paladugu, Anusha Immadisetty, M. Ramesh

Abstract: The project "Disease Prediction Model" focuses on predicting the type of skin cancer. It deals with constructing a Convolutional Neural Network(CNN) sequential model in order to find the type of a skin cancer which takes a huge troll on mankind well-being. Since development of programmed methods increases the accuracy at high scale for identifying the type of skin cancer, we use Convolutional Neural Network, CNN algorithm in order to build our model . For this we make use of a sequential model. The data set that we have considered for this project is collected from NCBI, which is well known as HAM10000 dataset, it consists of massive amounts of information regarding several dermatoscopic images of most trivial pigmented lesions of skin which are collected from different sufferers. Once the dataset is collected, cleaned, it is split into training and testing data sets. We used CNN to build our model and using the training data we trained the model , later using the testing data we tested the model. Once the model is implemented over the testing data, plots are made in order to analyze the relation between the echos and loss function. It is also used to analyse accuracy and echos for both training and testing data.

Keywords : sequential model, keras, skin cancer, cnn.

I. INTRODUCTION

Composed of epidermis, dermis, and subcutaneous tissues, skin is the largest organ of the human body, containing blood vessels, lymphatic vessels, nerves, and muscle groups, that can perspire, understand the outside temperature, and shield the body. overlaying the entire body, the skin can guard a couple of tissues and organs within the body from external breach such as artificial skin harm, chemical damage, adventitious viruses, and people' immune machine. Except, skin also can avoid the lack of lipids together with water within dermis and epidermis in order that pores and skin barrier features can be sustained. in spite of protection and barrier function, pores and skin isn't indestructible in that skin has a tendency to be continuously motivated by way of a selection of external and genetic aspects. Presently, there are three important sorts of skin diseases appearing in the human frame, which includes viral skin diseases, fungal skin illnesses, and allergic pores and skin

disorder. in spite of the reality that these types of skin diseases can be cured at present, those illnesses certainly have added hassle to patients' lifestyles. Nowadays, most of the people's conclusions at the patients' existing signs are drawn specifically primarily based on docs' years of revel in or their own subjective judgments, which may also cause misjudgments and therefore delay the remedy of those. Therefore, it's miles of wonderful theoretical significance and realistic fee to examine how to extract signs and symptoms of divergent pores and skin diseases on the premise of modern-day technology and era. Below this circumstance, powerful and accurate identity of the sorts of skin illnesses can be accomplished to decide remedies in keeping with sufferers' signs.

Deep Learning

Deep learning may be referred to as a deeper part of machine learning methods in synthetic intelligence that develops the concept of networks for studying unsupervised from facts that are unstructured or unlabeled.

CNN

CNN stand for Convolutional Neural network or sometimes can be referred as ConvNet and is one of the most important deep learning methods which takes in an input, assign importance (learnable weights and biases) to numerous aspects/gadgets inside the picture and have the ability to differentiate one from the opposite. The pre-processing required in a ConvNet is an entire lot decreased as compared to other kind algorithms. Even as in the primitive techniques filters are hand-engineered, with sufficient training, ConvNets have the capability to investigate the ones filters/traits.

Keras

Keras is a python package designed to enable fast experimentation with deep neural networks. Keras is one of the most important python libraries which is mainly used for implementing the deep learning algorithms that are capable of running on top of Theano or TensorFlow. It was developed to implement deep learning models as fast and easy as possible for research and development.

Keras Sequential Model

Sequential model in keras is a sequential stack of layer, which is appropriate for a plain stack of layers in which each layer has one tensor input and one tensor output.

Numpy

Numpy is one of the most probably used python libraries which handle the high level mathematical operations for large arrays or matrices.

Revised Manuscript Received on July 10, 2020.
* Correspondence Author
   Roopa Sri Paladugu*, Department of IT, V.R. Siddhartha Engineering College, Vijayawada, India. Email: roopa.paladugu9999@gmail.com
   Anusha Immadisetty, Department of IT, V.R. Siddhartha Engineering College, Vijayawada, India. Email: immadisettyanusha@gmail.com
   Dr. M. Ramesh, Department of IT, V.R. Siddhartha Engineering College, Vijayawada, India. Email: ramesh.welcome@gmail.com

Seaborne

Seaborne is also one of the most important python libraries which is dependent on matplotlib (which is also a python library used for plotting graphs). The seaborne library is used for creating or displaying informative statistical graphics.

Pandas

Pandas is also one of the python libraries, mainly used for manipulating large amounts of numerical tables which consists of data and performs effective mathematical operations.

Sklearn

Scikit-learn , which is formerly referred as scikits.learn and most often known as sklearn, is one of the python libraries that offers many supervised and unsupervised algorithms in machine learning through a consistent interface.

Conv2D

Conv2D refers to a two-dimensional Convolution Layer, which is used to create and develop a kernel that is constructed from several layers of inputs and produces a tensor output.

II. PROBLEM STATEMENT

In the previous research there is an outlook for the design of a classifier to detect the type of cancer which provides better and more reliable results for the patients, so that more patients can be diagnosed and cured. In line with this, skin cancer identification for humans is very useful in encouraging good quality in skin cancer diagnosis. There is a need for automation in recognition of human skin cancer systems so that the abuses during diagnosis and treatment can be minimized.

III. LITERATURE SURVEY

A. Convolutional Neural Networks - Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment[1] :

Based on convolutional neural networks in deep learning, a model has been developed which is used for accurate prediction of MCI-to-AD conversion with an MRI scan data precisely, mainly aiming for increasing the accuracy in prediction. Initially adta procession is done for the MRI images and then from these 2.5 dimensional local patches were being extracted. Then these are used to train the developed CNN model in order to identify features of MCI subjects. After completion of training the model, systematic brain image features were being extracted with FreeSurfer in order to help and develop CNN model. At last, the above features which are extracted are fed into the ML classifier to predict the conversion of AD. The proposed approach makes use of the standardized MRI datasets which are collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project. The above proposed approach achieves an accuracy of 79.9%.

B. Prediction of Heart disease using CNN algorithm[2]:

In this paper, they proposed a heart disease prediction model which is based on convolutional neural networks. This model is used to predict the heart diseases using CNN I.e., Convolutional Neural Network. This model is evaluated on the Cleveland data set. The model is given 13 clinical features obtained from Cleveland data set as input. The training of the proposed model is executed with the modified backpropagation algorithm. The outcome of the model is the presence of heart disease and with different degree of presence. To the best of our insight, none of the existing models is based on CNN. It is the first attempt where the CNN model is applied in this kind of model. When compared to the existing systems, the accuracy of our prediction system is 9% relative higher than them.

C. Summary Of Literature Survey

The above papers used convolutional neural networks . In the first paper they used MIR scanned data in order to predict . In the second paper, they had proposed an CNN model for predicting heart diseases. Here we used the images data sets rather than scanning which is simple when compared to the entire process of scanning for a user.

IV. PROCEDURE

### A. Design Methodology (UML Representation)
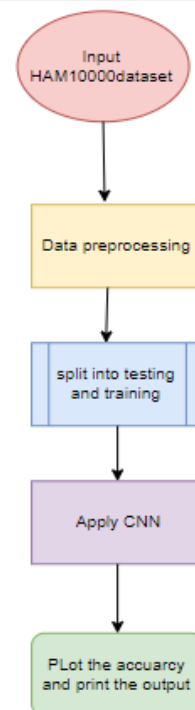
Fig. 1.UML Representation

Step 1: Take HAM100000 dataset[3] as an input.
Step 2: Next step is to clean and pre-process the data and removing the null values.
Step3: Next split them into training and testing data set.
Step 4: Apply CNN algorithm via sequential keras model.
Step 5: Plot the accuracy and print the output.

B. System Architecture

Initially we have used the HAM10000 data set i.e., HAM10000 data set is taken as input. Then the data is cleaned and processed in order to remove any null values, etc. Now the preprocessed data is split into testing and training. Then the CNN algorithm is applied. Now we'll build several layers to construct a sequential model.

Finally a plot is constructed in order to compare the relation between echos and accuracy of training and testing data set.
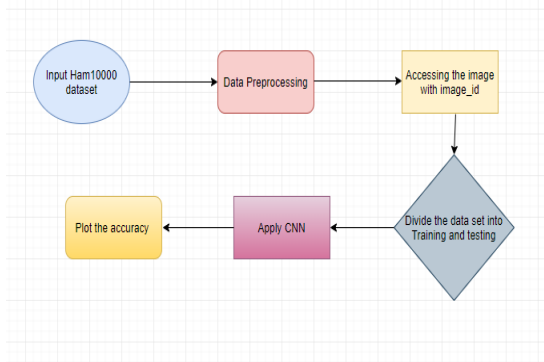
Fig. 2. System Architecture

C. Description of the Data Set[3]

The HAM10000("Human Against Machine with ten thousand training pictures") dataset, a huge assortment of multi-source dermatoscopic pictures of normal pigmented skin injuries has been used as a source data set for prediction. It includes attributes such as image_id, target variable(dx), age, gender. The attribute image_id is to read the image from its path.

D. Software Requirements:

- Any OS
- Python/Anaconda

E. Pseudo code of algorithm:

```
Data: T input images {X1,....XT},T target label 1-in-K vectors {y1,.....yT},
      number j of convolution layers, size Mw x Nw of each convolution layer,
      number of BP epochs R
while epoch r:1->R do
    while training image number t:1->T do
        -Compute the j hidden activation matrices Z1,.....Zj
        -Downsample the j hidden activation matrices Z1,....Zj by a factor of 2
        -Compute the K-D output soft-max activation vector a
        -Generate noise vector n:
        if n^T log(a)>=0 then
         Add noise:yt<-yt+n;
        else
         Do nothing
        end
        -Compute error yt-a
        -compute entropy gradient
        -Update network parameters using gradient descent
    end
end
```

In order to make a model, it is important to experience the accompanying stages.
Those phases are:
- ✓ Model Development
- ✓ Training The Model
- ✓ Testing The Model
- ✓ Model assessment

1) Model Development

Model development relies upon machine mastering algorithms. In this task case, it was neural networks. Such an algorithm seems like:
start with its object: model = Sequential()
then consist of layers with their types: model.add(type_of_layer())
after including a sufficient wide variety of layers the version is compiled.
At this moment Keras is connecting to TensorFlow to build the model.
Writing a loss function and an optimizer algorithm is essential during model compilation.
It looks like: model.compile(loss= 'name of loss function,' optimiser= 'name of opimazer alg')
The loss feature shows the accuracy of the model obtained from each prediction. .
Before model training it is vital to scale records for their similar use.
Training the model can be completed after the model development.

2) Training the model

In this stage, the preparation information and expected yield for this information is utilized for model preparing. It looks along these lines: model.fit(training_data, expected_output). Progress is obvious on the support when the content runs. Toward the end, the report will show the last precision of the model. When the model training is done it is conceivable to complete model testing.

3) Testing the Model

During this stage a second arrangement of information is stacked can be called as testing information. This testing informational index has never been seen by the model and along these lines it's actual precision will be checked. After the model testing is finished, it is comprehended that the model shows the correct outcome.

4) Model Assessment

Finally, the obtained model, after all the above phases have been completed, can be utilized in reality. The name of this stage is model assessment which implies now the model can be utilized to assess new information.

V. RESULTS AND OBSERVATIONS

A. Sequence of steps:

Step 1: To read the input file using pandas
Step 2: Clean the data in the input file. It involves-
Checking of null values and their sum for each column using data.isnull().sum()
Then using data.dropna() method , remove the records with missing or null   values
Step 3: For each row in the data:
   -identify the id of the image and add .jpg to   it
      -join the entire path using path manipulation method os.path.join()
   -Read the images into an array
   -Resize the images and convert their color as per needed
   -Then append the image value with their   respective target variable

47

Step 4: Randomize the image values and take features and label from it in separate variables

Step 5: Normalize the labels which converts non-numerical labels to numerical labels

Step 6: Now split the data into training and testing data

Step 7: Build the model which used for prediction of disease, here it is Convolutional Neural Networks

Step 8: Use the training data for fitting the model and use testing data for prediction of disease, finding accuracy

Step 9: Plot the graphs for accuracy and loss functions in order to analyse differences for both testing and training data.

B. Test Results:
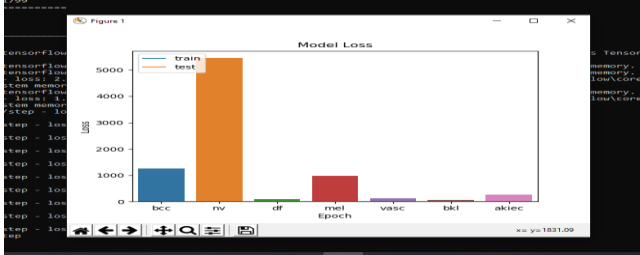
Fig. 3. Test Result1

The above figure represents the labels or target variable for the given data. The level of histogram represents the number of occurrences of that label in the given data.
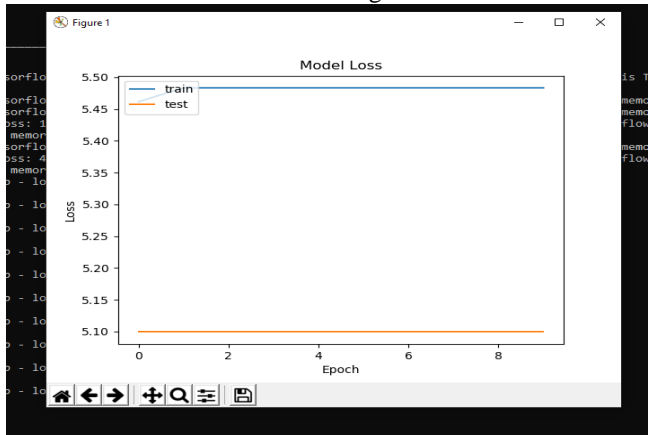
Fig. 4. Epoch VS Loss function

The above figure represents the plot between Epoch and Loss functions for both training and testing data. It implies that even though Epoch value increases loss for both training, testing has remained constant but loss for training data is more compared to testing data.
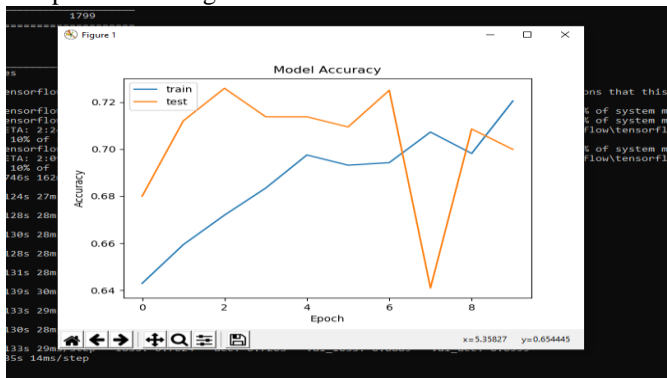
Fig. 5. Model Accuracy

The above figure represents the plot between the Model

accuracy and Epoch. As Epoch value increases accuracy has variations for both training and testing data. Testing data's accuracy is initially low increased later and then decreased, finally it reaches to median value

Fig. 6. Final Accuracy

The above figure represents the loss and accuracy values for each epoch and also the final accuracy of the evaluation of the model. Here we have obtained 70 percent of accuracy for this model.

Fig. 7. List of labels

The features in testing data can be given as input to the model in order to predict the disease label for the given image. The above figure represents the entire list of labels for images in testing data.

OBSERVATIONS FROM THE WORK:

From the whole process which we used for skin disease prediction by CNN, we learned how to make use of a sequential model to attain the required output. We have chosen sequential over functional models because it is simple and can be applicable for our project. However finally the major points that we observed are: high accuracy , simple algorithm.

VI. CONCLUSION

The main purpose of this project is to identify the type of skin cancer. We've built a sequential model with layers Conv2D, MapPooling2D, Dropout which enables us to detect the type of skin cancer accurately with high precision.

REFERENCES

1. Weiming Lin, Min Xioa, Min Du, "Convolutional Neural Networks Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment",sciencedirect.com, august.

2. Shubhanshi Singhal, Harish Kumar, Vishal Passricha, "Prediction of Heart Disease using CNN", American International Journal of Research in Science, Technology and mathematics.
3. Philipp Tschandi, Cliff Rosendahl, Harald Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", NCBI gov. articles.
4. Hagyeong Lee, Jongwoo Song, "Introduction to Convolutional Neural Network using Keras; an understanding from a statistician", Korean Statistical Society, 2019.
5. S. V. G. Reddy, K. Thammi Reddy, V. ValliKumari, "Optimization of Deep Learning using various Optimizers, Loss functions and Dropout, IJRTE.
6. Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gin Do, Kaori Toqashi," Convolutional Neural Networks: An Overview and Application in Radiology", Springer.
7. Sakshilndolia, Anil Kumar, Goswami, S. P. Mishra, Pooj Asopa, " Conceptual Understanding of Convolutional Neural Network - A Deep Learning Approach", ScienceDirect-volume132.
8. Rezoana Bente Arif, Md. Abu Bakr Siddique, Mohammad Mahmudur Rahman Khan, Mahjabin Rahma Oishe, " Study and Observation of the variations of accuracies for handwritten digits recognition with various hidden layers and epochs using convolutional neural network", IEEE Xplore.
9. *Introduction to convolutional neural network*, MIKE IRVINE https://sempwn.github.io/blog/2017/04/06/conv_net_intro(URL).

AUTHORS PROFILE

Roopa Sri Paladugu, final year student pursuing B.Tech in the stream of Information Technology in Velagapudi Ramakrishna Siddhartha Engineering College.

Anusha Immadisetty, final year student pursuing B.Tech in the stream of Information Technology in Velagapudi Ramakrishna Siddhartha Engineering College.

Dr. M. Ramesh is working as an Asst. Professor in Velagapudi Ramakrishna Siddhartha Engineering College.