EMR ANALYSIS TO PREDICT VARIOUS DISEASE

Dharmesh Chaudhary, Neha Dodake, Sakshi Mukkirwar

Electronics and Telecommunication Department

Pimpri Chinchwad College of Engineering and Research, Pune, India

(dharmesh.chaudhari_etc2017@pccoer.in), (neha.dodake_etc2017@pccoer.in), (sakshi.mukkirwar_etc2017@pccoer.in)

Guide by: - Assistant Professor Maithili Andhare

Electronics and Telecommunication Department

Pimpri Chinchwad College of Engineering and Research, Pune, India

Abstract

There is a need to study and make a system that will make it easy for end-users to predict diseases. To detect the various diseases through examining medical images of a patient using different techniques of deep learning models. To handle image data and structured data is no Proper method. The proposed system considers both the structure & unstructured data. The accuracy will Increase using deep learning and machine learning concept.

Keyword: *Convolutional Neural Network (CNN), Image data Augmentation*

1. INTRODUCTION

There is a tremendous growth in image acquisition devices, the data is moving to big data. This makes it very difficult for image analysis. The speedy growth in medical images requires a lot of efforts by the medical specialist that is irrational, liable to err. This may vary expert to expert. A machine learning algorithm, deep learning algorithm, and image processing techniques automate the diagnosis process by using data sets. This is a substitute solution.

1.1 DATA SET

In this, datasets are separated into test and train which sets a part of evaluating data mining models. Train/Test is a method is used to measure the accuracy of the model. There are two sets implemented in this model.

1. A training set is used to build up a model. 2. A test set is used to authenticate the model.

The training data helps to fit in the model whereas; testing data to test it. This model is built to predict the results which are present in test data sets. Here datasets are nothing but cell or organ images in which there are two different types of images known to be uninfected Images and infected images.

1.2 FEATURE EXTRACTION

Descriptive sentences written by experts for the medical images, such as Red blood cell images (RBCs), X-ray and magnetic resonance imaging (MRI) help in medical image diagnosis.

Deep learning provides high performance computing which deal with huge medical data for precise diagnosis. Deep learning and image processing are key parts of the project**.** These techniques are composed of conventional algorithms- Convolutional Neural Network (CNN); these algorithms indicate the behavior of the data by analyzing the information representation and abstraction spontaneously from a large set of images.

Image processing Techniques used:-

1. A medical image acquisition.

2. Image data Augmentation (pre-processing the layers)

3. Image diagnoses (Extraction of the data of images)

2. LITERATURE SURVEY

In December 2012, Edward Kennedy stated that the use of the electronic medical record (EMR) might increase extensively in future, but a lot of research needs to be conducted to analyze whether using EMR data is flexible, adaptive and if statistical methods could enhance medical disease prediction.

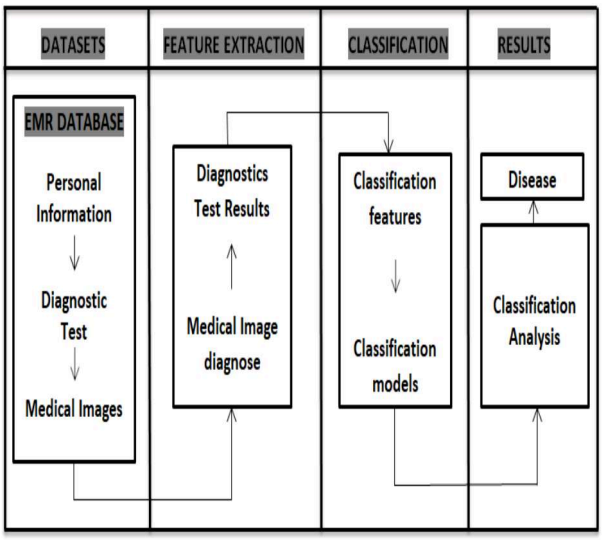
In March 2016, Riccardo Miotto stated the methodology that describes motivation for preventive medicine because of increase in healthcare cost. The main concern is recognizing the effects of the disease so that we can take preventive actions in the earlier stages. Hence they came up with an application of deep learning to obtain patient representations from EMR and to determine diseases.

In August 2018, POIS one group proposed a machine learning model for predicting patient mortality in coronary artery disease. This model primarily uses the electronic medical records of the patient that can overcome the shortcoming of the conventional survival models.

3. PROPOSED METHODOLOGY

This system is used to predict most diseases. It accepts the structured and image type of data as

input to the machine learning model. The end user will use this system. The system will predict the disease based on Blood microscopic images, x-ray, and other biological data. This system uses Machine Learning Technology. For predicting diseases Naïve Bayes algorithm, for clustering CNN algorithm and other deep learning algorithm, the final output will be in the form of labels like Infected for uninfected for which Logistic tree is used.



4. FLOWCHART

The first step is to collect the dataset which has images of Uninfected and infected blood cells or an infected organ by diagnosing. The next step is to get the data from given uninfected and infected cells or blood images by resizing and training the last layer to extract the data from the given image and preprocessing, again and again, using the CNN model. The step ahead is to plot the graph for both the uninfected and infected cells to get accuracy and loss. The second last step is to classify the plot and evaluate the classifier. The last step is to find the closest match. The flowchart of operation is as shown in the figure below.

5. ALGORITHM

1. Import and Initialize the Libraries.
2. Import the VGG16 library and add a preprocessing layer, get the summary of images.
3. Initiate the images from the training dataset.
4. Train the dataset.
   a) Resize all the images.
   b) Flatten the images.
   c) Create a model and get the summary.
   d) Create a model using CNN and get the summary by training the last layer again and again.
   e) Compile the model
   f) Use the Image data generator to read the images from Datasets.
   g) Apply the Image data Augmentation, fit the model and run the cell.
   h) Plot the loss and accuracy.
5. Evaluate the classifier using the test dataset and trained dataset.

6. Select the infected and uninfected images.
7. Represent the feature of images.
8. Find the closest match.
9. Display the string labels of the image.

6. SIMULATION RESULTS

This is the output after the classification of parasitic and uninfected images

```
if(a==1):
    print("Uninfected")
else:
    print("Infected")
```

```
Infected
```

7. THE CONCLUSION

Considering the above methodology, we can conclude that the newly designed models of image processing and deep learning technologies are highly suitable for accurate detection of diseases. These applications will reduce human efforts as well as human errors. These activities will be performed by automatic machines. Machine learning and deep learning facilitate and assist doctors to diagnose disease accurately. The risk of diseases can be understood faster and can be prevented earlier. We should look for more sophisticated deep learning methods due to the increase in data and the sensitivity of healthcare data. This will help to handle complex healthcare data with greater efficiency. So far deep learning and machine learning based algorithm have provided a positive feedback.

8. ADVANTAGES

1. The disease prediction model has the potential help interested parties such as the government and health insurance companies. It can identify the risk of disease or health conditions of a patient.
2. Tends to give fast results on execution
3. Scalable approach to large datasets.

9. APPLICATION

1. Faster detection using the image than the microscopic test.
2. Low cost.
3. The detection can be made more accurate than another form of detecting disease.

10. REFERENCES

[1] G.S. Birkhead, M. Klompas, N.R. Shah
Uses of electronic health records for public health surveillance to advance public health
Annu. Rev. Public Health, 36 (1) (2015), pp. 345-359,

[2] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, J.F. Hurdle
Extracting information from textual documents in the electronic health record

Yearbook Med. Inform., 17 (01) (2008), pp. 128-144

[3] A. Rajkomar, E. Oren, K. Chen, A.M. Dai, N. Hajaj, M. Hardt, P.J. Liu, X. Liu, Scalable and accurate deep learning with electronic health records
Med., 1 (1) (2018), p. 18, 10.1038/s41746-018-0029-1

[4] Yu Cheng, Fei Wang, Ping Zhang, and JianyingHu.2016.
Risk Prediction with Electronic Health Records: A Deep Learning Approach. In Proceedings of the 2016 SIAM International Conference on Data Mining (SDM'16). 432–440.

[5] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On ArXiv preprint arXiv: 1409.1259 (2014).

[6] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, (KDD'16). 1495–1504

[7] Masumi Shah, Khushbu Kawale, Divyansh Shah, Rahul Mapari, Santosh Randive. "Beneficial to humankind Malaria Parasite Detection Using Deep Learning": International Conference on Intelligent Computing and Control Systems (ICICCS), 2020