International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

Human Disease Detection using Machine and Deep Learning algorithms

Ritik Dixit[1], Rahul Verma[2], Pushkar Dhakad[3], Prof. Sushma Khatri[4]

[1,2,3] *Student, Department of Computer Science and Engineering,*
Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India

[4] *Associate Professor, Department of Computer Science and Engineering,*
Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India

---

Abstract – This research delineates the detection of human diseases with the help of x-ray images and symptoms. The main objective of the research is to ease, expedite, and economize the process of human disease diagnosis using intricate machine learning and deep learning algorithms. Detection of diseases like COVID-19, Lung cancer, Brain tumor, and Breast cancer is performed in this research and promising results have been obtained. Algorithms like Support Vector Machine, Logistic Regression, and Convolutional Neural Networks are employed to detect human diseases. The aim of this research is to surmount the disadvantages of traditional human disease detection methods and to compare machine and deep learning algorithms for the purpose of human disease detection.

Key Words: Machine Learning, Deep Learning, Support Vector Machine (SVM), Logistic Regression, Convolutional Neural Networks (CNN), Image Recognition

1. INTRODUCTION

Medical X-rays are images that are generally used to diagnose some sensitive physical body parts like bones, chest, teeth, skull, and so on. Doctors have used this system for several decades to explore and visualize fractures or abnormalities in body organs. This is often thanks to the very fact that X-rays are very effective diagnostic tools in revealing pathological alterations, additionally to their noninvasive characteristics and economic considerations. Chest diseases are often shown in CXR images within the sort of cavitations, consolidations, infiltrates, blunted costophrenic angles, and little broadly distributed nodules. By analyzing the chest X-ray image, the radiologists can diagnose many conditions and diseases like pleurisy, effusion, pneumonia, bronchitis, infiltration, nodule, atelectasis, pericarditis, cardiomegaly, pneumothorax, fractures, and lots of others.

Classifying the chest X-ray abnormalities is taken into account as a tedious task for radiologists; hence, many algorithms were proposed by researchers to accurately perform this task. Over the past decades, computer-aided diagnosis (CAD) systems are developed to extract useful information from X-rays to assist doctors in having a quantitative insight about an X-ray. However, these CAD systems couldn't have achieved a significant level to form decisions on the sort of conditions of diseases in an X-ray.

Various research works are administered on the diagnosis of chest diseases using AI methodologies. In, multilayer, probabilistic, learning vector quantization, and generalized regression neural networks are used for the diagnosis of chest diseases. The diagnosis of chronic obstructive pulmonary and pneumonia diseases was implemented using neural networks and artificial systems. In, the detection of lung diseases like TB, pneumonia, and carcinoma using chest radiographs is taken into account. The histogram equalization in image segmentation was applied for image preprocessing, and a feedforward neural network is employed for classification purposes. Deep learning-based systems are applied to extend the accuracy of image classification. These deep networks showed superhuman accuracies in performing such tasks. This success motivated the researchers to use these networks to medical images for diseases classification tasks, and therefore the results showed that deep networks can efficiently extract useful features that distinguish different classes of images. The most ordinarily used deep learning architecture is that the convolutional neural network (CNN). CNN has been applied to varied medical images classification thanks to its power of extracting different level features from images.

During this paper, a deep convolutional neural network (CNN) is used to enhance the performance of the diagnosis of chest diseases in terms of accuracy and minimum square error achieved. For this purpose, the CNN model is trained over a large x-ray image dataset and its performance is evaluated.

For the SVM implementation, we supply a series of symptoms as an input in SVM, using which we SVM makes decisions. The data collected are categorized as follows. The input variables are those symptoms that a patient observed before coming to the hospital, the test results, and a few biodatas of the patient. They include Breathing problems, Fever, Dry Cough, Sore throat, Hyper Tension, Abroad travel, Contact with COVID Patient, Attended Large Gathering, Visited Public Exposed Places, Family working in Public Exposed Places. After processing this data the prediction is made whether the person has that particular disease or not,

2. Related Work

There has been a considerable amount of research that has given impetus to machines and deep learning. Following are the related works that have been done in this realm:

[1] "Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks", Boran Sekeroglu, Ilker Ozsahin performed detection of COVID-19 using chest

X-ray images and obtained an accuracy of 98.50%. Their data consists of X-ray images (1583 healthy, 4292 pneumonia, and 225 confirmed COVID-19).

[2] "A novel approach to predict COVID-19 using support vector machine", by Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, and Jyoti Sekhar Banerjee. They performed the detection of COVID-19 by using a support vector machine classifier. They used the predominant set of symptoms and checked whether the person is COVID-19 positive or not. They achieved an accuracy of 87% in predicting the cases.

3. Methodology

The comparison of the algorithms (Support vector machines & Convolutional neural network) is predicated on the characteristic chart of every algorithm on common grounds like dataset, the complexity of the algorithm, accuracy of every algorithm, the runtime of the algorithm, under ideal conditions.

3.1 Dataset

Dataset is collected form various data providing websites like kaggle, Laboratório Visão Robótica e Imagem. We have collected thousands of X-ray images for image-based diagnoses and CSV files for symptom-based diagnoses. The COVID dataset contained 21 columns (each representing symptom) and 5434 rows (each representing patient no). The lung cancer dataset comprises seven columns (symptom) and 60 rows (patient no). COVID-19 x-ray dataset contains x-ray images of COVID positive patients and COVID negative patients.

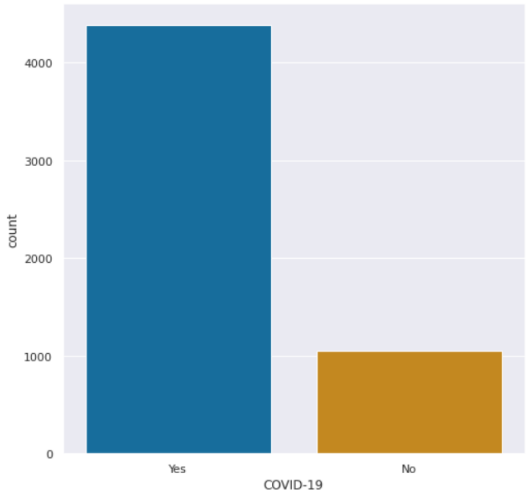

Fig 1: Picture depicts the count of COVID + and – in the symptom-based COVID dataset.

3.2 Support Vector Machines

Support Vector Machine (SVM) is a class of supervised machine learning algorithms used for classification. There's generally plotting of knowledge items in n-dimensional space where n is that the number of features, a specific coordinate represents the worth of a feature, we perform the classification by finding the hyperplane that distinguishes the 2 classes. It'll choose the hyperplane that separates the classes correctly. SVM chooses the acute vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed Support Vector Machine. There are mainly two sorts of SVMs, linear and non-linear SVM. We have used linear SVM.

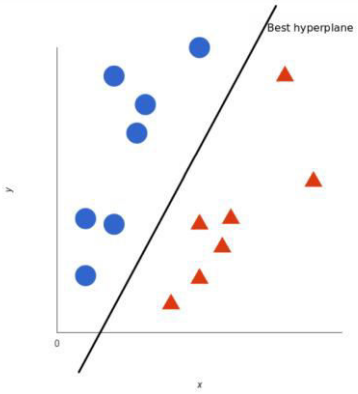

Fig 2: This figure shows SVM Classifier with supporting vectors and hyperplanes.

3.3 Convolution Neural Network

CNN may be a deep learning algorithm that's widely used for image recognition and classification. It's a category of deep neural networks that need minimum pre-processing. It inputs the image within sort of small chunks instead of inputting one pixel at a time. CNN an input layer, an output layer, convolutional layers, Pooling layers(Max and Average pooling), Fully connected layers (FC), and normalization layers. CNN uses a filter (kernel) and employs activation functions at each layer.

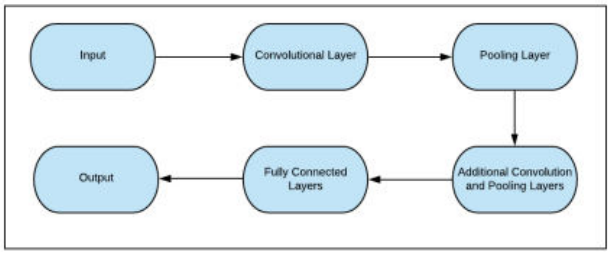

Fig 3: This figure shows the architectural design of CNN layers.

3.4 Visualization

In order to present model clear result, visualization is performed. This is done using the python library Matplotlib. Various bar graphs are plotted to delineate and compare the accuracies of these algorithms (SVM and CNN).

4. Implementation

To compare the algorithms based on accuracy for the detection of the human disease this paper uses: Support Vector Machine Classifier and Convolutional Neural Network Classifier. It provides a complete understanding of how each algorithm is implemented explicitly:

4.1 Pre-Processing:

Pre-processing is the inceptive process in the machine and deep learning. This process helps in ameliorating the input data which can help in obtaining better results and accuracy. Various steps are performed here like checking null values,

feature transformation, finding co-relation between features, dropping unwanted features. Also, the input images have been reshaped into dimensions (224,224,3).

4.2 Support Vector Machines

In the paper, Support Vector Machine is used to detect COVID-19 and lung cancer from symptoms. SVM is a supervised algorithm which means that the data we are using must have their respective output values. After pre-processing the data, the data is split into two parts: train and test. The kernel used here is a linear kernel. Training data is given to the SVM classifier to learn from the data and then the prediction is performed on testing data.

4.3 Convolution Neural Network

As Convolutional Neural Network is an image classifying algorithm, thus image detection is performed with the help of this. X-ray images are used to train and test the models. Detection of diseases like COVID-19, Breast cancer, and Brain tumor is performed using CNN. After splitting the data into test and train sets, the model is defined. Various layers like Convolutional Layer, Pooling Layer, dropout layer are used to define the model. Activation functions like Relu are also employed. After these sets of layers, fully connected layers are defined and lastly, the sigmoid function is used to predict outputs.

5. Result

The main objective of this paper is to compare the accuracies of the above-specified algorithms (SVM) & (CNN). This comparison is supported by experimental graphs in order to get more perspicuous results. We have compared both the training and testing accuracy of the models stated above. After executing all the models, the following results were obtained:

Table -1: Comparison of accuracy of various algorithms

S. no	Name of the algorithm	Name of the disease	Detected through	Accuracy (in %)
1	Support Vector Machine	COVID-19	Symptoms	97.3
2	Support Vector Machine	Lung Cancer	Symptoms	100
3	Convolutional Neural Network	COVID-19	X-ray image	98.33
4	Convolutional Neural Network	Brain Tumor	X-ray image	70.18
5	Convolutional Neural Network	Breast Cancer	X-ray image	60.94

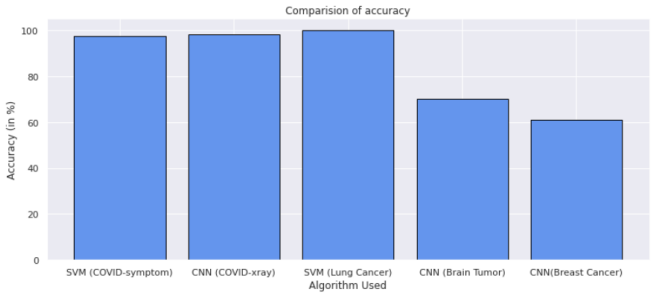

Fig 4: This figure shows the comparison of accuracy for different algorithms.

6. Conclusion

This research paper has implemented two models namely Support Vector Machine and Convolutional Neural Network. It has compared them on the basis of their working accuracy, execution time, complexity, to inform the best accurate model among them and concluded that:

1. Execution time for SVM was less as compared to CNN.
2. For COVID-19 detection it is better to use CNN (along with an x-ray image dataset) to obtain better accuracy.
3. SVM gave a promising result for lung cancer detection with an accuracy of 100%.

REFERENCES

1. **“**Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks", Boran Sekeroglu, Ilker Ozsahin.

2. "A novel approach to predict COVID-19 using support vector machine", by Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, and Jyoti Sekhar Banerjee.

3. COVID-19 X-ray dataset, http://cb.lk/covid_19

4. J. Bouvrie, Notes on Convolutional Neural Networks, http://cogprints.org/5869/1/cnn_tutorial.pdf.

5. “Support Vector Machine Algorithm”, https://www.javatpoint.com/machine-learning-supportvector-machine-algorithm

6. “Basic Overview of Convolutional Neural Network” https://medium.com/dataseries/basic-overview-ofconvolutional-neural-network-cnn-4fcc7dbb4f1

7. “Support vector machines”, https://scikitlearn.org/stable/modules/svm.html

8. “How Do Convolutional Layers Work in Deep Learning Neural Networks” available: https://machinelearningmastery.com/convolutional-layersfor-deep-learning-neural-networks

9. “Basic Overview of Convolutional Neural Network (CNN)”, https://medium.com/dataseries/basic-overviewof-convolutional-neural-network-cnn-4fcc7dbb4f17

10.“Pooling Layers for Convolutional Neural Networks”, https://machinelearningmastery.com/pooling-layers-forconvolutional-neural-networks/