International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Detecting Pneumonia from Chest X-Ray Images using Committee Machine

Sanjana Pawshe[1], Nivedita Prasad[2], Vaishnavi Phondekar[3], Prof. Rasika Shintre[4]

[1,2,3]Student, Computer Engineering, SIGCE, Navi Mumbai, Maharashtra, India
[4]Asst. Professor, Computer Engineering, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India

---

Abstract - *The medical field is the most sensitive of all the domains ever known, for a simple reason that it deals with humans and advances in this field is a matter of pride for the entire human race. The system designed in this research paper is one such attempt. Pneumonia is most important cause of death worldwide even though it is a vaccine preventable disease. It can be detected by analysing chest x-rays. Analysing chest x-rays is a difficult task and requires precision. We aim at designing a highly efficient system to predict if a user suffers from Pneumonia by analysing the patient's chest X-ray images and increasing the accuracy of the system by use of Committee Machine.*

Key Words: Image Processing, Machine Learning, Classification Algorithms, Committee Machine

1. INTRODUCTION

This project is about building a web application that can diagnose a patient with pneumonia correctly by analysing its X-ray image. Even though there are already several working and successful systems that have the same function, this project attempts to create a system that has a different approach from other systems, in terms of how the system is built. Pneumonia is a disease that causes the inflammation of the air sacs inside either of the lungs. This infection is most commonly diagnosed by taking the X-ray of the patient. The risk of pneumonia is immense for many, especially in developing nations where billions face energy poverty and rely on polluting forms of energy. The WHO estimates that over 4 million premature deaths occur annually from household air pollution-related diseases including pneumonia. Over 150 million people get infected with pneumonia on an annual basis especially children under 5 years old. In such regions, the problem can be further aggravated due to the dearth of medical resources and personnel. For example, in Africa's 57 nations, a gap of 2.3 million doctors and nurses exists. For these populations, accurate and fast diagnosis means everything. It can guarantee timely access to treatment and save much needed time and money for those already experiencing poverty.

People with infectious pneumonia often have a productive cough, fever accompanied by shaking chills, shortness of breath, sharp or stabbing chest pain during deep breaths, and an increased rate of breathing. In elderly people, confusion may be the most prominent sign.

The typical signs and symptoms in children under five are fever, cough, and fast or difficult breathing. Fever is not very specific, as it occurs in many other common illnesses and may be absent in those with severe disease, malnutrition or in the elderly. In addition, a cough is frequently absent in children less than 2 months old. More severe signs and symptoms in children may include blue-tinged skin, unwillingness to drink, convulsions, ongoing vomiting, extremes of temperature, or a decreased level of consciousness.

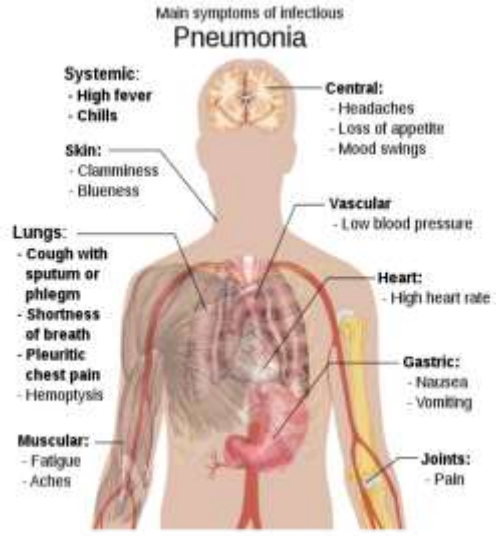

Fig -1: Main Symptoms of Infectious Pneumonia

Bacterial and viral cases of pneumonia usually result in similar symptoms. Some causes are associated with classic, but non-specific, clinical characteristics. Pneumonia caused by Legionella may occur with abdominal pain, diarrhea, or confusion. Pneumonia caused by Streptococcus pneumonia is associated with rusty colored sputum. Pneumonia caused by Klebsiella may have bloody sputum often described as "currant jelly". Bloody sputum (known as hemoptysis) may also occur with tuberculosis, Gram-negative pneumonia, lung abscesses and more commonly acute bronchitis. Pneumonia caused by Mycoplasma pneumonia may occur in association with swelling of the lymph nodes in the neck, joint pain, or a middle ear infection. Viral pneumonia presents more commonly with wheezing than bacterial pneumonia. Fig -1 shows the major symptoms of pneumonia patients.

The medical field is predicted to be the most benefitted field, after finance from the new age concepts of Artificial Intelligence. We aim at applying these concepts to the field of medical science and make use of these concepts to unleash new horizons in medical diagnosis. We aim at processing X-ray images of chest to predict if the patient is diagnosed with pneumonia. We have a set of 5,863 X-Ray images which will be used to predict if X-ray of chest is suffering from pneumonia or no. In summary of the system, the system consists of modules, were each module has smaller sub-modules that build-up to the whole system. The image processing component is for obtaining the data set and converting each image into a matrix of data using various algorithms. This is because machine learning algorithms can only accept numerical values. A component of this tabulated data will be used to train the machine learning algorithm. Once the algorithm has been trained, the more a component of the data set will be used to test the algorithm and obtain the system's level of efficiency and accuracy. However, each component has more in-depth steps which will be explained in further.

2. LITERATURE SURVEY

The predominant way hospitals diagnose pneumonia today is through getting a radiologist to create radiology reports for all the key pneumonia findings. The latest ai system can report key pneumonia findings are called chexpert. This is an ai system that uses Convolution Neural Networks (CNN) as a deep learning algorithm that can be used in a variety of image analysis and hence is befitting for the analysis of chest x-ray images. The reason for creating such a system that already has substitutes in the market that have had success is because this system may provide fresh insight. This insight might be useful when building more systems in the future. It is important to approach solutions to problems from all directions. In Table -1, Comparative study of different research papers.

In Paper [1], Here, it establishes a diagnostic system based on a deep-learning framework for detecting whether the patient has pneumonia or not. Only classification algorithm is used the Neural Network and there is no comparison between the accuracies of several algorithm. Automated detection of diseases from chest X-rays at the level of expert radiologists would not only have tremendous benefit in clinical settings, it would also be invaluable in delivery of health care to populations with inadequate access to diagnostic imaging specialists.

Okeke Stephen, Mangal Sain , Uchenna, and Do-Un Jeong In [2] proposed a convolutional neural network model trained from scratch to classify and detect the presence of pneumonia from a collection of chest X-ray image samples. Unlike other methods that rely solely on transfer learning approaches or traditional handcrafted techniques to achieve a remarkable classification performance, they constructed a convolutional neural network model from scratch to extract features from a given chest X-ray image and classify it to determine if a person is infected with pneumonia. This model could help mitigate the reliability and interpretability challenges often faced when dealing with medical imagery.

In IEEE Paper [3], It studies the differences in the results when neural network and support vector machines are used in studying the gait of various age groups. the document shows the accuracies derived from using different neural network concepts and accuracies using different kernels of SVM. In this case, all the SVM kernels outperform the Neural Network algorithms by a significant margin. Therefore, this research paper shows that SVM has better system performance compared the neural network algorithms.

In [4], that IEEE that will be examined to assist in deciding which is the more appropriate classifier model. The direct quote that has been picked up from the paper than supports the usage of SVM more than NN is "the SVCs extract more information from the training/input space"

By Daniel S. Kermany, Michael Goldbaum, Wenjia In paper [5] Here, It establish a diagnostic tool based on a deep-learning framework for the screening of patients with common treatable blinding retinal diseases. The performance of model depends highly on the weights of the pre-trained model. Therefore, the performance of this model would likely be enhanced when tested on a larger ImageNet dataset with more advanced deep-learning techniques and architecture.

In paper [6], The proposed paper presents a deep neural network based on convolutional neural networks and residual network along with techniques of identifying optimum differential rates using cosine annealing and stochastic gradient with restarts to achieve an efficient and highly accurate network which will help detect and predict the presence of pneumonia using chestx-rays.

Table -1: Comparative Study

Sr. No	Titles	Methods /Techniques	Advantages	Disadvantages
1	CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning	Deep Learning &Framework classification algorithm used is the Neural Network	Automated detection of diseases from chest X-rays at the level of expert radiologists	No comparison is made between the accuracies of several algorithm
2	An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare	Deep Learning Framework is used and the only classification algorithm used is the Neural Network	Convolutional neural network model is constructed from scratch to extract features from given chest X-ray image and classify it to determine if a person is infected pneumonia.	No comparison is made between the accuracies of several algorithm.
3	Comparison of Neural Networks and Support Vector Machines for Recognizing Young-Old Gait Patterns	Classification algorithms used are Neural Network and Support Vector Machine	Comparison is made between the accuracies of both algorithms.	The overall classification accuracy was found to be the same irrespective of the kernel types.
4	Estimation of multi-pattern-to-single-pattern functions by combining feedforward neural networks and support vector machines	Classification algorithms used are Neural Network and Support Vector Machine	Proposes 200 *200*3 model for better validation accuracy with minimal training loss	Occurrence of errors are more in single Feed Forward Neural Network with large no. of hidden neurons
5	Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning	Deep Learning Framework is used and the only classification algorithm used is the Neural Network	X-rays presenting as pneumonia versus normal, we achieved an accuracy of 92.8% with sensitivity of 93.2% and specificity of 90.1%	No comparison is made between the accuracies of several algorithm
6	Deep Learning Approach For Prediction Of Pneumonia	Deep Learning Framework is used and the classification algorithm used is the Convolutional Neural Network	Predicts Pneumonia in minimal time, high efficiency with 92.9% accuracy	Processing building the model requires fast and efficient processors which is cost consuming

3. PROBLEM STATEMENT

In the early stages when technology was yet to transform the prediction of diseases was mainly based on human knowledge and intuition. The probability of human error is high and thus the prediction of disease also becomes difficult. Hence the treatment may not help the patient. Also, when in a remote area, access to medical facilities is not always possible. Thus, doctors cannot be available at such occasions as well.

Thus we have created a system which will help the patients understand whether they are suffering from anything severe like pneumonia. If yes, they need to take the right actions pertaining the same. The chest X-ray is provided by the user as inputs, and the system will determine depending on the input whether the user is suffering from pneumonia or no. The chances of doctors scamming patients to conduct unnecessary test can also be avoided.

4. PROPOSED SYSTEM

4.1. Dataset

The original dataset consists of three main folders (i.e., training, testing, and validation folders) and two subfolders containing pneumonia (P) and normal (N) chest X-ray images, respectively. Total number of 5,863 X-ray images of anterior-posterior chests were carefully chosen from retrospective paediatric patients. Below fig- 2 and 3 shows

the sample x-ray images of pneumonia and images without pneumonia.

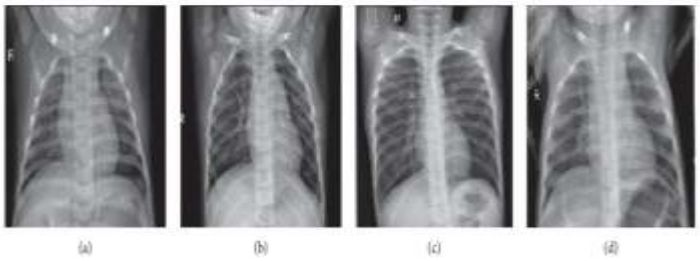

Fig -2: Sample images without pneumonia

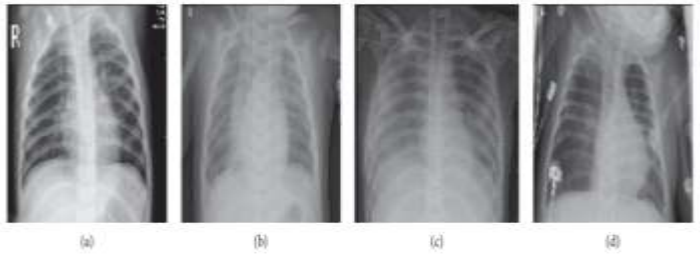

Fig -3: Sample images with pneumonia

4.2. Image Processing

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them.

Following steps are performed for implementing image processing module:

1. Take the dataset of 5,863 images.
2. Repeat the steps from for all the images in a particular folder.
3. Load the image in python. Perform image processing on it to get only the required part of the image.
4. This can include getting rid of the background, finding the outline, etc
5. Perform Feature Extraction on the image. Migrate the extracted features to an external storage location. These extracted features will be used for the further steps in the algorithm.

4.3. Classification Algorithms

For detecting pneumonia from chest x-ray images we aim at using three classifiers. By using different classification algorithm we can get different results in which we

4.3.1. Back Propagation Neural Network (BPNN)

BPNN is an Artificial Neural Network (ANN) based powerful technique which is used for detection of the intrusion activity. Basic component of BPNN is a neuron, which stores and processes the information. BPNN is a supervised algorithm in which error difference between the desired output and calculated output is back propagated. The procedure is repeated during learning to minimize the error by adjusting the weights thought the back propagation of error. As a result of weight adjustments, hidden units set their weights to represent important features of the task domain.

BPNN consists of three layers: 1) Input Layer 2) Hidden Layer and 3) Output Layer. Number of the hidden layers, and number of hidden units in each hidden layers depend upon the complexity of the problem.

4.3.2. Naïve Bayes (Nb) Classifier

Naïve Bayes is a relatively simple machine learning technique based on probability models - Bayesian theorem. Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. Naive Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence.

4.3.3. Support Vector Machine (Svm) Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification challenges. However, it is mostly used in classification problems. In this algorithm, plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.

Support Vectors are simply the coordinates of individual observation. In this paper mainly we will consider the input is based upon Support Vector Machine as training data, testing data is decision value. In this method we consider the following steps like Load Dataset, after loading the dataset will Classify Features (Attributes) based on class labels then estimate Candidate Support Value, like the condition is While (instances!=null), Do condition if Support Value=Similarity

between each instance in the attribute then finding the Total Error Value. Suppose if any instance < 0 then the estimated decision value = Support Value\Total Error, repeated for all points until it will be empty. Therefore mainly we have calculated the entropy and gini index.

Following steps are performed for classification process:

1. Load the data into Python for classification.
2. Pre-process the data to get it into the required form.
3. Split the data into training and testing data set. Form classifier models of the algorithms stated previously.
4. Train the classifier with the training data set obtained in step.
5. Test the classifier which has been created in the previous step.

4.4. Committee Machine And Ada Boosting

Ensemble boosting classifier which combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get high accuracy strong classifier. The basic concept behind Ada-boost is to set the weights of classifiers and training the data sample in each iteration such that it ensures the accurate predictions of unusual observations. Any machine learning algorithm can be used as base classifier if it accepts weights on the training set.

In Committee Machine we take outputs of all classifiers and show best optimized result. It is used to optimize the results by increasing accuracy.

Following are steps to perform committee machine:

1. Take outputs from all the classification algorithms provided.
2. Pass the outputs to a committee machine that makes use of Ada Boosting as inputs.
3. The committee machine will process all of those.
4. The committee machine will make use of Ada boosting algorithm to find the final result to the user's query.
5. This output will enhance the accuracy of the system.

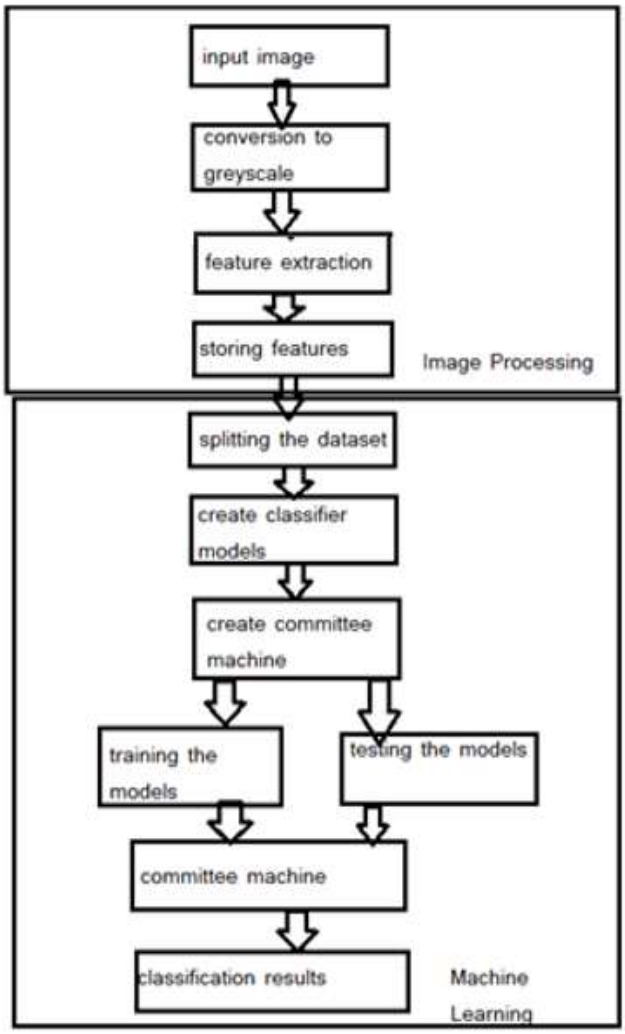

Fig -4: System Flow Diagram

Fig-4 shows whole system Flow of Pneumonia Detection which mainly consist of two modules which are Image Processing module and Machine Learning module. There will be simple User Interface be there so that user select x-ray image to detect the particular patient has pneumonia or not.

5. CONCLUSION

The proposed system planned after extensive research during a literature survey includes the implementation of Committee machine for predicting pneumonia based on the chest x- ray image provided as input by the user. The system will provide the final result that whether that person is having pneumonia or not. It will also provide highest accuracy from the three different classifiers. Easy user interface has been designed keeping the user's convenience in mind.

ACKNOWLEDGEMENT

The success and final outcome of this research required a lot of guidance and assistance and we are extremely privileged to have got this all. All that we have done is only due to such supervision and assistance and we would not forget to thank them.

We respect and thank Prof. Rasika Shintre, for providing us insight and expertise that greatly assisted the research. We are extremely thankful to her for providing such a nice support and guidance.

REFERENCES

[1] CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning —Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng PLoS Med 15(11): e1002686. https://doi.org/10.1371/journal.pmed.1002686

[2] Okeke Stephen , Mangal Sain , Uchenna Joseph Maduh , and Do-Un Jeong 'An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare' Hindawi Journal of Healthcare Engineering Volume 2019, ID 4180949

[3] Begg, R. 'A Comparison of Neural NetFvorks and Support Vector Machines for Recognizing Young-Old Gait Patterns' ISBN: 0-7803-8162-9

[4] Pakka, V.H., Thukararn, D.;\, Khincha H.P., 'Estimation of multi-pattern-to-single-pattern functions by combining feedforward neural networks and support vector machines' ISBN:0-7803-8547-0

[5] Daniel S. Kermany, Michael Goldbaum, Wenjia 'Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning' Cai Cell 172, 1122–1131

[6] Kalyani Kadam, Dr.Swati Ahirrao, Harbir Kaur ,Dr. Shraddha Phansalkar , Dr. Ambika Pawar' Deep Learning Approach For Prediction Of Pneumonia' INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 8, ISSUE 10, ISSN 2277-8616

[7] Aydogdu, M, Ozyilmaz, E, Aksoy, Handan, Gursel,G, and Ekim, Numan. Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores. Tuberk Toraks, 58(1):25–34, 2010

BIOGRAPHIES

Sanjana S. Pawshe, Pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi College of Engineering (SIGCE), Navi Mumbai. Her current research interests include Web Designing & Machine Learning

Nivedita C. Prasad, Pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi College of Engineering (SIGCE), Navi Mumbai. Her current research interests include Image Processing & IoT.

Vaishnavi S. Phondekar, Pursuing the Bachelor degree (B.E.) in Computer Engineering from Smt. Indira Gandhi College of Engineering (SIGCE), Navi Mumbai. Her current research interests include Image Processing & Web development.

Prof. Rasika Shintre, Obtained the Bachelor degree (B.E. Computer) in the year 2011 from Ramrao Adik Institute of Technology (RAIT), Nerul, and Master degree (M.E. Computer) from Bharati Vidyapeeth College of Engineering, Navi Mumbai. She is Asst. Professor in Smt. Indira Gandhi College of Engineering of Mumbai university and having about 8 yrs. of experience. Her area of interest includes Data mining & information retrieval.